

Research Article

Open Access



PINK: physical-informed machine learning for lattice thermal conductivity

Yujie Liu, Xiaoying Wang, Yuzhou Hao, Xuejie Li, Jun Sun, Turab Lookman , Xiangdong Ding, Zhibin Gao 

State Key Laboratory for Mechanical Behavior of Materials, School of Materials Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China.

*Correspondence to: Prof. Zhibin Gao, State Key Laboratory for Mechanical Behavior of Materials, School of Materials Science and Engineering, Xi'an Jiaotong University, No.28, West Xianning Road, Xi'an 710049, Shaanxi, China. E-mail: zhibin.gao@xjtu.edu.cn

How to cite this article: Liu, Y.; Wang, X.; Hao, Y.; Li, X.; Sun, J.; Lookman, T.; Ding, X.; Gao, Z. PINK: physical-informed machine learning for lattice thermal conductivity. *J. Mater. Inf.* 2025, 5, 12. <https://dx.doi.org/10.20517/jmi.2024.86>

Received: 13 Dec 2024 **First Decision:** 17 Jan 2025 **Revised:** 27 Jan 2025 **Accepted:** 11 Feb 2025 **Published:** 20 Feb 2025

Academic Editor: Sergei Manzhos **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Lattice thermal conductivity (κ_l) is crucial for efficient thermal management in electronics and energy conversion technologies. Traditional methods for predicting κ_l are often computationally expensive, limiting their scalability for large-scale material screening. Empirical models, such as the Slack model, offer faster alternatives but require time-consuming calculations for key parameters such as sound velocity and the Grüneisen parameter. This work presents a high-throughput framework, physical-informed kappa (PINK), which combines the predictive power of crystal graph convolutional neural networks (CGCNNs) with the physical interpretability of the Slack model to predict κ_l directly from crystallographic information files (CIFs). Unlike previous approaches, PINK enables rapid, batch predictions by extracting material properties such as bulk and shear modulus from CIFs using a well-trained CGCNN model. These properties are then used to compute the necessary parameters for κ_l calculation through a simplified physical formula. PINK was applied to a dataset of 377,221 stable materials, enabling the efficient identification of promising candidates with ultralow κ_l values, such as $\text{Ag}_3\text{Te}_4\text{W}$ and $\text{Ag}_3\text{Te}_4\text{Ta}$. The platform, accessible via a user-friendly interface, offers an unprecedented combination of speed, accuracy, and scalability, significantly accelerating material discovery for thermal management and energy conversion applications.

Keywords: Physical-informed machine learning, thermoelectrics, lattice thermal conductivity, phonon engineering



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



INTRODUCTION

Understanding the temperature dependence of lattice thermal conductivity (κ_L) is essential for assessing the thermal transport capabilities of a material. This property plays a crucial role in both scientific research and industrial applications, including thermal management in microelectronics^[1,2], energy conversion^[3], and temperature regulation^[4]. For example, materials exhibiting high κ_L , such as boron arsenide (BAs), are particularly suitable for heat dissipation in gallium nitride devices^[5]. Conversely, materials with low κ_L can improve thermoelectric conversion efficiency by enabling the effective transformation of waste heat into electrical energy^[6].

In recent years, significant theoretical advancements have been made in the theoretical prediction of κ_L in solid materials^[7-10]. A widely used approach for predicting κ_L involves solving the phonon Boltzmann transport equation (PBTE) within the framework of density functional theory (DFT)^[9], while classical molecular dynamics (MD) simulations are particularly useful for systems with complex crystal structures^[11]. However, identifying materials with exceptionally low or high κ_L remains a significant challenge, mainly due to the high computational costs and time-consuming synthesis processes^[12]. Moreover, calculations required to obtain interatomic force constants (IFCs) are especially demanding for large, low-symmetry primitive cells^[13]. Furthermore, the reliability of MD simulations is strongly dependent on the selection of interatomic potentials, limiting their broader applicability^[14]. Besides these challenges, significant progress has been made in accelerating material discovery and improving performance. Luo *et al.* reviewed the application of machine learning (ML) for predicting κ_L , emphasizing the potential of high-throughput predictions and ML potentials (MLPs) to overcome the limitations of traditional approaches^[13]. Liu *et al.* focused on active and reversible techniques for regulating κ_L , such as the use of ferroelectric, ferromagnetic, and nanomaterials, enabling dynamic control of thermal conductivity for efficient thermal management^[15]. Additionally, Shi *et al.* examined advancements in thermoelectric materials for multifunctional energy conversion and storage technologies, highlighting ongoing challenges related to scalability, material stability, and efficiency that must be addressed to fully realize their potential in practical applications^[16]. Consequently, rapid determination of κ_L is crucial for advancing these materials.

Alternatively, empirical models such as the Debye-Callaway model^[17,18] and the Slack model^[19,20] provide faster and more cost-effective approaches for estimating κ_L . The Slack model, in particular, has been widely applied to predict κ_L in a variety of materials^[21-26]. For instance, Qin *et al.* successfully employed the model to quickly predict thermal conductivity, offering valuable insights into thermal transport behavior^[27]. Cao *et al.* explored the n-type thermoelectric properties of ABO₃ cubic chalcogenides using a high-throughput method combined with Slack modeling^[28]. They screened 46 stable materials, identified four conduction band minima structures, investigated the influence of chemical bonding on transport properties, and shortlisted 13 candidates with high thermoelectric figure of merit (ZT) values. However, the model's reliance on experimental data or first-principles calculations for several parameters limits its scalability for large-scale, high-throughput screenings. Obtaining critical parameters, such as average sound velocity, acoustic Debye temperature, and the Grüneisen parameter, often requires considerable time and resources, posing a significant barrier.

In our previous work^[29], we proposed a refined formula based on the Slack model, which enables highly accurate predictions of κ_L with an 8.97% mean relative error. The formula utilizes only the shear modulus, average sound velocity, and Grüneisen parameter, all of which are relatively easy to obtain. For example, the bulk modulus (B) and shear modulus (G) can be used to derive the average sound velocity in a material^[30]. Additionally, significant research has been conducted to simplify the estimation of the Grüneisen parameter. Belomestnykh^[31] developed a method that links Poisson's ratio with sound velocity and elastic properties,

yielding results consistent with quasiharmonic lattice dynamics calculations. This work underscores the strong relationship between elastic modulus and κ_L ^[29]. The crystallographic information file (CIF) provides comprehensive data on crystal structures, including lattice constants, crystal systems, density, and other key parameters. However, existing approximation methods have not fully exploited this information to predict κ_L . Only a limited number of studies have directly connected CIF data with κ_L for fast, high-throughput predictions. For example, Ju *et al.* used a neural network that leveraged descriptors from a pre-trained model to establish a relationship between crystal information and thermal conductivity^[32]. Xie *et al.* introduced the crystal graph convolutional neural network (CGCNN) method^[33], which converts crystal structure data into graph representations, enabling convolutional neural networks to predict the relationship between crystal features and thermoelectric properties^[34,35]. Recently, Omeo *et al.* reviewed the performance of five out-of-distribution (OOD) test sets across eight graph neural network (GNN) models using elasticity datasets^[36]. Notably, the CGCNN model achieved the best mean absolute error (MAE) for both the Leave-One-Cluster-Out (LOCO^[37]) test [0.0585 log₁₀ (GPa)] and the SparseXsingle test [0.0499 log₁₀ (GPa)], targeting structures with the lowest density and surpassing the other seven GNN models. These findings highlight the excellent generalization capabilities of CGCNN models.

In this work, to enable rapid and high-throughput κ_L predictions, we integrate the physical interpretability of our derived formula with the predictive power of the CGCNN model. This study introduces a high-throughput framework that combines a trained modulus model with our formula, facilitating the fast estimation of κ_L directly from CIF files. Encapsulated in a custom-developed web application, physical-informed kappa (PINK), this process enables batch predictions of κ_L within seconds of uploading CIF files. Users can also customize inputs such as bulk modulus, shear modulus, and Grüneisen parameter.

Our framework begins by extracting crystallographic information from the CIF files and utilizing the trained CGCNN model to predict the bulk and shear modulus. Subsequently, physical models are applied to calculate the average sound velocity and Grüneisen parameters, which are then incorporated into a formula to calculate the κ_L . Using this approach, we predict κ_L for 377,221 stable materials identified by Merchant *et al.* through graph networks^[38]. Building on these high-throughput predictions, we develop an efficient method to accelerate the screening of materials with ultralow κ_L , applicable to any inorganic crystal structure with one or more CIF files. This method enabled the identification of thousands of promising materials with low κ_L from over 370,000 inorganic crystalline samples, with minimal computational cost. To validate our results, we confirm the ultralow κ_L values for Ag₃Te₄X (X = W, Ta) through first-principles calculations. The PINK application, powered by the CGCNN model, serves as a powerful tool for rapid material pre-screening. It provides researchers with an efficient, user-friendly platform for estimating κ_L , accelerating the discovery of materials with optimal thermal properties.

MATERIALS AND METHODS

CGCNN algorithms

Before presenting the framework of PINK, it is essential to clarify the method by which the CGCNN model predicts material properties based on crystal structures. CGCNN, an advanced ML algorithm, uses trained models to predict material properties with high efficiency^[33]. The crystal structure is converted into a graph representation, where nodes correspond to atoms, and edges represent the bonds between them. This format allows the model to capture the local chemical environment.

Through convolutional and pooling layers, CGCNN autonomously identifies critical features necessary for predicting various material properties, such as bulk modulus and shear modulus. These predictions are both accurate and interpretable, providing valuable insights for the rational design of new materials. Moreover,

the robust generalization capabilities of this model enable it to handle diverse crystal structures and compositions, significantly accelerating the material discovery process^[33,34,39]. In this study, the elastic modulus dataset was split into training, validation, and test sets with a ratio of 80%, 10%, and 10%, respectively. The model consisted of three convolutional layers and two hidden layers, and was trained for 30 epochs with a learning rate of 0.01.

The PINK framework

As illustrated in [Figure 1](#), we present a comprehensive workflow for calculating κ_L using our automated property prediction system. This workflow is designed to be user-friendly, requiring only CIF files as input. To ensure accurate calculations, the system automatically converts the uploaded crystal structure into its primitive cell format, which is essential for both CGCNN predictions and the parameters used in Equation (2). The process begins by extracting fundamental crystallographic data from the CIF file, including the primitive cell volume, number of atoms, and density. Next, our embedded CGCNN model, trained on extensive material data, predicts the bulk modulus and shear modulus. Using these predicted values, custom Python scripts calculate additional physical parameters crucial for estimating κ_L , including the longitudinal and transverse sound velocities, the average speed of sound, and the Grüneisen parameters. Finally, all of these calculated quantities are systematically incorporated into Equation (2) to compute κ_L . This automated workflow significantly streamlines the process of κ_L prediction, making it accessible to researchers without requiring in-depth expertise in each individual computational step.

The application provides comprehensive physical property data for 377,221 new materials, including 11,869 materials screened in this study. The modified open-source CGCNN code used for predicting bulk and shear modulus, as well as the Python scripts for CIF file processing and calculation execution (e.g., “app.py”), is also available. All of these data and codes are accessible via the following link: <https://github.com/Jack-Liu0227/AI4Kappa>.

Surrogate an interpretable formula for κ_L

Recently, Wang *et al.* proposed a simple and universal empirical formula that exhibits strong generalization ability and provides clear physical insights for κ_L of crystals, which is given as^[29]:

$$\kappa_L = \frac{Gv_s V^{\frac{1}{3}}}{nT\delta} \cdot e^{-\gamma}, \quad (1)$$

where G is the shear modulus, v_s represents the average sound velocity, V is the volume of the primitive cell, n is the number of atoms in the primitive cell, δ lies between 1 and 2 (with $\delta = 1$ for three-phonon scattering), T is the temperature in Kelvin, and γ denotes the Grüneisen parameter. It is important to note that κ_L and v_s in Equation (1) do not exhibit a conventional proportional correlation, as both G and γ are functionally dependent on v (see [Supplementary Materials](#) for details)^[29].

The theoretical basis of the power law is complex, involving competition between scattering processes driven by cubic and quartic anharmonic terms^[40,41]. For simplicity, we focus only on three-phonon scattering, assuming $\delta = 1$:

$$\kappa_L = \frac{Gv_s V^{\frac{1}{3}}}{nT} \cdot e^{-\gamma}, \quad (2)$$

Which, derived from Slack’s approach^[23] is useful for evaluating κ_L across various materials. A key aspect in evaluating κ_L involves determining the average speed of sound (v_s) and the Grüneisen parameter (γ). Jia *et al.* proposed that v_s can be accurately estimated from elastic properties [bulk modulus (B) and shear modulus

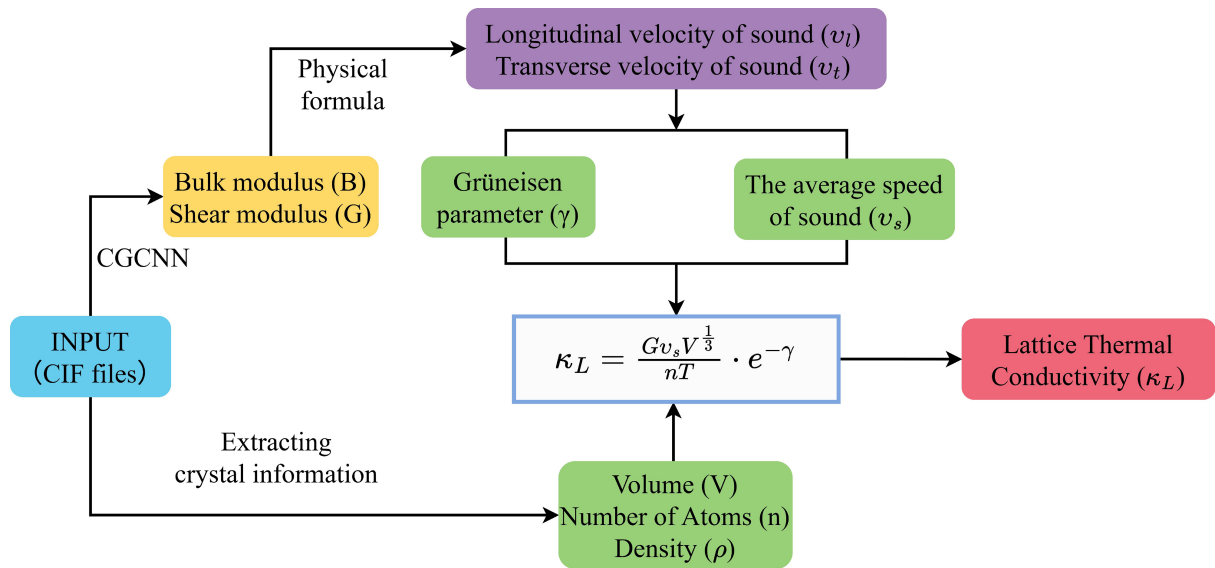


Figure 1. The workflow for calculating κ_L using PINK begins with the input of CIF files representing crystal structures. Starting with these CIF files, the framework utilizes CGCNN to predict the bulk and shear modulus, while also extracting crystal information such as volume, number of atoms, and density. These parameters are subsequently used to calculate both longitudinal and transverse sound velocities, which are essential for determining the Grüneisen parameter and the average speed of sound. All of these parameters are incorporated into Equation (2), which includes the Grüneisen parameter (γ), volume (V), temperature (T), and other variables necessary for predicting κ_L . PINK: Physical-informed kappa; CIF: crystallographic information file; CGCNN: crystal graph convolutional neural network.

(G)]^[42]. This approach is computationally more efficient than experimental methods or costly lattice dynamics simulations. The bulk modulus (B) and shear modulus (G) can both be extracted from our trained CGCNN model, providing an alternative means of estimating elastic properties and sound velocities, as demonstrated in^[27,42].

$$v_l = \sqrt{\frac{B + \frac{4}{3}G}{\rho}}, \quad (3)$$

$$v_t = \sqrt{\frac{G}{\rho}}, \quad (4)$$

$$v_s = \left\{ \frac{1}{3} \left[\frac{1}{v_l^3} + \frac{2}{v_t^3} \right] \right\}^{-\frac{1}{3}}, \quad (5)$$

where v_s , v_b , and v_t are the average sound velocity, longitudinal sound velocity, and transverse sound velocity, respectively, and ρ is the material density.

After estimating v_s from the bulk modulus (B) and shear modulus (G), the next step is to determine the Grüneisen parameter, which quantifies the anharmonicity of the material^[43]. The speed of sound serves as an indicator of the strength of atomic interactions, with weaker interactions generally leading to lower sound velocities. It has been shown that the relationship between Poisson's ratio (ν) and γ is as follows^[31,44]:

$$\nu = \frac{x^2 - 2}{2x^2 - 2}, \quad \gamma = \frac{3}{2} \left(\frac{1 + \nu}{2 - 3\nu} \right). \quad (6)$$

where x represents the ratio of longitudinal to transverse sound velocity, $x = v_l/v_t$.

Using the above method, both ν_s and γ can be estimated quickly from elastic properties, particularly the shear and bulk modulus. Previous studies indicate that this approach aligns well with experimental results for cubic, isotropic, and quasi-isotropic structures^[42,45,46].

Deployment of PINK for κ_L

Streamlit is an open-source Python library that simplifies the creation of custom web apps for data-driven applications. It facilitates rapid development of interactive apps by converting Python scripts into shareable web applications in just a few minutes. Figure 2 illustrates the process of deploying an app: first, upload the project code to GitHub, and then create the application on the Streamlit platform by selecting the relevant project branch and Python file (e.g., `app.py`) to run. Setting up the application environment can be challenging, as web applications typically require multiple Python packages with specific versions. Fortunately, our code streamlines this process by including a `requirements.txt` file to ensure all dependencies are installed correctly. This allows the application to be deployed entirely in Python without requiring front-end experience. By leveraging Equation (2) and the CGCNN model, we developed the κ_L calculation application based on this framework.

After deploying the application, users can quickly calculate a material's elastic properties, κ_L , and other relevant outputs by uploading CIF files. The results are displayed on the website in a DataFrame format, and users can download them as CSV files. An illustration of the program's interface is shown in Figure 3. Importantly, the app supports uploading single or multiple CIF files simultaneously, running the entire framework in parallel to provide results for all materials at once.

Our application, PINK, is easily accessible via the following link: <https://kappap-ai.streamlit.app>, which can be used both on your phone and on your local computer. If you prefer to deploy it locally or on Streamlit's server, please refer to the [README.md](#) for detailed instructions on setting up the software.

Calculating κ_L using *ab initio* study

We implemented *ab initio* study through the Vienna Ab-initio Simulation Package (VASP)^[47]. The calculations incorporated the projector augmented-wave (PAW) approach combined with the Perdew-Burke-Ernzerhof (PBE) functional for exchange-correlation^[48-50]. To achieve high computational precision, we selected a 520 eV planewave cutoff energy alongside a Monkhorst-Pack sampling with a $4 \times 4 \times 4$ k -mesh. The computational parameters were optimized with convergence thresholds of 10^{-8} eV for total energy and 10^{-4} eV/Å for atomic forces. In determining the second-order IFCs, our calculations utilized a supercell configuration of $2 \times 2 \times 2$, employing the finite displacement methodology with a $4 \times 4 \times 4$ k -point mesh. The third-order software package^[40] was subsequently used to extract the third-order IFCs. The κ_L calculations, which account for three-phonon scattering processes, were performed using ShengBTE with a dense $20 \times 20 \times 20$ q -point sampling^[51].

RESULTS AND DISCUSSION

To streamline the time-intensive process of learning CGCNN for predicting material properties and handling file processing, we developed a high-throughput framework encapsulated in a user-friendly application. The interface allows researchers to input single or multiple CIF files simultaneously, generating instantaneous κ_L predictions for the specified compounds. The efficiency of this framework arises from its

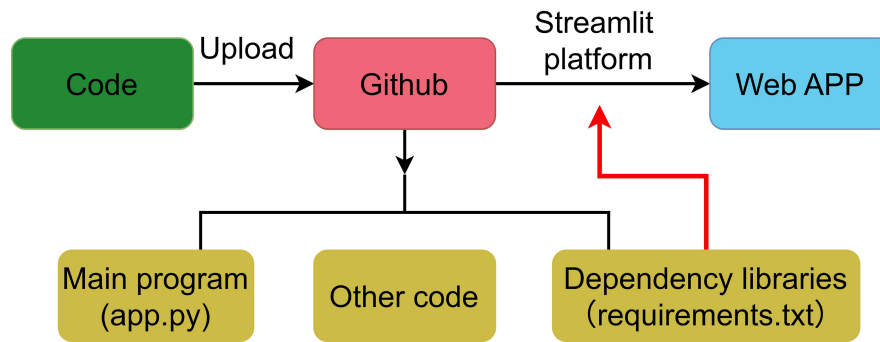


Figure 2. PINK code deployment process. To deploy and run the web application, one first uploads the code - along with the “app.py”, “requirements.txt”, and any other necessary files to GitHub. Then, we use the Streamlit platform to deploy the application online. PINK: Physical-informed kappa.

Lattice Thermal Conductivity APP

9 files uploaded successfully!

Physics-informed machine learning for lattice thermal conductivity (PINK)

	Number of atoms	Density (g/cm ³)	Volume (Å ³)	(the total atomic mass (amu))	Bulk modulus (GPa)	Shear modulus (GPa)	Sound velocity of the transverse wave (m/s)	Sound velocity of the longitudinal wave (m/s)	Speed of sound (m/s)	Poisson ratio	Grüneisen parameter	Acoustic Debye Temperature (K)	Kappa_Lat (W m ⁻¹ K ⁻¹)
mp-64	2	9.1572	55.7632	200.232	45.2534	22.7495	1.5753221	2.8722206	3.75627	0.2849	1.6527	141.2182	4.5828
mp-58	1	0.8209	78.5206	39.0983	1.8549	1.4672	1.3320908	2.2513622	3.4693399	0.231	1.2519	102.2464	0.8739
mp-67	2	2.9966	59.0245	89.9128	49.4707	26.314	3.090099	5.491363	5.423357	0.2997	1.5476	276.61	12.6848
mp-55	1	7.2065	27.3504	118.71	52.5951	19.6362	1.6245907	3.2990596	3.8241452	0.3386	1.0396	190.2464	4.5318
mp-62	1	18.8664	34.4054	228.056	119.4867	24.9357	1.2422969	3.1121017	3.4673219	0.463	1.0679	144.3365	2.3412
mp-64	2	9.1573	21.2059	117.8664	170.2307	108.656	4.1485648	4.2047162	4.226782	0.2214	1.2596	446.166	61.2885
mp-60	2	3.5341	13.2966	24.0214	394.4489	350.3993	13.4761915	12.8447096	13.5523018	0.104	0.7566	1,794.1444	1,256.5584
mp-53	1	3.5983	14.3465	30.9738	81.8358	61.6393	4.2465394	6.7441021	6.5774822	0.189	1.2818	505.8911	61.4466
mp-56	2	3.3993	135.3618	279.434	8.8932	6.4219	1.2611493	2.4441608	3.7343608	0.1401	1.0026	105.6007	4.2338

The file name of displaying crystalline is: mp-64

The information of uploaded crystal structure is:

Formula: CeO₂

Space group: *Cmcm* (63)

cell_length_a 3.46499582

cell_length_b 3.46499582

cell_length_c 5.22098210

cell_angle_alpha 90.00000000

cell_angle_beta 90.00000000

cell_angle_gamma 125.92428753

The volume of the crystal structure is 55.76117316494666 Å³.

The bulk modulus of the crystal structure is 45.2534977210391 GPa.

The Shear modulus of the crystal structure is 22.74950516087618 GPa.

The sound velocity of the crystal structure is 1.575240615702 m/s.

The Grüneisen parameter of the crystal structure is 1.6527220801381482.

The formula of the lattice thermal conductivity is $\kappa_L = \frac{C_V}{3} \cdot v \cdot \theta_D^{-3}$.

The calculated lattice thermal conductivity is 4.582765509028179 W/(m·K).

We strive to have clear documentation and examples to help everyone with using PINK on their own. We will happily fix issues in the documentation and examples should you find any, however, we will not be able to offer extensive user support and training, except for our collaborators.

Figure 3. The web page for our PINK app is divided into two panels. The left panel allows users to upload files, while the right panel displays the results. The output includes a DataFrame that lists various properties such as the number of atoms, density (g/cm³), volume (Å³), atomic mass (amu), bulk modulus (GPa), shear modulus (GPa), transverse and longitudinal wave sound velocities (m/s), speed of sound (m/s), Poisson’s ratio (ν), Grüneisen parameter (γ), acoustic Debye temperature (θ_D , K), and lattice thermal conductivity (W·m⁻¹·K⁻¹). For detailed instructions on using PINK, please refer to the [PINK_tutorial.mp4](#). Additionally, the app supports custom functions for calculating bulk modulus (GPa), shear modulus (GPa), and Grüneisen parameter, with a separate tutorial available in [PINK_Custom_Parameters_tutorial.mp4](#). PINK: Physical-informed kappa.

integration of pre-trained CGCNN models with Equation (2), providing rapid assessments of thermal transport properties.

Given that thermoelectric performance is strongly influenced by materials with low κ_L , we conducted a high-throughput screening across material dataset. This systematic evaluation successfully identified 11,869 potential candidates with promising thermal transport characteristics. To validate our screening approach, we selected Ag₃Te₄X (X = W, Ta) from a specific ternary system and performed detailed *ab initio*

calculations to verify their properties.

Data collection

For evaluating ML performance in materials science applications, we utilized the Matbench V0.1^[52]. Our analysis focused on two specific datasets within this collection that address elastic properties: “matbench_log_gvrh” and “matbench_log_kvrh”. These elastic modulus datasets contain identical material entries (10,987 in total) and are specifically designed to predict the logarithmic values of shear modulus (G) and bulk modulus (B) using the Voigt-Reuss-Hill (VRH) averaging methodology. The comprehensive nature of these standardized collections makes them ideal for training our ML models to predict key elastic characteristics. A detailed analysis of the dataset is presented in Figure 4, which illustrates the statistical distribution across three key aspects: the classification of crystal systems, the number of atoms in the primitive cell, and the distribution of chemical elements. The dataset exhibits remarkable diversity, incorporating materials from all seven fundamental crystal systems and 84 different elements in various structural arrangements.

For predicting κ_L , we utilized datasets obtained from the AFLOW database^[53] and relevant publications^[20,54-57], which include both experimentally measured values and computationally derived properties. The AFLOW database provides comprehensive information on crystal structures and thermal characteristics, with κ_L values calculated using the methodologies outlined in^[29,58]. To construct a test dataset for evaluating our application, we collected crystal structures, Grüneisen parameters (γ), and their corresponding κ_L values from AFLOW and other literature.

Obtaining high-throughput datasets for computational materials science can be challenging. However, recent advancements in ML have significantly enhanced the discovery of stable materials. Merchant *et al.* employed deep learning and GNNs to scale materials discovery, particularly for inorganic crystals^[38]. Their work expanded the known set of stable materials by adding 381,000 new entries to the convex hull, resulting in a total of 377,221 stable crystal structures - a tenfold increase over previous datasets.

We accessed their extensive dataset through GitHub: https://github.com/google-deepmind/materials_discovery. The repository includes 377,221 valid CIF files in the “by_composition” folder, compatible with CGCNN, and a summary CSV file containing bandgap, crystal symmetry, and decomposition energy data. These materials encompass compositions ranging from two to six elements, with atomic numbers spanning from 2 to 106. Additional data from the Materials Project further complements these datasets, enabling the targeted retrieval of material properties through its open-source API. Together, these resources provide a robust foundation for high-throughput computations and analyses.

Model evaluation of CGCNN

The interpretable formula, detailed in the methods section, elucidates the correlation between elastic modulus and κ_L . In ML, MAE and R^2 (R-squared) are standard metrics for evaluating regression models. MAE quantifies the average magnitude of prediction errors and is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (7)$$

where y_i denotes the actual values and \hat{y}_i the predicted values. Lower MAE values denote higher prediction accuracy. R^2 assesses the proportion of variance in the dependent variable explained by:

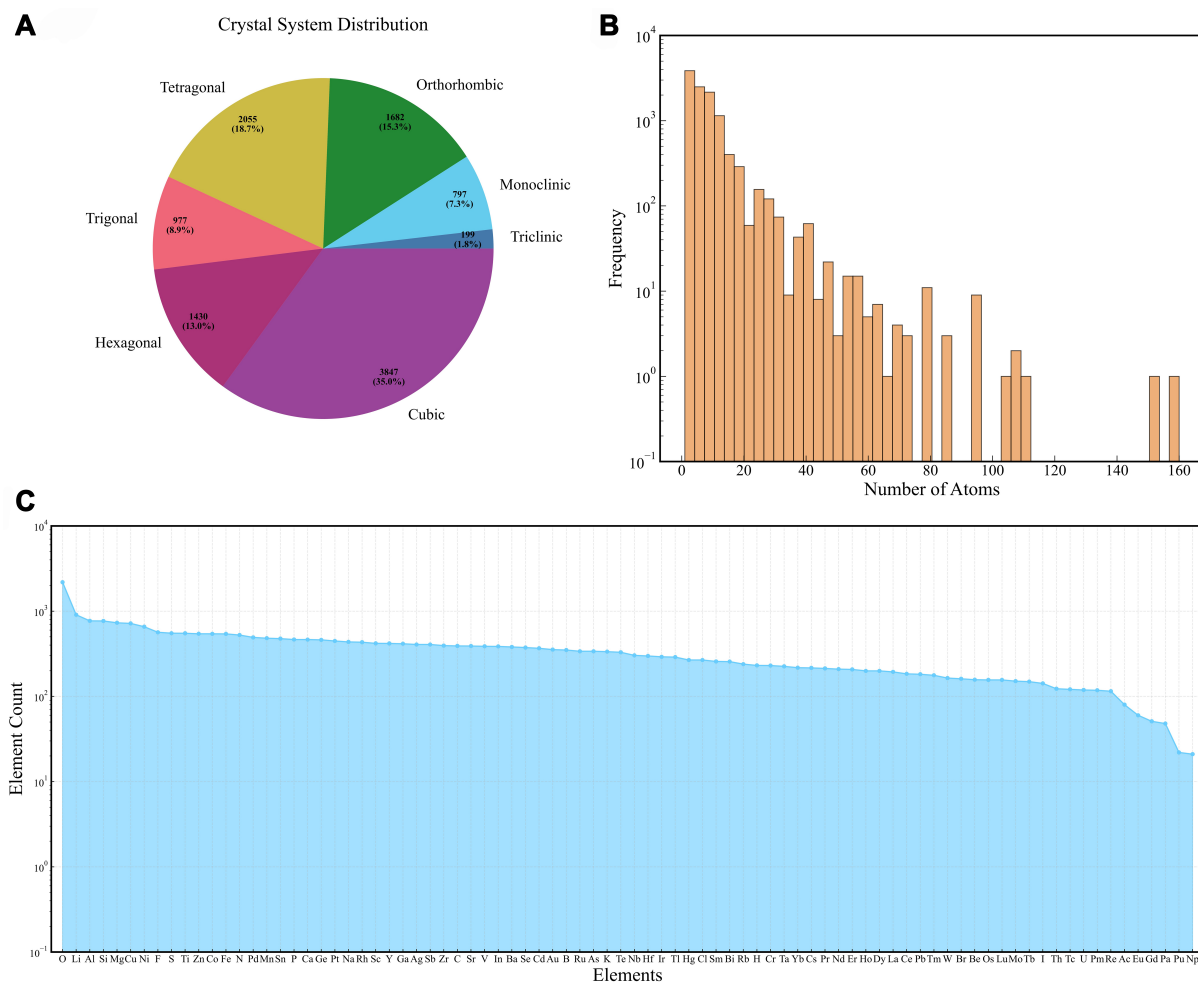


Figure 4. Statistical analysis of the training dataset. (A) The distribution of seven crystal systems, with cubic being the most common (3,847 structures), followed by tetragonal (2,055 structures), while triclinic is the least one (199 structures); (B) Distribution of range of number of atoms in the primitive cell (1-160 atoms) across the dataset; (C) Elemental distribution that illustrates the frequency of 84 distinct elements. The dataset encompasses transition metals, main group elements, and rare earth elements, with oxygen showing the highest frequency.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

where \bar{y} represents the mean of the actual values. R^2 values approaching 1 indicate a better model fit.

The performance of the embedded CGCNN model on the test dataset is illustrated in Figure 5. The MAE for both the shear and bulk moduli is below 13, with R^2 values approaching 1, indicating a strong correlation between the predicted and DFT-calculated elastic modulus. These results demonstrate the model's reliability and predictive accuracy.

Model evaluation of κ_L

After predicting the shear and bulk moduli using the trained CGCNN model, the approximate average speed of sound was estimated. Utilizing known crystal structure information, we applied Equation (2) to approximate the material's κ_L . To validate the PINK application, we compared its predictions with κ_L values calculated via DFT for 2,535 materials from the AFLOW database^[53] and 46 experimentally measured values

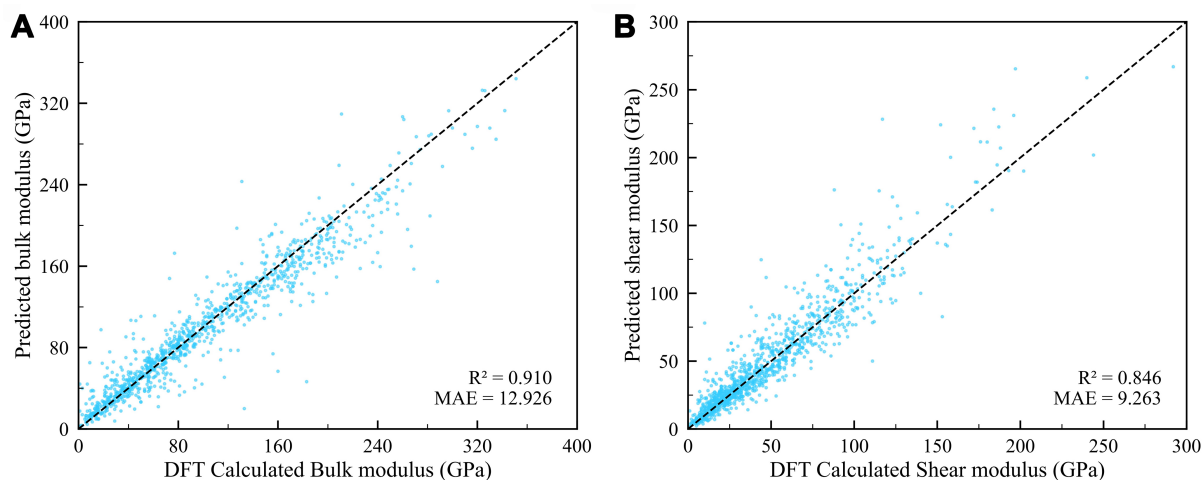


Figure 5. The comparison of predicted vs. DFT-calculated values for (A) bulk modulus and (B) shear modulus across the test dataset of 10,987 structures. DFT: Density functional theory.

from the literature^[20,54-57]. The Grüneisen parameters were obtained from the AFLOW database and experimental data, respectively.

Figure 6 presents scatter plots comparing κ_L predictions by PINK with calculated and experimentally measured values. In **Figure 6A**, each point represents a material, with the solid diagonal line indicating perfect agreement between predicted and calculated values. The dashed lines denote an acceptable range of deviation. Within the dataset, 2,415 points (95.27%) fall within this range, highlighting the model's high accuracy. The clustering of points near the diagonal line further confirms a strong correlation between PINK predictions and DFT calculations. Deviations are likely attributable to the inapplicability of certain materials to the Slack model or inaccuracies in the elastic modulus predictions^[43].

In **Figure 6B**, κ_L predictions from PINK are compared with experimental values, with each point labeled by the corresponding material. Similar to **Figure 6A**, the solid diagonal line represents perfect agreement, and the dashed lines denote acceptable deviation boundaries. The model achieves a MAE of 0.526 and an R^2 value of 0.881, indicating a close correspondence between PINK predictions and experimental results. The clustering of data points near the diagonal line demonstrates that PINK effectively predicts κ_L across diverse materials and crystal symmetries.

For additional validation, **Table 1** presents the predicted κ_L values alongside their experimental counterparts for 46 materials. These results further substantiate the reliability and effectiveness of PINK in predicting κ_L , showing close alignment with both DFT-calculated and experimentally measured values. Notably, the accuracy of κ_L predictions could be significantly enhanced with more precise Grüneisen parameters and improved predictions of shear and bulk moduli^[29].

Comparison of calculation time for κ_L

To demonstrate κ_L the efficiency of prediction application, PINK, we compared its computational time against other commonly used methods. Traditional approaches, such as solving the PBTE with second- and third-order force constants^[60] or employing the equilibrium MD Green-Kubo, typically require several hours for simple systems and days to weeks for complex ones.

Table 1. Predicted and experimental room-temperature κ_L values for compounds from the literature^[20,54-57] are presented

Materials	ID-number	<i>n</i>	ρ (g·cm ³)	<i>V</i> (Å ³)	<i>G</i> (GPa)	<i>v_s</i> (m·s ⁻¹)	γ	κ_{exp} (W·m ⁻¹ ·K ⁻¹)	κ_{PINK} (W·m ⁻¹ ·K ⁻¹)
AgCl ^[54]	mp-22922	2	5.583	42.631	8.801	1,423.597	1.900	1	1.091
AlAs ^[20]	mp-2172	2	3.591	47.126	40.510	3,733.087	0.660	98	47.054
AlSb ^[20]	mp-2624	2	4.078	60.561	28.808	2,959.047	0.600	56	30.620
BN ^[20]	mp-1639	2	3.458	11.919	350.203	10,995.529	0.700	760	727.981
BP ^[20]	mp-1479	2	2.953	23.500	201.292	9,011.073	0.923 [by Equation (6)]	350	344.207
C ^[20]	mp-66	2	3.496	11.410	547.436	13,613.398	0.750	3,000	1,320.871
CaO ^[20]	mp-2605	2	3.287	28.332	63.316	4,863.622	1.570	27	32.552
CdTe ^[20]	mp-406	2	5.473	72.827	14.349	1,818.434	0.520	7.5	10.797
GaAs ^[20]	mp-2534	2	5.053	47.532	45.152	3,291.363	0.750	45	42.381
GaP ^[20]	mp-3490	2	4.006	41.737	48.296	3,846.005	0.750	100	50.725
GaSb ^[20]	mp-1156	2	5.288	60.133	28.033	2,557.748	0.750	40	22.115
Ge ^[20]	mp-32	2	5.042	47.847	47.134	3,362.109	1.060	65	33.219
InAs ^[20]	mp-20305	2	5.336	59.050	23.662	2,355.122	0.570	30	20.454
InP ^[20]	mp-20351	2	4.582	52.840	27.554	2,742.762	0.600	93	25.940
InSb ^[20]	mp-20012	2	5.384	72.965	17.668	2,026.668	0.560	20	14.244
KBr ^[20]	mp-23251	2	2.624	75.294	6.156	1,709.446	1.450	3.4	1.737
KBr ^[20]	mp-570891	2	2.989	66.111	8.391	1,885.943	1.450	3.4	2.502
KCl ^[20]	mp-23193	2	1.904	65.033	6.387	2,050.803	1.450	7.1	2.059
KI ^[20]	mp-22898	2	2.972	92.743	5.790	1,546.914	1.450	2.6	1.585
LiF ^[20]	mp-1009009	2	2.569	16.768	58.174	5,236.933	1.500	17.6	28.999
LiH ^[20]	mp-23703	2	0.825	16.002	39.346	7,537.234	1.280	15	34.630
MgO ^[20]	mp-1265	2	3.471	19.279	123.811	6,564.689	1.440	60	86.060
NaBr ^[20]	mp-22916	2	3.121	54.749	14.495	2,392.521	1.500	2.8	4.897
NaCl ^[20]	mp-22862	2	2.105	46.096	16.650	3,123.706	1.560	7.1	6.531
NaF ^[20]	mp-682	2	2.693	25.894	21.929	3,171.392	1.500	18.4	7.651
NaI ^[20]	mp-23268	2	3.572	69.675	6.823	1,547.000	1.560	1.8	1.521
PbS ^[20]	mp-21276	2	7.334	54.174	26.485	2,111.102	2.000	2.9	4.772
PbSe ^[20]	mp-2201	2	7.886	60.254	22.618	1,876.743	1.500	2	6.189
RbBr ^[20]	mp-22867	2	3.164	86.781	6.905	1,651.509	1.450	3.8	1.974
RbCl ^[20]	mp-23295	2	2.672	75.148	7.337	1,854.222	1.450	2.8	2.244
RbI ^[20]	mp-22903	2	3.360	104.957	6.228	1,517.862	1.410	2.3	1.814
Si ^[20]	mp-149	2	2.281	40.888	55.828	5,480.852	1.060	166	60.869
SiC ^[20]	mp-8062	2	3.227	20.635	222.879	9107.212	0.750	490	438.312
SrO ^[20]	mp-2472	2	4.878	35.277	47.864	3,468.152	1.520	12	19.845
TePb ^[20]	mp-19717	2	7.857	70.758	17.511	1,674.996	2.009 [by Equation (6)]	2.5	2.711
ZnS ^[20]	mp-10695	2	3.999	40.476	32.768	3,191.504	0.750	27	28.268
ZnSe ^[20]	mp-1190	2	5.064	47.338	31.048	2,763.063	0.750	19	24.432
ZnTe ^[20]	mp-2176	2	5.419	59.146	25.383	2,416.126	0.970	18	15.097
AlN ^[20]	mp-661	4	3.201	42.527	135.549	7,197.895	0.700	350	140.931
BeO ^[20]	mp-2542	4	2.967	27.992	123.309	7,116.682	0.750	370	104.885
CdS ^[20]	mp-672	4	4.576	104.863	16.885	2,166.455	0.750	16	6.790
GaN ^[20]	mp-804	4	5.924	46.943	110.843	4,794.507	0.700	210	79.333
ZnO ^[20]	mp-2133	4	5.438	49.719	33.376	2,790.196	0.750	60	13.479
Bi ₂ Te ₃ ^[55]	mp-34202	5	7.315	181.767	10.489	1,339.685	1.490	1.6	1.196
Al ₂ O ₃ ^[56]	mp-1143	10	3.873	87.420	132.798	6,501.426	1.340	30	33.445
ZnSb ^[57]	mp-753	16	6.347	391.759	32.387	2,518.332	1.681 [by Equation (6)]	3.5	2.315

The CIFs for these materials were obtained from the Materials Project. Here, *n*, ρ , *V* and *G* are the number of atoms in the primitive cell, the density, the volume of the primitive cell and the shear modulus predicted by CGCNN, respectively. *v_s* represents the average speed of sound

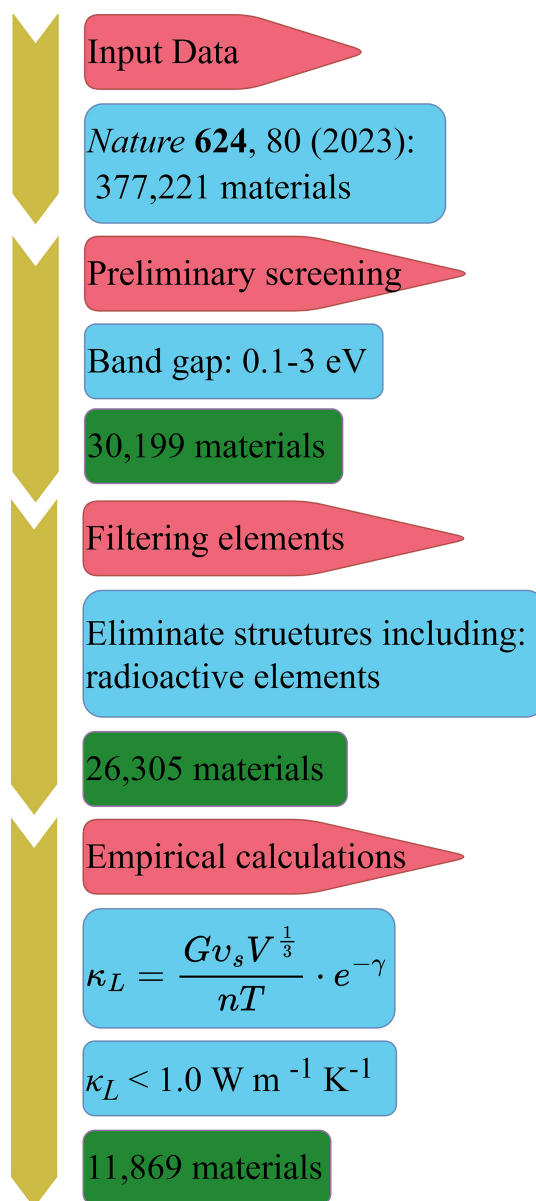


Figure 7. Flowchart of the high-throughput screening process, illustrating steps from data acquisition to filtering and empirical calculations for κ_L prediction.

Subsequently, the CGCNN model was utilized to predict shear and bulk moduli, which were then employed to estimate κ_L using the Slack model at 300 K. Materials with κ_L values below $1 \text{ W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$ were identified as promising candidates for thermoelectric applications. This filtering yielded 11,869 materials, documented in [Nature-filtered-low-Kappa.csv](#).

Additionally, using the Materials Project API, 54,359 structures with band gaps between 0.1-3.0 eV and no radioactive elements were extracted. PINK was employed to predict κ_L , resulting in a refined dataset of 21,001 low κ_L materials, detailed in [MP-semiconductor-low-kappa.csv](#).

Statistical analysis of screening results

The process of screening 11,869 materials with low κ_L values ($\kappa_L \leq 1 \text{ W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$) has been detailed in a separate CSV file (Nature-filtered-low-kappa.csv), which also includes the associated material data. Statistical results for materials that have passed this screening are shown in Figure 8. The histogram in Figure 8A shows the distribution of κ_L values, with the majority between 0.1 and 0.5, highlighting a promising subset of high-performance thermoelectric materials. Meanwhile, Figure 8B represents the distribution of crystal structures among these screened materials, specifying the number of space groups for cubic systems. The analysis reveals an inverse trend between symmetry and material count - fewer materials are found with higher symmetry, as in cubic systems (262 materials), while lower symmetry, such as in triclinic systems, is associated with a larger count. For cubic structures, the relevant space group numbers include 198, 205, 214, 215, 216, 217, 225, 227, 229, and 230. Low κ_L values are widely acknowledged as critical for improving the efficiency of thermoelectric materials in converting waste heat to electrical energy.

To analyze the compositional distribution of promising thermoelectric materials, a histogram was generated to display the elemental distribution within the screened materials, as shown in Figure 8C. This diagram underscores the importance of the 20 most common elements in compounds with κ_L values less than 1. Cesium bromine, rubidium, and adenosine oxide emerge as the elements most frequently encountered. Additionally, elements such as oxygen (O) and selenium (Se) are also prevalent in materials with low κ_L values. The corresponding electronegativity values of these elements are provided at the top of each column in Figure 7C. A thorough examination of the electronegativity data reveals that elements with higher electronegativity are more likely to form stronger ionic bonds. Among the ten elements that occur most frequently, the majority exhibit electronegativity values greater than 2.5. Interestingly, elements often associated with low thermal conductivity, such as cesium and selenium, are part of this group. Moreover, fluorine, characterized by its high electronegativity, readily forms ionic compounds with alkali metals, including cesium, rubidium, potassium, and sodium.

First-principle validation

On the basis of the results from our high-throughput screening and prior experience, we observed that compounds containing heavy elements and Group VIA elements generally exhibit lower thermal conductivity. Given the structural feasibility and computational efficiency, we selected the cubic structure for our study. Consequently, $\text{Ag}_3\text{Te}_4\text{W}$ and $\text{Ag}_3\text{Te}_4\text{Ta}$ were chosen as validation targets. To calculate κ_L for a given material with a specific structure, a series of DFT calculations are performed within the volume of the primitive cell. To validate the materials screened by the PINK, including those with low κ_L , we have selected $\text{Ag}_3\text{Te}_4\text{W}$ and $\text{Ag}_3\text{Te}_4\text{Ta}$ as case studies for detailed analysis. Both crystals belong to space group 215. As illustrated in Figure 9A, X (W, Ta) atoms are tetrahedrally coordinated by four Te atoms, while Ag atoms occupy interstitial sites between neighboring tetrahedra. Phonon spectra calculations for these materials [Figure 9B and C] reveal no imaginary frequencies, confirming their dynamic stability and theoretical viability.

The κ_L values of these compounds were calculated using the 3-phonon (3ph) method. Notably, $\text{Ag}_3\text{Te}_4\text{X}$ (X = W, Ta) exhibits ultralow κ_L , comparable to benchmark thermoelectric materials such as PbQ (Q = Te, Se) and SnSe^[46,63,64]. Figure 10A compares the temperature-dependent κ_L of $\text{Ag}_3\text{Te}_4\text{X}$ with state-of-the-art systems including SnSe, Tl_9SbTe_6 ^[62], and PbQ. At 300 K, $\text{Ag}_3\text{Te}_4\text{W}$ and $\text{Ag}_3\text{Te}_4\text{Ta}$ demonstrate κ_L values of 0.267 and 0.478 $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$, respectively. By comparison, PbTe, PbSe, SnSe, and Tl_9SbTe_6 exhibit κ_L values of 2.3, 2.64, 0.62, and 0.143 $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$ at the same temperature, respectively [Table 2]. The exceptionally low κ_L of $\text{Ag}_3\text{Te}_4\text{X}$ positions these materials as promising candidates for thermoelectric applications.

Table 2. The predicted κ_L values from PINK were compared with those computed using DFT at 300 K

Materials	n	V (\AA^3)	G (GPa)	B (GPa)	v_s ($\text{m}\cdot\text{s}^{-1}$)	γ	κ_{PINK} ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	κ_{DFT} ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)
PbTe ^[46]	2	69.9891	24	38	1,927.318	2.180 ^[63]	3.5914	2.300 (Experiment)
PbSe ^[46]	2	59.0544	27	24	2,035.793	2.660	2.493	2.640 (Experiment)
SnSe ^[64]	8	226.257	15	24	1,783.133	2.300	0.681	0.680 (Experiment)
Ag ₃ Te ₄ Ta	8	250.951	13.776	45.133	3,074.950	2.231	0.628	0.478 (DFT)
Ag ₃ Te ₄ W	8	248.111	12.406	42.316	2,939.400	2.276	0.507	0.267 (DFT)

In this comparison, n stands for the number of atoms in the primitive cell, V is the volume of the primitive cell, and G and B are the shear modulus and bulk modulus from Materials Project database for PbTe, PbSe and SnSe or our CGCNN model for Ag₃Te₄Ta and Ag₃Te₄W. Additionally, v_s refers to the average speed of sound, as defined in Equation (5), while γ signifies the Grüneisen parameter, as given in Equation (6). PINK: Physical-informed kappa; DFT: density functional theory; CGCNN: crystal graph convolutional neural network.

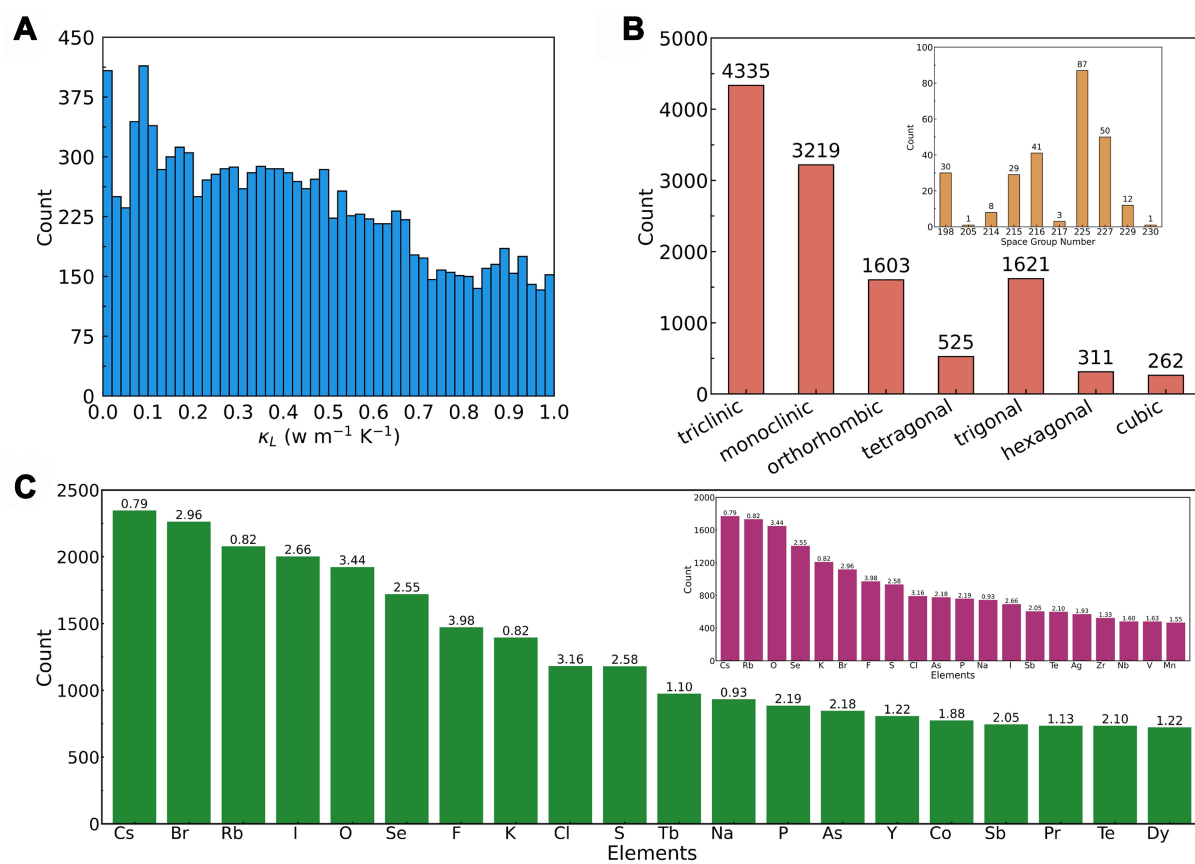


Figure 8. Statistical results of 11,869 screened candidates. (A) Distribution of κ_L and corresponding count; (B) Distribution of crystal symmetry, with space group details for cubic symmetry shown in the inset; (C) Histogram of elemental distribution in 11,869 compounds, with electronegativity values indicated at the top of each column. The electronegativity results are as follows: Cs: 0.79, Br: 2.96, Rb: 0.82, I: 2.66, O: 3.44, Se: 2.55, F: 3.98, K: 0.82, Cl: 3.16, S: 2.58, Tb: 1.10, Na: 0.93, P: 2.19, As: 2.18, Y: 1.22, Co: 1.88, Sb: 2.05, Pr: 1.13, Te: 2.1, Dy: 1.22. The inset in the top-right corner is the counting number excluding lanthanide-containing materials.

Acoustic phonons typically serve as the dominant contributors to thermal transport in materials. As illustrated in Figures 9B and C and Figure 10B, acoustic phonon branches predominantly occupy low-frequency regimes, with low-frequency acoustic modes dominating the contribution to κ_L . To unravel the microscopic mechanisms underlying the ultralow κ_L , we systematically examined key parameters governing thermal conductivity - including heat capacity, phonon group velocities, weighted phase space, and scattering rates - for Ag₃Te₄X (X = W, Ta) and Tl₅SbTe₆^[62]. These analyses, presented in Figure 10B-F,

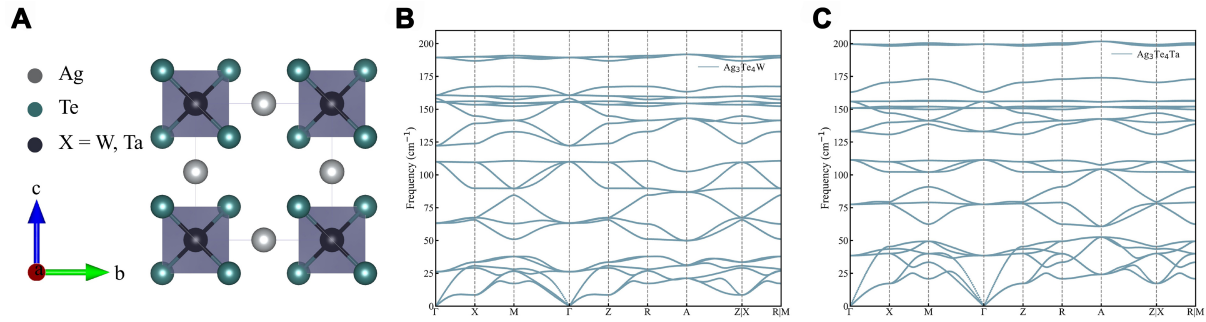


Figure 9. (A) The primitive crystal structures of $\text{Ag}_3\text{Te}_4\text{X}$ ($\text{X} = \text{W}, \text{Ta}$). Phonon dispersions for (B) $\text{Ag}_3\text{Te}_4\text{W}$ and (C) $\text{Ag}_4\text{Se}_4\text{Ta}$, respectively, along the high-symmetry points which are defined as Γ (0, 0, 0), X (0, 0.5, 0), M (0.5, 0.5, 0.0), Z (0, 0, 0.5), R (0, 0.5, 0.5), A (0.5, 0.5, 0.5).

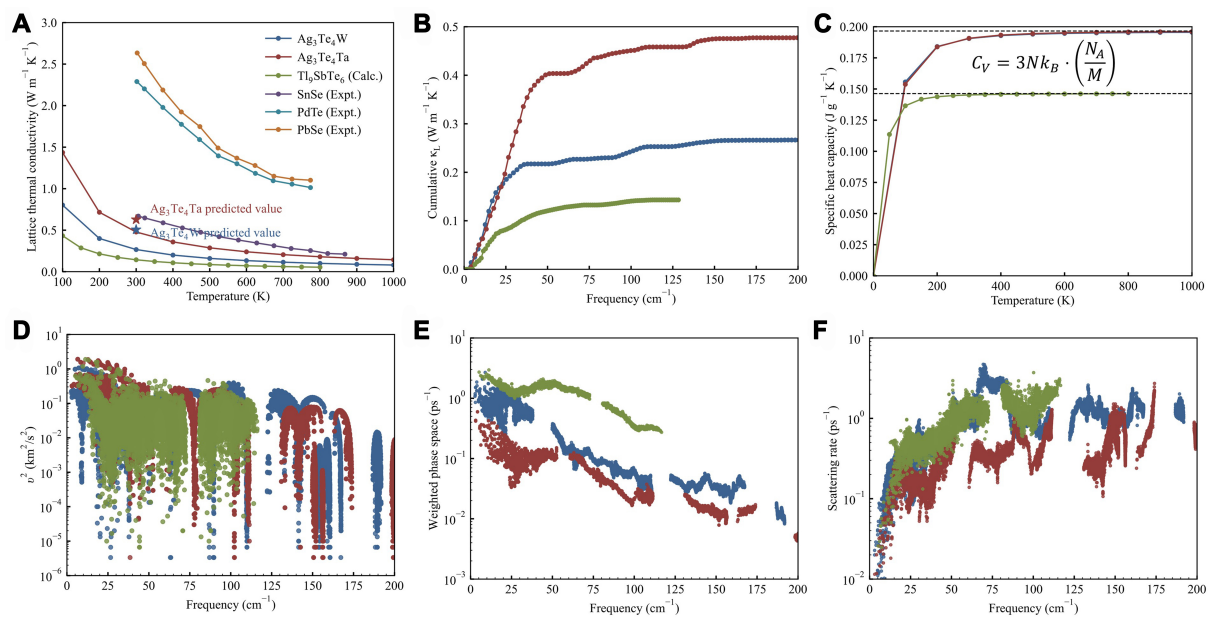


Figure 10. (A) κ_L as a function of temperature for $\text{Ag}_3\text{Te}_4\text{X}$ ($\text{X} = \text{W}, \text{Ta}$), Tl_9SbTe_6 ^[62], SnSe ^[46], and PbQ ($\text{Q} = \text{Te}, \text{Se}$)^[46]. Comparison of microscopy heat transport parameters for $\text{Ag}_3\text{Te}_4\text{X}$ and Tl_9SbTe_6 ^[62] at 300 K; (B) Cumulative κ_L using 3ph methods; (C) Specific heat capacity (C_V) at constant volume; (D) Squared phonon group velocities (v^2) in the harmonic approximation; (E) Weighted phonon scattering phase space of 3ph; (F) Phonon scattering rates of 3ph.

provide critical insights into the interplay of phonon dynamics and thermal transport.

The specific heat capacity (C_V) of solids at elevated temperatures approximates the Dulong–Petit limit, defined as $3Nk_B$ (N_A/M), where N denotes the number of atoms per formula unit, k_B is the Boltzmann constant, N_A the Avogadro constant, and M the molar mass. DFT calculations yield C_V values of 0.196, 0.197, and 0.146 $\text{J}\cdot\text{g}^{-1}\cdot\text{K}^{-1}$ for $\text{Ag}_3\text{Te}_4\text{W}$, $\text{Ag}_3\text{Te}_4\text{Ta}$, and Tl_9SbTe_6 , respectively, closely aligning with Dulong–Petit predictions. Notably, the C_V values for $\text{Ag}_3\text{Te}_4\text{X}$ ($\text{X} = \text{W}, \text{Ta}$) exceed those of Tl_9SbTe_6 and marginally surpass the 0.156 $\text{J}\cdot\text{g}^{-1}\cdot\text{K}^{-1}$ reported for PbTe ^[65].

Given the proportionality $\kappa_L \propto v^2$, we analyzed the frequency-dependent squared phonon group velocities (v^2) of $\text{Ag}_3\text{Te}_4\text{X}$ ($\text{X} = \text{W}, \text{Ta}$) and Tl_9SbTe_6 , as illustrated in **Figure 10D**. Within the frequency range dominant for κ_L , $\text{Ag}_3\text{Te}_4\text{Ta}$ displays the highest v^2 (2 $\text{km}^2\cdot\text{s}^{-2}$), an order of magnitude lower than PbTe 's

reported v^2 of $14 \text{ km}^2 \cdot \text{s}^{-2}$ ($\kappa_L = 2 \text{ W} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$ at 300 K)^[66]. $\text{Ag}_3\text{Te}_4\text{W}$ exhibits intermediate values, while Tl_9SbTe_6 shows the lowest v^2 .

To elucidate phonon scattering mechanisms, we calculated the weighted phonon scattering phase space (W_3), which quantifies available phonon-phonon interaction pathways. As shown in [Figure 10E](#), $\text{Ag}_3\text{Te}_4\text{Ta}$ has the smallest W_3 , contrasting sharply with the largest W_3 observed in Tl_9SbTe_6 . Similarly, phonon scattering rates [[Figure 10F](#)] are significantly higher for Tl_9SbTe_6 than for $\text{Ag}_3\text{Te}_4\text{Ta}$. These results collectively underpin the ultralow κ_L of $\text{Ag}_3\text{Te}_4\text{X}$ ($X = \text{W}, \text{Ta}$).

Discussions

In this work, we present PINK, a high-throughput computational framework designed to enhance the prediction of κ_L across diverse materials. Building on this platform, several strategic directions emerge for future refinement and application. First, integrating PINK with experimental databases and materials informatics platforms could accelerate the discovery of novel materials for thermoelectrics, thermal management in microelectronics, and energy conversion systems. Coupling high-throughput predictions^[67] with experimental validation would enable rapid identification of high-performance materials, narrowing the gap between computational insights and functional material synthesis. Additionally, synergizing PINK with tools such as BoltzTraP^[68] and TransOpt^[69] could enable concurrent optimization of thermal and electrical transport properties in semiconductors.

A second critical opportunity lies in the precise engineering of multifunctional materials. By investigating the interaction between κ_L , mechanical properties (e.g., strength, elasticity), and environmental stability, researchers could design materials that simultaneously achieve thermal, mechanical, and operational demands in sectors such as aerospace, renewable energy, and advanced electronics. Such integration of properties would advance applications requiring both efficient heat regulation and structural resilience.

CONCLUSIONS

We have developed a high-throughput framework, packaged as an application named PINK, designed to rapidly predict the κ_L of materials based on the CIF files. The material space for κ_L was expanded significantly, increasing by an order of magnitude, through predictions for 377,221 newly reported materials^[38]. Through high-throughput screening, several materials with ultralow κ_L were identified, and their predictions were validated using first-principles calculations.

Although first-principle calculations of κ_L require significant computational resources, especially for phonon spectra and third-order force constant matrices, there are existing databases related to phonons and κ_L . For instance, Togo developed an automated workflow interfaced with Phonopy, which calculated phonon spectra, density of states, entropy, and heat capacity for over 11,000 materials, creating a phonon database available at <https://github.com/atztogo/phonondb/blob/main/mdr/phonondb/README.md>. AFLOW^[53], a comprehensive database, includes thermal property data for 5,664 materials, though this represents only a small fraction of the total material space. PINK addresses this gap by extending κ_L predictions to hundreds of thousands of materials, with accuracy contingent on the performance of its embedded CGCNN for elastic modulus prediction.

To enhance prediction accuracy, future advancements could integrate advanced crystal graph convolutional networks. Examples include the orbital graph convolutional neural network (OGCNN), which considers orbital roles^[70]; the materials graph network (MEGNet), incorporating outfield information^[71]; the geometric-information-enhanced crystal graph neural network (GeoCGNN), which integrates topological

and geometric structure data^[72]; and the atomistic line graph neural network (ALIGNN), which includes bond angle details^[73]. Other notable approaches include the graph-attention graph neural network (GATGNN), utilizing attention mechanisms^[74], and the scalable global graph attention neural network model DeeperGATGNN, featuring differentiable group normalization (DGN) and skip connections^[75]. Moreover, powerful descriptors such as SOAP^[76] and Voronoi tessellations^[77] could be employed to further elucidate the link between crystal structures and material properties. Ruff *et al.* introduced a connection-optimized crystal graph network (coGN/coNGN), leveraging message passing and line graph templates^[78]. Their model demonstrated exceptional performance on the MatBench benchmark dataset^[52], outperforming other models and establishing itself as the leading general-purpose model in the benchmark.

DECLARATIONS

Authors' contributions

Conceptualization, methodology, software, data curation, visualization, writing-original draft preparation: Liu, Y.

Writing-review and editing, supervision, project administration, funding acquisition: Gao, Z.

Performed data analysis and interpretation: Wang, X.; Hao, Y.; Li, X.

Investigation, discussion: Ding, X.; Lookman, T.; Sun, J.

Availability of data and materials

All data are available at https://github.com/JackLiu0227/AI4Kappa/tree/master/JMI_Supporting_Information.

Financial support and sponsorship

We acknowledge the support from the National Natural Science Foundation of China (No.12104356 and No.52250191). This work is sponsored by the Key Research and Development Program of the Ministry of Science and Technology (No.2023YFB4604100). We also acknowledge the support of the HPC Platform, Xi'an Jiaotong University.

Conflicts of interest

Ding, X. is an Associate Editor of *Journal of Materials Informatics*. He was not involved in any steps of the editorial process, including reviewer selection, manuscript handling, or decision-making. The other authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Li, N.; Ren, J.; Wang, L.; Zhang, G.; Hänggi, P.; Li, B. *Colloquium: Phononics: manipulating heat flow with electronic analogs and beyond. Rev. Mod. Phys.* **2012**, *84*, 1045. [DOI](#)
2. van Erp R, Soleimanzadeh R, Nela L, Kampitsis G, Matioli E. Co-designing electronics with microfluidics for more sustainable cooling. *Nature* **2020**, *585*, 211-6. [DOI](#) [PubMed](#)
3. Kim, H. S.; Liu, W.; Chen, G.; Chu, C. W.; Ren, Z. Relationship between thermoelectric figure of merit and energy conversion efficiency. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 8205-10. [DOI](#) [PubMed](#) [PMC](#)

4. Ouyang, H.; Gu, Y.; Gao, Z.; et al. Kirigami-inspired thermal regulator. *Phys. Rev. Appl.* **2023**, *19*, L011001. DOI
5. Kang, J. S.; Li, M.; Wu, H.; Nguyen, H.; Aoki, T.; Hu, Y. Integration of boron arsenide cooling substrates into gallium nitride devices. *Nat. Electron.* **2021**, *4*, 416-23. DOI
6. Xu, L.; Wang, X.; Wang, Y.; Gao, Z.; Ding, X.; Xiao, Y. Enhanced average power factor and *ZT* value in PbSe thermoelectric material with dual interstitial doping. *Energy. Environ. Sci.* **2024**, *17*, 2018-27. DOI
7. Broido, D. A.; Malorny, M.; Birner, G.; Mingo, N.; Stewart, D. A. Intrinsic lattice thermal conductivity of semiconductors from first principles. *Appl. Phys. Lett.* **2007**, *91*, 231922. DOI
8. Stackhouse, S.; Stixrude, L. Theoretical methods for calculating the lattice thermal conductivity of minerals. *Rev. Mineral. Geochem.* **2010**, *71*, 253-69. DOI
9. Tadano, T.; Tsuneyuki, S. Self-consistent phonon calculations of lattice dynamical properties in cubic SrTiO₃ with first-principles anharmonic force constants. *Phys. Rev. B.* **2015**, *92*, 054301. DOI
10. Carbogno, C.; Ramprasad, R.; Scheffler, M. Ab initio green-kubo approach for the thermal conductivity of solids. *Phys. Rev. Lett.* **2017**, *118*, 175901. DOI PubMed
11. Fan, Z.; Pereira, L. F. C.; Wang, H.; Zheng, J.; Donadio, D.; Harju, A. Force and heat current formulas for many-body potentials in molecular dynamics simulations with applications to thermal conductivity calculations. *Phys. Rev. B.* **2015**, *92*, 094301. DOI
12. Yusuf, A.; Ballikaya, S. Electrical, thermomechanical and cost analyses of a low-cost thermoelectric generator. *Energy* **2022**, *241*, 122934. DOI
13. Luo, Y.; Li, M.; Yuan, H.; Liu, H.; Fang, Y. Predicting lattice thermal conductivity via machine learning: a mini review. *npj. Comput. Mater.* **2023**, *9*, 964. DOI
14. Mishin, Y. Machine-learning interatomic potentials for materials science. *Acta. Mater.* **2021**, *214*, 116980. DOI
15. Liu, C.; Wu, C.; Zhao, Y.; et al. Actively and reversibly controlling thermal conductivity in solid materials. *Phys. Rep.* **2024**, *1058*, 1-32. DOI
16. Shi, X. L.; Wang, L.; Lyu, W.; et al. Advancing flexible thermoelectrics for integrated electronics. *Chem. Soc. Rev.* **2024**, *53*, 9254-305. DOI
17. Callaway, J. Model for lattice thermal conductivity at low temperatures. *Phys. Rev.* **1959**, *113*, 1046-51. DOI
18. Zhang, Y. First-principles Debye–Callaway approach to lattice thermal conductivity. *J. Materiomics.* **2016**, *2*, 237-47. DOI
19. Slack, G. Nonmetallic crystals with high thermal conductivity. *J. Phys. Chem. Solids.* **1973**, *34*, 321-35. DOI
20. Morelli, D. T.; Slack, G. A. High lattice thermal conductivity solids. In: Shindé SL, Goela JS, editors. High thermal conductivity materials. New York: Springer-Verlag; 2006. pp. 37-68. DOI
21. Nielsen, M. D.; Ozolins, V.; Heremans, J. P. Lone pair electrons minimize lattice thermal conductivity. *Energy. Environ. Sci.* **2013**, *6*, 570-8. DOI
22. Chang, C.; Zhao, L. Anharmonicity and low thermal conductivity in thermoelectrics. *Mater. Today. Phys.* **2018**, *4*, 50-7. DOI
23. Virkar, A. V.; Cutler, R. A. Fabrication of high-thermal-conductivity polycrystalline aluminum nitride: thermodynamic and kinetic aspects of oxygen removal. In: Shindé SL, Goela JS, editors. High thermal conductivity materials. New York: Springer-Verlag; 2006. pp. 143-66. DOI
24. Chen, S.; Chen, Y.; Yan, W.; Gao, T. First-principles investigation of elastic anisotropy and thermal transport property of transition metal monosilicides CrSi, TiSi, and ZrSi under pressure. *Mater. Today. Commun.* **2024**, *39*, 108958. DOI
25. Cao, Y.; Dai, S.; Wang, X.; et al. High-throughput screening of potentially ductile and low thermal conductivity ABX₃ (X = S, Se, Te) thermoelectric perovskites. *Appl. Phys. Lett.* **2024**, *124*, 092101. DOI
26. Li, R.; Li, X.; Xi, L.; Yang, J.; Singh, D. J.; Zhang, W. High-throughput screening for advanced thermoelectric materials: diamond-like ABX₂ compounds. *ACS. Appl. Mater. Interfaces.* **2019**, *11*, 24859-66. DOI
27. Qin, G.; Huang, A.; Liu, Y.; et al. High-throughput computational evaluation of lattice thermal conductivity using an optimized Slack model. *Mater. Adv.* **2022**, *3*, 6826-30. DOI
28. Cao, Y.; Zhu, Z.; Li, X.; et al. Unraveling the relationships between chemical bonding and thermoelectric properties: n-type ABO₃ perovskites. *J. Mater. Chem. A.* **2022**, *10*, 11039-45. DOI
29. Wang, X.; Shu, G.; Zhu, G.; et al. An interpretable formula for lattice thermal conductivity of crystals. *Mater. Today. Phys.* **2024**, *48*, 101549. DOI
30. Anderson, O. L. A simplified method for calculating the debye temperature from elastic constants. *J. Phys. Chem. Solids.* **1963**, *24*, 909-17. DOI
31. Belomestnykh, V. N. The acoustical Grüneisen constants of solids. *Tech. Phys. Lett.* **2004**, *30*, 91-3. DOI
32. Ju, S.; Yoshida, R.; Liu, C.; et al. Exploring diamondlike lattice thermal conductivity crystals via feature-based transfer learning. *Phys. Rev. Mater.* **2021**, *5*, 053801. DOI
33. Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301. DOI PubMed
34. Zhu, T.; He, R.; Gong, S.; et al. Charting lattice thermal conductivity for inorganic crystals and discovering rare earth chalcogenides for thermoelectrics. *Energy. Environ. Sci.* **2021**, *14*, 3559-66. DOI
35. Vaitesswar, U. S.; Bash, D.; Huang, T.; et al. Machine learning based feature engineering for thermoelectric materials by design. *Digit. Discov.* **2024**, *3*, 210-20. DOI
36. Omeo, S. S.; Fu, N.; Dong, R.; Hu, M.; Hu, J. Structure-based out-of-distribution (OOD) materials property prediction: a benchmark

- study. *npj. Comput. Mater.* **2024**, *10*, 1316. DOI
37. Meredig, B.; Antono, E.; Church, C.; et al. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* **2018**, *3*, 819-25. DOI
 38. Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80-5. DOI PubMed PMC
 39. Griesemer, S. D.; Xia, Y.; Wolverton, C. Accelerating the prediction of stable materials with machine learning. *Nat. Comput. Sci.* **2023**, *3*, 934-45. DOI PubMed
 40. Wang, X.; Gao, Z.; Zhu, G.; et al. Role of high-order anharmonicity and off-diagonal terms in thermal conductivity: a case study of multiphase CsPbBr₃. *Phys. Rev. B.* **2023**, *107*, 214308. DOI
 41. Herring, C. Role of low-energy phonons in thermal conduction. *Phys. Rev.* **1954**, *95*, 954-65. DOI
 42. Jia, T.; Chen, G.; Zhang, Y. Lattice thermal conductivity evaluated using elastic properties. *Phys. Rev. B.* **2017**, *95*, 155206. DOI
 43. Knoop, F.; Purcell, T. A. R.; Scheffler, M.; Carbogno, C. Anharmonicity measure for materials. *Phys. Rev. Mater.* **2020**, *4*, 083809. DOI
 44. Belomestnykh, V. N.; Tesleva, E. P. Interrelation between anharmonicity and lateral strain in quasi-isotropic polycrystalline solids. *Tech. Phys.* **2004**, *49*, 1098-100. DOI
 45. Chung, D. H.; Buessem, W. R. The Voigt-Reuss-Hill approximation and elastic moduli of polycrystalline MgO, CaF₂, β-ZnS, ZnSe, and CdTe. *J. Appl. Phys.* **1967**, *38*, 2535-40. DOI
 46. Xiao, Y.; Chang, C.; Pei, Y.; et al. Origin of low thermal conductivity in SnSe. *Phys. Rev. B.* **2016**, *94*, 125203. DOI
 47. Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B. Condens. Matter.* **1996**, *54*, 11169-86. DOI PubMed
 48. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B. Condens. Matter.* **1994**, *50*, 17953-79. DOI PubMed
 49. Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B.* **1999**, *59*, 1758-75. DOI
 50. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-8. DOI PubMed
 51. Li, W.; Carrete, J.; A Katcho N, Mingo N. ShengBTE: A solver of the Boltzmann transport equation for phonons. *Comput. Phys. Commun.* **2014**, *185*, 1747-58. DOI
 52. Dunn, A.; Wang, Q.; Ganose, A.; Dopp, D.; Jain, A. Benchmarking materials property prediction methods: the Matbench test set and Automater reference algorithm. *npj. Comput. Mater.* **2020**, *6*, 406. DOI
 53. AFLOW. <https://afLOWlib.org/>. (accessed 2025-02-17).
 54. Beasley, J. D. Thermal conductivities of some novel nonlinear optical materials. *Appl. Opt.* **1994**, *33*, 1000-3. DOI PubMed
 55. Toberer, E. S.; Zevalkink, A.; Snyder, G. J. Phonon engineering through crystal chemistry. *J. Mater. Chem.* **2011**, *21*, 15843. DOI
 56. Slack, G. A. Thermal conductivity of MgO, Al₂O₃, MgAl₂O₄, and Fe₃O₄ crystals from 3° to 300° K. *Phys. Rev.* **1962**, *126*, 427-41. DOI
 57. Bjerg, L.; Iversen, B. B.; Madsen, G. K. H. Modeling the thermal conductivities of the zinc antimonides ZnSb and Zn₄Sb₃. *Phys. Rev. B.* **2014**, *89*, 024304. DOI
 58. Toher, C.; Oses, C.; Plata, J. J.; et al. Combining the AFLOW GIBBS and elastic libraries to efficiently and robustly screen thermomechanical properties of solids. *Phys. Rev. Mater.* **2017**, *1*, 015401. DOI
 59. Curtarolo, S.; Setyawan, W.; Hart, G. L.; et al. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **2012**, *58*, 218-26. DOI
 60. Wang, Y.; Gao, Z.; Wang, X.; et al. Anomalous thermal conductivity in 2D silica nanocages of immobilizing noble gas atom. *Appl. Phys. Lett.* **2024**, *124*, 122205. DOI
 61. Zhou, F.; Nielson, W.; Xia, Y.; Ozoliņš, V. Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations. *Phys. Rev. Lett.* **2014**, *113*, 185501. DOI PubMed
 62. Hao, Y.; Zuo, Y.; Zheng, J.; et al. Machine learning for predicting ultralow thermal conductivity and high ZT in complex thermoelectric materials. *ACS Appl. Mater. Interfaces.* **2024**, *16*, 47866-78. DOI
 63. Zhang, Y.; Ke, X.; Chen, C.; Yang, J.; Kent, P. R. C. Thermodynamic properties of PbTe, PbSe, and PbS: first-principles study. *Phys. Rev. B.* **2009**, *80*, 024304. DOI
 64. Zhao, L. D.; Lo, S. H.; Zhang, Y.; et al. Ultralow thermal conductivity and high thermoelectric figure of merit in SnSe crystals. *Nature* **2014**, *508*, 373-7. DOI
 65. Pashinkin, A. S.; Mikhailova, M. S.; Malkova, A. S.; Fedorov, V. A. Heat capacity and thermodynamic properties of lead selenide and lead telluride. *Inorg. Mater.* **2009**, *45*, 1226-9. DOI
 66. Tian, Z.; Garg, J.; Esfarjani, K.; Shiga, T.; Shiomi, J.; Chen, G. Phonon conduction in PbSe, PbTe, and PbTe_{1-x}Se_x from first-principles calculations. *Phys. Rev. B.* **2012**, *85*, 184303. DOI
 67. Zhou, Z.; Cao, G.; Liu, J.; Liu, H. High-throughput prediction of the carrier relaxation time via data-driven descriptor. *npj. Comput. Mater.* **2020**, *6*, 417. DOI
 68. Madsen, G. K.; Singh, D. J. BoltzTraP. A code for calculating band-structure dependent quantities. *Comput. Phys. Commun.* **2006**, *175*, 67-71. DOI
 69. Li, X.; Zhang, Z.; Xi, J.; et al. TransOpt. A code to solve electrical transport properties of semiconductors in constant electron-phonon

- coupling approximation. *Comput. Mater. Sci.* **2021**, *186*, 110074. DOI
70. Karamad, M.; Magar, R.; Shi, Y.; Siahrostami, S.; Gates, I. D.; Barati, F. A. Orbital graph convolutional neural network for material property prediction. *Phys. Rev. Mater.* **2020**, *4*, 093801. DOI
 71. Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **2019**, *31*, 3564-72. DOI
 72. Cheng, J.; Zhang, C.; Dong, L. A geometric-information-enhanced crystal graph network for predicting properties of materials. *Commun. Mater.* **2021**, *2*, 194. DOI
 73. Choudhary, K.; Decost, B. Atomistic line graph neural network for improved materials property predictions. *npj. Comput. Mater.* **2021**, *7*, 650. DOI
 74. Louis, S. Y.; Zhao, Y.; Nasiri, A.; et al. Graph convolutional neural networks with global attention for improved materials property prediction. *Phys. Chem. Chem. Phys.* **2020**, *22*, 18141-8. DOI
 75. Omea, S. S.; Louis, S. Y.; Fu, N.; et al. Scalable deeper graph neural networks for high-performance materials property prediction. *Patterns* **2022**, *3*, 100491. DOI PubMed PMC
 76. Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **2017**, *8*, 15679. DOI PubMed PMC
 77. Ward, L.; Liu, R.; Krishna, A.; et al. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B.* **2017**, *96*, 024104. DOI
 78. Ruff, R.; Reiser, P.; Stühmer, J.; Friederich, P. Connectivity optimized nested line graph networks for crystal structures. *Digit. Discov.* **2024**, *3*, 594-601. DOI