ADVANCED MATERIALS
www.advmat.de

A Machine Learning Study on High Thermal Conductivity Assisted to Discover Chalcogenides with Balanced Infrared Nonlinear Optical Performance

Qingchen Wu, Lei Kang, and Zheshuai Lin**

Exploration of novel nonlinear optical (NLO) chalcogenides with high laser-induced damage thresholds (LIDT) is critical for mid-infrared (mid-IR) solid-state laser applications. High lattice thermal conductivity (κ_L) is crucial to increasing LIDT yet often neglected in the search for NLO crystals due to lack of accurate κ_L data. A machine learning (ML) approach to predict κ_L for over 6000 chalcogenides is hereby proposed. Combining ML-generated κ_L data and first-principles calculation, a high-throughput screening route is initiated, and ten new potential mid-IR NLO chalcogenides with optimal bandgap, NLO coefficients, and thermal conductivity are discovered, in which Li_2SiS_3 and $AlZnGaS_4$ are highlighted. Big-data analysis on structural chemistry proves that the chalcogenides having dense and simple lattice structures with low anisotropy, light atoms, and strong covalent bonds are likely to possess higher κ_L. The four-coordinated motifs in which central cations show the bond valence sum of +2 to +3 and are from IIIA, IVA, VA, and IIB groups, such as those in diamond-like defect-chalcopyrite chalcogenides, are preferred to fulfill the desired structural chemistry conditions for balanced NLO and thermal properties. This work provides not only an efficient strategy but also interpretable research directions in the search for NLO crystals with high thermal conductivity.

1. Introduction

Mid-infrared (mid-IR) nonlinear optical (NLO) crystals play an essential role in laser frequency conversion, which has extensive utilities in various fields such as remote communications, environmental monitors, and medical diagnosis.[1] To date, chalcogenides have been considered as one of the most studied and preferred material families for mid-IR NLO applications due to their wide transmission ranges and strong second harmonic generation (SHG) effects, among which the benchmark $AgGaS_2$ (AGS) and $AgGaSe_2$ (AGSe) are the most widely used commercial mid-IR NLO crystals.[1c,2] However, they both suffer from low laser-induced damage thresholds (LIDTs), which hinders their applications involving high-power laser. The achievement of high LIDT has become an important issue in the exploration of novel mid-IR NLO materials.[3]

Intrinsically, the LIDT for a perfect crystal (irrespective to the presence of impurities and defects) can be increased by enlarging the energy bandgap E_g, as it can effectively suppress the occurrence of two- or multi-photon absorption which is usually a major cause of crystal damage under laser.[4] However, E_g is not the only factor that affects LIDT. In principle, the laser can also induce damage to a crystal from thermal effects, which are closely related to local laser-induced heating and heat transfer to surrounding media.[5] Thus, high thermal conductivity is advantageous for achieving good heat dissipation performance and consequently large LIDT. Moreover, increasing the thermal conductivity helps to overcome the problem of thermal lens effects[6] at high-power pumping that occurs in, e.g., AGS.[7] Overall, high thermal conductivity is of great practical value in the field of mid-IR NLO crystals and is receiving increasing interest as one of the key performance metrics.[8]

Apart from high thermal conductivity, a good mid-IR NLO chalcogenide needs to simultaneously satisfy several stringent conditions for optical performance, including[1b–f,9] i) a large SHG coefficient d_{ij} of at least >10× KDP (KH_2PO_4, $d_{36} = 0.39$ pm V^{-1}), and preferably >1× AGS ($AgGaS_2$, $d_{36} = 13.40$ pm V^{-1}); ii) a wide bandgap E_g of preferably >3.00 eV; and iii) a moderate birefringence Δn of 0.03–0.10 to achieve the phase-matching condition (**Figure 1**a). It is a great challenge to balance all the required properties in the discovery of good mid-IR NLO chalcogenides, as these requirements are often contradictory. For example, increasing the bandgap would result in a reduction of the SHG coefficient.[10] A powerful method for balancing multiple properties is high-throughput screening and

---

Q. Wu, L. Kang, Z. Lin
Functional Crystals Lab, Key Laboratory of Functional Crystals and Laser Technology
Technical Institute of Physics and Chemistry
Chinese Academy of Sciences
Beijing 100190, China
E-mail: kanglei@mail.ipc.ac.cn; zslin@mail.ipc.ac.cn
Q. Wu, Z. Lin
Center of Materials Science and Optoelectronics Engineering
University of Chinese Academy of Sciences
Beijing 100049, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adma.202309675

DOI: 10.1002/adma.202309675

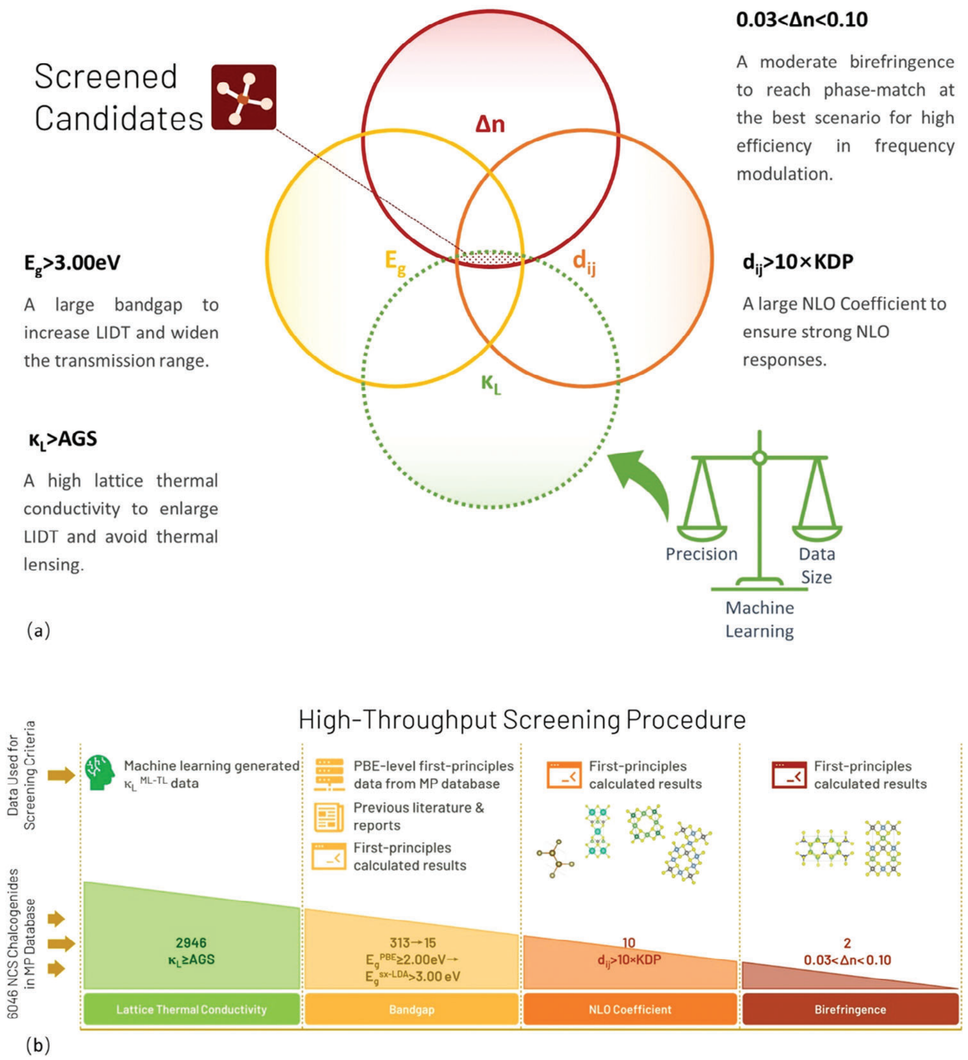

Figure 1. a) Illustration of the strategy to discover novel mid-IR NLO materials with the most balanced properties. Traditionally, E_g, d_{ij}, and Δn are the three major aspects that need to be balanced in the search for NLO materials. In this study, we add κ_L as another primary criterion. b) The high-throughput screening procedure for the discovery of balanced mid-IR NLO chalcogenides.

preexperimental design. However, although the energy bandgap E_g, SHG coefficient d_{ij}, and birefringence Δn can be accurately determined by first-principles calculations[1c] at relatively low cost and have been widely investigated in high-throughput searches for new NLO crystals,[1a,11] the thermal conductivity is often missing in such studies owing to the limited availability of data, despite its crucial role in improving LIDT.

For mid-IR NLO crystals with E_g of several eVs, their total thermal conductivities are dominated by the lattice thermal conductivity (κ_L). Experimental measurements of κ_L need to be performed on large-sized crystals which are usually unavailable in the preliminary design and discovery of new materials, while accurate first-principles calculations on κ_L by solving Boltzmann's transport equation require too tremendous computational resources for high-throughput investigations. Another solution is to compute κ_L through semiempirical models, which can save a lot of time and resources. For example, a high-throughput κ_L database, the TE Design Lab,[12] was obtained using semiempirical methods based on a simplified Debye–Callaway model. A recent study by Chu et al.[13] used thermal conductivities calculated by the semiempirical Slack model[14] to screen IR NLO materials with high κ_L. Nevertheless, semiempirical methods still need to obtain certain results either from experimental measurements or from first-principles calculations such as elastic modulus or velocity of sound, and often suffer from limited accuracy. For instantaneous and high-throughput study of κ_L with a small cost, machine learning (ML) provides another effective path by adopting a well-trained algorithm to learn from existing κ_L data and then predict the κ_L directly from material structures.

ML is playing an increasingly important role in high-throughput screening of functional materials in many areas,[15] including the field of NLO materials for the determinations of optical properties.[16] Till now, although ML has shown initial success in large-scale prediction and screening of thermal properties,[17] most of the κ_L prediction models focus on the thermoelectric materials area[18] which favors lower κ_L of

<1.0 W m^{-1} K^{-1}, contrary to the requirements of NLO crystals, and thus tend to underperform in NLO systems. For instance, in 2021, aiming at identifying new thermoelectric materials, Zhu et al.[17b] established several ML models to predict κ_L of over 90 000 synthesized structures, but their predicted κ_L values for typical mid-IR NLO crystals exhibit relatively large deviations from experimental measurements, especially with the predicted κ_L values of AGS and AGSe being over 4 W m^{-1} K^{-1} (given that the experimental values are 1.4 and 1.1 W m^{-1} K^{-1} for AGS and AGSe, respectively).[19] Therefore, obtaining a large κ_L dataset with satisfactory efficiency and accuracy simultaneously is a main bottleneck in ML discovery of good mid-IR NLO chalcogenides.

In this work, we build a ML model to predict κ_L for 6046 non-centrosymmetric (NCS) chalcogenides from the Materials Project (MP) database.[20] Combined with the ML study, a high-throughput screening procedure is performed to search for potential NLO chalcogenides with high thermal conductivity, thermodynamically stable structure, and suitable bandgap. After that, the first-principles NLO calculations are performed to evaluate the optical properties for the screen candidates. The structure-property relationship in κ_L is analyzed, and certain structural features are pointed out to achieve a proper balance in thermal and NLO properties. Accordingly, two chalcogenides, i.e., Li_2SiS_3 and $AlGaZnS_4$, are identified as the most promising crystals with balanced mid-IR NLO performance, of which the optical properties are reported for the first time. Our work provides not only an efficient strategy but also interpretable research directions in the discovery of crystals with balanced mid-IR NLO performance including high thermal conductivity.

2. Results and Discussion

2.1. High-Throughput Screening Pipeline

A high-throughput screening pipeline is launched to find potential mid-IR NLO materials in the popular chalcogenide family. In all, there are 15 393 chalcogenides from the MP database, in which 6046 are NCS excluding some systems with significant differences in properties compared with ordinary chalcogenides, such as hydro-chalcogenides, chalcogen-nitrides, and carbon-chalcogenides. To discover new mid-IR NLO crystals with most balanced properties from these candidates, we design a screening procedure by successively taking κ_L, E_g, d_{ij}, and Δn into consideration in the order of their importance, as described in Figure 1b. κ_L is relevant to the actual durability under laser and causes the major bottleneck in existing materials, so that it becomes our first concern and the primary criterion in this work. E_g and d_{ij} are vital indexes of NLO performances in evaluating a new NLO material and come secondly. Δn that determines phase-matching properties comes last since it can be compensated by techniques like quasi-phase-matching.[21]

2.2. Machine Learning Prediction of κ_L

We start by training a ML model to predict κ_L of 6046 material candidates. To retrieve data for the model to learn from, 2410 κ_L values under 300 K are obtained, including 2243 calculated κ_L (κ_L^{cal}) and 167 experimental κ_L (κ_L^{exp}), from the MP database. To achieve our goal in training the ML model to effectively predict the actual κ_L in materials, a dilemma between the size and precision of the training dataset must be solved. On the one hand, the scale of existed κ_L^{exp} database is small. A ML model trained on such small datasets often suffers from overfitting, which limits its generalization ability (i.e., the predictive ability for unreported materials). On the other hand, the database for κ_L^{cal} is vast but has inherent inaccuracy compared with experimentally data as they depend on computational methods, thus the ceiling for accuracy of a ML model training on the κ_L^{cal} set is even lower than in the calculated dataset itself.

Adopting the advantage of both databases while overcoming their weak points, an efficient and verified method in κ_L prediction is to implement the transfer learning (TL)[17b,22] technique into a state-of-the-art ML algorithm, e.g., the crystal graph convolutional neural network, or crystal graph convolutional neural network (CGCNN).[23] An advantage of the CGCNN algorithm is that it only requires the input of structured CIF documents, which can be directly obtained from online databases. By utilizing the TL-CGCNN technique, both κ_L^{cal} and κ_L^{exp} datasets can be simultaneously used to reach credible κ_L estimation (see the Experimental Section). First, a CGCNN model (with five convolutional layers and one following fully connected layer) is trained on the dataset with 2243 κ_L^{cal} values. The result for $\log_{10}\kappa_L^{cal}$ prediction can be considered accurate with the mean average error (MAE) of 0.15 and the coefficient of determination R^2 of 0.80 on the test set (**Figure 2**a), which means the predicted κ_L is 0.7–1.4 times of the calculated values in average. However, when extended to fit the 167 κ_L^{exp} values, the CGCNN model shows a large error with the MAE and R^2 for $\log_{10}\kappa_L^{exp}$ prediction equal to 0.42 and 0.55, respectively (Figure 2b). Since the CGCNN model is trained to predict κ_L^{cal}, which has inherent error compared to κ_L^{exp},[12] it is not surprising for this model to show an increased error in κ_L^{exp} prediction. When the TL technique is incorporated, layers from this model are transferred to another CGCNN model with an additional new fully connected layer which is further trained and optimized on the κ_L^{exp} dataset. By adopting this TL-CGCNN technique, the MAE of 0.25 and R^2 of 0.86 for $\log_{10}\kappa_L^{exp}$ prediction are achieved for the κ_L^{exp} training dataset (Figure 2c). This means that the $\kappa_L^{ML\text{-}TL}$ obtained from this model is 0.56–1.78 times of actual κ_L^{exp} in average, displaying significant improvement from the original CGCNN method. Note that the accuracy of our κ_L prediction by TL-CGCNN model is improved to the same level with the semiempirical computational approach, but with much lower requisition for computing power and no requisition of experimental values. In particular, the predicted $\kappa_L^{ML\text{-}TL}$ values for AGS and AGSe are 2.27 and 1.07 W m^{-1} K^{-1}, respectively, which match well with the experimental values and have better quality than those predicted in previous ML study[17b] (see Table S1, Supporting Information). Clearly, the thermal conductivity $\kappa_L^{ML\text{-}TL}$ shows good consistency with available experimental values.

2.3. Structural Chemistry for High κ_L from ML Model

Using our TL-CGCNN model, the high-precision datasets of predicted $\kappa_L^{ML\text{-}TL}$ can be generated in large scale. The predicted $\kappa_L^{ML\text{-}TL}$ data for 6046 NCS chalcogenides in the MP database are

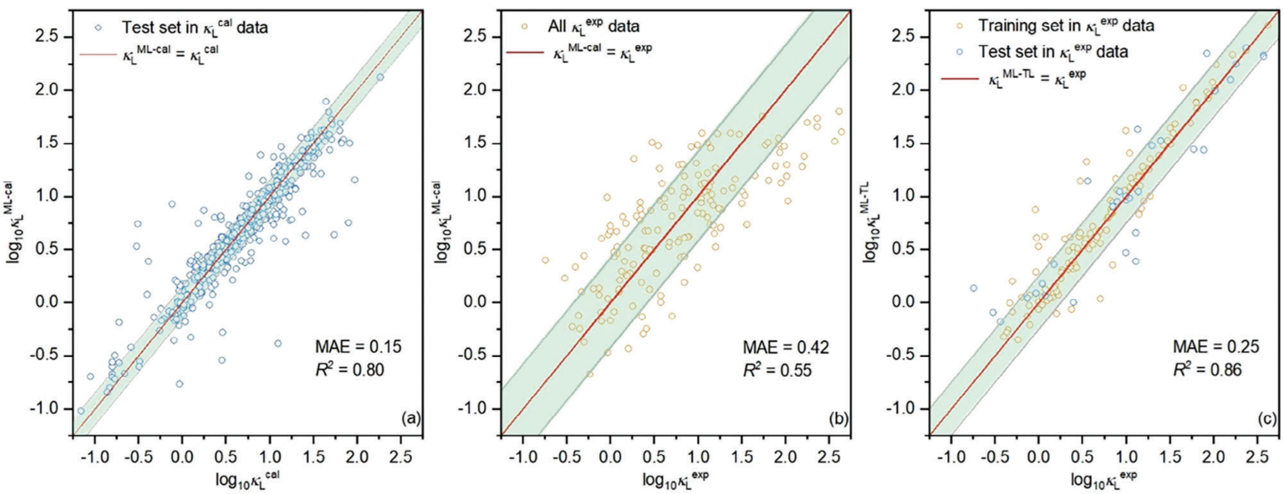

Figure 2. a) Test results of CGCNN trained on $\kappa_L{}^{cal}$ set. b) $\kappa_L{}^{ML\text{-}cal}$ predicted by CGCNN trained on $\kappa_L{}^{cal}$ set compared with $\kappa_L{}^{exp}$. c) $\kappa_L{}^{ML\text{-}TL}$ predicted by TL-CGCNN trained on $\kappa_L{}^{exp}$ set. The green shadow area denotes the test error smaller than MAE. Clearly, the prediction accuracy on κ_L is greatly improved after TL is applied.

listed in Table S2 (Supporting Information). We use AGS as a benchmark for further $\kappa_L{}^{ML\text{-}TL}$ screening. Accordingly, $\kappa_L{}^{ML\text{-}TL} > 2.27$ W m^{-1} K^{-1} is chosen to figure out the materials with possibly higher thermal conductivity than AGS, from which 2946 materials are screened out.

Intrinsically, our ML method is based on graph neural network with structural CIF documents as the only input information and thus has the "black box" characteristics, meaning that the underlying mechanism cannot be directly interpreted from the model itself. However, the big data generated by our predictive ML model can reflect the trends in κ_L and help to investigate the structure–property relationship by linking with certain microscopic structural descriptors. Here, we choose several fundamental physical parameters to be the descriptors from two considerations: i) they are relevant to κ_L in previous literatures,[12a,17b,c,24] and ii) they can be obtained directly from the CIF documents and contain basic structural and elemental information of the material without further first-principles calculations. These descriptors include the average atomic mass of cations (m_A), maximum atomic mass difference of cations (Δm_A), largest electronegativity difference ($\Delta\chi_A$), average cell volume per atom (V_A), bond length (l_B), maximum distance difference of neighboring atoms (ΔD_n), bond valence sum (BVS) of cations, and Voronoi coordination number (VCN) of cations. The consideration for such choices is not only from the general lattice and elemental properties of a material (m_A, Δm_A, and l_B), but also from the coordination environment of cations (VCN and BVS), bond polarity ($\Delta\chi_A$), and lattice anisotropy (Δm_A and ΔD_n). The dependence of $\kappa_L{}^{ML\text{-}TL}$ on these descriptors is displayed in **Figure 3**.

From the viewpoint of phonon transportation, κ_L is mainly determined by the interactions among the phonons via anharmonic Umklapp processes and can be expressed as $\kappa_L = A \cdot \frac{\bar{M} v_m^3}{\gamma^2 V^{\frac{2}{3}} T N^{\frac{1}{3}}}$, where A is a collection of physical constants, $\bar{M}$ is the average mass of the atoms in the crystal, V is the volume per atom, v_m is the mean speed of sound which has negative correlation with $\bar{M}$, N is the number of atoms in the primitive cell, and γ is the high temperature limit of the Grüneisen parameter which characterizes the phonon anharmonicity of materials.[24,25] Therefore, it is commonly accepted that high lattice thermal conductivity requires light constituent element, simple crystal structure with a small unit cell, stronger bond strength, and low phonon anharmonicity.[14,24,26] This general knowledge is also consistent with some structure–property relationships of high κ_L revealed in our ML model, i.e., the chalcogenides with dense cell (small V_A) composed of light atoms (small m_A) confined by stronger bonds (small l_B) are more likely to show higher κ_L (Figure 3a–c).

The descriptors concerning the coordination environment within the lattice (VCN and BVS) would present a more complicated association with κ_L that either a too large or too small value would be harmful (Figure 3d). For a certain cation, VCN tells the possible number of bonding anions, while BVS indicates the possible valence state on the cation.[27] A too large VCN or BVS suggests a complex bonding environment and may introduce higher phonon anharmonicity, resulting in the reduced κ_L.[24] Meanwhile, a too small VCN or BVS indicates large interatomic distances and consequently weak chemical bonds which is also unfavorable to increasing κ_L. The optimized VCN and BVS for high κ_L is 4–5 and 2–3, respectively (Figure 3d). In other words, chalcogenides composed of tetrahedral and penta-coordinated motifs in which the central cations have the oxidation states of +2 and +3 are promising to possess large κ_L.

The polarity of chemical bonds and lattice anisotropy are also influential to κ_L. As an indicator of bond polarity, the effect of $\Delta\chi_A$ on κ_L is double-edged. On the one hand, κ_L would be decreased owing to the enlargement of electronegativity difference ($\Delta\chi_A$) in a material by either inducing large lattice distortion[28] or introducing strong electron–phonon interaction like Fröhlich electron–phonon coupling which can scatter heat-carrying phonons.[29] On the other hand, the strongly polarized bonds with large $\Delta\chi_A$ can prompt the longitudinal optical (LO)–transverse optical (TO) phonon splitting which would slow down the decrease of κ_L over temperature, leading to high κ_L at a certain temperature.[30] As clearly indicated in Figure 3e, the $\Delta\chi_A$

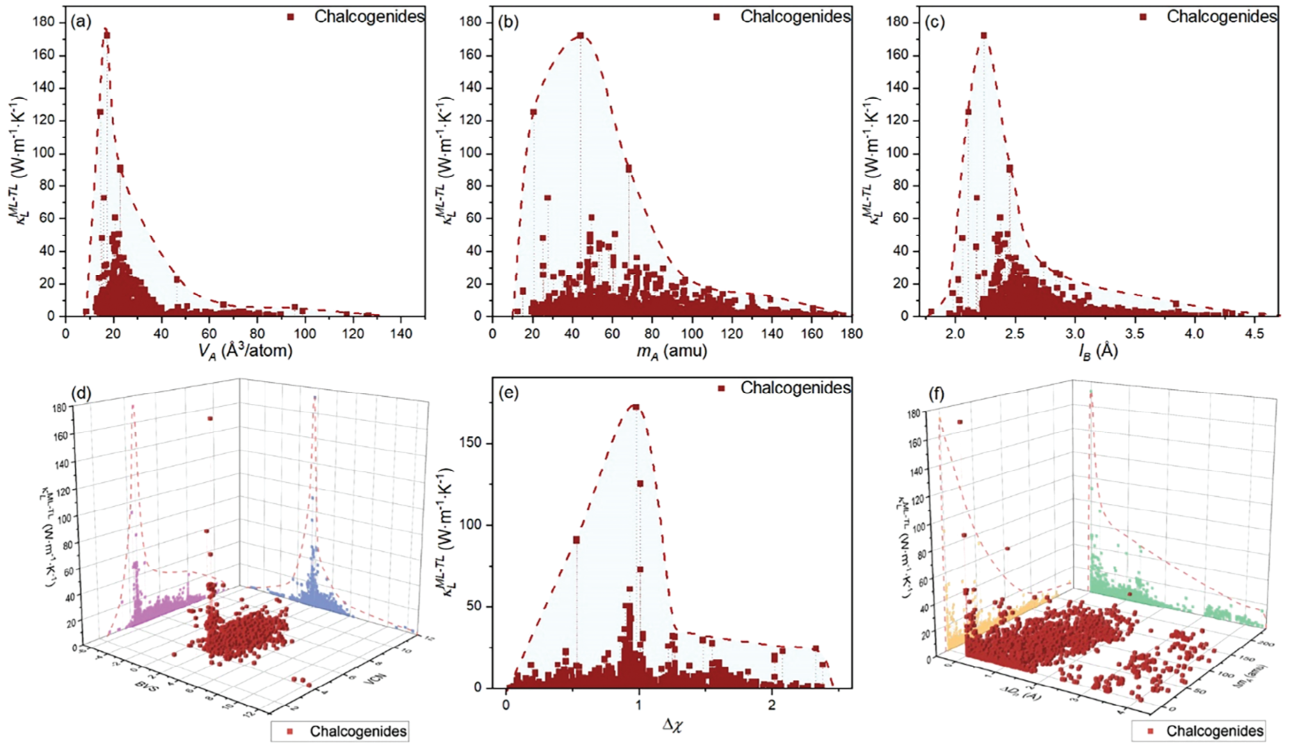

Figure 3. Illustration of $\kappa_L^{\text{ML-TL}}$ with different a) V_A, b) m_A, c) l_B, d) BVS (blue) and VCN (violet), e) $\Delta\chi_A$, f) Δm_A (gold) and ΔD_n (green). Clearly, small values of V_A, m_A, l_B, Δm_A, and ΔD_n are beneficial to obtaining large κ_L, while the effects of BVS, VCN, and $\Delta\chi_A$ are double-edged. The red dashed line and blue area denote the approximate envelopes of all data points.

of 0.8–1.0 in chalcogenides is optimal to achieve high κ_L. With the χ_A (i.e., electronegativity) of major chalcogen elements being 2.5–2.6, the χ_A of cations should be around 1.5–1.8. This means that the choice of cations prefers to IIB, IIIA, IVA, and VA groups. At the same time, the lattice anisotropy described by ΔD_n and Δm_A has an obviously negative correlation with κ_L (Figure 3f), since high anisotropy will add to the complexity of a structure and increases the chance of phonon scattering and is disadvantageous to promoting κ_L. Indeed, the structures predicted with extremely large κ_L at an order of $\geq 10^2$ are mainly binary compounds with uniform chemical groups.[24]

2.4. Screening of Mid-IR NLO Chalcogenides with High κ_L and Balanced Properties

In order to further screen the mid-IR NLO chalcogenides with balanced properties, the properties including PBE-calculated bandgap (E_g^{PBE}), number of inequivalent atom sites in the primitive cell (N_{site}), and energy above hull (E_{hull}) are extracted from the 2946 candidates with $\kappa_L^{\text{ML-TL}} > 2.27$ W m^{-1} K^{-1}. The E_g^{PBE} and E_{hull} indices are first-principles (density functional theory [DFT]) values by PBE functional and incorporated in the MP database. By considering the calculated E_g^{PBE} usually smaller than the experimental bandgap by about 30%,[4] the compounds with $E_g^{\text{PBE}} \geq 2.00$ eV are selected. In the 2946 materials, 313 have a E_g^{PBE} no less than 2.00 eV. Among these 313 candidates, we further choose the materials with $E_{\text{hull}} \leq 0.01$ eV to ensure stability and synthesizability of candidates and $N_{\text{site}} < 30$ to save computation resources for further investigation. After this coarse screening step, 78 chalcogenides are retained, in which 39 candidates have been reported for the NLO properties, certainly including those with large LIDT such as $LiAlS_2$,[3e] $Ba(BS_2)_2$,[31] Y_3GaS_6,[3a] $Na_4MgSi_2Se_6$,[3c] Na_4SnS_4,[32] Na_6ZnS_4,[33] and Li_2MGeS_4 (M = Cd, Zn),[3b,34] and some other candidates such as Sr_2SnS_4 and Li_4TiS_4 considered promising with high κ_L by other field researchers.[13] This confirms the validity of our screening approach. The relevant data are listed in Table S3 (Supporting Information). For the remaining 39 candidates without reported NLO properties, 21 compounds are further ditched as they have uncertain sites, are duplicate with known NLO materials, or contain lanthanide elements like Nd, Dy, and Er that cannot be handled correctly in our first-principles method (including $Yb(DyS_2)_2$, Tb_2YbS_4, $Sr(PrS_2)_2$, $Yb(SmS_2)_2$, $Sr(LaS_2)_2$, $Sr(NdS_2)_2$, $Yb(PrS_2)_2$, $Yb(NdS_2)_2$, La_2YbS_4, and $Dy_6Mg(GeS_7)_2$). After that, we perform the first-principles calculations on the bandgaps $E_g^{\text{sx-LDA}}$ (by using the more accurate but more expensive hybrid sx-LDA functional) and optical properties for the 18 remaining structures, and the results are given in **Table 1**.

Using $E_g^{\text{sx-LDA}} > 3.00$ eV as standard, 15 out of the 18 candidates are screened, of which 10 compounds with $d_{ij} > 10\times$ KDP are finally screened out as potential mid-IR NLO materials with preferable optical properties and large κ_L. All these 10 chalcogenides are either existing compounds or dynamically stable with no imaginary phonon modes (Figure S2, Supporting Information). Among them, Li_2SiS_3 and $AlZnGaS_4$ fulfill the most optimized phase matching condition $0.03 < \Delta n < 0.10$ with a moderate birefringence and are chosen as the most balanced ones. In

Table 1. Calculated results for 18 potential mid-IR NLO chalcogenides.

Candidate	Space group	$\kappa_L^{\text{ML-TL}}$ [W m^{-1} K^{-1}]	$E_g^{\text{sx-LDA}}$ [eV]	d_{ij} [pm V^{-1}]	Δn @1064 nm
BeTe	$P6_3mc$	91.25	2.42	$d_{33} = 8.23$	0.001
MgS	$P6_3mc$	12.85	4.49	$d_{33} = 0.91$	0.012
*MgTe	$P6_3mc$	12.42	3.41	$d_{33} = 13.00$	0.010
*DL-Al_2ZnS_4	$I\bar{4}$	10.68	3.10	$d_{36} = -7.80$	0.023
*$Al_2ZnSe_2S_2$	$C2$	10.48	3.29	$d_{36} = 10.91$	0.014
MgSe	$P6_3mc$	10.04	3.76	$d_{33} = 3.70$	0.009
*$AlZnGaS_4$	$I\bar{4}$	9.69	3.66	$d_{36} = -11.19$	0.032
$KLiZnS_2$	$I\bar{4}m2$	7.40	2.98	$d_{24} = -d_{15} = 5.33$	0.013
$MgCdTe_2$	$P\bar{4}m2$	6.14	2.45	$d_{24} = -d_{15} = 22.69$	0.057
*$RbLiZnS_2$	$I\bar{4}m2$	4.06	3.08	$d_{24} = -d_{15} = 5.16$	0.014
*P-Li_3NbS_4	$P\bar{4}3m$	3.93	4.35	$d_{36} = 7.17$	0.000
*$CsLiZnS_2$	$I\bar{4}m2$	3.68	3.25	$d_{24} = -d_{15} = 5.10$	0.016
$Y_6Mg(GeS_7)_2$	$P3$	3.26	3.16	$d_{24} = d_{15} = -2.30$ $d_{33} = -1.05$	0.008
$Y_6Mg(SiS_7)_2$	$P3$	3.05	3.16	$d_{24} = d_{15} = -1.20$ $d_{33} = -1.72$	0.014
*β-Na_3PS_4	$I\bar{4}3m$	3.01	3.34	$d_{36} = 8.32$	0.000
$Y_6Al_2SiS_{14}$	$P3$	2.92	3.26	$d_{16} = -d_{22} = 1.24$ $d_{33} = -1.83$	0.050
*Al_2S_3	Cc	2.67	3.79	$d_{11} = -6.26$ $d_{15} = -2.86$	0.020
*Li_2SiS_3	$Cmc2_1$	2.46	4.23	$d_{15} = -4.05$ $d_{24} = -2.84$	0.066

Note: The 10 compounds with $E_g^{\text{sx-LDA}} > 3.00$ eV and $d_{ij} > 10\times$ KDP are marked by star symbols.

addition, diamond-like (DL)-Al_2ZnS_4, $Al_2ZnSe_2S_2$, $ALiZnS_2$ (A = Rb or Cs), and Al_2S_3 (*Cc*) also have promising d_{ij} and $E_g^{\text{sx-LDA}}$ but are probably non-phase-matchable due to their small theoretical Δn. MgTe, β-Na_3PS_4, and P-Li_3NbS_4 are intrinsically not phase-matchable despite their large d_{ij} and $E_g^{\text{sx-LDA}}$, because these three materials possess extremely small Δn values since they are composed of only one type of tetrahedral NLO motif and/or crystallizes in cubic lattices.

In order to investigate the structural chemistry conditions for balancing d_{ij} and κ_L, we focus on the known NLO materials from the 6046 chalcogenides in MP database. There are 27 materials in which the d_{ij} data have been reported in previous literatures that satisfy the conditions of $N_{\text{site}} < 30$, $E_{\text{hull}} \leq 0.01$ eV, and $E_g^{\text{PBE}} \geq 2.00$ eV but with small $\kappa_L^{\text{ML-TL}}$ ($\kappa_L^{\text{ML-TL}} \leq 2.27$ W m^{-1} K^{-1}), see Table S4 (Supporting Information). So, combining the data listed in Table 1 and Figures S3 and S4 (Supporting Information), totally a set of 84 NLO materials can be established. **Figure 4**a shows the distribution of d_{ij} and $\kappa_L^{\text{ML-TL}}$ data from the 84 NLO-active chalcogenides. Based on this database, one may further conclude the structural chemistry conditions to achieve optimized performance of high κ_L and balanced NLO properties in chalcogenides from the element and coordinate levels as follows. i) On the element level, a statistical analysis is conducted on the element selection of cations that tend to enlarge κ_L, E_g, and d_{ij} at the same time. Figure 4b plots the dependence of the numbers (N) of chalcogenides containing specific elements that $\kappa_L^{\text{ML-TL}}$ > AGS and $E_g^{\text{PBE}} \geq 2.00$ eV ($N_{(\kappa_L^{\text{ML-TL}} > \text{AGS})}$ and $N_{(E_g^{\text{PBE}} \geq 2.00\text{ eV})}$) on the d_{ij} magnitude in all 6046 chalcogenides. Clearly, the chalcogenides with main-group short period (especially IIIA, IVA, and VA families) elements like Al, Ge and P, along with IIB group elements such as Zn and rare earth elements from La family, are more likely to simultaneously provide prominent κ_L, E_g, and d_{ij} in a material. ii) On the level of coordination structure, using VCN and BVS as the main indices, it is revealed that the four- or five-coordination structures and +2 to +3 oxidation states for cations have large d_{ij} (Figure 4c), which is actually overlapped with the optimal structural characteristics for high κ_L. This also accords with our high-throughput screening results where all of the major functional groups in the good candidates are the IIIA, IVA, VA, and IIB cation-centered tetrahedrons including [AlS_4], [ZnS_4], [$ZnSe_4$], [GaS_4], [SiS_4], [PS_4], etc. In addition, to increase the birefringence Δn, the large lattice anisotropy is preferable, and the structures with $\Delta D_n > 0.5$ Å have more chance of showing large Δn (Figure 4d), which is consistent with other reported

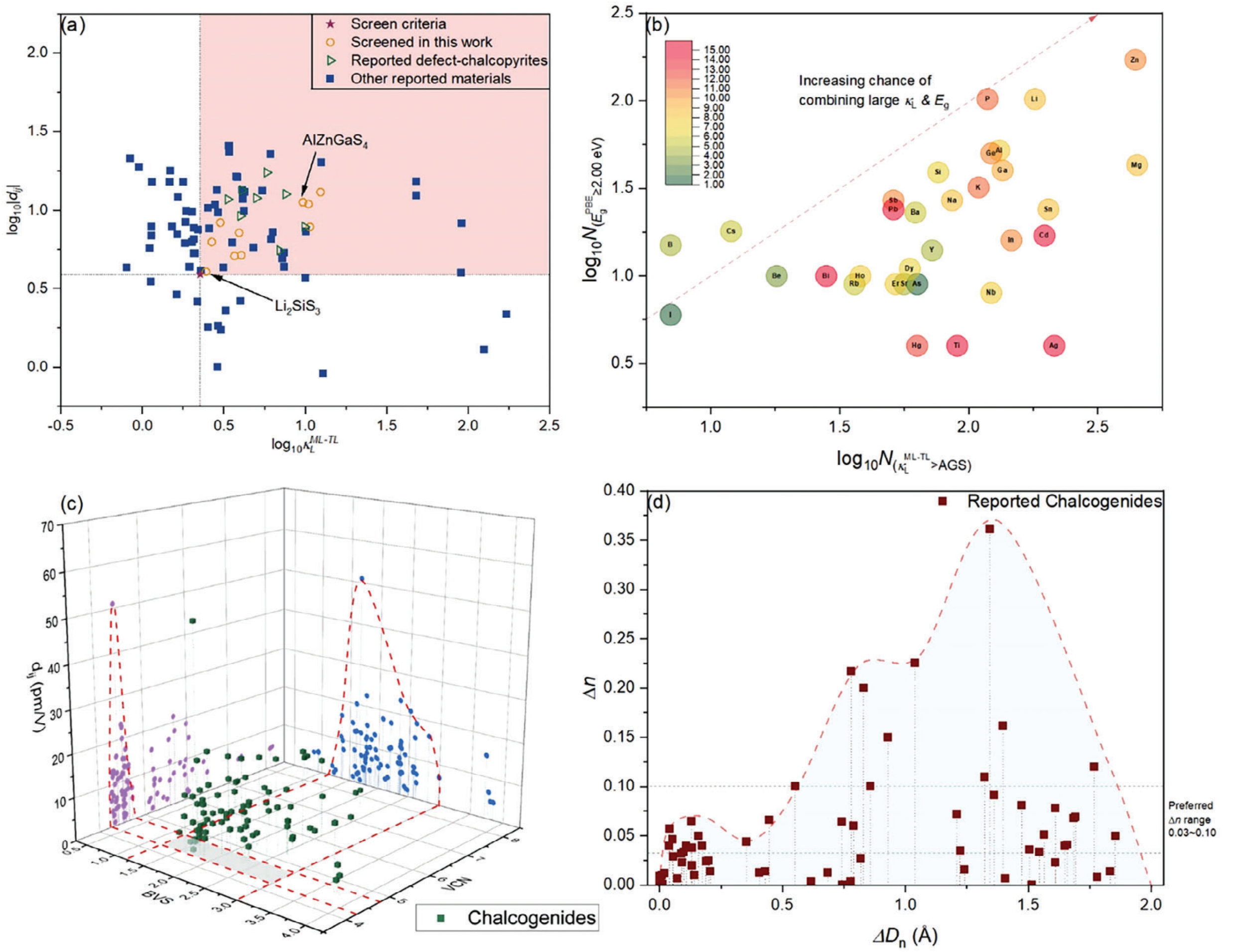

Figure 4. a) Distribution of d_{ij} and $\kappa_L^{ML\text{-}TL}$ data from 84 NLO-active chalcogenides that satisfy the conditions of $N_{site} < 30$, $E_{hull} \leq 0.01$ eV, and $E_g^{PBE} \geq 2.00$ eV. Compounds falling into the red shaded area on the upper right are the candidates that satisfy the optimal criteria of $\kappa_L^{ML\text{-}TL}$ > AGS and d_{ij} > 10× KDP. b) Evaluation on the number of chalcogenides with different elements by considering κ_L, E_g, and d_{ij}. The elements on the upper right are more likely to have high κ_L and large E_g in chalcogenides. The color of the circles represents the magnitude of average d_{ij} of chalcogenides containing the specific elements in pm V^{-1}. c) Relationship between d_{ij}, BVS, and VCN for the NLO-active chalcogenides satisfying $N_{site} < 30$, Ehull ≤ 0.01 eV and $E_g^{PBE} \geq 2.00$ eV. Red dashed lines indicate the approximate distribution for large d_{ij}. d) Relationship between Δn and ΔD_n for the NLO-active chalcogenides satisfying $N_{site} < 30$, $E_{hull} \leq 0.01$ eV, and $E_g^{PBE} \geq 2.00$ eV. The red dashed line and blue area denote the approximate envelopes of all data points.

studies.[16c] However, large lattice anisotropy is harmful for improving the thermal conductivity. For mid-IR NLO applications, since the preferred Δn is within 0.03–0.10, a relatively small $\Delta D_n \approx 0.25$ Å is acceptable for maintain moderate Δn while maximizing κ_L.

2.5. First-Principles Characterization of Selected NLO Chalcogenides with High κ_L

As obtained from the above screening procedures, Li$_2$SiS$_3$ and AlZnGaS$_4$ are predicted to have very balanced mid-IR NLO performance (see Table 1). Here, we make the first-principles calculations to confirm the high thermal conductivity in these two chalcogenides.

Li$_2$SiS$_3$ crystallizes in *Cmc*2$_1$ (36) space group with orthorhombic enargite-like structure consisting of corner-sharing [LiS$_4$] and [SiS$_4$] tetrahedrons (**Figure 5**a). Based on the linearized phonon Boltzmann equation (LBTE),[35] the first-principles κ_L values at 300 K are 3.19, 4.70, and 4.10 W m^{-1} K^{-1} along *xx*, *yy*, and *zz* directions, respectively (Figure 5b), basically consistent with the predicted $\kappa_L^{ML\text{-}TL}$ (2.46 W m^{-1} K^{-1}). Density functional theory calculations show that Li$_2$SiS$_3$ has a large indirect bandgap of 4.23 eV and wide IR transmission range up to ≈17.1 µm (Figures S3 and S4, Supporting Information), and the birefringence Δn is ≈0.066 (Figure S5, Supporting Information), indicating the capability of this crystal to achieve phase match in the mid-IR region.

The further SHG-weighted electron density analysis reveals that, the electron orbital near Si and S is the primary contributor of the largest component of SHG coefficient (d_{15}) while almost no SHG-weighted electron cloud is found near Li (Figure 5c), confirming [SiS$_4$] tetrahedron as the main SHG-active group in Li$_2$SiS$_3$. This is also supported by band-resolved SHG analysis and partial density of states (PDOS), see Figure S3 (Supporting Information). Note that, although the structure of Li$_2$SiS$_3$ was first reported in 1989,[36] it has only been synthesized in powder

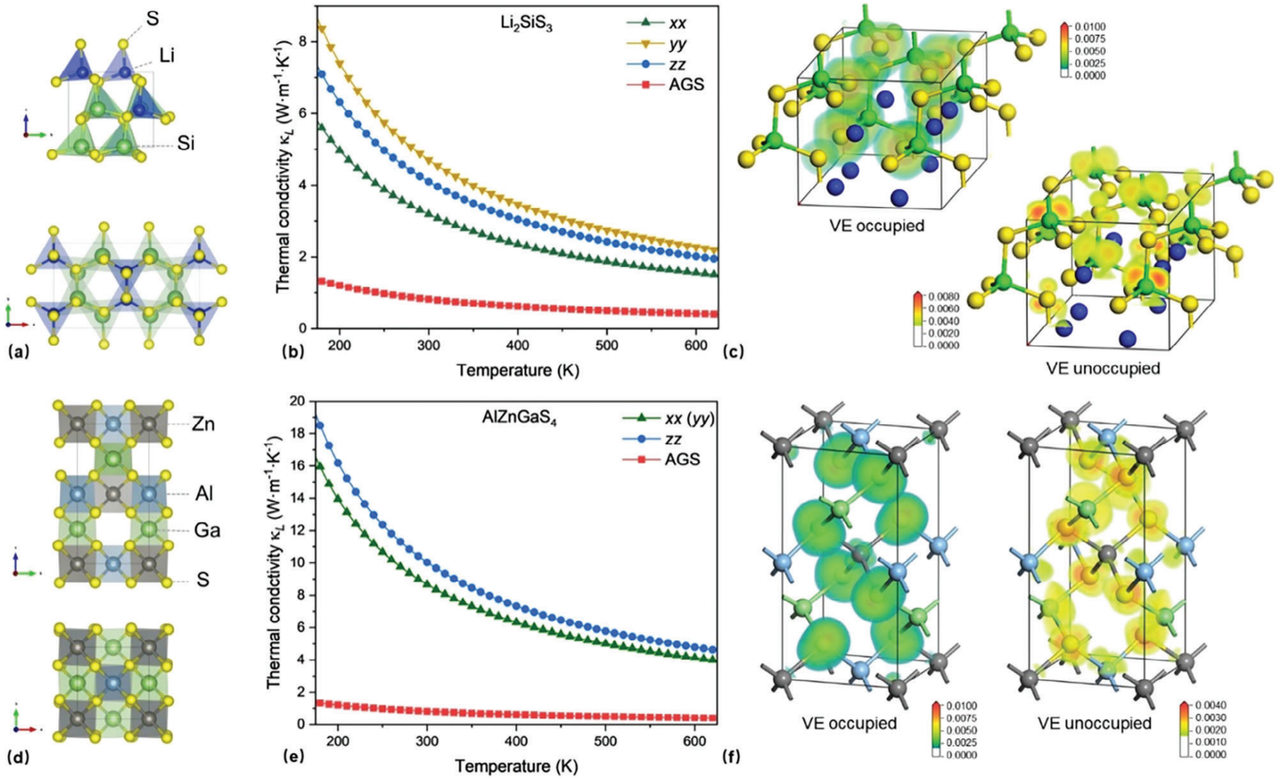

Figure 5. a) Structure of Li_2SiS_3. b) First-principles κ_L calculated by LBTE method in Li_2SiS_3 compared to the benchmark AGS. c) SHG-weighted electron density of Li_2SiS_3. Only the virtual electron (VE) process is displayed since it dominates d_{15} (see Figure S3, Supporting Information). d) Structure of $AlZnGaS_4$. e) First-principles κ_L calculated by LBTE method in $AlZnGaS_4$ compared to the benchmark AGS. f) SHG-weighted electron density of $AlZnGaS_4$. Only the VE process is displayed since it dominates d_{36} (see Figure S6, Supporting Information).

and may contain impurities such as centrosymmetric Li_4SiS_4 and metastable n-Li_2SiS_3 phase.[37] It requires further investigations to obtain large and pure crystals for experimental characterizations on the NLO and thermal properties.

$AlZnGaS_4$ crystallizes in $I\bar{4}$ (82) space group where Zn is bonded to four equivalent S atoms to form [ZnS_4] tetrahedrons that share corners with four equivalent [GaS_4] tetrahedrons and four equivalent [AlS_4] tetrahedrons (Figure 5d). The LBTE calculated κ_L values of $AlZnGaS_4$ at 300 K are 8.67 W m^{-1} K^{-1} in the *xx* (*yy*) direction and 10.04 W m^{-1} K^{-1} along the *z*-axis, respectively (Figure 5e), which is in excellent agreement with our ML prediction (9.69 W m^{-1} K^{-1}). This compound is predicted to have a direct $E_g^{\text{sx-LDA}}$ of 3.66 eV and the good IR transparent range and birefringence for mid-IR NLO applications (Figures S4–S6, Supporting Information). As shown by SHG-weighted electron density analysis, the major SHG coefficient d_{36} in $AlZnGaS_4$ is mainly contributed by the occupied S orbitals and unoccupied orbitals from all three cations (Al^{3+}, Zn^{2+}, and Ga^{3+}), meaning that all three tetrahedral groups have considerable contribution to the NLO responses (Figure 5f and Figure S6, Supporting Information).

It should be emphasized that $AlZnGaS_4$ can be obtained by element substitution from the reported Ga_2ZnS_4 structure,[1c,38] a member in the A_2BC_4 (*A* from IIIA group; *B* from IIB group; *C* = S, Se) defect-chalcopyrite family which shares structural similarity with the chalcopyrite-type AGS and has long been recognized as a competitive NLO material family.[1c,39] Our data analysis confirms that this family contains many NLO materials with higher $\kappa_L^{\text{ML-TL}}$ than the AGS (also shown in Figure 4a), thanks to the tetrahedral coordinated structures and the cation compositions from IIB and IIIA groups. The detailed comparison between $AlZnGaS_4$ and the benchmark AGS in thermal conductivity is presented in Figure S7 (Supporting Information). As the NLO properties in $AlZnGaS_4$ have never been reported previously, experimental investigations on these chalcogenides are deserved to be carried out in further research.

3. Conclusion

In conclusion, we present in this work an effective machine learning approach to predict the lattice thermal conductivity of materials and integrate it into a high-throughput screening process using the first-principles calculations to identify novel mid-IR NLO crystals with high thermal conductivity in the chalcogenide domain. Ten out of 6046 NCS chalcogenides are selected as potential high-κ_L mid-IR NLO materials with calculated $d_{ij} > 10\times$ KDP, $E_g^{\text{sx-LDA}} > 3.00$ eV and predicted $\kappa_L^{\text{ML-TL}} >$ AGS, among which two chalcogenides, namely, Li_2SiS_3 and $AlZnGaS_4$, are phase-matchable with $0.03 < \Delta n < 0.10$.

With the first-principles level of accuracy, the big database of κ_L generated by machine learning is utilized to investigate the structure–property relationships in κ_L as well as the optimal structural chemistry conditions to achieve a balance between NLO and thermal properties in non-centrosymmetric chalcogenides. We found that a dense and simple lattice structure with low anisotropy, light atoms and strong covalent bonds is beneficial for higher κ_L. Structures composed of four-coordinated motifs where central cations show a bond valence sum of +2 to +3 and are from IIIA, IVA, VA, and IIB groups are advantageous to combine good NLO properties and high κ_L, such as the defect-chalcopyrite chalcogenides represented by $AlZnGaS_4$. Our ML-boosted screening strategy provides not only a powerful tool to discover potential high-κ_L NLO crystals in a high-throughput manner, but also a novel approach to understand the mechanism in balancing NLO and thermal properties of materials through big data, which opens a new horizon in the search for high-κ_L and NLO-active material genes and is expected to be extensible for more NLO materials systems other than chalcogenides. Note that in the present work some important physicochemical properties of NLO crystals, e.g., hygroscopes and mechanical stability, are not considered. It is anticipated that the more comprehensive investigations on the balanced NLO crystals will be fulfilled in future studies with the developments of computational and experimental means.

4. Experimental Section

Data Retrieving: All κ_L values at 300 K used for ML modeling were obtained via Citrination platform (https://citrination.com/)[40] which provides access to data from various reported works. Computed κ_L ($\kappa_L{}^{cal}$) values were retrieved from TE Design Lab database[12b] as well as data deposited from other studies,[41] and experimental κ_L ($\kappa_L{}^{exp}$) values were from different literatures.[7,12a,17b,41] The MP database[20] was used to crawl their corresponding structure data, along with $E_g{}^{PBE}$, N_{site}, and E_{hull}, since this database contains not only structure CIF documents but also certain first-principles material properties calculated by PBE functional[42] which can be conveniently accessible with an API interface. Structures with the same MP id were considered identical as one material. If there were multiple reported $\kappa_L{}^{cal}$ or $\kappa_L{}^{exp}$ values for one material via different data sources, only the smallest value was reserved. In total, 2633 sets of $\kappa_L{}^{cal}$ value and 411 sets of $\kappa_L{}^{exp}$ value were originally found. After discarding duplicate data and materials not included in the MP database, 2243 nonduplicate $\kappa_L{}^{cal}$ and 167 nonduplicate $\kappa_L{}^{exp}$ were left. The structure-related information (m_A, Δm_A, $\Delta\chi_A$, V_A, l_B, ΔD_n, BVS, and VCN) was obtained using Pymatgen package.[43]

Machine Learning Algorithm: The crystal graph convolutional neural network (CGCNN)[23] was used as the basis of ML modeling. This algorithm enjoys the advantage of using merely the structural CIF documents as the input without demanding any measured or calculated physical indexes. CGCNN is a neural network architecture in which the major components are composed of convolutional layers followed with hidden fully connected layers. In CGCNN, the crystal structure is transformed into a multigraph G, in which nodes and edges i indicate atoms and chemical bonds between them, respectively. Each node i (edge k) is represented by a corresponding atom (bond) feature vector $\boldsymbol{\nu}_i$ ($\boldsymbol{u}_k$) encoding the property of the atom (bond). The multigraph G is transformed into continuous representations and input to the convolutional layers, where the atom feature vector $\boldsymbol{\nu}_i$ is iteratively updated by "convolution" with surrounding atoms and bonds with a nonlinear graph convolution function. The output by the convolutional layers then produces an overall feature vector $\boldsymbol{\nu}_c$ for the crystal after pooling. Finally, $\boldsymbol{\nu}_c$ used as input to a network of fully connected layers to predict the target property. The more details on CGCNN refer to the work by Xie and Grossman.[23]

Since the κ_L data are distributed in a wide range that spans several orders of magnitude depending on the corresponding material, $\log_{10}\kappa_L$ was used instead of κ_L as the output label. Mean average error (MAE) and coefficient of determination R^2 were used as the main accuracy indexes to evaluate the forecasting ability of the machine learning model

$$\mathrm{MAE} = \frac{\sum_i^n \left|\log_{10}\kappa_{L,i}^{ML} - \log_{10}\kappa_{L,i}^{R}\right|}{n} \tag{1}$$

$$R^2 = 1 - \frac{\sum_i^n \left(\log_{10}\kappa_{L,i}^{ML} - \log_{10}\kappa_{L,i}^{R}\right)^2}{\sum_i^n \left(\log_{10}\kappa_{L,i}^{R} - \overline{\log_{10}\kappa_{L,\,i}^{R}}\right)^2} \tag{2}$$

For number i material in total n samples, $\kappa_{L,i}{}^{ML}$ represents the machine learning generated value of itsκ_L, and $\kappa_{L,i}{}^{R}$ represents the reported (either computational or experimental in different cases) value of its κ_L. $\overline{\log_{10}\kappa_{L,i}^{R}}$ means the average value of $\log_{10}\kappa_{L,i}^{R}$ in all samples. Since $\log_{10}a - \log_{10}b = \log_{10}\frac{a}{b}$, the physical meaning of MAE here is to measure the average difference between machine learning values and actual values of κ_L in ratio. The machine learning results should be $\frac{1}{10^{MAE}}$ to 10^{MAE} times of the actual value. As for R^2, a higher value (closer to 1) indicates better fitting capability of the ML model on the test data. A training/test set ratio of 4:1 is chosen in all ML modeling processes, and 20% of the training set is reserved for validation in every training cycle.

To accurately predict experimental κ_L with a comparably small $\kappa_L{}^{exp}$ dataset, the CGCNN code was modified to implement TL. TL was realized by first establishing a ML model on a large dataset and then extracting part of information in this model as basis to structure a new model trained on a small dataset. In this way, the second model can utilize knowledge learned from both the large and the small datasets while accounting for the difference between the two datasets at the same time. Several published studies have successfully performed TL on the CGCNN model with different implementation styles.[17b,22a] Here, a "layer freezing" method was used to implement TL. The idea was to copy the earlier layers of a pretrained model to a second TL model and freeze them, meaning that the parameters of these layers were fixed without additional optimizations in the TL model. The optimization of parameters was only performed for the newly built latter layers of the second TL model, as illustrated in Figure S1 (Supporting Information). Specifically in this study, the first CGCNN was trained on $\kappa_L{}^{cal}$ dataset with 2243 samples. Then its convolutional layers and the following first fully connected layer (namely, the "conv_to_fc" layer in the CGCNN method) were frozen and transferred to the second TL-CGCNN to be trained on $\kappa_L{}^{exp}$ dataset while the output layer and other fully connected layers (if there exists any) were ditched. When training TL-CGCNN on the $\kappa_L{}^{exp}$ set with 167 samples, additional hidden fully connected layers were added after the transferred layers and before the output layer, which was iterated to account for the difference between the computational and experimental datasets. All parameters in the transferred layers remain unchanged. The TL-CGCNN code can be found at https://github.com/wuqc970716/TL-CGCNN. The super parameters were optimized using Optuna package[44] by fivefold cross-validation realized with scikit-learn package.[45] The optimized super parameters for the CGCNN and TL-CGCNN model in this study are provided in Table S5 (Supporting Information).

First-Principles Calculations on Electronic, Optical, and Phonon Properties: The first-principles calculations were performed by the plane wave pseudopotential method using CASTEP[46] based on DFT. The lattice parameters and atomic positions in the unit cells of all first-screened candidates were fully optimized using the BFGS method.[47] The exchange correlation (XC) functional is described by the local density approximation (LDA).[48] Norm-conserving pseudopotentials were used to describe the interactions between the valence electrons and the ionic cores.[49] A Monkhorst–Pack[50] k-point mesh spanning less than 0.04 Å^3 in the Brillouin zone was chosen for geometry optimization as well as optical calculations. The optimization convergence thresholds and kinetic energy cutoff were set as default in the ultrafine option of the software. The SHG coefficient d_{ij} is one-half of the second-order NLO susceptibility $\chi_{ijk}{}^{(2)}$, which was calculated using the expression developed by Lin et al. as shown below:[51]

$$\chi_{ijk}^{(2)} = \chi_{ijk}^{(2)}\ (\mathrm{VE}) + \chi_{ijk}^{(2)}\ (\mathrm{VH}) + \chi_{ijk}^{(2)}\,(\text{two bands}) \tag{3}$$

$$\chi_{ijk}^{(2)}\ (\mathrm{VH}) = \frac{\mathrm{e}^3}{2\hbar^2 m^3} \sum_{\nu\nu' c} \int \frac{\mathrm{d}^3 k}{4\pi^3} P\,(ijk)\ \mathrm{Im}\left[p_{\nu\nu'}^{i} p_{\nu' c}^{j} p_{c\nu}^{k}\right] \left(\frac{1}{\omega_{c\nu}^{3}\omega_{\nu' c}^{2}} + \frac{2}{\omega_{\nu c}^{4}\omega_{c\nu'}}\right) \tag{4}$$

$$\chi_{ijk}^{(2)}\ (\mathrm{VE}) = \frac{\mathrm{e}^3}{2\hbar^2 m^3} \sum_{\nu cc'} \int \frac{\mathrm{d}^3 k}{4\pi^3} P\,(ijk)\ \mathrm{Im}\left[p_{\nu c}^{i} p_{cc'}^{j} p_{c'\nu}^{k}\right] \left(\frac{1}{\omega_{c\nu}^{3}\omega_{\nu c'}^{2}} + \frac{2}{\omega_{\nu c}^{4}\omega_{c'\nu}}\right) \tag{5}$$

$$\chi_{ijk}^{(2)}\ (\text{two bands}) = \frac{\mathrm{e}^3}{2\hbar^2 m^3} \sum_{\nu c} \int \frac{\mathrm{d}^3 k}{4\pi^3} P\,(ijk)\ \frac{\mathrm{Im}\left[p_{\nu c}^{i} p_{c\nu}^{j}\left(p_{\nu\nu}^{k} - p_{cc}^{k}\right)\right]}{\omega_{\nu c}^{5}} \tag{6}$$

Here, *i*, *j*, and *k* are Cartesian components. ν and ν' denote valence bands, while *c* and *c′* denote conduction bands. *P*(*ijk*) denotes full permutation. At zero frequency, the total second-order NLO susceptibility can be described to virtual electron (VE), virtual hole (VH), and two-band processes. The contribution from two-band process is extremely small and can be neglected. The band-resolved SHG analysis was performed to identify the contribution of SHG from VE and VH processes originating from the occupied or unoccupied electronic states in a level-by-level basis.[52] In the SHG-weighted electron density analysis, by further applying the percentage of SHG contribution onto the electronic charge densities of corresponding orbitals, the occupied or unoccupied electronic states irrelevant to SHG are not shown, whereas those vital to SHG are highlighted in real space.[4]

Owing to the underestimation of bandgap by LDA functional, the screened exchange LDA (sx-LDA) functional[53] was used to calculate bandgaps in high accuracy. In the optical calculations, a scissors operator[54] which represents bandgap difference between the sx-LDA and LDA bandgaps was adopted to correct the optical properties. Past studies proved that these parameters were sufficiently accurate.[1c,1d] Phonon calculation was performed by linear response method to evaluate the dynamic stabilities of screened candidates.[55]

First-Principles Calculations on Thermal Conductivity: The high-accuracy calculation on lattice thermal conductivity was implemented by solving the LBTE[35] with single-mode relaxation time (SMRT) approximation[56] which can be described as

$$\kappa_{\mathrm{L}} = \frac{1}{VN_0} \sum_{\lambda} C_{\lambda} \nu_{\lambda} \otimes \nu_{\lambda} \tau_{\lambda}^{\mathrm{SMRT}} \tag{7}$$

where V and N_0 are volume and number of unit cell, respectively. C_λ, ν_λ, and ${\tau_\lambda}^{\mathrm{SMRT}}$ are the mode-dependent heat capacity, phonon group velocity, and single-mode relaxation time of mode λ. The ${\tau_\lambda}^{\mathrm{SMRT}}$ can be approximated by phonon lifetime τ_λ here.[56]

The tetrahedron method was used to perform SMRT calculation with Phono3py package.[41b,57] For all cases in this work, the third-order force constants were calculated on 2 × 2 × 2 supercells of the primitive cells using Monkhorst–Pack grids of 3 × 3 × 3. All κ_{L} values were calculated with 8 × 8 × 8 q-point meshes. A 5 Å cutoff for interatomic interactions was chosen to save computational resources. The calculated κ_{L} for $\mathrm{AgGaS_2}$ is 0.82 W m^{-1} K^{-1} at 300 K, which is smaller than the measured value in literature but close to the calculated results in other reports (e.g., 0.88 W m^{-1} K^{-1} by Chu et al.)[13] under similar computation parameters, thus proving the validity of the methods.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 22133004). The authors thank Prof. Ludmila Isaenko and Mr. Linfeng Dong for helpful discussions.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords

balanced infrared nonlinear optical performance, high-throughput screening, machine learning, nonlinear optical chalcogenides, thermal conductivity

Received: September 19, 2023
Revised: October 24, 2023
Published online: December 6, 2023

[1] a) J. Yu, B. Zhang, X. Zhang, Y. Wang, K. Wu, M.-H. Lee, *ACS Appl. Mater. Interfaces* **2020**, *12*, 45023; b) P. Gong, F. Liang, L. Kang, Z. Lin, *Chem. Mat.* **2022**, *34*, 5301; c) F. Liang, L. Kang, Z. Lin, Y. Wu, C. Chen, *Coord. Chem. Rev.* **2017**, *333*, 57; d) L. Kang, M. Zhou, J. Yao, Z. Lin, Y. Wu, C. Chen, *J. Am. Chem. Soc.* **2015**, *137*, 13049; e) P. Gong, F. Liang, L. Kang, X. Chen, J. Qin, Y. Wu, Z. Lin, *Coord. Chem. Rev.* **2019**, *380*, 83; f) S.-P. Guo, Y. Chi, G.-C. Guo, *Coord. Chem. Rev.* **2017**, *335*, 44; g) C. Li, X. Meng, Z. Li, J. Yao, *Coord. Chem. Rev.* **2022**, *453*, 214328; h) H. Chen, M.-Y. Ran, W.-B. Wei, X.-T. Wu, H. Lin, Q.-L. Zhu, *Coord. Chem. Rev.* **2022**, *470*, 214706.
[2] I. Chung, M. G. Kanatzidis, *Chem. Mat.* **2013**, *26*, 849.
[3] a) M. J. Zhang, B. X. Li, B. W. Liu, Y. H. Fan, X. G. Li, H. Y. Zeng, G. C. Guo, *Dalton Trans.* **2013**, *42*, 14223; b) J. A. Brant, D. J. Clark, Y. S. Kim, J. I. Jang, J.-H. Zhang, J. A. Aitken, *Chem. Mat.* **2014**, *26*, 3045; c) K. Wu, Z. Yang, S. Pan, *Inorg. Chem.* **2015**, *54*, 10108; d) G. Li, K. Wu, Q. Liu, Z. Yang, S. Pan, *J. Am. Chem. Soc.* **2016**, *138*, 7422; e) G. Li, Z. Yang, S. Pan, *Sci. China Mater.* **2022**, *66*, 1189; f) H.-D. Yang, M.-Y. Ran, W.-B. Wei, X.-T. Wu, H. Lin, Q.-L. Zhu, *Mater. Today Phys.* **2023**, *35*, 101127.
[4] Z. Lin, X. Jiang, L. Kang, P. Gong, S. Luo, M.-H. Lee, *J. Phys. D: Appl. Phys.* **2014**, *47*, 253001.
[5] a) M. Currie, J. D. Caldwell, F. J. Bezares, J. Robinson, T. Anderson, H. Chun, M. Tadjer, *Appl. Phys. Lett.* **2011**, *99*, 211909; b) A. A. Manenkov, *Opt. Eng.* **2014**, *53*, 010901.
[6] S. G. Sabouri, S. C. Kumar, A. Khorsandi, M. Ebrahim-Zadeh, *IEEE J. Sel. Top. Quantum Electron.* **2014**, *20*, 563.
[7] L. I. Isaenko, A. P. Yelisseyev, *Semicond. Sci. Technol.* **2016**, *31*, 123001.
[8] a) J. Huang, S. Shu, G.-M. Cai, *J. Phys. Chem. C* **2022**, *126*, 7047; b) J. Huang, C. Xie, L. Wei, Q. Bian, Z. Yang, S. Pan, *Chem. Mat.* **2022**, *34*, 10059; c) A. P. Yelisseyev, V. A. Drebushchak, A. S. Titov, L. I. Isaenko, S. I. Lobanov, K. M. Lyapunov, V. A. Gruzdev, S. G. Komarov, V. Petrov, J.-J. Zondy, *J. Appl. Phys.* **2004**, *96*, 3659; d) A. P. Yelisseyev, A. S. Titov, K. M. Lyapunov, V. A. Drebushchak, L. I. Isaenko, S. I. Lobanov, *J. Cryst. Growth* **2005**, *275*, e1679; e) D. Hou, A. S. Nissimagoudar, Q. Bian, K. Wu, S. Pan, W. Li, Z. Yang, *Inorg. Chem.* **2019**, *58*, 93.

[9] H. Chen, W.-B. Wei, H. Lin, X.-T. Wu, *Coord. Chem. Rev.* **2021**, *448*, 214154.

[10] A. G. Jackson, M. C. Ohmer, S. R. Leclair, *Infrared Phys. Technol.* **1997**, *38*, 233.

[11] a) L. Gao, J. Huang, S. Guo, Z. Yang, S. Pan, *Coord. Chem. Rev.* **2020**, *421*, 213379; b) B. Zhang, X. Zhang, J. Yu, Y. Wang, K. Wu, M.-H. Lee, *Chem. Mat.* **2020**, *32*, 6772.

[12] a) J. Yan, P. Gorai, B. Ortiz, S. Miller, S. A. Barnett, T. Mason, V. Stevanovic, E. S. Toberer, *Energy Environ. Sci.* **2015**, *8*, 983; b) P. Gorai, D. Gao, B. Ortiz, S. Miller, S. A. Barnett, T. Mason, Q. Lv, V. Stevanovic, E. S. Toberer, *Comput. Mater. Sci.* **2016**, *112*, 368.

[13] D. Chu, Y. Huang, C. Xie, E. Tikhonov, I. Kruglov, G. Li, S. Pan, Z. Yang, *Angew. Chem., Int. Ed.* **2023**, *62*, e202300581.

[14] G. A. Slack, *Solid State Physics*, Vol. *34*, Academic, New York **1979**.

[15] a) M. Iwayama, S. Wu, C. Liu, R. Yoshida, *J. Chem. Inf. Model.* **2022**, *62*, 4837; b) J. Behler, *J. Chem. Phys.* **2016**, *145*, 170901; c) V. Gupta, K. Choudhary, F. Tavazza, C. Campbell, W.-K. Liao, A. Choudhary, A. Agrawal, *Nat. Commun.* **2021**, *12*, 6595; d) S. Zeng, Y. Zhao, G. Li, R. Wang, X. Wang, J. Ni, *NPJ Comput. Mater.* **2019**, *5*, 84; e) Z. Rao, P.-Y. Tung, R. Xie, Y. Wei, H. Zhang, A. Ferrari, T. P. C. Klaver, F. Körmann, P. T. Sukumar, A. Kwiatkowski Da Silva, Y. Chen, Z. Li, D. Ponge, J. Neugebauer, O. Gutfleisch, S. Bauer, D. Raabe, *Science* **2022**, *378*, 78.

[16] a) W. Cai, A. Abudurusuli, C. Xie, E. Tikhonov, J. Li, S. Pan, Z. Yang, *Adv. Funct. Mater.* **2022**, *32*, 2200231; b) R. Wang, F. Liang, Z. Lin, *Sci. Rep.* **2020**, *10*, 3486; c) M. Wu, E. Tikhonov, A. Tudi, I. Kruglov, X. Hou, C. Xie, S. Pan, Z. Yang, *Adv. Mater.* **2023**, *35*, 2300848; d) Z. Fan, Z. Sun, A. Wang, Y. Yin, G. Jin, C. Xin, *Electron. Mater.* **2023**, *52*, 4162; e) Z.-Y. Zhang, X. Liu, L. Shen, L. Chen, W.-H. Fang, *J. Phys. Chem. C* **2021**, *125*, 25175.

[17] a) J. J. Plata, V. Posligua, A. M. Márquez, J. Fernandez Sanz, R. Grau-Crespo, *Chem. Mat.* **2022**, *34*, 2833; b) T. Zhu, R. He, S. Gong, T. Xie, P. Gorai, K. Nielsch, J. C. Grossman, *Energy Environ. Sci.* **2021**, *14*, 3559; c) G. Qin, Y. Wei, L. Yu, J. Xu, J. Ojih, A. D. Rodriguez, H. Wang, Z. Qin, M. Hu, *J. Mater. Chem. A* **2023**, *11*, 5801; d) Y. Sheng, Y. Wu, J. Yang, W. Lu, P. Villars, W. Zhang, *NPJ Comput. Mater.* **2020**, *6*, 171.

[18] a) X. Qian, J. Zhou, G. Chen, *Nat. Mater.* **2021**, *20*, 1188; b) J. Carrete, W. Li, N. Mingo, S. D. Wang, S. Curtarolo, *Phys. Rev. X* **2014**, *4*, 011019; c) T. Wang, C. Zhang, H. Snoussi, G. Zhang, *Adv. Funct. Mater.* **2019**, *30*, 1906041.

[19] V. G. Dmitriev, G. G. Gurzadyan, D. N. Nikogosyan, *Handbook of Nonlinear Optical Crystals*, Springer-Verlag, Heidelberg **1997**.

[20] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, *1*, 011002.

[21] M. M. Fejer, G. A. Magel, D. H. Jundt, R. L. Byer, *IEEE J. Quantum Electron.* **1992**, *28*, 2631.

[22] a) J. Lee, R. Asahi, *Comput. Mater. Sci.* **2021**, *190*, 110314; b) S. Ju, R. Yoshida, C. Liu, S. Wu, K. Hongo, T. Tadano, J. Shiomi, *Phys. Rev. Mater.* **2021**, *5*, 053801.

[23] T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301.

[24] W. G. Zeier, A. Zevalkink, Z. M. Gibbs, G. Hautier, M. G. Kanatzidis, G. J. Snyder, *Angew. Chem., Int. Ed.* **2016**, *55*, 6826.

[25] E. S. Toberer, A. Zevalkink, G. J. Snyder, *J. Mater. Chem.* **2011**, *21*, 15843.

[26] a) D. T. Morelli, J. P. Heremans, *Appl. Phys. Lett.* **2002**, *81*, 5126; b) T. Zhu, Y. Liu, C. Fu, J. P. Heremans, J. G. Snyder, X. Zhao, *Adv. Mater.* **2017**, *29*, 1605884.

[27] a) J. P. Naskar, S. Hati, D. Datta, *Acta Crystallogr., Sect. B* **1997**, *53*, 885; b) L. Pauling, *J. Am. Chem. Soc.* **1947**, *69*, 542.

[28] J.-K. Lee, S.-M. Choi, W.-S. Seo, Y.-S. Lim, H.-L. Lee, I.-H. Kim, *Renew Energy* **2012**, *42*, 36.

[29] J.-Y. Yang, G. Qin, M. Hu, *Appl. Phys. Lett.* **2016**, *109*, 242103.

[30] G. Qin, Z. Qin, H. Wang, M. Hu, *Phys. Rev. B* **2017**, *95*, 195416.

[31] H. Li, G. Li, K. Wu, B. Zhang, Z. Yang, S. Pan, *Chem. Mat.* **2018**, *30*, 7428.

[32] L. Gao, G. Bian, Y. Yang, B. Zhang, X. Wu, K. Wu, *New J. Chem.* **2021**, *45*, 12362.

[33] Y. Huang, L. Gao, H. Yu, Z. Yang, J. Li, S. Pan, *Chemistry* **2021**, *27*, 6538.

[34] J.-H. Zhang, D. J. Clark, J. A. Brant, K. A. Rosmus, P. Grima, J. W. Lekse, J. I. Jang, J. A. Aitken, *Chem. Mat.* **2020**, *32*, 8947.

[35] G. P. Srivastava, *Pramana* **1974**, *3*, 209.

[36] B. T. Ahn, R. A. Huggins, *Mater. Res. Bull.* **1989**, *24*, 889.

[37] a) B. T. Ahn, R. A. Huggins, *Mater. Res. Bull.* **1990**, *25*, 381; b) B. T. Ahn, R. A. Huggins, *Solid State Ionics* **1991**, *46*, 237; c) W. Huang, N. Matsui, S. Hori, K. Suzuki, M. Hirayama, M. Yonemura, T. Saito, T. Kamiyama, Y. Sasaki, Y. Yoon, S. Kim, R. Kanno, *J. Am. Chem. Soc.* **2022**, *144*, 4989.

[38] S. I. Radautsan, I. M. Tiginyanu, *Jpn. J. Appl. Phys.* **1993**, *32*, 5.

[39] a) T. Ouahrani, R. Khenata, B. Lasri, A. H. Reshak, A. Bouhemadou, S. Bin-Omran, *Physica B: Condens. Matter* **2012**, *407*, 3760; b) A. H. Reshak, S. A. Khan, *J. Alloys Compd.* **2014**, *595*, 125; c) J. Cheng, M.-H. Lee, J. Zhang, *J. Alloys Compd.* **2018**, *768*, 883; d) A. Yalikun, M.-H. Lee, M. Mamat, *RSC Adv.* **2019**, *9*, 41861.

[40] J. O'mara, B. Meredig, K. Michel, *JOM* **2016**, *68*, 2031.

[41] a) 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, E. S. Toberer, *Chem. Mat.* **2017**, *29*, 2494; b) A. Togo, L. Chaput, I. Tanaka, *Phys. Rev. B* **2015**, *91*, 094306; c) C. Toher, J. J. Plata, O. Levy, M. De Jong, M. Asta, M. B. Nardelli, S. Curtarolo, *Phys. Rev. B* **2014**, *90*, 174107.

[42] H. Li, J. Min, Z. Yang, Z. Wang, S. Pan, A. R. Oganov, *Angew. Chem., Int. Ed.* **2021**, *60*, 10791.

[43] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Comput. Mater. Sci.* **2013**, *68*, 314.

[44] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, presented at *KDD'19: Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Anchorage, AK USA, August **2019**.

[45] 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, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.

[46] J. C. Stewart, D. S. Matthew, J. P. Chris, J. H. Phil, I. J. P. Matt, R. Keith, C. P. Mike, *Z. Kristallogr. - Cryst. Mater.* **2005**, *220*, 567.

[47] B. G. Pfrommer, M. Côté, S. G. Louie, M. L. Cohen, *J. Comput. Phys.* **1997**, *131*, 233.

[48] a) D. M. Ceperley, B. J. Alder, *Phys. Rev. Lett.* **1980**, *45*, 566; b) J. P. Perdew, A. Zunger, *Phys. Rev. B* **1981**, *23*, 5048.

[49] A. M. Rappe, K. M. Rabe, E. Kaxiras, J. D. Joannopoulos, *Phys. Rev. B* **1990**, *41*, 1227.

[50] H. J. Monkhorst, J. D. Pack, *Phys. Rev. B* **1976**, *13*, 5188.

[51] J. Lin, M.-H. Lee, Z.-P. Liu, C. Chen, C. J. Pickard, *Phys. Rev. B* **1999**, *60*, 13380.

[52] M.-H. Lee, C.-H. Yang, J.-H. Jan, *Phys. Rev. B* **2004**, *70*, 235110.

[53] D. M. Bylander, L. Kleinman, *Phys. Rev. B* **1990**, *41*, 7868.

[54] R. W. Godby, M. Schlüter, L. J. Sham, *Phys. Rev. B* **1988**, *37*, 10159.

[55] S. Baroni, S. De Gironcoli, A. Dal Corso, P. Giannozzi, *Rev. Mod. Phys.* **2001**, *73*, 515.

[56] T. Ouyang, M. Hu, *Phys. Rev. B* **2015**, *92*, 235204.

[57] A. Togo, *J. Phys. Soc. Jpn.* **2023**, *92*, 012001.