ARTICLE OPEN

High-throughput prediction of the carrier relaxation time via data-driven descriptor

Zizhen Zhou^{1,2}, Guohua Cao^{1,3}, Jianghui Liu¹ and Huijun Liu¹✉

It has been demonstrated that many promising thermoelectric materials, such as tetradymite compounds are also three-dimensional topological insulators. In both cases, a fundamental question is the evaluation of carrier relaxation time, which is usually a rough task due to the complicated scattering mechanisms. Previous works using the simple deformation potential theory or considering complete electron-phonon coupling are, however, restricted to small systems. By adopting a data-driven method named SISSO (Sure Independence Screening and Sparsifying Operator) with the training data obtained via deformation potential theory, we propose an efficient and physically interpretable descriptor to evaluate the relaxation time, using tetradymites as prototypical examples. Without any input from first-principles calculations, the descriptor contains only several elemental properties of the constituent atoms, and could be utilized to quickly and reliably predict the carrier relaxation time of a substantial number of tetradymites with arbitrary stoichiometry.

npj Computational Materials (2020)6:149; <https://doi.org/10.1038/s41524-020-00417-0>

INTRODUCTION

Substantial advances have been made in screening or discovering functional materials via high-throughput computational method, which involves physics, statistics, computer science, and artificial intelligence¹. Over the last decade, such an approach has been extensively applied in materials science, such as identifying potential photovoltaic absorbers², predicting the stability of compounds³, and screening topological insulators (TIs)^{4–6}. Among various applications, the search for good thermoelectric (TE) materials^{7–9} is a particularly intriguing one due to the increasingly serious energy crisis and environmental pollution. Presently, there have been a few works focusing on the high-throughput prediction of compounds with low thermal conductivity^{10–12}, which is essential for high TE conversion efficiency. Another critical transport coefficient closely related to the TE performance is the so-called power factor⁷, where a troublesome issue is suitable treatment of the relaxation time owing to the complex scattering mechanisms involved.

Generally speaking, the electron-acoustic phonon interaction is the dominant scattering mechanism in the operational temperature range of TE materials¹³. Within the effective mass approximation, Bardeen and Shockley¹⁴ derived an analytical method named deformation potential (DP) theory for the relaxation time around the band edges. Furthermore, the energy dependent *k*-resolved relaxation time can be obtained by a complete electron-phonon coupling (EPC) calculation^{15,16} based on the Wannier interpolation techniques¹⁷. However, both approaches involve quite complex first-principles calculations and huge computational efforts are needed. To date, it is still a challenging task to evaluate the relaxation time, which even becomes prohibitive for system with large unit cell.

In this work, a high-throughput investigation combined with the compressed sensing approach^{18,19} is employed to quickly predict the carrier relaxation time of the tetradymite compounds.

Using first-principles calculations, we first compute the electronic band structures of a small set of tetradymites with integer stoichiometry. For simplicity, we use the DP theory to calculate the corresponding relaxation times (τ_{cal}) considering the dominant electron-acoustic phonon scattering. It should be mentioned that many compounds in the tetradymite family were demonstrated to be TIs²⁰, where the electronic band edges exhibit distinct energy dispersions and electron-phonon transition probabilities compared with those of normal insulators (NIs)^{21,22}. We thus divide the calculated τ_{cal} into two training sets according to their different topological natures. For each set, we then adopt the compressed sensing method to construct a physically intuitive descriptor for the prediction of the relaxation time (τ_{pre}). We shall see that both NIs and TIs exhibit good quantitative agreement between the τ_{cal} and τ_{pre} , with Pearson correlation coefficient higher than 90%. Importantly, generalizing the proposed descriptors allows us to rapidly predict the relaxation time of a vast number of tetradymites with fractional stoichiometry, and selective cross-checks find good agreement between our predictions and explicit first-principles calculations.

RESULTS

Reliable training data

The candidate tetradymite compounds investigated in this work are constructed by combining group-VA elements (As, Sb, and Bi) with group-VIA elements (S, Se, and Te), which respectively occupy the A/B and C/D/E sites in the rhombohedral primitive cell shown in Fig. 1a. The hexagonal unit cell is plotted in Fig. 1b, where the covalently bonded quintuple layers (QLs) are held together via van der Waals (vdW) interactions. Among a total of 243 obtained tetradymites with integer stoichiometry, 85 NIs and 67 TIs were predicted to be mechanically stable²⁰. By performing accurate first-principles calculations, the carrier relaxation time of

¹Key Laboratory of Artificial Micro- and Nano-Structures of Ministry of Education and School of Physics and Technology, Wuhan University, Wuhan 430072, China. ²Center for Quantum Materials and Devices, College of Physics, Chongqing University, Chongqing 401331, China. ³International Center for Quantum Design of Functional Materials (ICQD), Hefei National Laboratory for Physical Sciences at the Microscale, and Synergetic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China. ✉email: phljh@whu.edu.cn

the 152 tetradymites can be obtained. Compared with those of NIs, there are obvious band inversion around the Fermi level of TIs, which can lead to distinct effective mass and electron-phonon scattering phase space²³. It is thus necessary to separate the calculated relaxation times into NIs and TIs parts, which are respectively summarized in Supplementary Tables 1 and 2. As nearly all the tetradymite compounds exhibit better TE performance with hole doping^{24–26}, we focus on the *p*-type relaxation time at room temperature in the present work.

Data-driven descriptors with strong predictive power

By performing SISSO training (see details in the “Methods”) for the τ_{cal} of 85 NIs, an optimized descriptor is obtained for predicting the relaxation time:

$$\tau_{\text{pre}} = 7.8 \times 10^{-3} \times [(m_A + m_B + m_E - m_{\text{ave}}) \sqrt{m_C m_D}] - 0.6 \times \frac{(\chi_{\text{ave}} - \chi_C)(\chi_{\text{ave}} - \chi_D)(m_C - m_D)}{(\chi_{\text{ave}} - \chi_E)} - 52.9 \quad (1)$$

Here m is the atomic mass and χ denotes the Pauling electronegativity of the constituent elements. The subscripts A/B/C/D/E and “ave” indicate the occupied site and the average value, respectively. For the case of TIs, the best descriptor is given by:

$$\tau_{\text{pre}} = 79.0 \times [|(\ln r_{\text{mse}})(\chi_A - \chi_B)| - \frac{\chi_E(r_C - r_D)}{r_E}] - 3.0 \times 10^3 \times \left[\left(\frac{m_{\text{mse}}}{\chi_{\text{ave}}} - \frac{m_{\text{ave}}}{\chi_{\text{mse}}} \right) (r_{\text{mse}}^3 |\chi_{\text{ave}} - \chi_E|) \right] + 20.3 \quad (2)$$

where the extra feature r indicates the radius of the *p*-orbital and the subscript “mse” refers to the mean square error. Figure 2a, b show the quantitative comparison of the τ_{pre} and τ_{cal} for the 85 NIs

Fig. 1 The crystal structure of tetradymite compounds. **a** Primitive cell. **b** Unit cell. Note that the A/B and C/D/E sites can be occupied by the As/Sb/Bi and S/Se/Te atoms, respectively.

Fig. 2 The intuitive linear correlation between the τ_{pre} and τ_{cal} . **a** 85 NIs. **b** 67 TIs. The insets in **a** and **b** are the orbital-decomposed band structures of the representative NI (Sb_2S_3) and TI (Bi_2Te_3), respectively.

and 67 TIs, respectively. It is clear that the relaxation times predicated by the optimized descriptors agree very well with those obtained from first-principles calculations. The Pearson correlation coefficient is as high as 94% (92%) for the NIs (TIs), corroborating the strong predictive power of the obtained descriptors. Besides, it is quite necessary to compare the calculated and predicted carrier relaxation time with the experimental results. However, very few works reported the direct measurement of relaxation time, which can be alternatively derived by fitting the experimental transport properties such as the electrical conductivity or carrier mobility. In Supplementary Fig. 1, the τ_{cal} and τ_{pre} of five tetradymites in the training data are compared with those obtained experimentally (τ_{exp})^{27–30}. We see that τ_{cal} and τ_{pre} exhibit good quantitative agreement, but are somewhat larger than τ_{exp} . This is reasonable since τ_{cal} in the training data are calculated around the band edges, where the carrier concentrations are relatively lower (10^{17} cm^{-3} – 10^{18} cm^{-3}). In contrast, the five τ_{exp} obtained from limited experimental works correspond to higher carrier concentrations (10^{18} cm^{-3} – 10^{20} cm^{-3}) with optimized TE performance, which lead to enhanced scattering and thus smaller relaxation time. Besides, the impurity and grain boundary scattering in the experiments could also reduce the relaxation time. On the other hand, we find that τ_{cal} and τ_{pre} exhibit almost the same behavior with τ_{exp} for the five investigated tetradymites, which further confirms the reliability of our training data and the strong predictive power of the proposed descriptors. Since it is a delicate combination of the elemental properties of the constituent atoms, the descriptor allows us to pinpoint the determinant factors governing the value of relaxation time, as further demonstrated below.

With an intensive analysis of Eq. (1), it is found that the first term (mass term) dominate the predicted relaxation time of NIs. In particular, we see from Fig. 3 that the average atomic mass (m_{ave}) exhibits the strongest correlation with the predicted relaxation time. To have a better understanding, the mass term is rewritten as:

$$7.8 \times 10^{-3} \times [(4m_{\text{ave}} - m_C - m_D) \sqrt{m_C m_D}] \quad (3)$$

It is obvious that the relaxation time of NIs increases with m_{ave} . As evidenced by our additional first-principles calculations, we find that the tetradymites containing light elements usually exhibit relatively flat band structures around the Fermi level, which correspond to bigger effective mass (m^*) and thus lower relaxation time ($\tau_{\text{cal}} \propto m^{*-3/2}$). Indeed, we find that the hole effective mass of the 85 NIs decrease obviously with increasing average atomic mass, as shown in Supplementary Fig. 2a. Such a fact can be understood by a simple tight-binding analysis, where the effective mass is believed to be inversely related to the lattice constant (or bond length). As heavier mass of the constituent

Fig. 3 The correlation between τ_{pre} and input features. The vertical axis gives the absolute values of the Pearson correlation coefficients and the horizontal axis indicates the individual atomic properties of both NIs and TIs.

atom usually corresponds to larger orbital radius and thus longer bond length, the inverse relation between m_{ave} and m^* can be qualitatively convinced, which strongly confirms the delicate physical insight of the descriptor. In comparison, the second term of Eq. (1) is less important, which can be rewritten as:

$$-0.6 \times \frac{(\chi_C - \chi_{\text{ave}})(\chi_D - \chi_{\text{ave}})(m_C - m_D)}{(\chi_{\text{ave}} - \chi_E)} \quad (4)$$

According to previous work²⁰, the χ_C and χ_D of NIs are usually higher than χ_{ave} while the χ_E is relatively lower. For the NIs with different elements in the C and D sites, we thus find lower relaxation time in systems with larger χ_C and χ_D . The physical origin is that larger Pauling electronegativity of the anions usually have smaller atomic radii, which leads to smaller lattice constants. According to a simple tight-binding analysis, this would result in larger effective mass, and consequently give rise to lower relaxation time.

For the case of TIs, we see from Fig. 3 that the relaxation time is mostly related to the r_{mse} . Owing to the band inversions induced by the SOC, the valance band edges of the TIs are mainly contributed by the p -orbitals of both the cations and anions³¹, which greatly affect the energy dispersions and the hole effective mass. As r_{mse} is less than 1, we find from the first term of Eq. (2) that lower relaxation time is obtained in the topological tetradymites with larger r_{mse} . This is consistent with the fact that the effective mass (m^*) of the 67 TIs in the training data is roughly inversely related to the $|\ln r_{\text{mse}}|$, as shown in Supplementary Fig. 2b. This effect becomes pronounced when separately comparing the results of $\text{Bi}_2\text{Se}_2\text{S}$ with other TIs, where the former one exhibits the highest r_{mse} corresponding to the largest effective mass ($2.3 m_e$) and the lowest relaxation time (6.8 fs). Such an observation is consistent with the established understanding that bigger r_{mse} means larger discrepancy between the p -orbitals of the anions and cations, causing weaker orbital hybridization and thus narrower band width near the Fermi level. Consequently, larger effective mass and lower relaxation time can be expected. On the other hand, we find that the effect of the second term of Eq. (2) is somewhat hidden due to the ultrasmall value of $r_{\text{mse}}^3 |\chi_{\text{ave}} - \chi_E|$ (less than 10^{-4}). For instance, we see that the m_{ave} in the second term is positively related to the predicted relaxation time. However, the corresponding Pearson correlation coefficient is very small as shown in Fig. 3. Of course, the second term cannot be ignored, since the χ_{mse} is also strongly related to the predicted relaxation time as shown in Fig. 3. In principle, the TIs with larger χ_{mse} tend to be polar covalent or even ionic compounds. Hence, the valance electrons are largely localized which lead to relatively flat bands with bigger effective mass and thus lower relaxation time. We see from the second term of Eq. (2) that the τ_{pre} is also positively related to the m_{ave} . Compared with that of the NIs, the

descriptor of the TIs also correctly captures the essential physical factor (m^*) governing the relaxation time, whereas the delicate mechanism is quite different suggesting the necessity in classifying the investigated tetradymites as mentioned above.

Predictions beyond the training data

Beyond the above-mentioned 152 tetradymites with integer stoichiometry, there are more doped systems with fractional stoichiometry used for favorable TE applications^{24–26}. In practice, doping at various concentrations is a typical approach to enhance the TE performance by achieving the band convergence³² or reducing the lattice thermal conductivity³³. Unfortunately, it is quite difficult or even impossible to obtain the carrier relaxation time of these systems using first-principles calculations since prohibitively large supercells are needed to model the structures. Such a problem can be effectively solved by utilizing the optimized descriptors given by Eqs. (1) and (2). In general, the doped tetradymites can be represented by a nominal formula of $\text{As}_x\text{Sb}_y\text{Bi}_{2-x-y}\text{S}_{x'}\text{Se}_{y'}\text{Te}_{3-x'-y'}$ ($0 \leq x, y, x+y \leq 2$ and $0 \leq x', y', x'+y' \leq 3$). According to the virtual crystal approximation³⁴, the site-specific properties of the randomly doped systems should be weighted averages for first-principles modeling. For any tetradymites, the site-specific atomic mass can be defined as:

$$\begin{aligned} m_A &= x_1 m_{\text{As}} + y_1 m_{\text{Sb}} + (1 - x_1 - y_1) m_{\text{Bi}} \\ m_B &= x_2 m_{\text{As}} + y_2 m_{\text{Sb}} + (1 - x_2 - y_2) m_{\text{Bi}} \\ m_C &= x_3 m_{\text{S}} + y_3 m_{\text{Se}} + (1 - x_3 - y_3) m_{\text{Te}} \\ m_D &= x_4 m_{\text{S}} + y_4 m_{\text{Se}} + (1 - x_4 - y_4) m_{\text{Te}} \\ m_E &= x_5 m_{\text{S}} + y_5 m_{\text{Se}} + (1 - x_5 - y_5) m_{\text{Te}} \end{aligned} \quad (5)$$

Here the coefficients ($0 \leq x_i, y_i \leq 1$, $1 \leq i \leq 5$) refer to the fractional occupancy of each site, which define the stoichiometry via $x = x_1 + x_2$, $y = y_1 + y_2$, $x' = x_3 + x_4 + x_5$, and $y' = y_3 + y_4 + y_5$. With the site-specific Pauling electronegativity and the site-specific radius of the p -orbital given by the same way, our data-driven descriptors can thus be used to rapidly predict the relaxation time of any tetradymites with arbitrary stoichiometry. As an example, we list the descriptor predicted relaxation time of 13 NIs and 13 TIs with fractional stoichiometry in Supplementary Table 3. Note that all of them have medium supercells and thus first-principles calculations are still reasonably tractable. We see that τ_{pre} shows good agreement with first-principles results for both NIs and TIs. Indeed, as evidenced from Supplementary Fig. 3a, b, the Pearson correlation coefficient between the τ_{pre} and τ_{cal} of the 13 NIs (TIs) are 96% (93%). Moreover, we compare in Supplementary Fig. 4 the τ_{pre} and τ_{exp} for five additional tetradymites with fractional stoichiometry^{27,29,35}, where they exhibit the same behaviors as discussed above for the systems with integer stoichiometry. Once again, such an observation demonstrates that the optimized descriptors are robust and reliable for evaluating the relaxation time of the tetradymites with either integer or fractional stoichiometry. It should be emphasized that we focus on the relaxation time around the band edge of tetradymites and the carrier concentration is usually in the range of 10^{17} cm^{-3} – 10^{18} cm^{-3} , which is lower than those found at optimized TE performance as mentioned above. Since the relaxation time usually decreases with increasing carrier concentration, the τ_{pre} is found to be smaller than the τ_{exp} , which is similar to those found for systems with integer stoichiometry. Such an observation also indicates that our descriptors may not be suitable for predicting the relaxation time of heavily doped systems.

By varying each stoichiometry (x_i and y_i) at an interval of 1/6, we can obtain over sixteen million different tetradymites and quickly predict their carrier relaxation time using the generalized data-driven descriptors. Figure 4a, b show the distributions of the room temperature τ_{pre} for over eight million NIs and eight million TIs,

Fig. 4 Distribution of sixteen million tetradymite compounds with respect to their values of relaxation time. **a** NIs. **b** TIs.

respectively. It is found that most NIs have the relaxation time in the range of 60–70 fs, whereas the corresponding region for the TIs is 30–40 fs. Remarkably, we see from Fig. 4a that more than 2×10^5 NIs exhibit large τ_{pre} exceeding 120 fs, which are very desirable for achieving promising TE performance²³. On the other hand, Liang et al.³⁶ theoretically suggested that the decreased relaxation time of the bulk state can lead to enhanced TE performance for both *p*- and *n*-type systems in the TI films of tetradymites. Interestingly, we see from Fig. 4b that $\sim 5 \times 10^5$ TIs have small values of τ_{pre} (0–10 fs). For instance, the SISSO prediction indicates that the topological $\text{Sb}_{0.167}\text{Bi}_{1.833}\text{S}_{0.5}\text{Te}_{2.5}$ exhibits very small bulk relaxation time of 0.55 fs. As the surface state exhibit very larger relaxation time of 550 fs³⁷, a large relaxation time ratio of 1000 between the surface and bulk states can be expected, which can lead to a considerable *ZT* value of ~ 2.0 for both *p*- and *n*-type systems. In particular, more than 20,000 TIs exhibit ultralow τ_{pre} less than 0.5 fs, such as $\text{Bi}_2\text{S}_{1.667}\text{Se}_{0.5}\text{Te}_{0.833}$ (0.3 fs), $\text{Sb}_{0.5}\text{Bi}_{1.5}\text{S}_{1.5}\text{Se}_{0.833}\text{Te}_{0.667}$ (0.2 fs), and $\text{BiSbS}_{2.167}\text{Te}_{0.833}$ (0.2 fs). Such a result offers a good opportunity in the exploration of the topological tetradymites for superior thermoelectric applications.

DISCUSSION

In summary, by employing the data-driven SISSO approach, we report two simple descriptors for predicting the carrier relaxation time of NIs and TIs in the tetradymite family, where the training sets are obtained from accurate first-principles calculations. Unlike those from the DP theory or EPC calculations which is forbidden for systems with large unit cell, our descriptors enable an effective prediction for the relaxation time of the tetradymite compounds with either integer or fractional stoichiometry at negligible computational cost. The robust predictive power originates from the fact that the descriptors could perfectly capture the essential physical mechanism governing the value of relaxation time. Additionally, such physically meaningful descriptors also clarify how to efficiently tune the value of the relaxation time so that maximum TE performance can be realized in both NIs and TIs of the tetradymite family. It should be also emphasized that the obtained descriptors contain only the intrinsic properties of the constituent elements, which can be essentially adapted to fast predict the carrier relaxation time of many other TE materials such as the half-Heusler³⁸ and rock-salt-type³⁹ compounds.

METHODS

Accurate first-principles calculations

The electronic band structures of these 152 tetradymites were calculated within the framework of density functional theory^{40,41} using the Vienna ab-initio simulation package⁴². The Perdew–Burke–Ernzerhof functional with

Table 1. The atomic mass (*m*, relative to ^{12}C), the radius of the *p*-orbital (*r*, in unit of Å), and the Pauling electronegativity (χ , in unit of eV) of the group-VA elements (As, Sb and Bi) and group-VIA elements (S, Se, and Te).

Elements	<i>m</i>	<i>r</i>	χ
As	74.92	1.05	2.18
Sb	121.80	1.24	2.05
Bi	208.98	1.30	2.02
S	32.07	0.85	2.58
Se	78.96	0.96	2.55
Te	127.60	1.15	2.12

generalized gradient approximation⁴³ was adopted, and the spin-orbit coupling was explicitly included in the calculations. The van der Waals interaction between adjacent quintuple layers was taken into account by means of the optB86b-vdW functional⁴⁴. We adopted a plane-wave cutoff energy of 550 eV and a $13 \times 13 \times 13$ Monkhorst-Pack *k*-mesh in the Brillouin zone. The energy convergence threshold was 10^{-6} eV and the relaxed structure was determined until the magnitude of the force acting on each atom is less than $0.01 \text{ eV } \text{\AA}^{-1}$. By considering the dominant acoustic phonon scattering, we used the DP theory¹⁴ to evaluate the relaxation time, which can be expressed as

$$\tau_{\text{cal}} = \frac{2\sqrt{2}\pi\hbar^4}{3(k_B T m^*)^{3/2} E^2} \quad (6)$$

Here *C* is the elastic modulus, $\hbar$ is the reduced Planck constant, k_B is the Boltzmann constant, *T* is the absolute temperature, and *E* is the DP constant represents the shift of band edge per unit strain. In addition to the case of optimized structure, four reference calculations with lattice deformation of $\pm 1\%$ and $\pm 2\%$ were performed to evaluate *C* and *E*. Such a method has been successfully used to predict the relaxation time of many known materials^{45–47}.

SISSO training based on the compressed sensing technique

Using the calculated τ_{cal} of both NIs and TIs as two independent training sets, we could obtain the simple and physically meaningful descriptors based on the SISSO method^{48,49}. In our case, three elemental properties of the constituent atoms (the atomic mass, the radius of the *p*-orbital, and the Pauling electronegativity, see the details in Table 1) as well as their average value and mean square error were considered as input features. It should be mentioned that the band edges of tetradymites are dominated by the *p*-orbitals of both the cations and anions, which is one of the crucial factors in determining the relaxation time. The intrinsically linear relationship observables in the compressed-sensing formalism was made nonlinear by equipping the features with nonlinear operators $H \equiv \{I, +, -, \times, /, \exp, \log, | - |, \sqrt{\cdot}, ^{-1}, ^2, ^3\}$. At each iteration, *H* operates on all available combinations, and more than 10^{10} features were constructed up to a complexity cutoff of 3. The sure independence screening (SIS) scores each feature with a correlation magnitude and keeps only the top ranked. The

subset extracted by the SIS algorithm was set to 80,000. After the reduction, the sparsifying operators (SO) was used to pinpoint the optimal descriptor. In contrast to previous machine learning methods^{50,51}, SISSO can identify descriptors and capture the underlying mechanisms of physical properties, which is highly desirable for the accelerative discovery of functional materials. Meanwhile, SISSO shows superior advantages compared with other established method which suffers with large and highly correlated features spaces. In particular, SISSO is built to work even when only relatively small training sets are available. As a step forward with respect to previous algorithms, SISSO is an effective tool for high-throughput investigation in materials science.

DATA AVAILABILITY

All relevant data are included in the paper and/or its Supplementary Information files.

CODE AVAILABILITY

The calculations were implemented using the VASP and SISSO packages.

Received: 26 March 2020; Accepted: 16 September 2020;

Published online: 08 October 2020

REFERENCES

1. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
2. Castelli, I. E. et al. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy Environ. Sci.* **5**, 5814–5819 (2012).
3. Bartel, C. J. et al. New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **5**, eaav0693 (2019).
4. Yang, K., Setyawan, W., Wang, S., Nardelli, M. B. & Curtarolo, S. A search model for topological insulators with high-throughput robustness descriptors. *Nat. Mater.* **11**, 614–619 (2012).
5. Zhang, P. F., Shen, H. T. & Zhai, H. Machine learning topological invariants with neural networks. *Phys. Rev. Lett.* **120**, 066401 (2018).
6. Tang, F., Po, H. C., Vishwanath, A. & Wan, X. G. Comprehensive search for topological materials using symmetry indicators. *Nature* **566**, 486–489 (2019).
7. Yang, J. et al. On the tuning of electrical and thermal transport in thermoelectrics: an integrated theory-experiment perspective. *npj Comput. Mater.* **2**, 15015 (2016).
8. Gorai, P., Stevanović, V. & Toberer, E. S. Computationally guided discovery of thermoelectric materials. *Nat. Rev. Mater.* **2**, 17053 (2017).
9. Hao, S. Q. et al. Design strategy for high-performance thermoelectric materials: the prediction of electron-doped KZrCuSe₃. *Chem. Mater.* **31**, 3018–3024 (2019).
10. Carrete, J., Li, W., Mingo, N., Wang, S. D. & Curtarolo, S. Finding unprecedentedly low-thermal-conductivity half-Heusler semiconductors via high-throughput materials modeling. *Phys. Rev. X* **4**, 011019 (2014).
11. Seko, A. et al. Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization. *Phys. Rev. Lett.* **115**, 205901 (2015).
12. Roekeghem, A. V., Carrete, J., Oses, C., Curtarolo, S. & Mingo, N. High-throughput computation of thermal conductivity of high-temperature solid phases: the case of oxide and fluoride perovskites. *Phys. Rev. X* **6**, 041061 (2016).
13. Xi, J. Y., Long, M. Q., Tang, L., Wang, D. & Shuai, Z. G. First-principles prediction of charge mobility in carbon and organic nanomaterials. *Nanoscale* **4**, 4348–4369 (2012).
14. Bardeen, J. & Shockley, W. Deformation potentials and mobilities in non-polar crystals. *Phys. Rev.* **80**, 72–80 (1950).
15. Noffsinger, J. et al. EPW: a program for calculating the electron-phonon coupling using maximally localized wannier functions. *Comput. Phys. Commun.* **181**, 2140–2148 (2010).
16. Liu, T. H., Zhou, J. W., Liao, B. L., Singh, D. J. & Chen, G. First-principles mode-by-mode analysis for electron-phonon scattering channels and mean free path spectra in GaAs. *Phys. Rev. B* **95**, 075206 (2017).
17. Giustino, F., Cohen, M. L. & Louie, S. G. Electron-phonon interaction using wannier functions. *Phys. Rev. B* **76**, 165108 (2007).
18. Donoho, D. L. Compressed sensing. *IEEE Trans. Inform. Theory* **52**, 1289–1306 (2006).
19. Nelson, L. J., Hart, G. L. W., Zhou, F. & Ozoliņš, V. Compressive sensing as a paradigm for building physics models. *Phys. Rev. B* **87**, 035125 (2013).

20. Cao, G. H. et al. Artificial intelligence for high-throughput discovery of topological insulators: the example of alloyed tetradymites. *Phys. Rev. Mater.* **4**, 034204 (2020).
21. Zhang, H. J. et al. Topological insulators in Bi₂Se₃, Bi₂Te₃ and Sb₂Te₃ with a single dirac cone on the surface. *Nat. Phys.* **5**, 438–442 (2009).
22. Shi, H. L., Parker, D., Du, M.-H. & Singh, D. J. Connecting thermoelectric performance and topological-insulator behavior: Bi₂Te₃ and Bi₂Te₂Se from first principles. *Phys. Rev. Appl.* **3**, 014004 (2015).
23. Park, J., Xia, Y. & Ozoliņš, V. High thermoelectric power factor and efficiency from a highly dispersive band in Ba₂BiAu. *Phys. Rev. Appl.* **11**, 014058 (2019).
24. Poudel, B. et al. High-thermoelectric performance of nanostructured bismuth antimony telluride bulk alloys. *Science* **320**, 634 (2008).
25. Ma, Y. et al. Enhanced thermoelectric figure-of-merit in p-type nanostructured bismuth antimony tellurium alloys made from elemental chunks. *Nano Lett.* **8**, 2580–2584 (2008).
26. Lan, Y. C. et al. Structure study of bulk nanograin thermoelectric bismuth antimony telluride. *Nano Lett.* **9**, 1419–1422 (2009).
27. Grauer, D. C., Hor, Y. S., Williams, A. J. & Cava, R. J. Thermoelectric properties of the tetradymite-type Bi₂Te₂S-Sb₂Te₂S solid solution. *Mater. Res. Bull.* **44**, 1926–1929 (2009).
28. Aminoroaya-Yamini, S., Zhang, C., Wang, X. L. & Nevirkovets, I. Crystal structure, electronic structure and thermoelectric properties of n-type BiSbTe₂. *J. Phys. D: Appl. Phys.* **45**, 125301 (2012).
29. Chasapis, T. C. et al. Two-band model interpretation of the *p*- to *n*-transition in ternary tetradymite topological insulators. *APL Mater.* **3**, 083601 (2015).
30. Morikawa, S., Inamoto, T. & Takashiri, M. Thermoelectric properties of nanocrystalline Sb₂Te₃ thin films: experimental evaluation and first-principles calculation, addressing effect of crystal grain size. *Nanotechnology* **29**, 075701 (2018).
31. Cao, G. H. et al. Rhombohedral Sb₂Se₃ as an intrinsic topological insulator due to strong van der Waals interlayer coupling. *Phys. Rev. B* **97**, 075147 (2018).
32. Pei, Y. Z. et al. Convergence of electronic bands for high performance bulk thermoelectrics. *Nature* **473**, 66–69 (2011).
33. Li, J. Q. et al. Extremely low thermal conductivity in thermoelectric Ge_{0.55}Pb_{0.45}Te solid solutions via Se substitution. *Chem. Mater.* **28**, 6367–6373 (2016).
34. Nordheim, L. Zur elektronentheorie der metalle. *I. Ann. Phys. Leipz.* **401**, 607–640 (1931).
35. Wu, B. et al. Oscillating planar Hall response in bulk crystal of topological insulator Sn doped Bi_{1.1}Sb_{0.9}Te₂S. *Appl. Phys. Lett.* **113**, 011902 (2018).
36. Liang, J. H., Cheng, L., Zhang, J., Liu, H. J. & Zhang, Z. Y. Maximizing the thermoelectric performance of topological insulator Bi₂Te₃ films in the few-quintuple layer regime. *Nanoscale* **8**, 8855–8862 (2016).
37. Qu, D.-X., Hor, Y. S., Xiong, J., Cava, R. J. & Ong, N. P. Quantum oscillations and hall anomaly of surface states in the topological insulator Bi₂Te₃. *Science* **329**, 821–824 (2010).
38. Uher, C., Yang, J., Hu, S., Morelli, D. T. & Meisner, G. P. Transport properties of pure and doped Mn₂Sn (M=Zr, Hf). *Phys. Rev. B* **59**, 8615–8621 (1999).
39. Hong, M., Chen, Z.-G., Pei, Y. Z., Yang, L. & Zou, J. Limit of *zT* enhancement in rocksalt structured chalcogenides by band convergence. *Phys. Rev. B* **94**, 161201 (R) (2016).
40. Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
41. Kresse, G. & Hafner, J. *Ab initio* molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).
42. Kresse, G. & Furthmüller, J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
43. Perdew, J. P. & Burke, K. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
44. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
45. Long, M. Q., Tang, L., Wang, D., Wang, L. J. & Shuai, Z. G. Theoretical predictions of size-dependent carrier mobility and polarity in graphene. *J. Am. Chem. Soc.* **131**, 17728–17729 (2009).
46. Long, M. Q., Tang, L., Wang, D., Li, Y. L. & Shuai, Z. G. Electronic structure and carrier mobility in graphdiyne sheet and nanoribbons: theoretical predictions. *ACS Nano* **5**, 2593–2600 (2011).
47. Chen, J. M., Xi, J. Y., Wang, D. & Shuai, Z. G. Carrier mobility in graphyne should be even larger than that in graphene: a theoretical prediction. *J. Phys. Chem. Lett.* **4**, 1443–1448 (2013).
48. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big data of materials science: critical role of the descriptor. *Phys. Rev. Lett.* **114**, 105503 (2015).

49. Ouyang, R. H., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiringhelli, L. M. SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2**, 083802 (2018).
50. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **3**, 54 (2017).
51. Bassman, L. et al. Active learning for accelerated design of layered materials. *npj Comput. Mater.* **4**, 74 (2018).

ACKNOWLEDGEMENTS

We thank financial support from the National Natural Science Foundation (Grant Nos. 51772220 and 62074114). The numerical calculations in this work have been done on the platform in the Supercomputing Center of Wuhan University.

AUTHOR CONTRIBUTIONS

H.J.L. initiated and supervised the project. Z.Z.Z. performed the first-principles and SISSO calculations. G.H.C. and J.H.L. analyzed the data. Z.Z.Z. and H.J.L. prepared the manuscript. All authors contributed to the interpretation of the results.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41524-020-00417-0>.

Correspondence and requests for materials should be addressed to H.L.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, 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. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020