Science and Technology of Advanced Materials

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsta20

Data-driven analysis of electron relaxation times in PbTe-type thermoelectric materials

Yukari Katsura, Masaya Kumagai, Takushi Kodani, Mitsunori Kaneshige, Yuki Ando, Sakiko Gunji, Yoji Imai, Hideyasu Ouchi, Kazuki Tobita, Kaoru Kimura & Koji Tsuda

To cite this article: Yukari Katsura, Masaya Kumagai, Takushi Kodani, Mitsunori Kaneshige, Yuki Ando, Sakiko Gunji, Yoji Imai, Hideyasu Ouchi, Kazuki Tobita, Kaoru Kimura & Koji Tsuda (2019) Data-driven analysis of electron relaxation times in PbTe-type thermoelectric materials, Science and Technology of Advanced Materials, 20:1, 511-520, DOI: 10.1080/14686996.2019.1603885

To link to this article: https://doi.org/10.1080/14686996.2019.1603885

Published online: 04 Jun 2019.

SCIENCE AND TECHNOLOGY OF ADVANCED MATERIALS
2019, VOL. 20, NO. 1, 511–520
https://doi.org/10.1080/14686996.2019.1603885

Data-driven analysis of electron relaxation times in PbTe-type thermoelectric materials

Yukari Katsura[a,b], Masaya Kumagai[c,d], Takushi Kodani[a,b], Mitsunori Kaneshige[e], Yuki Ando[b], Sakiko Gunji[a,b], Yoji Imai[b,c], Hideyasu Ouchi[a,b,c], Kazuki Tobita[a,b], Kaoru Kimura[a] and Koji Tsuda[a,b,c]

[a]Department of Advanced Materials Science, Graduate School of Frontier Sciences, SAKURA Research Center, The University of Tokyo, Chiba, Japan;
[b]Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki, Japan;
[c]Molecular Informatics Team, Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan;
[d]SAKURA Research Center, SAKURA Internet Inc., Osaka, Japan;
[e]Scientific Computing Division, X-Ability Co. Ltd, Tokyo, Japan

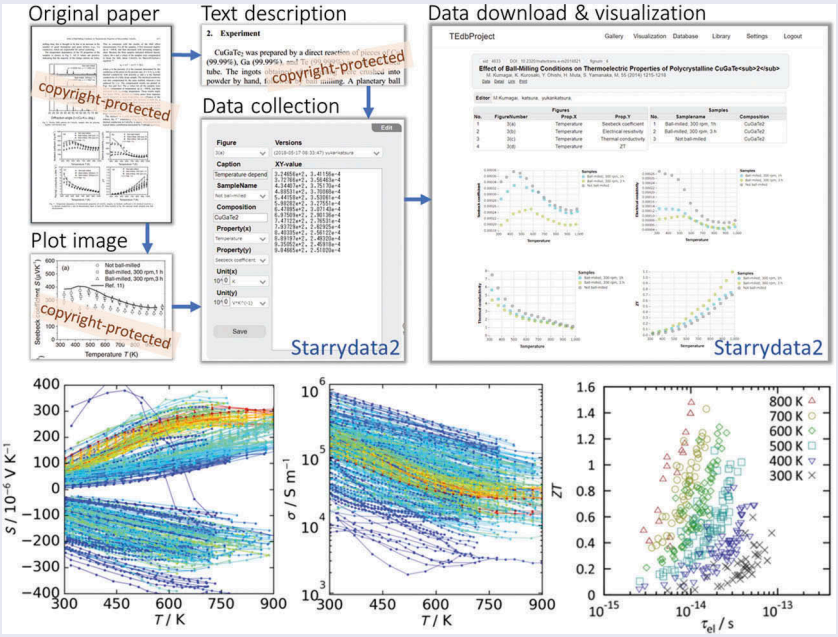
ABSTRACT

Data mining from published papers can generate large experimental datasets that have been overlooked in computational materials informatics. We developed an open web system *Starrydata2* to accelerate a comprehensive digitization of data of materials from as-reported plot images in published papers, without sample selection based on performance. By plotting results obtained from our dataset on experimental thermoelectric properties of 434 samples of rock-salt-type (PbTe-type) thermoelectric materials, we revealed differences in electronic structure of parent compounds PbTe, PbSe, PbS, and SnTe from just experimental data. We observed that the calculated Seebeck coefficients were fairly consistent with experimental data for n-type PbTe but not for p-type PbTe, indicating possible modifications in its valence-band electronic structure. We evaluated the electron relaxation time τ_{el} from 207 reported samples of n-type PbTe by combining calculations and experimental data. We found that τ_{el} is not a constant but varies by at least two orders of magnitude. Achieving long τ_{el} was suggested to be critical in increasing the thermoelectric figure of merit *ZT*.

ARTICLE HISTORY
Received 22 January 2019
Revised 8 March 2019
Accepted 1 April 2019

KEYWORDS
Materials informatics; database; data curation; thermoelectric materials; first-principles calculation; electron relaxation time

CLASSIFICATION
40 Optical, magnetic and electronic device materials; 210 Thermoelectronics / Thermal transport / insulators; 404 Materials informatics / Genomics; 401 1st principle calculations



Reproduced with permission from Thermoelectrics Society of Japan.

Introduction

Materials informatics based on large-scale first-principles calculations is rapidly developing in functional materials science [1]. However, the properties of functional materials are not only determined by the electronic structures of the parent compounds but also determined by various experimental factors such as impurity doping and microstructural control. To

CONTACT Yukari Katsura katsura@phys.mm.t.u-tokyo.ac.jp Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; Masaya Kumagai masaya.kumagai@riken.jp Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan

include these effects in materials informatics, we need to use a large number of digitized experimental data.

Thermoelectric materials are an example of such functional materials [2–4]. They are studied for applications in compact cooling and power-generation devices that interconvert heat and electricity. For more than a century, experimental searches for efficient thermoelectric materials have been conducted. Because there are many material families, the maximum reported value of the thermoelectric figure of merit

$$ZT = \frac{S^2\sigma T}{\kappa} = \frac{S^2\sigma T}{\kappa_{\mathrm{el}} + \kappa_{\mathrm{ph}}} \quad (1)$$

is widely used in selecting a good material family. Thermoelectric conversion efficiency increases with increasing ZT, and ZT>1 has been considered as the criterion for applications. Equation (1) implies that Seebeck coefficient S and electrical conductivity σ need to be increased to increase ZT; thermal conductivity κ, comprising electron thermal conductivity κ_{el} and phonon (lattice) thermal conductivity κ_{ph}, needs to be decreased to increase ZT. However, S, σ, and κ_{el} have strong dependence on the carrier doping density n. As a result, these parameters cannot be controlled independently, and a balance is needed to maximize ZT. Introducing impurities and defects of various scales [2–5] is effective in decreasing κ_{ph}. The phonon-glass-electron-crystal (PGEC) is a concept that describes the best materials for high ZT, in which phonons are scattered as in amorphous glass whereas electrons pass freely as in crystals [6].

Although first-principles calculation is a powerful tool to select many candidate thermoelectric materials [7,8], various kinds of uncertainties arise in predicting actual thermoelectric properties [9]. Calculations usually examine idealistically clean crystals, whereas experimental high-ZT samples are much dirtier. The errors in band gap values create huge errors in calculating Seebeck coefficients in many high-ZT compounds, which are narrow-gap semiconductors. Calculations employing Boltzmann's transport equation cannot calculate σ and κ_{el} directly; they can only evaluate $\sigma/\tau_{\mathrm{el}}$ and $\kappa_{\mathrm{el}}/\tau_{\mathrm{el}}$, where τ_{el} is an unknown variable called the electron relaxation time. As τ_{el} is the time between electron scattering, long τ_{el} is expected in PGEC materials. Although many studies assume constant τ_{el} (often at 10^{-14} s), a large sample dependence of τ_{el} is reported in a few studies [9,10].

Recently, by combining first-principles calculations and experimental data, materials informatics has emerged. The leading example is *Citrination thermoelectrics recommendation engine* [11], which by machine learning assists the users in the selection of good parent compounds of thermoelectric materials. As calculation data, they used the *TE Design Lab* database [7], which contains electronic structure parameters of over 2300 parent compounds. As experimental data, they used several experimental databases including *UCSB Thermoelectrics Data* (MRL Datamining Chart/Energy Materials Datamining) [12], which contains experimental data of about 300 samples (over 1000 data points at 4 different T) of thermoelectric materials reported in over 100 publications.

To empower such experimental materials informatics, we attempted to collect experimental data on a greater scale. During the history of thermoelectric materials, thousands of experimental samples were fabricated, and their properties were published as papers. Unfortunately, most of the precious data is buried in plot images, and not accessible in digital form. We attempted to recycle such experimental data by plot digitization, termed 'plot mining'. Unlike text and images, digitized scientific data extracted from plot images can be shared without violating the copyrights of publishers [13], if the necessary citation is indicated. Plot mining is a high-speed and costless way to obtain experimental data, rather than repeating similar experiments.

In this study, we developed a web system named *Starrydata2*, which can efficiently collect and share digital experimental data, extracted from plot images in published papers. We developed this web system from scratch, employing a completely different architecture and user interface from our prototype web system [14]. By embedding the plot digitization and data upload interface in a typical reference manager like Mendeley [15], we attempted to speed up the human-based comprehensive data collection. This web system enables community-based collections of published data by worldwide researchers, and the collected data can be downloaded freely.

A classical approach in theoretical materials science have been to use a state-of-the-art calculation derived from theory, and supporting it with a few experimental data of selected samples. However, in this approach, it was difficult to know what are the critical factors that determine the properties of the majority samples, which may be dirty or low in performance or missing some information. We present an alternative approach to use a large dataset pf experimental samples, and test for guiding principles that work for most of the experimental data. The name of our web system reflects our concept to treat every datum as important as top data, like a starry sky that contains numerous dwarf stars.

As an example, we used our dataset for experimental samples of rock-salt-type thermoelectric materials. Rock-salt-type thermoelectric materials, which are often referred as PbTe-type thermoelectric materials, are a family of thermoelectric materials

that crystallize in NaCl-type crystal structure. The typical parent compounds are PbTe, PbSe, PbS, and SnTe. Various other metals can substitute the cation sites, and non-metals can substitute the anion sites. Some samples in this family have been reported to possess high ZT values over 2 [16,17].

The phonon properties of the real samples of rock-salt-type thermoelectric materials is difficult to predict from first-principles. Large atomic mass and weak bonding results in low phonon velocity, and strong phonon anharmonicity related to the lone pairs of the divalent cations are related to the high phonon scattering rate in intrinsic PbTe [18,19]. The chemical properties of the parent compounds accept doping of non-stoichiometry, various elements and solid solutioning in wide ranges of compositions. Such site vacancies, impurity elements, intra-grain nanoprecipitates and grain boundaries in the polycrystalline samples all scatter phonons of various wavelengths [16] to achieve low κ_{ph} [17,20,21] around 1 $Wm^{-1}K^{-1}$.

The electronic structures of rock-salt-type thermoelectric materials are also complex, making theoretical prediction of transport properties extremely difficult. The valence bands are composed of the p-bands of the anions have hole pockets along at Σ and L points [22,23], requiring complex treatment of inter-band scattering for transport properties. The calculated hole Fermi surfaces of PbTe are complex in shapes hardly expressed in parabolic band model [22,23]. The band degeneracy changes with T and lattice parameters [16,24]. Spin-orbit interaction splits the bands to reconstruct the band structure around the Fermi level. The width of this splitting increases in the order Sn<Pb and S<Se<Te, and further variation in band structure is expected when these elements are randomly mixed. The conduction band, which is composed of the sp-bands of the cation, is relatively simpler than the valence band. There are electron pockets at L-point, forming a direct band gap at L. The site vacancies and impurity elements may further modify the band structure around Fermi level. The calculation results strongly depend on the selection of exchange-correlation potentials [23,25].

In such a complex material family, it is important to reveal the trend that is applicable for most of the experimental samples. We plotted the thermoelectric properties of 434 reported samples in single plots, to reveal the common trends across the samples. As a candidate parameter to enable prediction of ZT, we evaluated τ_{el} for constant relaxation time approximation by combination with a first-principles calculation for 207 samples of PbTe.

Methods

A list of papers with keyword 'thermoelectric' was retrieved from the *Scopus* web system [26]. From this list, we collated possible papers on material properties, by automatically detecting characteristic words of material names in the titles. We manually downloaded the full-text PDF (Portable Document Format) files for these selected papers from the publisher's web sites. Then we manually checked each for content and classified them into categories based on the parent compounds. The papers on bulk samples composed of PbTe, PbSe, PbS, and SnTe with less than 10% impurity elements were classified as rock-salt-type thermoelectric materials.

From each full-text PDF, we manually captured the plot images on the experimental T-dependence of the thermoelectric properties S, σ (or electrical resistivity $\rho = 1/\sigma$), $\kappa = \kappa_{el}+\kappa_{ph}$, power factor $P = S^2\sigma$, and ZT. We opened these images in WebPlotDigitizer [27], and extracted the original numerical data by semi-automatic colour detection or by manual mouse-clicking. The numerical data were fit using polynomial functions of up to 5th order, to evaluate the thermoelectric properties at T = 300, 400, …, 800 K. The missing parameters were estimated by mathematical operations between known parameters. If reported, the experimental Hall carrier density $n_{H,exp}$ at room temperature was also recorded. The possible chemical composition of each sample was extracted by comprehending the text and identifying the starting compositions.

To make the above data collection process more efficient, we developed a web system named *Starrydata2* on a cloud server at http://www.starrydata2.org. When a Digital Object Identifier (DOI) is supplied by a user, the web system automatically retrieves bibliographic information such as author names and journal names from CrossRef.org [28], and records those to our database. The web system automatically generates links to the publisher's website, the data-collection page, and the data-browsing page. The data-collection page contains the interface of WebPlotDigitizer [27], and a data-upload text box with an automatic unit convertor. The collected datasets can be freely downloaded as text files in Comma Separated Variables (CSV) and JavaScript Object Notation (JSON) formats. This web system is accessible to the public free of charge.

The example first-principles calculations of PbTe was performed using the Full-potential Linearized Augmented Plane Wave (FLAPW) method implemented in the WIEN2k code [29]. We employed the Generalized Gradient Approximation (GGA) correlation-exchange potential [30] with spin-orbit interaction. Core/valence cut-off energy was set at −6.0 Ry, and the calculation was performed on a 50 × 50 × 50 k-mesh. Thermoelectric properties were calculated from the Boltzmann transport equations, using the BoltzTraP code [31], which was modified [9] to include second-order terms for κ_{el}. The chemical potential (μ) dependences of additional charges per unit cell N, S, σ/τ_{el}, and κ_{el}/τ_{el}, at T = 300, 400, 500, 600, 700, and 800 K were

obtained from output files generated by BoltzTraP. The values of μ were converted to carrier doping densities $n = -N/V_{cell}$ [cm^{-3}], where V_{cell} is the unit cell volume.

We evaluated τ_{el} for each T in each sample using experimental Seebeck coefficient S_{exp} and experimental electrical conductivity σ_{exp}. From the calculated S–n curve, we estimated n, the carrier doping density that corresponds to S_{exp}. If the S–n curve is bell-shaped due to the bipolar effect, we selected the solution with higher n, unless bipolar effects were obvious in the experimental data. From the n and the calculated n-dependence of σ/τ_{el}, we evaluated $(\sigma/\tau_{el})_{calc}$ to estimate τ_{el} from

$$\tau_{el} = \frac{\sigma_{exp}}{(\sigma/\tau_{el})_{calc}}. \quad (2)$$

With this τ_{el}, we estimated κ_{ph} using

$$\kappa_{ph} = \kappa_{exp} - \tau_{el}(\kappa_{el}/\tau_{el})_{calc}. \quad (3)$$

Results and discussion

From Scopus [26], we retrieved with the keyword 'thermoelectric' a list of 47,936 papers published between 1875 and 2015. Our original material-name detection script selected 18,585 papers from the list, and among them we accessed the full-text of 14,835 papers to select those that contain plots of interest and to classify them into material families.

The screenshot of our original web system *Starrydata2* is shown in Figure 1. For each record of a paper, *Starrydata2* stores the bibliographic information, the numerical data extracted from the plots, and the chemical compositions of the corresponding samples. The system only shows the numerical data and the replots, without storing the original full-text and the plot images, which are often protected by publisher's copyright. The users can generate lists of publications of interest, and browse the data collected by all users. They can download them as a data file, either in spreadsheet-like format (CSV and JSON) or in a relational-database-like format (JSON only). Our data visualization system can display the data files in various formats including line plots, heat maps, and multiple scatter plots.

Using *Starrydata2*, we succeeded in attaining a considerable improvement in the speed of manual data collection. We rejected the selection of papers and samples in previous data collections to increase both the number of recorded samples and the speed of data collection. Currently, we have collected the data for 11,506 samples in 9509 figures published in 1957 papers. About 500–1000 samples are added each month. Since the experimental data for a sample usually appear in multiple figures, we manually related these data from an identical sample by comprehension of the paper. By using the recent version of our web system, a single data collector succeeded to process 166 papers (806 plots, 1148 samples, 3251 datasets and 89,210 data points) in 25 working days. On average, 1.03 papers (5.00 plots, 7.12 samples, 20.2 datasets, 553 data points) were processed per hour. This time includes the time to read the text to identify the chemical composition of each sample. By increasing the number of data collectors, much more experimental data on other material families will be uploaded in our database.

Figure 2 shows a part of our experimental dataset on rock-salt-type thermoelectric materials in a comparison with the UCSB Thermoelectric data [12], the largest literature-based experimental dataset on thermoelectric properties. Each data point, which corresponds to one experimental sample, is entered in a plot of P against κ. Our dataset contained 434 samples of PbTe, PbSe, PbS, SnTe and their solid solutions from 64 publications [33–86], whereas UCSB Thermoelectric data contained 8 such samples. The large diversity of the scatter plot is a result of the non-selective character of our dataset, which accepted many samples with bad properties. In contrast, the

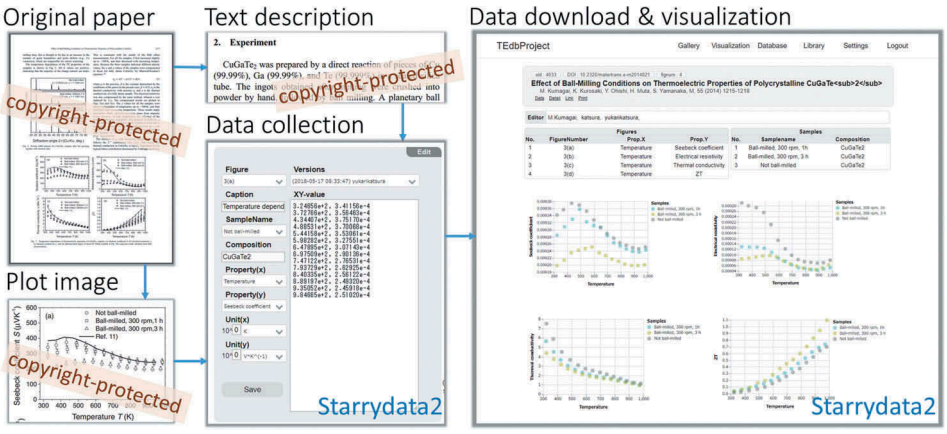


Figure 1. Concept of plot mining in the *Starrydata2* web system. An example paper [32] and the screenshots of *Starrydata2* web system are presented. Reproduced with permission from Thermoelectrics Society of Japan.

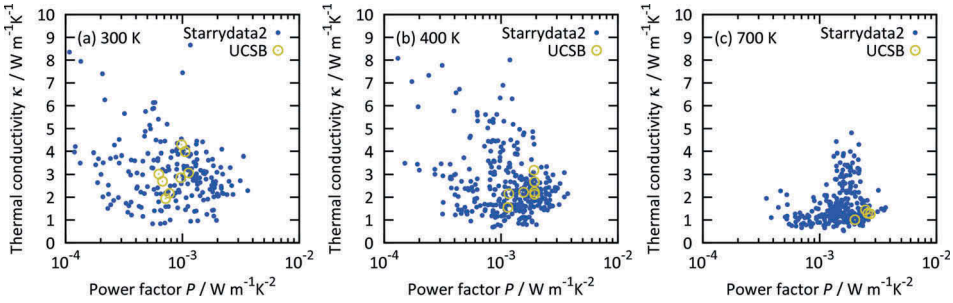


Figure 2. Scatter plot of the power factor P and thermal conductivity κ of rock-salt-type thermoelectric materials including PbTe, PbSe, PbS, SnTe, and their solid solutions, recorded in *Starrydata2* and *UCSB Thermoelectric data* (displayed as UCSB), at (a) 300 K, (b) 400 K, and (c) 700 K.

samples from UCSB thermoelectric data were distributed in the right-bottom corner of Figure 2(c), implying that the UCSB Thermoelectric data selectively collected the samples that possess high ZT.

Figure 3 shows the raw data of temperature dependences of S, σ, and κ of the 434 samples of rock-salt-type thermoelectric material. The colours indicate the maximum ZT of each sample. We observed that samples with wide ranges of S, σ, and κ can be fabricated in this type of thermoelectric material. Simultaneously, it can be said that it is difficult to select a single sample to represent the overall properties of rock-salt-type thermoelectric materials. Most of the samples possessed S values between ±300 μV/K that monotonically increased with T. Two samples undergoing sign changes in S were typical of low-n (bipolar) samples. Most of the samples underwent monotonic decreases in σ and κ with increasing T. The range of σ was between 10^3 and 10^6 S/m, and the high-ZT samples were distributed in the middle. The range for κ was below 10 W/mK, and the high-ZT samples were observed near the bottom of the distribution. Such a direct comparison of many experimental data is helpful to reveal the non-calculable characteristics of each material family from the confusion inherent in strong sample dependences of thermoelectric properties.

The diversity of our dataset enabled us to reveal the characteristics of the parent compounds, from only simple scatter plots. Figure 4 is a Jonker plot showing the relationship between S and log σ. Among p-type samples, the Jonker curve of p-type PbTe was higher than those of PbSe and PbS. In contrast, the Jonker curve of n-type PbTe was not distinguishable from those of PbSe and PbS. These differences are consistent with calculated electronic structures, where the valence bands are composed of p-bands of Te, Se, and S, and the conduction bands are composed of Pb/Sn sp-bands. From the plot, all reported samples of SnTe were seen to be p-type and high in σ, suggesting the presence of a strong hole-doping mechanism.

Figure 5 shows a comparison of theoretical S–n curves obtained from a typical first-principles

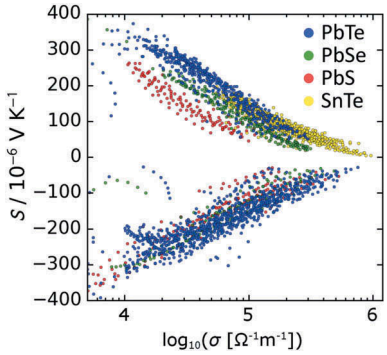


Figure 4. A Jonker plot of rock-salt-type thermoelectric materials with experimental values of the Seebeck coefficient S against electrical resistivity σ for PbTe, PbSe, PbS, and SnTe at T = 300, 400, 500, 600, 700, and 800 K. The parent compounds of the samples are displayed in different colours.

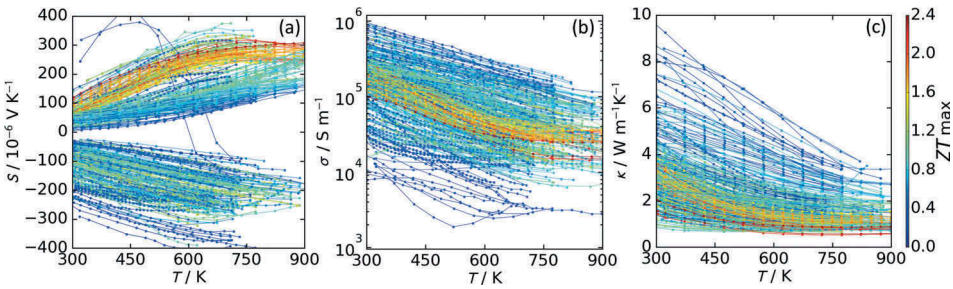


Figure 3. Experimental transport properties of 434 experimental samples of rock-salt-type thermoelectric materials including PbTe, PbSe, PbS, SnTe, and their solid solutions. Temperature dependences of (a) the Seebeck coefficient S, (b) electrical conductivity σ, and (c) the total thermal conductivity κ. Colours indicate the maximum ZT of each sample.

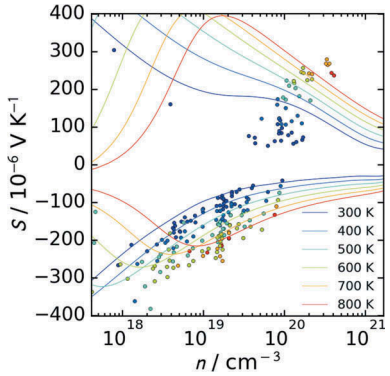


Figure 5. Calculated carrier doping level (*n*) dependences of the Seebeck coefficient (*S*) from first-principles calculation using Boltzmann transport equations, in the range 300–800 K, against the experimental Hall carrier concentration at room temperature, for 38 samples of n-type and p-type PbTe. Experimental data of *S* are also plotted against *n*.

calculation, to the experimental *S*–*n* curve of PbTe samples. Only 38 samples out of 207 samples of PbTe are shown in this plot, because the values of experimental *n* above room temperature were not reported for the other samples. Our transport calculation for n-type PbTe were close to the experimental data at for each *T*, especially in high-*n* region over 10^{19} cm^{-3}. In contrast, the *S*-*n* curve of our calculation was far from the most of reported p-type samples, especially at high *T*. This showed that the model of our transport calculation is not valid for most samples of p-type PbTe. Several other theoretical papers [38,87] have successfully reproduced the experimental *S*-*n* curve of n-type PbTe, however, they did not show *S*-*n* curve for p-type PbTe. These implied the possible difficulty in the modelling of transport properties of p-type PbTe, despite the high *ZT* values compared to n-type PbTe.

By combining first-principles calculations and our experimental data of S_{exp}, σ_{exp}, and κ_{exp}, we attempted to evaluate τ_{el}, an unknown parameter in first-principles calculations of σ and κ_{el}. During the evaluation of τ_{el}, we need the value of calculated σ/τ_{el}. As this value is given as a function of *n*, this analysis could be done only for samples with known experimental *n*. However, for most of the reported samples of thermoelectric materials, Hall measurement to determine experimental *n* has not been carried out. So in this study, we attempted to estimate *n* by using the reported values of experimental *S* and a calculated *S*-*n* curve. Since this analysis is applicable only when there is a consistency between calculated and experimental *S*–*n* curves, we carried out this analysis only for n-type samples.

Figure 6(a) shows the values of *n* and τ_{el} of the 207 samples of *n*-type PbTe, estimated from S_{exp} and σ_{exp}. The values of τ_{el} were between 10^{-15} s and 10^{-13} s, exhibiting a two-orders-of-magnitude variation. These τ_{el} values are composite values of intrinsic τ_{el} and extrinsic τ_{el}, which can be expressed by Matthiesen's rule such that $f_{el,total} = f_{el,intrinsic} + f_{el,extrinsic}$, by using electron scattering rate $f_{el} = \tau_{el}^{-1}$. The intrinsic f_{el} due to electron-phonon interaction of PbTe has been calculated to be around 10^{12}–10^{13} Hz by a DFPT calculation considering screening effects by electrons [87]. Our total f_{el} were in a range between 10^{13}–10^{15} Hz. The samples with f_{el} ~10^{13} Hz (τ_{el} ~10^{-13} s) are expected to be clean samples, in which intrinsic electron-phonon interaction is one of the dominant electron scattering mechanisms. The samples with f_{el} ~10^{15} Hz (τ_{el} ~10^{-15} s) were expected to be dirty samples, where extrinsic electron scattering mechanisms are dominant. The candidates of such extrinsic electron scattering centers in n-type PbTe include atomic vacancies, impurity atoms, intra-grain nanoprecipitates, dislocations, grain boundaries and impurity phases. The trend that short τ_{el} is observed more frequently in high-*n* samples was consistent with the expectation that the carrier-doping vacancies and impurities also scatter electrons. The trend that τ_{el} decreases with increasing *T* is observed especially in low-*n* samples, and this was consistent with the that phonon scattering rate increases with increasing *T*.

Our analysis also revealed that the popular transport calculations that use a fixed value of $\tau_{el} = 10^{-14}$ s for constant relaxation time approximation is not valid for most of the reported samples of n-type PbTe.

Figure 6(b) shows the relationship between τ_{el} and *ZT*. For each *T*, we observed a trend whereby an increase in τ_{el} results in a monotonic increase in *ZT*. The surprising thing here is not that *ZT* increased with τ_{el}, but that the τ_{el} of almost all the reported experimental samples clearly followed a common

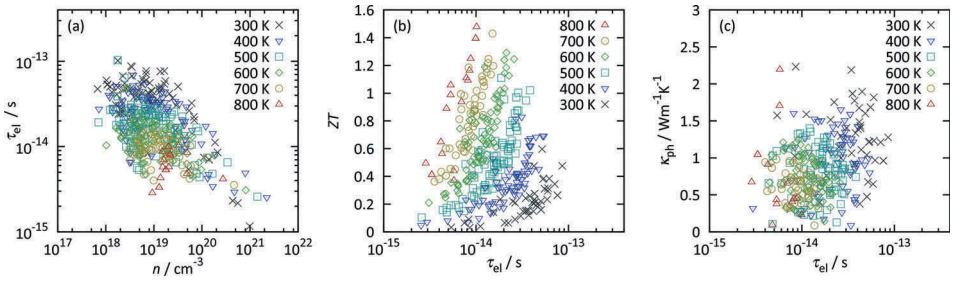


Figure 6. Relationship between (a) carrier doping level *n* and electron relaxation time τ_{el}, (b) τ_{el} and thermoelectric figure of merit *ZT*, and (c) τ_{el} and phonon thermal conductivity κ_{ph}, estimated for 207 experimental samples of *n*-type PbTe.

curve, which can be drawn for each T = 300, 400, …, 800 K. This showed that the ZT values can be expected only from T and τ_{el}, for these real samples of PbTe-type thermoelectric materials. In researches of thermoelectric materials, this is quite a rare case that ZT values can be predicted only from one physical parameter other than T. This result gives researchers a strong guiding principle that an increase in τ_{el} directly enhance ZT of n-type PbTe.

Figure 6(c) shows the relationship between τ_{el} and κ_{ph}. If the dominant electron scattering centers acted as the dominant phonon scattering centers, we would have seen a correlation in this plot. However, we could not observe any correlation between τ_{el} and κ_{ph}. This indicated that in most samples of n-type PbTe, there are dominant phonon scattering mechanisms other than electrons, or there are dominant electron scattering mechanisms other than phonons. Even though electrons in ideal PbTe crystals are reported to be scattered mainly by phonons [87], this does not mean that phonons are mainly scattered by electrons. The advantage of n-type PbTe may be the availability of the phonon scattering mechanisms without heavy electron scattering, as to realize the concept of PGEC.

Conclusion

We developed an original web system *Starrydata2* as an open database, to let researchers gather and share experimental data from published plot images. So far, we have succeeded to collect experimental data from more than 11,500 samples of thermoelectric materials. Our web system enabled collective analysis of published samples, to discover the guiding principles that work for most of the published samples.

By combining our partial dataset on rock-salt-type thermoelectric materials with first-principles calculations, we analysed thermoelectric properties of rock-salt-type thermoelectric materials. The differences in valence band electronic structures of PbTe, PbSe, PbS, and SnTe were revealed in the Jonker plots.

For n-type PbTe, we found that the effective τ_{el} for constant relaxation time approximation varied by more than two orders of magnitude, indicating that τ_{el} is not a constant but a very important determinant of ZT. This analysis is applicable for other thermoelectric materials, whose S-n curves were successfully reproduced by first-principles calculations and transport calculations. As our database is growing to cover more and more families of thermoelectric materials, evaluation of τ_{el} of various experimental samples of thermoelectric materials will be a strong guide for experimental researchers to design high-ZT samples of thermoelectric materials.

Acknowledgments

We thank all the people who gave us valuable suggestions to improve our system and our data collection project. We thank the members of MI^{2}I, the members of 'Thermoelectric Database Working Group' and the committees of Thermoelectrics Society of Japan, and the members of SAKURA Internet Inc. for supports of our project. We thank Shinji Nagashiro from X-Ability Co. Ltd. for technical advice on the development of the data visualizer. We thank Richard Haase, Ph.D, from Edanz Group (www.edanzediting.com/ac) for editing the first draft of this manuscript.

Data Availability

The experimental data obtained from the publications can be freely downloaded from our *Starrydata2* web system at http://www.starrydata2.org. The datasets generated in this study are available at *GitHub* repository, https://github.com/starrydata.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was financially supported by a KAKENHI grant [No. 16K14379] from the Ministry of Education, Culture, Sports, Science and Technology Japan, Watanabe memorial foundation for the advancement of new technology, and 'Materials Research by Information Integration' Initiative (MI2I) project of the Support Program for Starting Up Innovation Hub from the Japan Science and Technology Agency (JST).

References

[1] Jain A, Hautier G, Ong SP, et al. New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships. J Mater Res. 2016;31:977–994.
[2] Snyder GJ, Toberer ES. Complex thermoelectric materials. Nat Mater. 2008;7:105–114.
[3] Liu W, Yan X, Chen G, et al. Recent advances in thermoelectric nanocomposites. Nano Energy. 2012;1:42–56.
[4] Gayner C, Kar KK. Recent advances in thermoelectric materials. Prog Mater Sci. 2016;83:330–382.
[5] Zhao LD, Dravid VP, Kanatzidis MG. The panoscopic approach to high performance thermoelectrics. Energy Environ Sci. 2014;7:251–268.
[6] Slack GA. New materials and performance limits for thermoelectric cooling. CRC Handb Thermoelectr. 1995;2:407–440.
[7] Gorai P, Gao D, Ortiz B, et al. TE design lab: a virtual laboratory for thermoelectric material design. Comput Mater Sci. 2016;112:368–376.
[8] Chen W, Pöhls J-H, Hautier G, et al. Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment. J Mater Chem C. 2016;4:4414–4426.

[9] Katsura Y, Takagi H, Kimura K. Roles of carrier doping, band gap, and electron relaxation time in the Boltzmann Transport Calculations of a Semiconductor's Thermoelectric Properties. Mater Trans. 2018;59:1013–1021.
[10] Tan XJ, Liu W, Liu HJ, et al. Multiscale calculations of thermoelectric properties of n-type Mg_2Si_{1-x}. Phys Rev B. 2012;85:205212.
[11] Gaultois MW, Oliynyk AO, Mar A, et al. Perspective: web-based machine learning models for real-time screening of thermoelectric materials properties. APL Mater. 2016;4:53213.
[12] Gaultois MW, Sparks TD, Borg CKH, et al. Data-driven review of thermoelectric materials: performance and resource considerations. Chem Mater. 2013;25:2911–2920.
[13] Elliot R. Who owns scientific data? The impact of intellectual property rights on the scientific publication chain. Learn Publ. 2005;18:91–94.
[14] Katsura Y, Kumagai M, Gunji S, et al. Development of "Starry data" web system for data curation of published experimental thermoelectric properties. J Jpn Soc Powder Metall. 2017;64:467–470.
[15] Zaugg H, West RE, Tateishi I, et al. Mendeley: creating communities of scholarly inquiry through research collaboration. TechTrends. 2011;55:32–36.
[16] Pei Y, Tan G, Feng D, et al. Integrating band structure engineering with all-scale hierarchical structuring for high thermoelectric performance in PbTe system. Adv Energy Mater. 2017;7:1–11.
[17] Fu T, Yue X, Wu H, et al. Enhanced thermoelectric performance of PbTe bulk materials with figure of merit zT >2 by multi-functional alloying. J Mater. 2016;2:141–149.
[18] Waghmare UV, Spaldin NA, Kandpal HC, et al. First-principles indicators of metallicity and cation off-centricity in the IV-VI rocksalt chalcogenides of divalent Ge, Sn, and Pb. Phys Rev B - Condens Matter Mater Phys. 2003;67:10.
[19] Delaire O, Ma J, Marty K, et al. Giant anharmonic phonon scattering in PbTe. Nat Mater. 2011;10:614–619.
[20] Koh YK, Vineis CJ, Calawa SD, et al. Lattice thermal conductivity of nanostructured thermoelectric materials based on PbTe. Appl Phys Lett. 2009;94:2007–2010.
[21] Su C-H. Experimental determination of lattice thermal conductivity and Lorenz number as functions of temperature for n-type PbTe. Mater Today Phys. 2018;5:58–63.
[22] Singh DJ. Doping-dependent thermopower of PbTe from Boltzmann transport calculations. Phys Rev B - Condens Matter Mater Phys. 2010;81:1–6.
[23] Svane A, Christensen NE, Cardona M, et al. Quasiparticle self-consistent GW calculations for PbS, PbSe, and PbTe: band structure and pressure coefficients. Phys Rev B - Condens Matter Mater Phys. 2010;81:1–10.
[24] Pei YL, Liu Y. Electrical and thermal transport properties of Pb-based chalcogenides: PbTe, PbSe, and PbS. J Alloys Compd. 2012;514:40–44.
[25] Goyal A, Gorai P, Toberer ES, et al. First-principles calculation of intrinsic defect chemistry and self-doping in PbTe. NPJ Comput Mater. 2017;3:1–9.
[26] Elsevier Science Publishers. Scopus quick reference guide. 2015. CRC Press; p. 14.
[27] Rohatgi A. WebPlotDigitizer User Manual Version 3.9. 2015. p. 1–18.
[28] Gravim TB, Union I, Bgi T, et al. Annual report 2007–2008. CrossRef. 2008.
[29] Blaha P, Schwarz K, Madsen GKH, et al. WIEN2k - an augmented planewave + local orbitals program for calculating crystal properties. WIEN2k. 2010.
[30] Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. Phys Rev Lett. 1996;77:3865–3868.
[31] Madsen GKH, Singh DJ. BoltzTraP. A code for calculating band-structure dependent quantities. Comput Phys Commun. 2006;175:67–71.
[32] Kumagai M, Kurosaki K, Ohishi Y, et al. Effect of Ball-Milling conditions on thermoelectric properties of polycrystalline $CuGaTe_2$. Mater Trans. 2014;55:1215–1218.
[33] Ohta M, Biswas K, Lo S-H, et al. Enhancement of thermoelectric figure of merit by the insertion of MgTe nanostructures in p-type PbTe doped with Na_2Te. Adv Energy Mater. 2012;2:1117–1123.
[34] Wang H, Schechtel E, Pei Y, et al. High thermoelectric efficiency of n-type PbS. Adv Energy Mater. 2013;3:488–495.
[35] Paul B, Rawat PK, Banerji P. Dramatic enhancement of thermoelectric power factor in PbTe: Crco-doped with iodine. Appl Phys Lett. 2011;98:262101.
[36] Sootsman JR, He J, Dravid VP, et al. Microstructure and thermoelectric properties of mechanically robust PbTe-Si Eutectic composites. Chem Mater. 2010;22:869–875.
[37] Banik A, Shenoy US, Anand S, et al. Mg alloying in SnTe facilitates valence band convergence and optimizes thermoelectric properties. Chem Mater. 2015;27:581–587.
[38] LaLonde AD, Pei Y, Snyder GJ. Reevaluation of $PbTe_{1-x}I_x$ as high performance n-type thermoelectric material. Energy Environ Sci. 2011;4:2090.
[39] Wu D, Zhao L-D, Tong X, et al. Superior thermoelectric performance in PbTe–PbS pseudo-binary: extremely low thermal conductivity and modulated carrier concentration. Energy Environ Sci. 2015;8:2056–2068.
[40] Li JQ, Li SP, Wang QB, et al. Synthesis and thermoelectric properties of the $PbSe_{1-x}Te_x$ alloys. J Alloys Compd. 2011;509:4516–4519.
[41] Li XX, Li JQ, Liu FS, et al. Enhanced thermoelectric properties of $(PbTe)_{0.88}(PbS)_{0.12}$ composites by Bi doping. J Alloys Compd. 2013;547:86–90.
[42] Guch M, Raj Sankar C, Salvador JR, et al. Improvements of the thermoelectric properties of PbTe via simultaneous doping with indium and iodine. J Appl Phys. 2012;111:63706.
[43] Bali A, Royanian E, Bauer E, et al. Thermoelectric properties of PbTe with encapsulated bismuth secondary phase. J Appl Phys. 2013;113:123707.
[44] Han MK, Zhou X, Uher C, et al. Increase in the figure of merit by Cd-substitution in $Sn_{1-x}Pb_xTe$ and effect of Pb/Sn ratio on thermoelectric properties. Adv Energy Mater. 2012;2:1218–1225.
[45] Pei Y, Lalonde A, Iwanaga S, et al. High thermoelectric figure of merit in heavy hole dominated PbTe. Energy Environ Sci. 2011;4:2085–2089.
[46] Cohen I, Kaller M, Komisarchik G, et al. Enhancement of the thermoelectric properties of n-type PbTe by Na and Cl co-doping. J Mater Chem C. 2015;3:9559–9564.
[47] Kudman I. Thermoelectric properties of p-type PbTe-PbSe alloys. J Mater Sci. 1972;7:1027–1029.
[48] Chen Y, Zhu TJ, Yang SH, et al. Thermal and electrical transport properties of VA-element doped

$Pb_{9.6}M_{0.2}Te_{10-x}Se_x$ (M = Sb, Bi) thermoelectric materials. J Phys D Appl Phys. 2010;43:35405.
[49] Androulakis J, Lin CH, Kong HJ, et al. Spinodal decomposition and nucleation and growth as a means to bulk nanostructured thermoelectrics: enhanced performance in $Pb_{1-x}Sn_xTe$-PbS. J Am Chem Soc. 2007;129:9780–9788.
[50] Ahn K, Han MK, He J, et al. Exploring resonance levels and nanostructuring in the PbTe-CdTe system and enhancement of the thermoelectric figure of merit. J Am Chem Soc. 2010;132:5227–5235.
[51] Johnsen S, He J, Androulakis J, et al. Nanostructures boost the thermoelectric performance of PbS. J Am Chem Soc. 2011;133:3460–3470.
[52] Androulakis J, Todorov I, He J, et al. Thermoelectrics from abundant chemical elements: high-performance nanostructured PbSe-PbS. J Am Chem Soc. 2011;133:10920–10927.
[53] Zhao LD, He J, Hao S, et al. Raising the thermoelectric performance of p-type PbS with endotaxial nanostructuring and valence-band offset engineering using CdS and ZnS. J Am Chem Soc. 2012;134:16327–16336.
[54] Lee Y, Lo SH, Androulakis J, et al. High-performance tellurium-free thermoelectrics: all-scale hierarchical structuring of p-type PbSe-MSe systems (M = Ca, Sr, Ba). J Am Chem Soc. 2013;135:5152–5160.
[55] Korkosz RJ, Chasapis TC, Lo SH, et al. High ZT in p-type $PbTe_{1-2x}PbSe_xPbS_x$ thermoelectric materials. J Am Chem Soc. 2014;136:3225–3227.
[56] Wu HJ, Zhao LD, Zheng FS, et al. Broad temperature plateau for thermoelectric figure of merit ZT>2 in phase-separated $PbTe_{0.7}S_{0.3}$. Nat Commun. 2014;5:4515.
[57] Gelbstein Y, Dashevsky Z, Dariel MP. Powder metallurgical processing of functionally graded p-$Pb_{1-x}Sn_xTe$ materials for thermoelectric applications. Phys B Condens Matter. 2007;391:256–265.
[58] Rawat PK, Paul B, Banerji P. Impurity-band induced transport phenomenon and thermoelectric properties in Yb doped $PbTe_{1-x}I_x$. Phys Chem Chem Phys. 2013;15:16686–16692.
[59] Wu C-F, Wei T-R, Li J-F. Electrical and thermal transport properties of $Pb_{1-x}Sn_xSe$ solid solution thermoelectric materials. Phys Chem Chem Phys. 2015;17:13006–13012.
[60] Androulakis J, Lee Y, Todorov I, et al. High-temperature thermoelectric properties of n-type PbSe doped with Ga, In, and Pb. Phys Rev B - Condens Matter Mater Phys. 2011;83:195209.
[61] Androulakis J, Chung DY, Su X, et al. High-temperature charge and thermal transport properties of the n-type thermoelectric material PbSe. Phys Rev B – Condens Matter Mater Phys. 2011;84:155207.
[62] Zhang Q, Yang S, Zhang Q, et al. Effect of aluminum on the thermoelectric properties of nanostructured PbTe. Nanotechnology. 2013;24:345705.
[63] Jaworski CM, Heremans JP. Thermoelectric transport properties of the n-type impurity Al in PbTe. Phys Rev B – Condens Matter Mater Phys. 2012;85:033204.
[64] Zhang Q, Liao B, Lan Y, et al. High thermoelectric performance by resonant dopant indium in nanostructured SnTe. Proc Natl Acad Sci. 2013;110:13261–13266.
[65] Yamini SA, Wang H, Ginting D, et al. Thermoelectric performance of n-type $(PbTe)_{0.75}(PbS)_{0.15}(PbSe)_{0.1}$ composites. ACS Appl Mater Interfaces. 2014;6:11476–11483.
[66] Zhang Y, Wu L, Zhang J, et al. Eutectic microstructures and thermoelectric properties of MnTe-rich precipitates hardened PbTe. Acta Mater. 2016;111:202–209.
[67] Pei Y, LaLonde AD, Heinz NA, et al. High thermoelectric figure of merit in PbTe alloys demonstrated in PbTe-CdTe. Adv Energy Mater. 2012;2:670–675.
[68] He J, Girard SN, Kanatzidis MG, et al. Microstructure-lattice thermal conductivity correlation in nanostructured $PbTe_{0.7}S_{0.3}$ thermoelectric materials. Adv Funct Mater. 2010;20:764–772.
[69] Lo S-H, He J, Biswas K, et al. Phonon scattering and thermal conductivity in p-Type nanostructured PbTe-BaTe Bulk thermoelectric materials. Adv Funct Mater. 2012;22:5175–5184.
[70] Wang H, Pei Y, Lalonde AD, et al. Heavily doped p-type PbSe with high thermoelectric performance: an alternative for PbTe. Adv Mater. 2011;23:1366–1370.
[71] Sootsman JR, Kong H, Uher C, et al. Large enhancements in the thermoelectric power factor of bulk PbTe at high temperature by synergistic nanostructuring. Angew Chemie Int Ed. 2008;47:8618–8622.
[72] Wu CF, Wei TR, Li JF. Enhancing average ZT in pristine PbSe by over-stoichiometric Pb addition. APL Mater. 2016;4:104801.
[73] Zhou M, Gibbs ZM, Wang H, et al. Thermoelectric performance of co-doped SnTe with resonant levels. Appl Phys Lett. 2016;109:042102.
[74] Al Rahal Al Orabi R, Mecholsky NA, Hwang J, et al. Band degeneracy, low thermal conductivity, and high thermoelectric figure of merit in SnTe-CaTe alloys. Chem Mater. 2016;28:376–384.
[75] Wang H, Hwang J, Snedaker ML, et al. High thermoelectric performance of a heterogeneous PbTe nanocomposite. Chem Mater. 2015;27:944–949.
[76] Biswas K, He J, Wang G, et al. High thermoelectric figure of merit in nanostructured p-type PbTe-MTe (M = Ca, Ba). Energy Environ Sci. 2011;4:4675–4684.
[77] Zhao LD, Wu HJ, Hao SQ, et al. All-scale hierarchical thermoelectrics: MgTe in PbTe facilitates valence band convergence and suppresses bipolar thermal transport for high performance. Energy Environ Sci. 2013;6:3346–3355.
[78] Tan G, Shi F, Doak JW, et al. Extraordinary role of Hg in enhancing the thermoelectric performance of p-type SnTe. Energy Environ Sci. 2015;8:267–277.
[79] Wu H, Chang C, Feng D, et al. Synergistically optimized electrical and thermal transport properties of SnTe via alloying high-solubility MnTe. Energy Environ Sci. 2015;8:3298–3312.
[80] He J, Tan X, Xu J, et al. Valence band engineering and thermoelectric performance optimization in SnTe by Mn-alloying via a zone-melting method. J Mater Chem A. 2015;3:19974–19979.
[81] Zhou Z, Yang J, Jiang Q, et al. Multiple effects of Bi doping in enhancing the thermoelectric properties of SnTe. J Mater Chem A. 2016;4:13171–13175.
[82] Basu R, Bhattacharya S, Bhatt R, et al. Improved thermoelectric properties of Se-doped n-type $PbTe_{1-x}Se_x$ ($0 \le x \le 1$). J Electron Mater. 2013;42:2292–2296.

[83] Girard SN, He J, Zhou X, et al. High performance Na-doped PbTe-PbS thermoelectric materials: electronic density of states modification and shape-controlled nanostructures. J Am Chem Soc. 2011;133:16588–16597.
[84] Kang C, Wang H, Kim H, et al. Effect of excess Na on the morphology and thermoelectric properties of $Na_xPb_{1-x}Te_{0.85}Se_{0.15}$. J Electron Mater. 2014; 43:353–358.
[85] Levin EM, Besser MF, Hanus R. Electronic and thermal transport in GeTe: a versatile base for thermoelectric materials. J Appl Phys. 2013;114:83713.
[86] Papageorgiou C, Giapintzakis J, Kyratsi T. Low-temperature synthesis and thermoelectric properties of n-type PbTe. J Electron Mater. 2013;42:1911–1917.
[87] Song Q, Liu T-H, Zhou J, et al. Ab initio study of electron mean free paths and thermoelectric properties of lead telluride. Mater Today Phys. 2017;2:69–77.