

Manual of the TEXplorer platform(<https://www.TEXplorer.org>)

The web-based data platform TEXplorer consists of the following parts: 1) Deposition of raw data files of experimental and computational datasets with parsing and standardizing the datasets on data storage; 2) visualization of the synthesis information, measurement results, and calculated electronic structures with tables and figures; and 3) ML toolkits for predicting thermoelectric materials properties. We describe the usage and functionalities of each application below.

1.1. Data deposition and process: 'Upload'

A major function of TEXplorer is to collect and reposit thermoelectric materials data from experiments and computations. The experimental data consist of all information regarding synthesis, post-process, characterization, and measurement of transport properties. The computational data include the crystal structure of doped/alloyed materials and their electronic structures calculated using Vienna Ab initio Simulation Package (VASP).

For experimental datasets, we collected all parameters for the synthesis process and details for each sample were written in a spreadsheet file with pre-set templates, (which are downloadable at landing page) and the raw data files from measuring instruments were collected for the database. (The available raw data types are listed at the landing page.)

For examples, we synthesized doped/alloyed SnSe and Bi₂Te₃ compounds with the chemical composition of Sn_{x₁}Se_{x₂}:A_{x₃}B_{x₄} and Bi_{x₁}Te_{x₂}:A_{x₃}B_{x₄}C_{x₅}, where A, B, and C are doping/alloying atoms, and x₁₋₅ are the corresponding ratios of host and doped/alloyed elements. The solid-state/solution reaction was used for the synthesis. We additionally pulverized the product by hand/mechanical grinding, and consolidated the obtained powders

using spark plasma sintering (SPS). We selectively performed post-processing, such as ball-milling, oxygen reduction, and annealing, to enhance the thermoelectric performance. The crystal structure of the sample was then characterized with powder X-ray diffraction (XRD) using a Bruker D8 Advance diffractometer, and electrical and thermal transport were measured in a full range of temperatures [Fig. 1 (c)]. Electrical conductivities (σ) and Seebeck coefficient (S) were measured using Netzsch SBA458 and Ulvac-Riko ZEM3, and thermal diffusivity (α) was measured using Netzsch LFA457. Using values of σ , S , and α , thermoelectric figure of merit ZT is automatically evaluated in the platform. All the parameters and raw data files of XRD, SBA, ZEM3, and LFA measurements are deposited on the platform. To ensure systematic data collection, uploads are currently limited to authorized members.

* The thermoelectric figure of merit (ZT) is determined as

$$ZT = \frac{\sigma S^2}{\kappa_{\text{ele}} + \kappa_{\text{lat}}} T, \quad (1)$$

which is a combination of electrical conductivity (σ), Seebeck coefficient (S), absolute temperature (T), and electronic and lattice contribution to thermal conductivity (κ_{ele} , κ_{lat}).

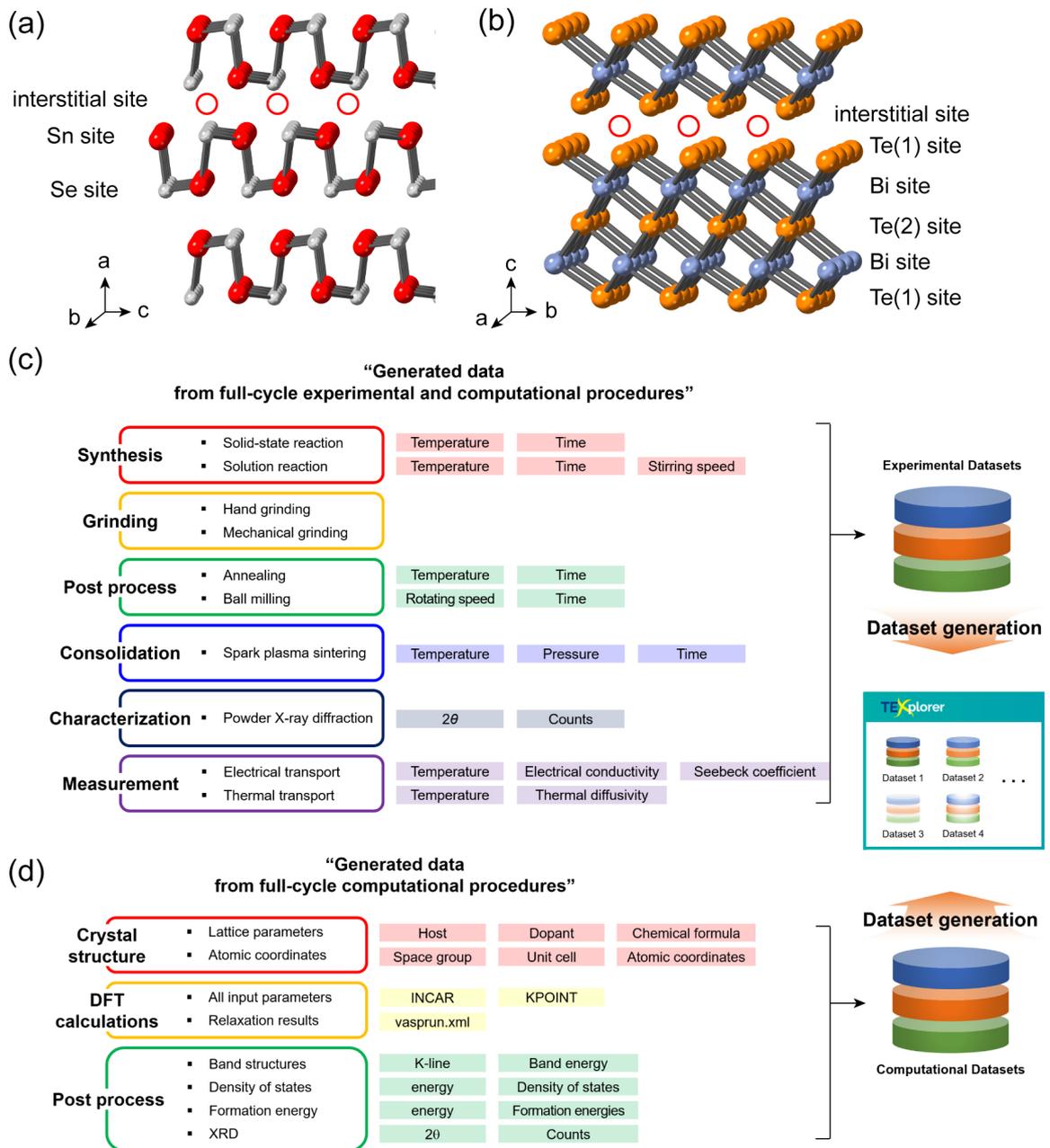


Fig. 1. (a) Crystal structures of SnSe and Bi₂Te₃. Doping elements substitute host elements or locate at interstitial sites. Red, grey, blue, and orange balls represent Sn, Se, Bi, and Te atoms. Generated data from full-cycle (c) experimental and (d) computational procedures.

Coupled with the experimental dataset, computational datasets of electronic band structures can be deposited on the platform in a raw data form. We performed the DFT calculations on SnSe and Bi₂Te₃ systems with different doping elements within the VASP. We computed each system's electronic band structure, density of states, formation energy profiles, bulk modulus, and simulated XRD patterns. [Fig. 1(d)] Each calculation step is organized into subfolders, and the raw data files, necessarily including the “vasprun.xml” file, are collected for the database. The whole folder with some subfolders is uploaded for a given composition.

Data deposition on the TEXplorer platform is designed to be user-friendly with a drag-and-drop interface on a webpage or by using the application programming interface (API). The pre-designed parser works automatically to give the ID number for a new dataset and extracts key/value pairs for data ingestion. For computational data, we used the pymatgen library for the data parsing. If some of the required data files are missing or inputted incorrectly, an error message will pop up, and the upload process will stop. The collected data were standardized to match and unify keywords across multiple datasets created by different users. For example, many users have saved the ball-milling process with their own keywords, such as ballmill, Ball mill, ball-milling, and BM, and thus, we unified the keywords as BM in the database, to remove inaccuracies and to improve the data quality. All ingested data extracted from the spreadsheet and the raw data files were stored in the MongoDB database in Javascript object notation (JSON) format, and can be downloaded in JSON format.

1.2. Searching and displaying data: 'DataExplorer'

The DataExplorer page provides a search function for each experimental and computational dataset. The experimental section allows users to search for target data by selecting elements from the periodic table and by advanced queries consisting of logical operators and additional keywords, such as molar mass, measuring device, and synthesis methods. When a sample is selected from the retrieved list, details of the synthesis process are displayed, and a new window with all experimental information for a given sample appears via a hyperlink on the ID number.

The new window displays a table of sample information and synthesis processes, as shown in Fig. 2, and the data types are listed in Table 1. The 'Data Information' section provides the name of the data creator, updated date, file-downloadable links, and additional information marked by a hashtag. 'Process' refers to primary and post-processing parameters in experiments such as nominal composition, synthesis method with corresponding details (temperature, time, milling speed, cooling method), and SPS conditions (temperature with heating rate, time). If ball milling is performed, the milling speed and time are also provided. The 'Results' shows the transport properties for the chosen sample measured by ZEM3, SBA, and LFA, and a corresponding powder XRD pattern. Plots of transport properties such as σ , S , Lorenz number (L), power factor, α , total thermal conductivity (κ_{tot}), κ_{lat} , and ZT are provided as a function of temperature. The sample size thickness and diameter, sample density, and heat capacity are also represented. The XRD pattern is presented as a function of 2θ to analyze the crystal structure of the sample.

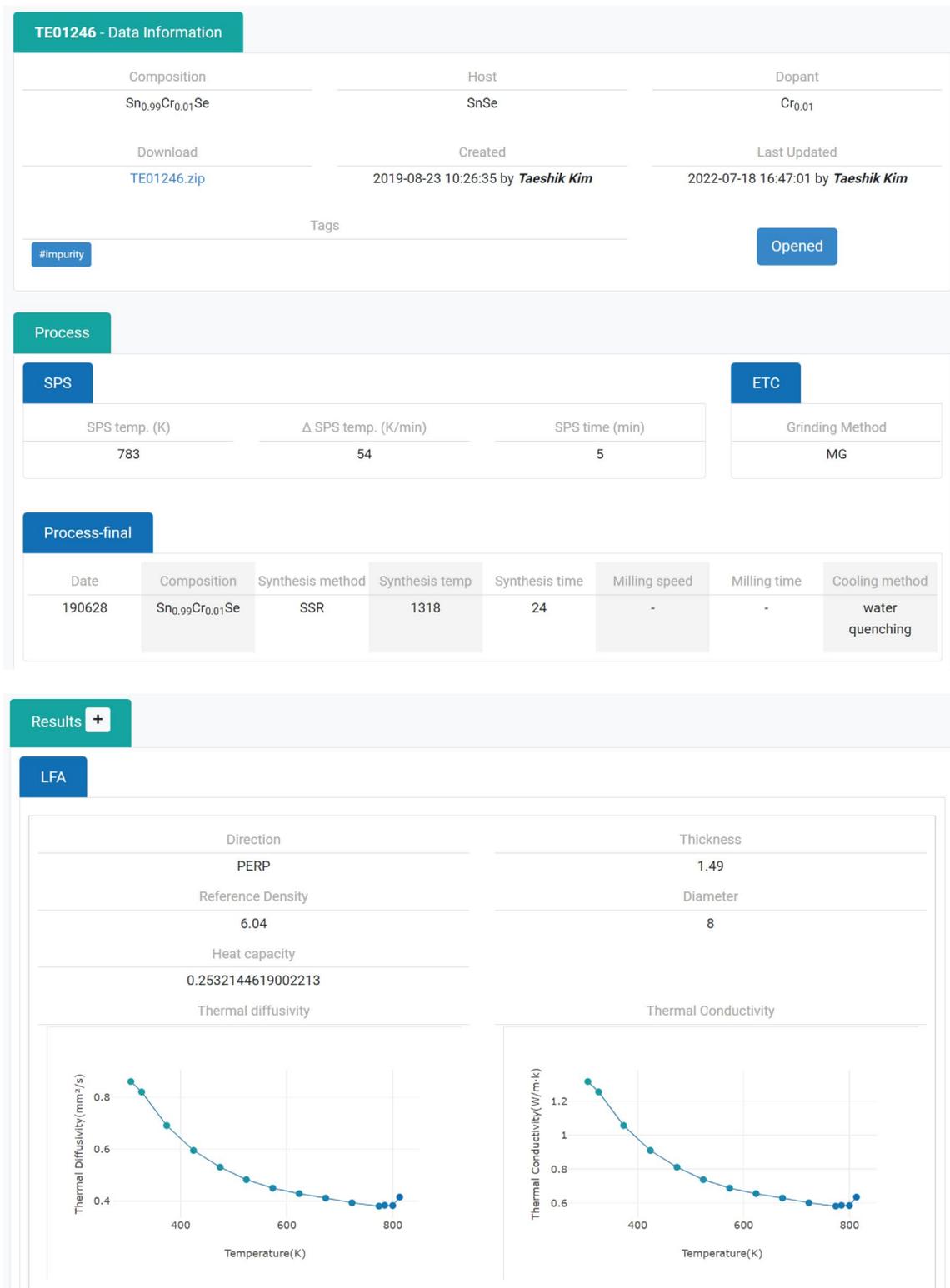


Fig. 2. ‘DataExplorer’ for an experimental dataset of Sn_{0.99}Cr_{0.01}Se (TE01246). The data information and synthesis process details are represented. Thermal diffusivity and thermal conductivity obtained from the LFA measurement are plotted among the measurement results.

Collected data	Data type	Collected data	Data type
Composition	List of objects ^a	SPS	
Host	List of objects ^a	Temperature (K)	Integer
Dopant	List of objects ^a	ΔSPS temperature (K/min)	Integer
Created time and person's name	String/integer	Time (min)	Integer
Uploaded time and person's name	String/integer	SBA/ZEM3 measurement	
Synthesis process		Direction	String
Date	Integer	Diameter	Float
Composition	List of objects ^a	Sample geometry	String
Synthesis method	String	Thickness	Float
Synthesis temperature (K)	Integer	Electrical conductivity (S/cm)	List of objects ^b
Synthesis time (min)	Integer	Seebeck coefficient (μV/K)	List of objects ^b
Milling speed	Integer	Power factor (μW/cm·K ²)	List of objects ^b
Milling time (min)	Integer	Lorenz number (W·Ω/10 ⁸ K ²)	List of objects ^b
Cooling method	String	ZT	
Grinding method	String	Direction	String
LFA measurement		ZT	List of objects ^b
Direction	String	XRD	
Thickness (mm)	Integer	Direction	String
Reference density	Float	Intensity (counts)	List of objects ^c
Diameter	Float	Lattice thermal conductivity	
Heat capacity	Float	Direction	String
Thermal diffusivity (mm ² /s)	List of objects ^b	Lattice thermal conductivity	List of objects ^b
Thermal conductivity (W/m·K)	List of objects ^b	(W/m·K)	

Table 1. Data format for the experimental dataset. ^a[“element”: string, “amount”: float],
^b[“temperature”: float, “target”: float], ^c[“2θ”: list of float, “target”: list of float].

The automatic calculation procedure was embedded to evaluate the specific heat capacity (C_p), κ_{tot} , κ_{lat} , L , and ZT values from the measured raw data from the equipment. κ_{tot} is simply evaluated by multiplying the measured α by specific heat capacity and density, ($\kappa_{tot} = \alpha\rho C_p$, where α is thermal diffusivity, ρ is density, and C_p is specific heat capacity), where C_p is automatically calculated from the atomic composition of a given sample according to the Dulong-Petit law. The lattice contribution to the thermal conductivity is evaluated by the

equation $\kappa_{\text{lat}} = \kappa_{\text{tot}} - \kappa_{\text{ele}}$, where κ_{ele} is estimated via the Wiedemann-Franz relationship ($\kappa_{\text{ele}} = L\sigma T$). L is calculated by employing a single parabolic band model within the acoustic phonon scattering. ZT at a given T was obtained according to Eq. (1), where each value is linearly interpolated at T because the measurements cannot be performed at the very same temperature with different instruments for the transport properties.

In the computational data section, the periodic table based-searches are available, and each hyperlink on the ID number leads to a new window showing the electronic structures and computational details for a given sample [Fig. 4]. The collected data formats are shown in Table 2. This page is specially optimized for doped/alloyed systems where host and dopant atoms are distinguished. The atomic coordinates with unit cell structures are represented by a table of xyz coordinates and a picture of a ball-and-stick model, and hovering the mouse over the table of xyz coordinates highlights the selected atom in the figure of the crystal structures. The band structures, density of states, formation energies of possible rich conditions, and simulated XRD patterns are displayed, and user-interactive zooming is a useful tool that allows users to visualize a subset of the displayed data. The user-customized plot can be downloaded as an image file. The computational details, such as the types of van der Waals (vdW) functionals and pseudopotentials, k -points, and energy cutoff, are provided.

Structure Information

Dopant

Host	Dopant	Ratio	Site
SnSe	Pb	1.56	Sn

Structure

Formula	Space Group	Energy (per atom)	Band gap
Sn31 Se32 Pb1	symbolCm number8	-1845.526 (-28.836) eV	0.75310eV

Visualizer

Sn: 31, Se: 32, Pb: 1
 Lattice (Å): 12.15, 12.15, 11.53
 angles (°): 89.99, 90.01, 93.69
 Density: 6.2324 g/cm³
 Molar mass: 1627.600 g/mol

Host Dopant Off

Sn Se Pb

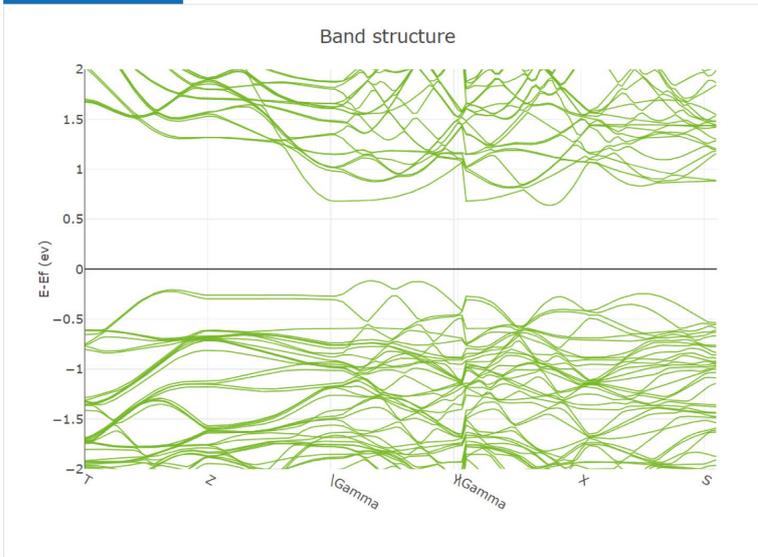
Lattice Parameters			
Vector	Length(Å)	Angle	Value(°)
a	12.14847	α	89.99070
b	12.14847	β	90.00930
a	11.52666	Γ	93.69473

Volume(Å³) 1697.62995

Atoms				
Index	Element	x(Å)	y(Å)	z(Å)
1	Sn	5.19956	4.00396	1.36598
2	Sn	9.34749	4.00618	1.37399
3	Sn	1.03861	-0.42182	1.37033
4	Sn	5.19033	-0.42397	1.37143

Analysis

Band structure



DOS PDOS

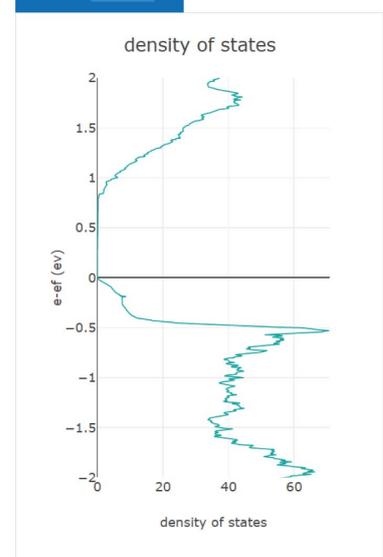


Fig. 3. ‘DataExplorer’ for a computational dataset of Sn₃₁Se₃₂Pb (TC00543). Atomic structures with unit cell information are visualized with a ball-and-stick model and xyz coordinates. The band structures and density of states are illustrated, and details of data information and calculation details are also given on the TEXplorer webpage.

Collected data	Data type	Collected data	Data type
Created time and person’s name	<i>String/integer</i>	Updated time and person’s name	<i>String/integer</i>
Calculation information			
Solver	<i>String</i>	K-points	<i>List of integer</i>
XC type	<i>String</i>	K-points shift	<i>List of float</i>
Pseudopotentials	<i>String</i>	Energy cutoff	<i>Float</i>
Structural information		DFT results	
Host/dopant	<i>String</i>	Total energy (eV)	<i>Float</i>
Substitution site	<i>String</i>	Band gap (eV)	<i>Float</i>
Chemical formula	<i>List of objects^a</i>	Band structure	<i>using pymatgen⁸</i>
Space group	<i>String</i>	Density of states	<i>using pymatgen⁸</i>
Lattice parameters (Å)	<i>Float</i>	Formation energy (eV)	<i>List of objects^b</i>
Unit cell volume (Å ³)	<i>Float</i>	Seebeck coefficient	<i>jpg image</i>
Atomic coordinates (Å)	<i>List of objects</i>	XRD Intensity (counts)	<i>List of objects^c</i>

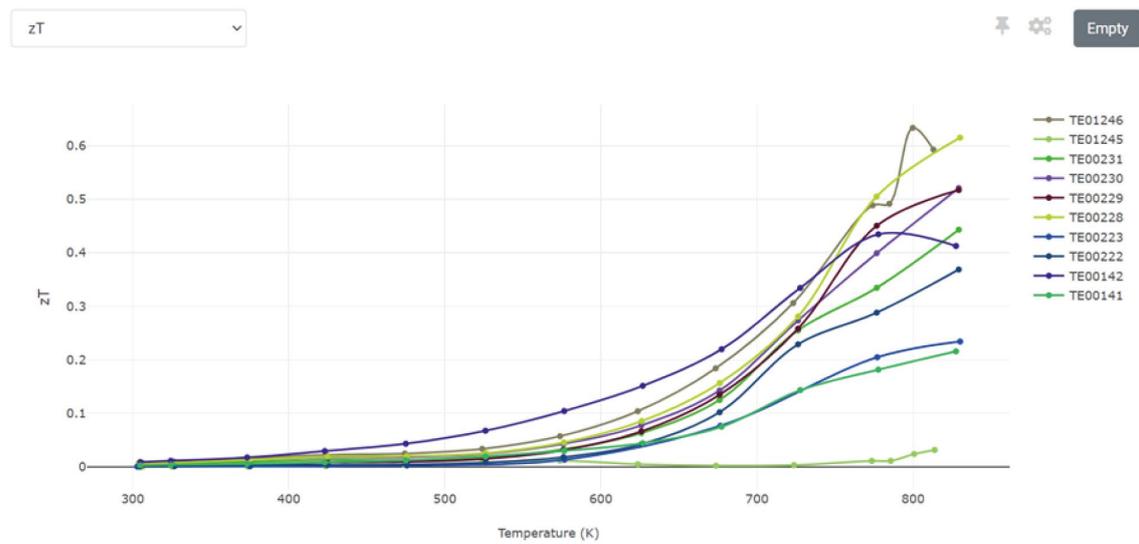
Table 2. Data format for the computational dataset. ^a[“element”: string, “amount”: list of float], ^b[rich_environment:[“energies”: list of float, “formation energy”: list of float]], and ^c[“2 θ ”: list of float, “target”: list of float].

1.3. Useful tools for data analysis: ‘Visualization’

The ‘Visualization-Graph’ page allows comparison of thermoelectric properties between different samples through well-designed plots, as shown in Fig. 4. If we retrieve and select experimental samples on the list, the corresponding thermoelectric properties are plotted as a function of temperature. The following thermoelectric properties of σ , S , α , κ_{tot} , κ_{lat} , power factor, and ZT can be drawn with the dynamic zooming function. Hovering the mouse over the ID number on the legend highlights the graph line of the selected sample, and clicking an ID number hides or shows its graph line on the figure. The manipulated figure is also downloadable as an image file.

For computational datasets, a user-designed three-dimensional plot is available along the x - and y - axes and the size of the points with the following parameters: ID number, uploader name, number of atomic sites, number of elements, incorporation of vdW interaction in the calculation, VASP version, space group of initial and final structures, cell volume, band gap, fermi energy, final total energy, and final total energy per atom. It is designed for practical statistical analysis of computational datasets.

Comparisons between electronic band structures and density of states for computational datasets are provided separately in the ‘Visualization-List’ section. The figures of multiple samples are arranged in parallel, as shown in Fig. 5, and the changes in electronic structures by doping/alloying elements can be easily identified on this page.



Search Show ▾

Data Table Rows per page
10 Select all Deselect all

	ID	Composition	Author	Synthesis Method	Measurement Method	Density	Heat Capacity	Carrier Concentration	SPS	Parsed Data List
Selected	TE01246	Sn _{0.99} Cr _{0.01} Se Dopant -Cr Nominal- Sn _{0.99} Cr _{0.01} Se	Taeshik Kim	SSR	LFA SBA XRD	6.040	2.532e-1		783K 54K/min 5min	density diameter electrical_conductivity
Selected	TE01245	Sn _{0.97} Cr _{0.03} Se Dopant -Cr Nominal- Sn _{0.97} Cr _{0.03} Se	Taeshik Kim	SSR	LFA SBA XRD	5.794	2.549e-1		783K 54K/min 5min	density diameter electrical_conductivity
Selected	TE00231	Sn _{0.8} Pb _{0.2} Se Dopant -Pb Nominal- Sn _{0.8} Pb _{0.2} Se	Virtual Lab	SSR	SBA HALL LFA XRD	6.548	2.316e-1	6.106e16	783K 80K/min 5min	carrier_concentration density electrical_conductivity

Fig. 4. ‘Visualization-Graph’. Comparison of thermoelectric properties (here, ZT) among different samples in experimental and computational datasets. With the search function, users can manipulate the plots with sample selections and user-interactive zooming.

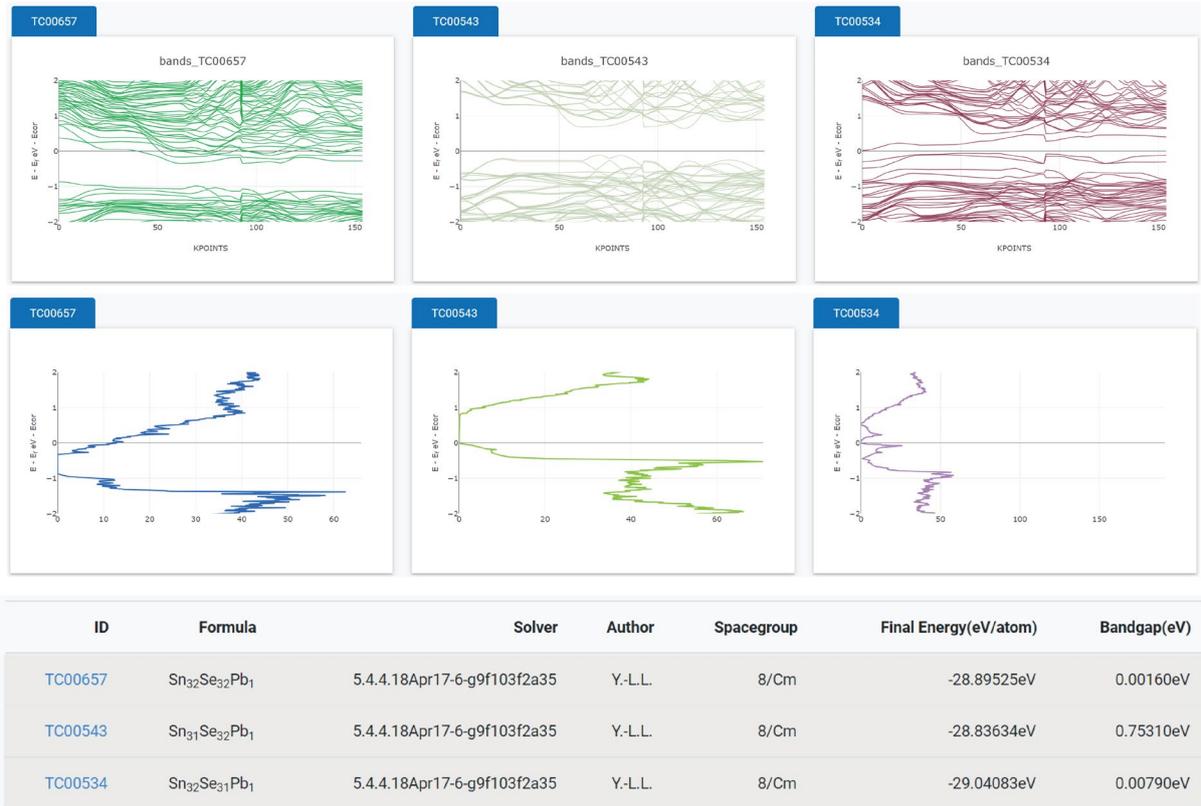


Fig. 5. ‘Visualization-List’. The electronic band structures and the density of states for selected samples are drawn in parallel for comparison.

1.4. Data-driven research for thermoelectric materials design: ‘ML’ toolkit

In our previous work, we constructed ML models based on our thermoelectric database to predict the thermoelectric properties of SnSe-based materials with arbitrary doping. We trained the relationship between material compositions and thermoelectric properties of σ , S , and κ_{tot} within the experimental datasets. The computational datasets were used as feature vectors to describe the electronic structures of a given sample. Using the ML models and the solubility limit of dopants, we searched for optimal doping compositions with a fast screening of 2,832 compositions and analyzed the physical mechanisms of the selected high- ZT compounds. Using our pre-trained ML model, the composition of Na_{0.01}(Sn_{0.96}Ge_{0.04})_{0.99}Se gives the

maximum ZT value of 2.393 at 800K among 2,832 compositions. Similarly, we also constructed ML models to predict the thermoelectric properties of Bi_2Te_3 systems with various doping/alloying.

TEXplorer provides ML toolkits for predicting thermoelectric properties of doped/alloyed $\text{SnSe}/\text{Bi}_2\text{Te}_3$ materials using pre-trained ML models. Once given an arbitrary doped/alloyed $\text{SnSe}/\text{Bi}_2\text{Te}_3$ composition with element types, molar ratios, and vacancy ratios, feature vectors are automatically generated in the background, and the ML models predict the thermoelectric properties of the given composition. The predicted results of σ , S , κ_{tot} , and ZT are plotted as a function of temperature, and the maximum value of ZT is displayed with the corresponding temperature [Fig. 6]. Users can predict the thermoelectric properties of various compositions in $\text{SnSe}/\text{Bi}_2\text{Te}_3$ systems on TEXplorer, and download the full results in Excel format.

(a) Input

[Hide Periodic Table](#)

Alloying atom:

Sn vacancy (0~2) %:

(b) Result

Formula : $\text{Na}_{0.01}(\text{Sn}_{0.96}\text{Ge}_{0.04})_{0.99}\text{Se}_{1.00}$
Max ZT : 2.393 at 800 K

[Download](#)

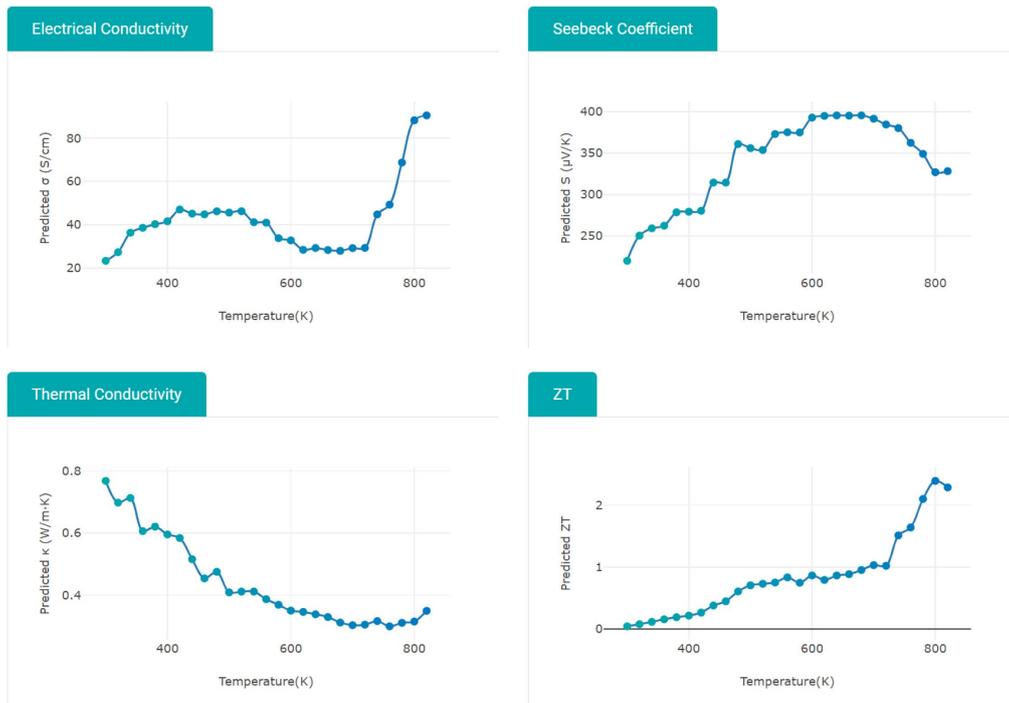


Fig. 6. 'ML'. The predicting toolkit for thermoelectric properties of electrical conductivity, Seebeck coefficient, thermal conductivity, and ZT values. (a) By entering the doping element type, ratio, and amount of Sn vacancy, (b) the ML models predict the thermoelectric properties. The chemical formula, the maximum ZT with temperature, and corresponding plots are displayed. The predicted results of $\text{Na}_{0.01}(\text{Sn}_{0.96}\text{Ge}_{0.04})_{0.99}\text{Se}$ are displayed.

[References]

TEXplorer.org: Thermoelectric material properties data platform for experimental and first-principles calculation results, Yea-Lee Lee, Hyungseok Lee, Seunghun Jang, Jeongho Shin, Taeshik Kim, Sejin Byun, In Chung, Jino Im, and Hyunju Chang, *APL Mater.* **11**, 041111 (2023) [<https://doi.org/10.1063/5.0137642>]

Data-driven enhancement of ZT in SnSe-based thermoelectric systems, Yea-Lee Lee, Hyungseok Lee, Taeshik Kim, Sejin Byun, Yong Kyu Lee, Seunghun Jang, In Chung, Hyunju Chang, and Jino Im, *J. Am. Chem. Soc.* **144**(30), 13748-13763 (2022) [<https://doi.org/10.1021/jacs.2c04741>]