Supplementary Information

Prediction of Thermoelectric Performance for Layered IV-V-VI Semiconductors by High-Throughput *Ab-Initio* Calculations and Machine Learning

Yu Gan[1#], Guanjie Wang[1#], Jian Zhou[1] and Zhimei Sun[1,2,*]

[1] *School of Materials Science and Engineering, Beihang University, Beijing 100191, China*

[2] *Center for Integrated Computational Materials Engineering, International Research Institute for Multidisciplinary Science, Beihang University, Beijing 100191, China*

[#] *These authors contributed equally*

[*] Email: zmsun@buaa.edu.cn (Z. M. Sun)

Supplementary Table 1 The definition of initial 31 descriptors. The content in parentheses of descriptors represents the unit of the corresponding physical quantity.

Descriptors	Definition
N_A, N_B, N_C	Number of atoms of A, B and C elements in the primitive cell
N_T	Total number of atoms
V_A, V_B, V_C	Valence electronic configuration
R_A, R_B, R_C	Row number in the periodic table
Z_A, Z_B, Z_C	Atomic number in the periodic table
r_A, r_B, r_C (pm)	Atomic radius
r_{CA}, r_{CB}, r_{CC} (pm)	Covalent radius
P_A, P_B, P_C	Pauling electronegativity
P_{Av}	Average Pauling electronegativity
P_{Sd}	Standard deviation of Pauling electronegativity
a/b, c (Å)	Lattice constant
M_{Av} (g/mol)	Average atomic weight
L_{AC}, L_{BC} (Å)	Bond length
L_{Av} (Å)	Average bond length $L_{Av} = 1/2(L_{AC} + L_{BC})$
T (K)	Temperature

Supplementary Note 1

- V_A, V_B, V_C

It should be noted that the same group elements show the same number of valence electrons, namely, 4 (s^2p^2) for IV (A = Si, Ge, Sn Pb), 5 (s^2p^3) for V (B = As, Sb, Bi) and 6 (s^2p^4) for VI atoms (C = S, Se, Te), but they may have different outmost valence electron layers, for example, $3s^2p^2$ for Ge and $5s^2p^3$ for Sb. Here we set a rule to convert the valence electronic configurations into a specified number, that is

$3s^2p^2 = 0$, $3s^2p^4 = 1$, $4s^2p^2 = 2$, $4s^2p^3 = 3$, $4s^2p^4 = 4$, $5s^2p^2 = 5$, $5s^2p^3 = 6$, $5s^2p^4 = 7$, $6s^2p^2 = 8$, $6s^2p^3 = 9$, which is beneficial to machine learning model training.

Supplementary Note 2

- P_{Av} and S_{Dv}

  The average Pauling electronegativity P_{Av} is defined as:

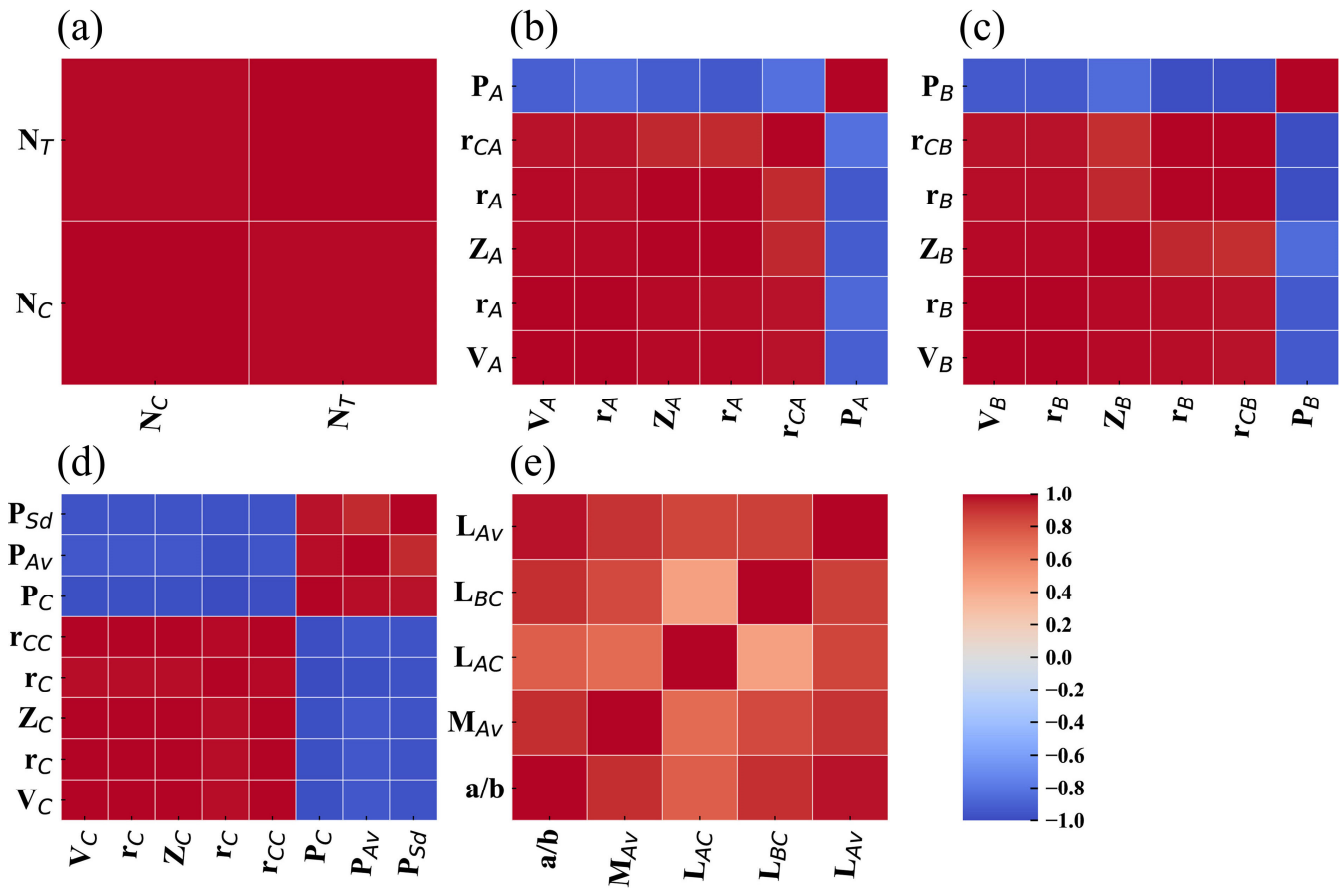
$$P_{Av} = \frac{N_A P_A + N_B P_B + N_C P_C}{N_A + N_B + N_C}. \quad (1)$$

Then the standard deviation of Pauling electronegativity can be extracted from P_{Av} using the formula:

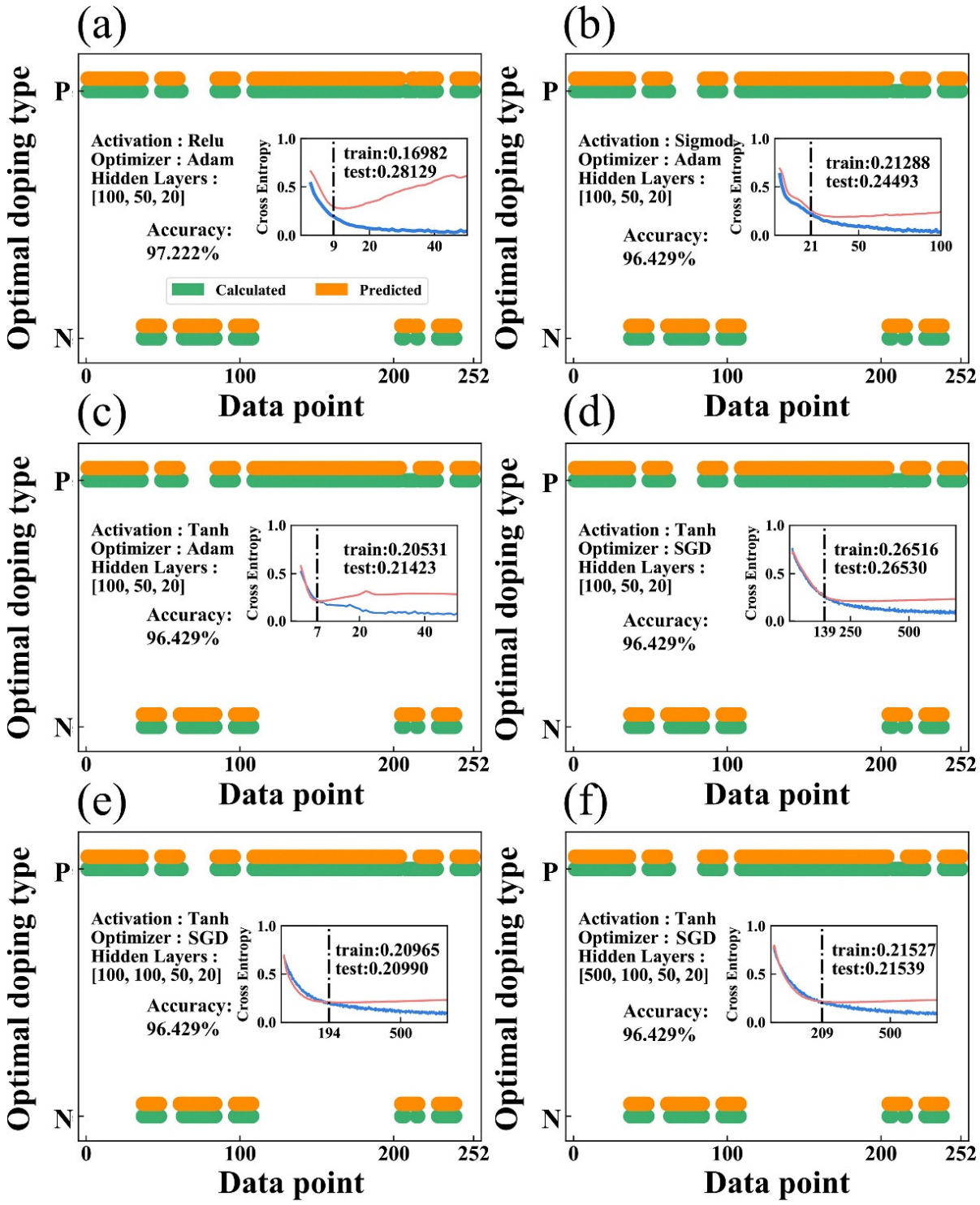
$$S_{Dv} = \sqrt{\frac{N_A \left(P_A - P_{Av}\right)^2 + N_B \left(P_B - P_{Av}\right)^2 + N_C \left(P_C - P_{Av}\right)^2}{N_A + N_B + N_C}}. \quad (2)$$

Supplementary Table 2 The used *k*-point grids for the calculations of density of states (DOS), electrical transport and phonon thermal transport properties. The parentheses indicate the scalebroad parameter.

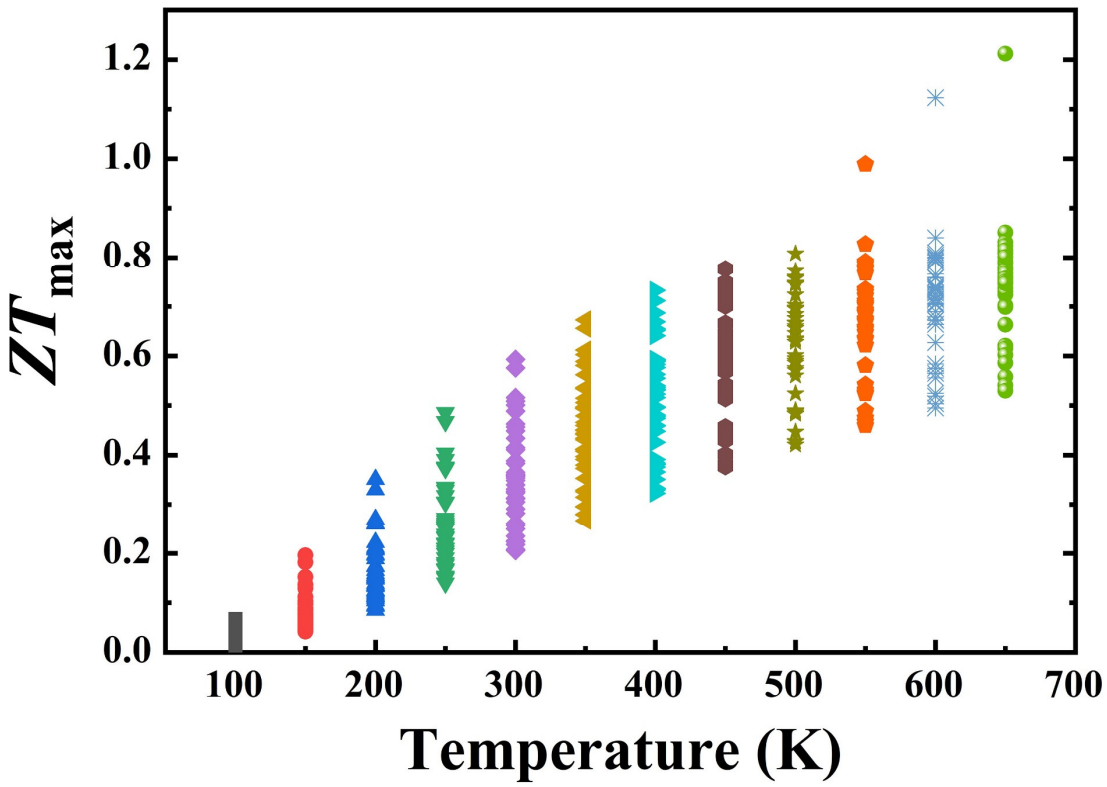
System	DOS	Electrical transport properties	Phonon transport properties
$A_1B_2C_4$	15×15×15	31×31×5	17×17×2 (0.2)
$A_1B_4C_7$	21×21×13	31×31×13	15×15×4 (0.5)
$A_2B_2C_5$	21×21×13	31×31×13	15×15×4 (0.5)
$A_3B_2C_6$	15×15×15	31×31×3	15×15×2 (0.2)



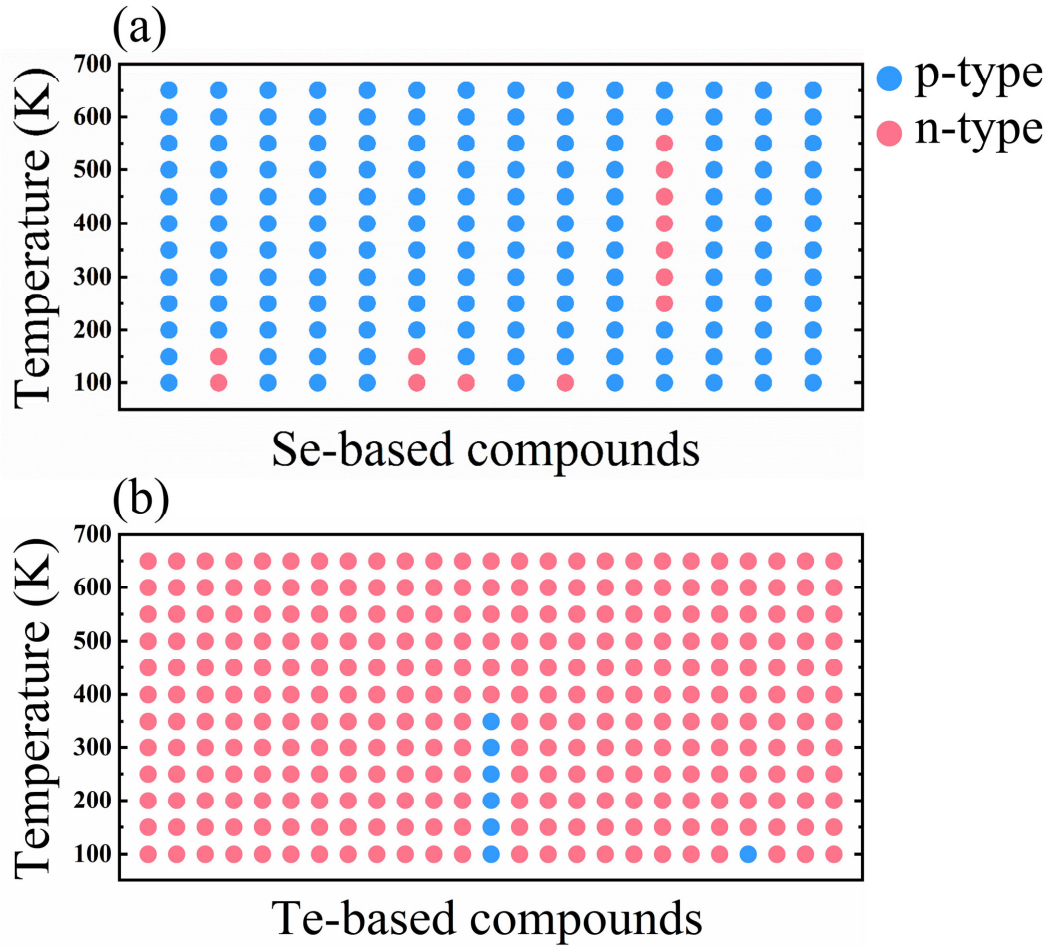
Supplementary Figure 1 Correlation coefficients. Excluded features with correlation coefficients above the absolute 0.90.



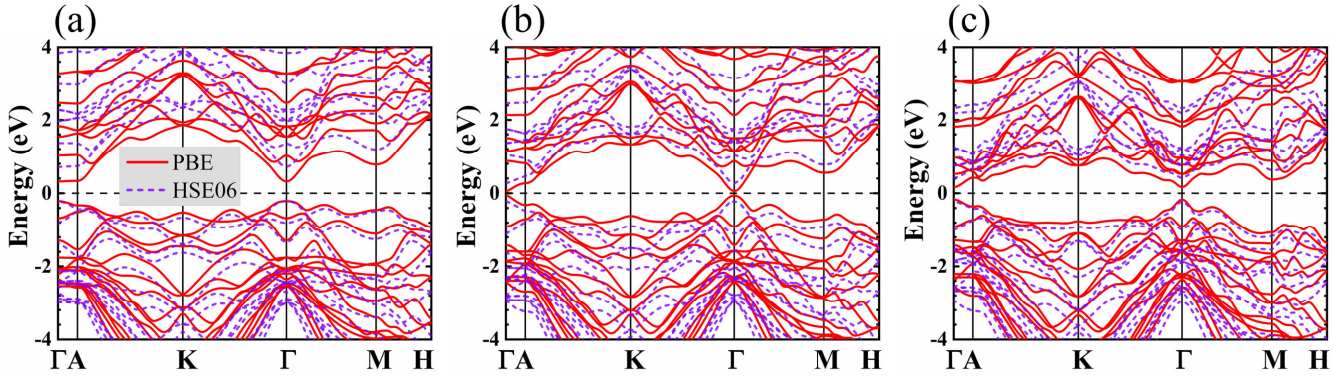
Supplementary Figure 2 Machine learning model. ML-predicted optimal doping type based on the deep neural networks with different activation functions, optimizers and hidden layers, in comparison with the DFT-calculated results. The insets indicate the convergence process and cross entropy during training (blue lines) and testing (red lines), in which the ordinate of the intersections of the black vertical line and the training and test learning curves represent their respective cross entropy values.



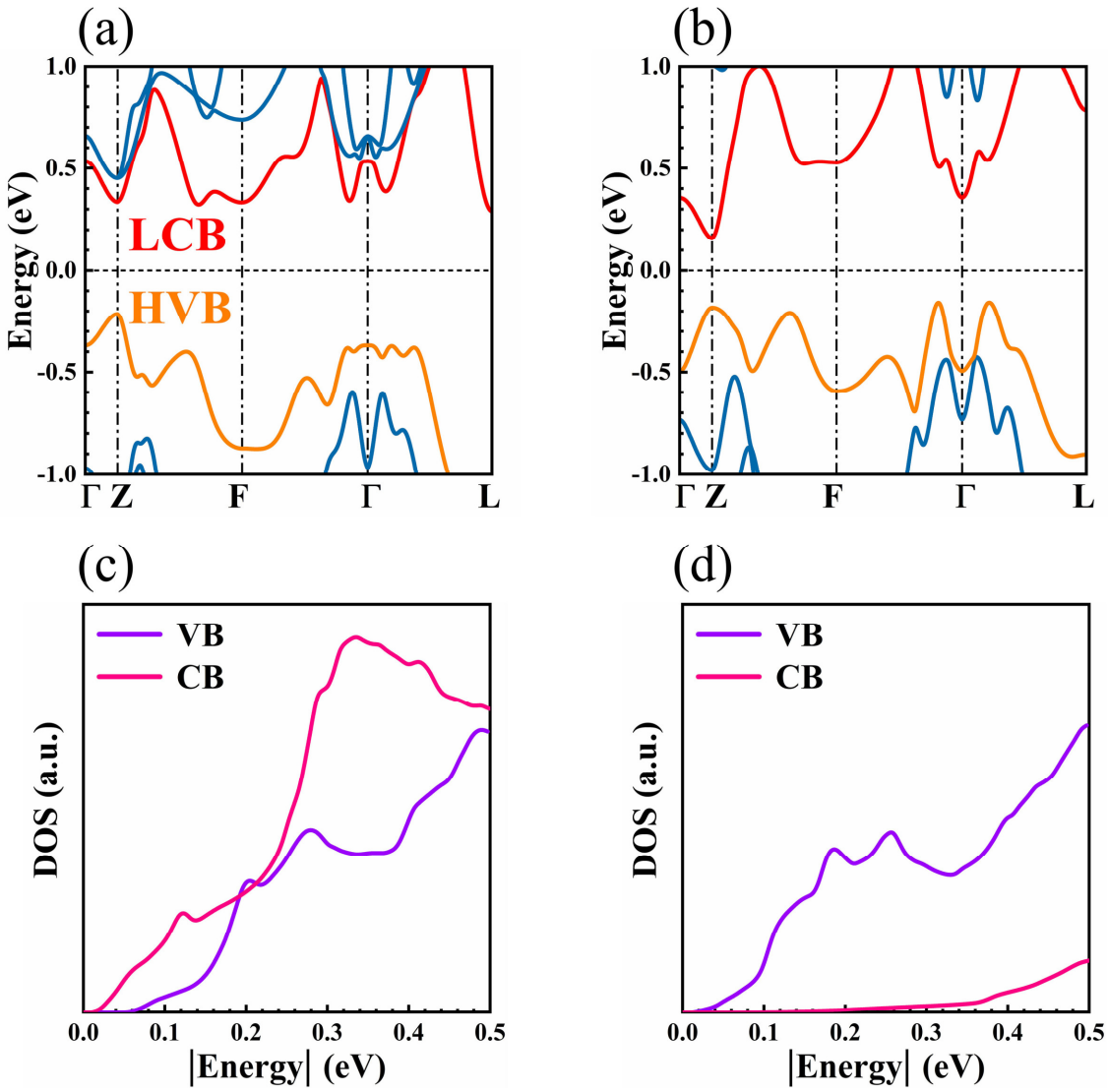
Supplementary Figure 3 DFT-calculated ZT_{max}. Calculated ZT_{max} values at different temperatures for all 40 input compounds.



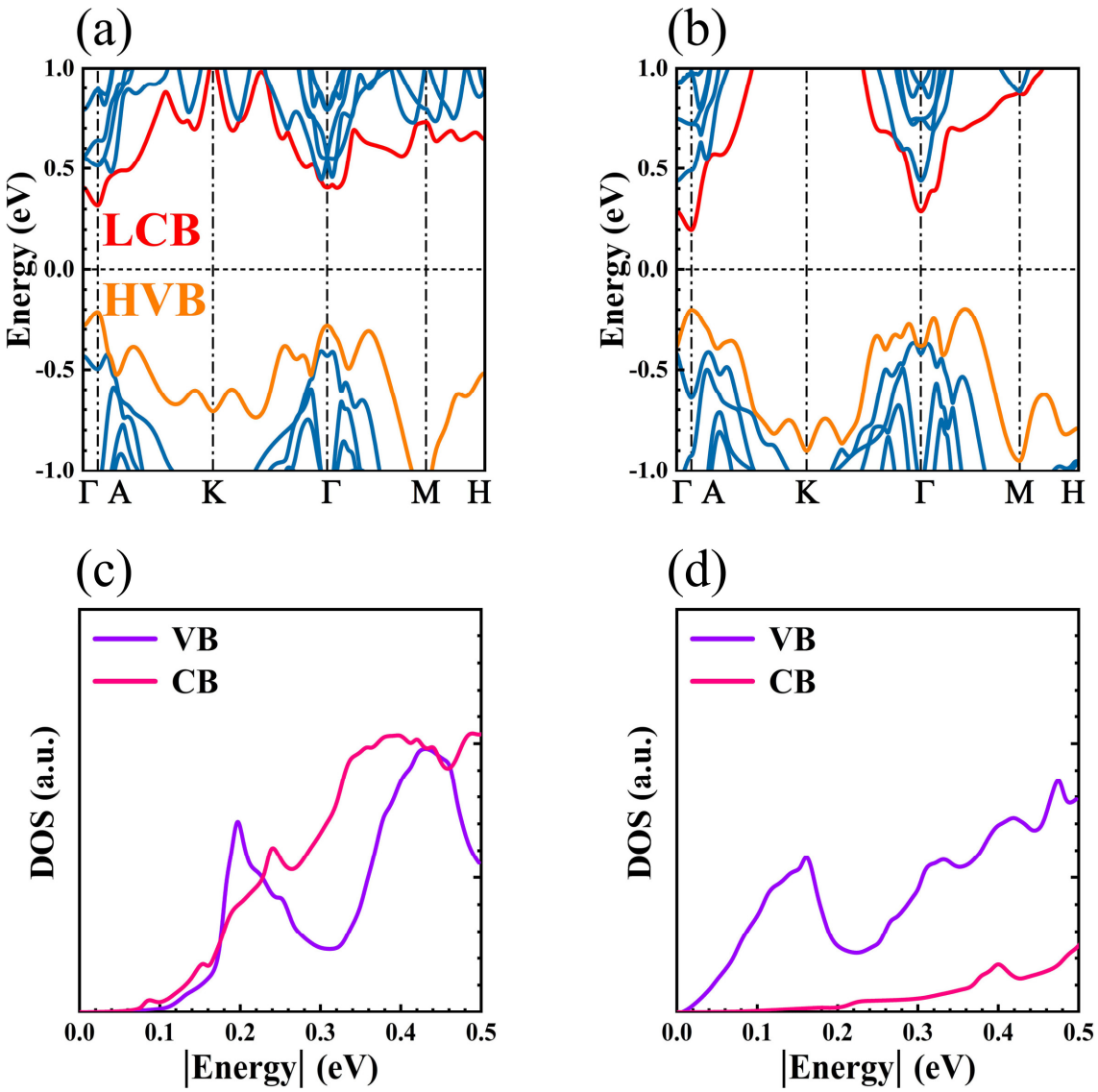
Supplementary Figure 4 DFT-calculated optimal doping type. Calculated optimal doping type at different temperatures for **a** Se-based and **b** Te-based input compounds.



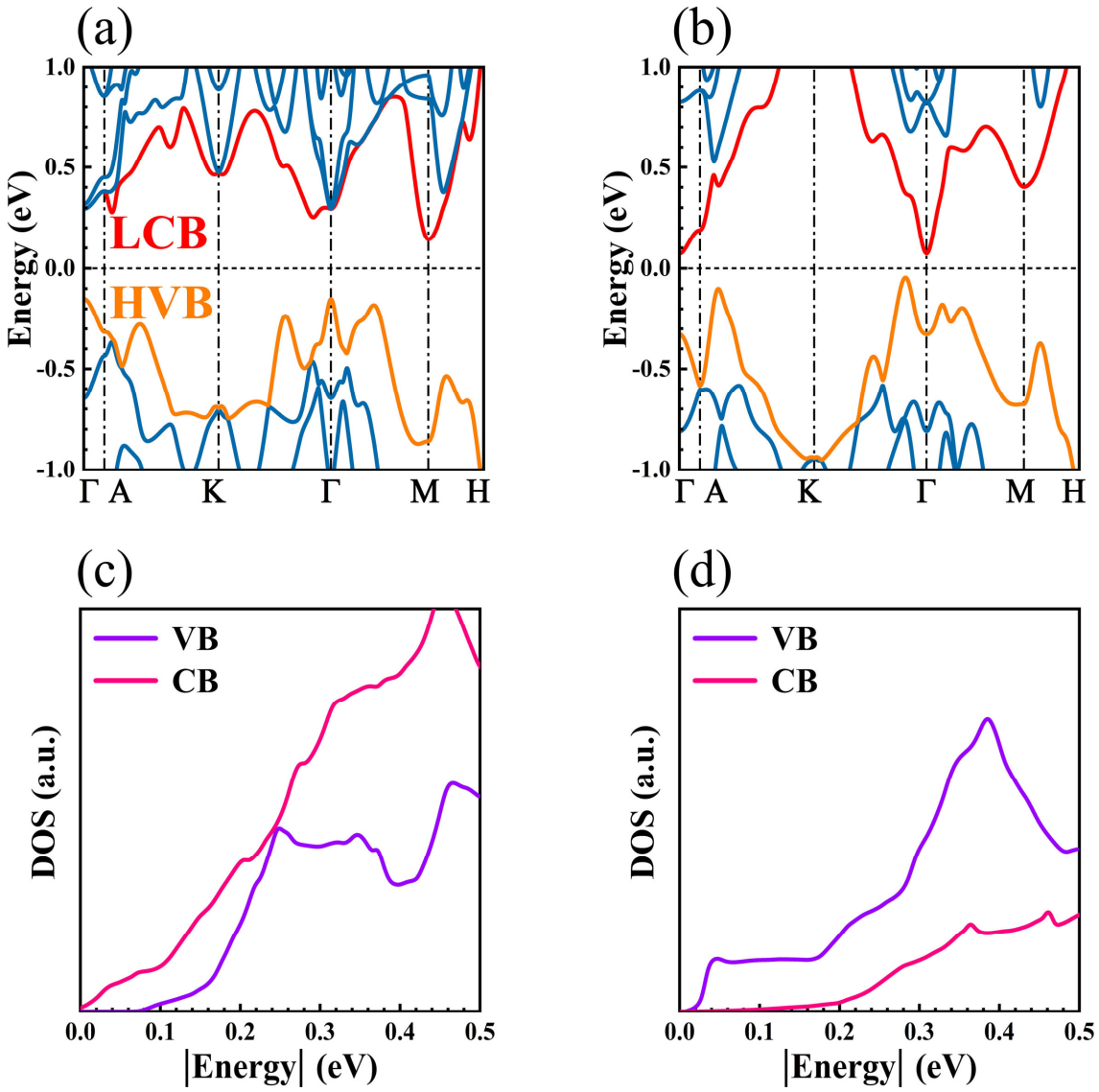
Supplementary Figure 5 Electronic band structures. Calculated band structures of **a** $Pb_2Sb_2S_5$, **b** $Pb_2Sb_2Se_5$ and **c** $Pb_2Sb_2Te_5$ using GGA-PBE (red lines) and HSE06 functionals (purple dotted lines), respectively.



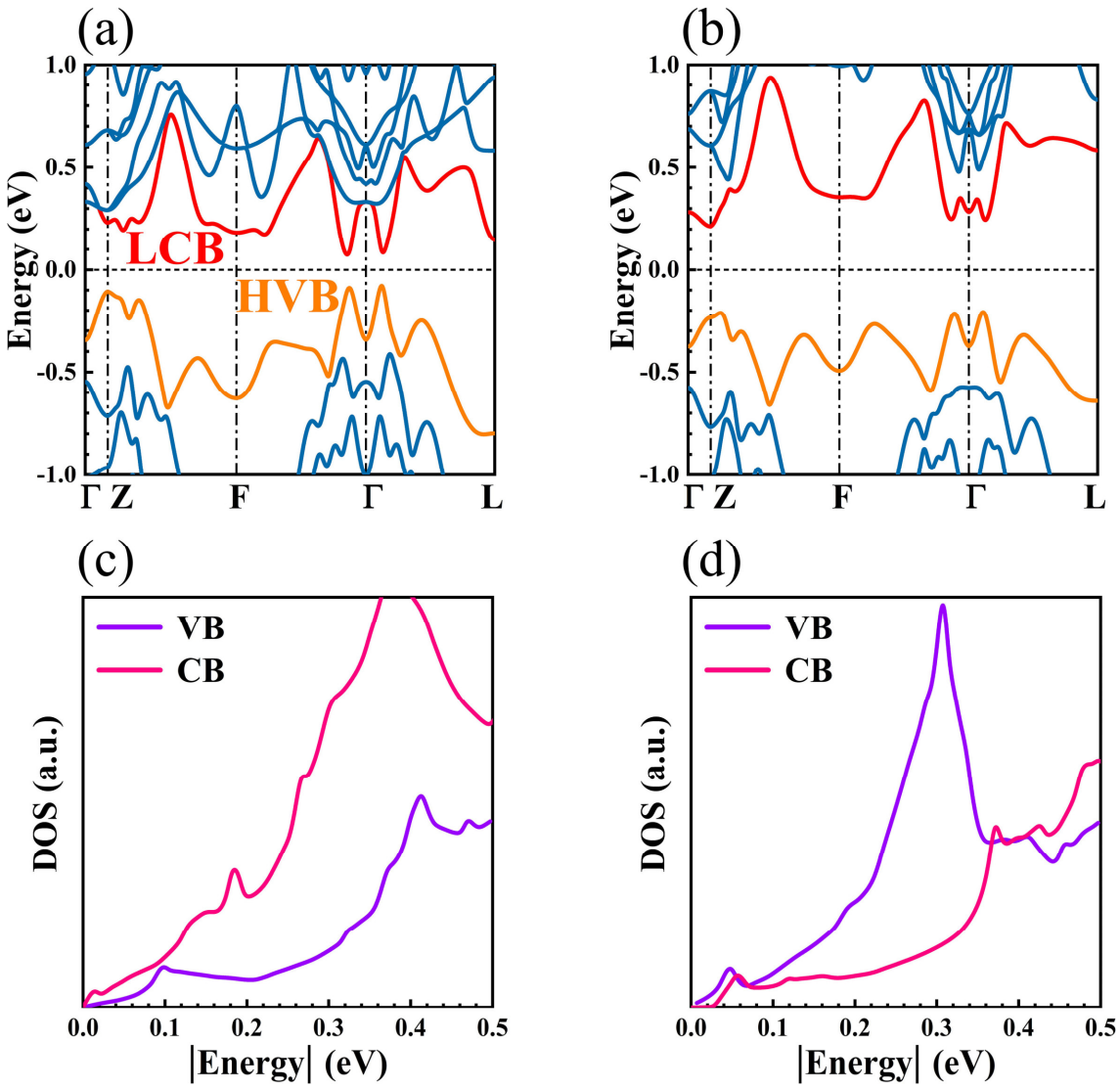
Supplementary Figure 6 Electronic structures around Fermi energy. a, b Electronic band structures and **c, d** total DOS near the VBM and CBM for **a, c** $Si_1Bi_2Te_4$ and **b, d** $Sn_1As_2Se_4$.



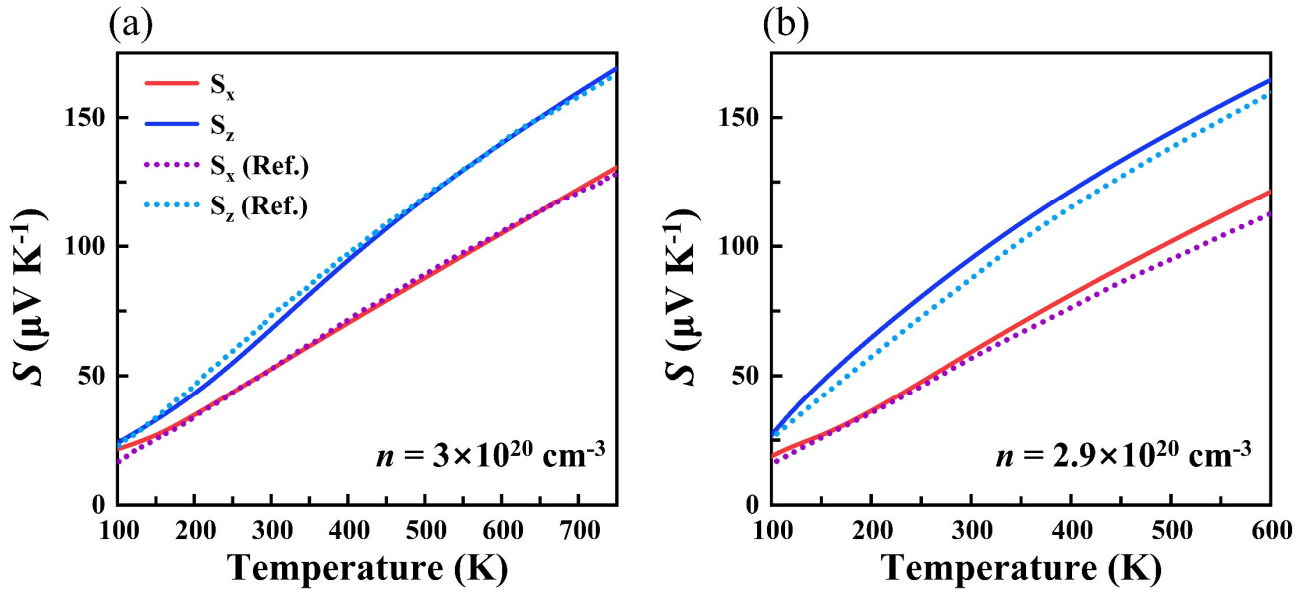
Supplementary Figure 7 Electronic structures around Fermi energy. Band structures and total DOS of **a, c** $Si_1Bi_4Te_7$ and **b, d** $Ge_1As_4Se_7$.



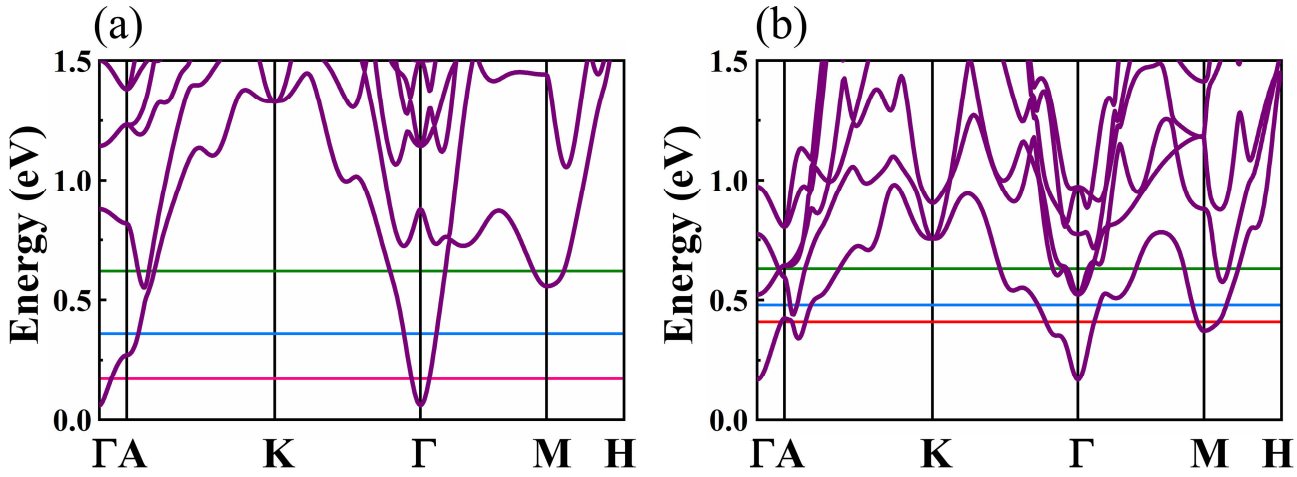
Supplementary Figure 8 Electronic structures around Fermi energy. Band structures and total DOS of **a, c** $Si_2Bi_2Te_5$ and **b, d** $Sn_2As_2Se_5$.



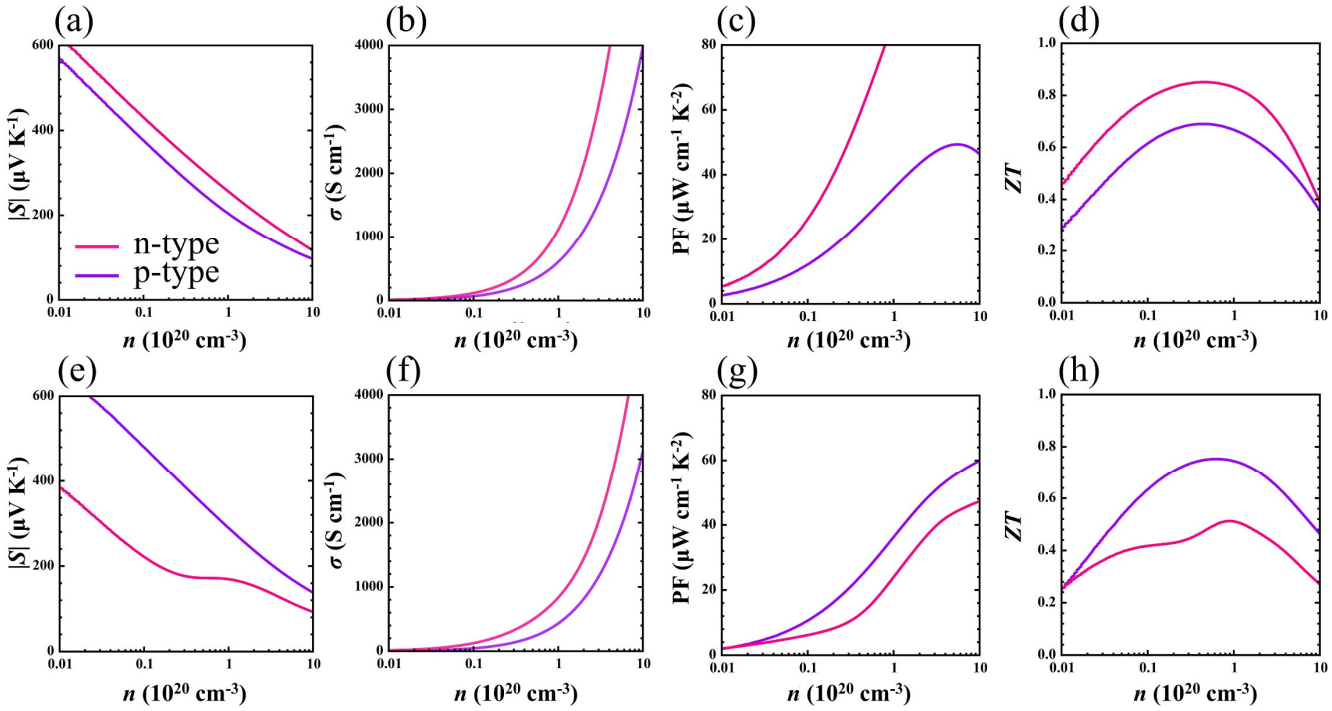
Supplementary Figure 9 Electronic structures around Fermi energy. Band structures and total DOS of top compounds **a, c** $Sn_3As_2Te_6$ and **b, d** $Ge_3As_2Se_6$.



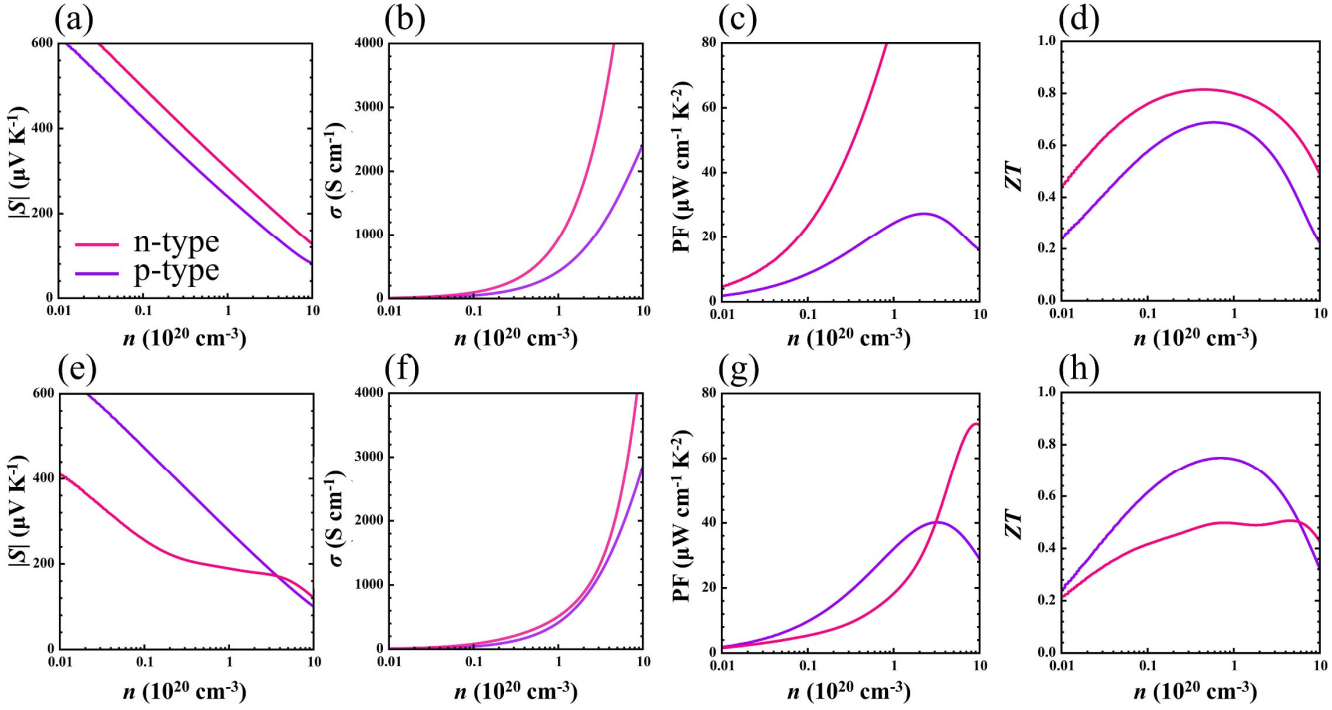
Supplementary Figure 10 Calculated Seebeck coefficients. Temperature-dependent Seebeck coefficient of **a** $Ge_1Sb_2Te_4$ and **b** $Ge_2Sb_2Te_5$, comparing with the previous theoretical simulated results at fixed hole carrier concentrations.[1]



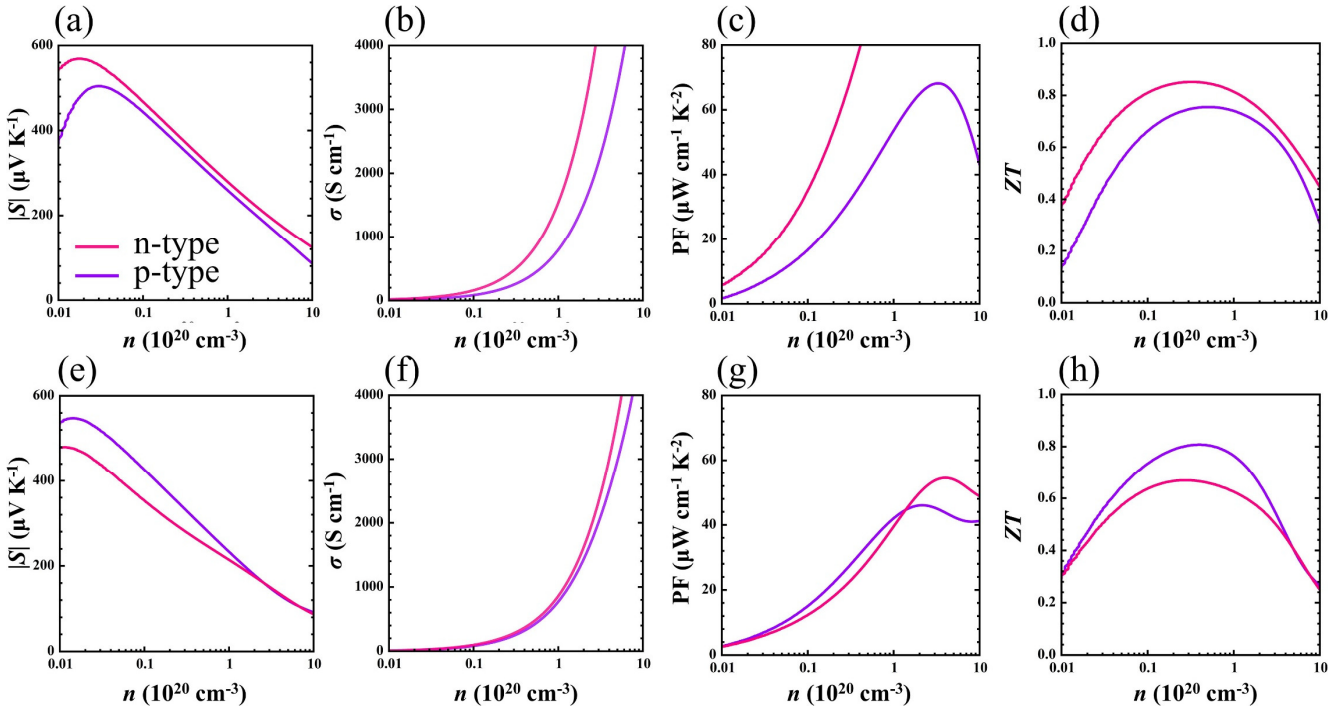
Supplementary Figure 11 Fermi energy change. Fermi energy change at the electron carrier concentrations of 1×10^{19} cm^{-3} (red lines), 1×10^{20} cm^{-3} (blue lines) and 1×10^{21} cm^{-3} (green lines) for **a** $Pb_2Sb_2Se_5$ and **b** $Pb_2Sb_2Te_5$, respectively.



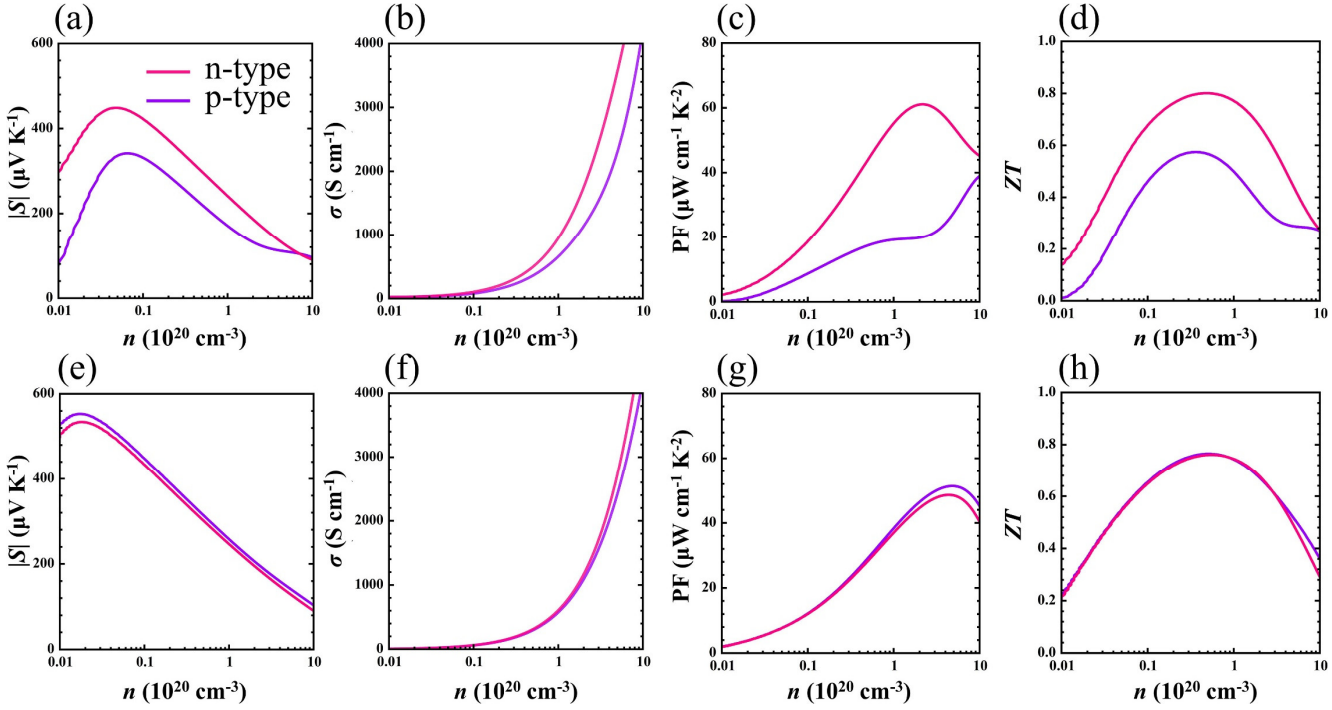
Supplementary Figure 12 Electronic transport properties. a, e Absolute Seebeck coefficient |*S*|, **b, f** electrical conductivity σ, **c, g** power factor PF, and **d, h** figure of merit *ZT* of **a-d** $Si_1Bi_2Te_4$ and **e-h** and $Sn_1As_2Se_4$ as a function of carrier concentration for p-type and n-type at 650 K.



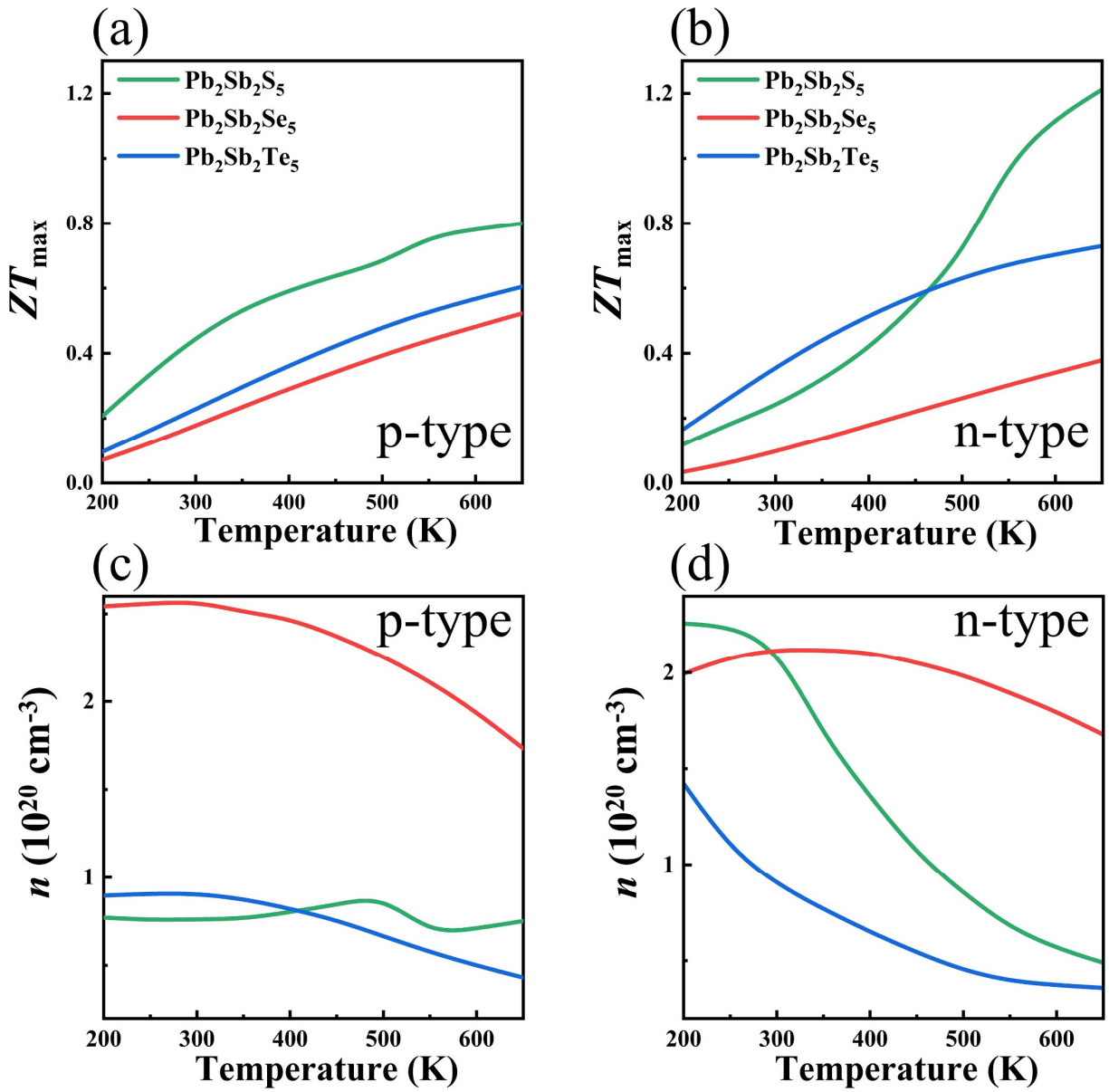
Supplementary Figure 13 Electronic transport properties. Electronic transport properties and *ZT* of **a-d** $Si_1Bi_4Te_7$ and **e-h** $Ge_1As_4Se_7$.



Supplementary Figure 14 Electronic transport properties. Electronic transport properties and *ZT* of **a-d** $Si_2Bi_2Te_5$ and **e-h** $Sn_2As_2Se_5$.



Supplementary Figure 15 Electronic transport properties. Electronic transport properties and *ZT* of **a-d** $Sn_3As_2Te_6$ and **e-h** $Ge_3As_2Se_6$.



Supplementary Figure 16 ZT_{max} and optimal carrier concentration. Temperature dependence of **a, b** ZT_{max} and **c, d** optimal carrier concentration under p-type and n-type dopings for $Pb_2Sb_2S_5$, $Pb_2Sb_2Se_5$ and $Pb_2Sb_2Te_5$.

SUPPLEMENTARY REFERENCES

1. Ibarra-Hernández, W. & Raty, J.-Y. Ab initio density functional theory study of the electronic, dynamic, and thermoelectric properties of the crystalline pseudobinary chalcogenide $(GeTe)_x/(Sb_2Te_3)$ (x = 1, 2, 3). *Phys. Rev. B* **97**, 245205 (2018).