


 Cite this: *RSC Adv.*, 2025, 15, 42034

Data-driven exploration of Na–Bi compounds: a first-principles and machine learning approach to topological thermoelectrics

 Souraya Goumri-Said,^a Houssam Eddine Hailouf,^b Khalid Reggab,^b Kingsley Onyebuchi Obodo^{cd} and Mohammed Benali Kanoun^{de}

We conduct an in-depth investigation of the structural, electronic, vibrational, thermodynamic, and thermoelectric characteristics of Na–Bi-based compounds, specifically tetragonal NaBi, hexagonal NaBi₃, and cubic Na₃Bi, using advanced first-principles calculations in conjunction with machine learning (ML) models. We used density functional theory (DFT) with spin–orbit coupling (SOC) to figure out the electronic structure, phonon dispersions, and thermoelectric transport using Boltzmann transport theory. Our findings validate the Dirac semimetal nature of cubic Na₃Bi and demonstrate varied topological and thermodynamic properties within the Na–Bi family. To speed up the prediction of the thermoelectric figure of merit (*ZT*) while enhancing interpretability to understand at the feature level, we trained supervised ML models [Random Forest (RF) and Neural Network (NN)] on thermoelectric results from DFT. It is possible to directly compute the figure of merit (*ZT*) from DFT-derived transport coefficients such as the Seebeck coefficient, electrical conductivity, and thermal conductivity. However, machine learning (ML) models serve as powerful surrogate predictors, enabling rapid screening of derivative compounds and quantitative assessment of feature importance through SHAP (SHapley Additive exPlanations) analysis. At low temperatures, RF models consistently outperformed NN models, but both performed well at high temperatures. SHAP analysis showed that the Seebeck coefficient has the biggest effect on *ZT* in all regimes. This integrated, physics-informed, and data-driven methodology demonstrates that machine learning can significantly augment first-principles approaches. It accelerates predictions, guides feature prioritisation, and enhances design capabilities. The developed workflow provides a generalizable and interpretable framework for the predictive modeling of advanced topological thermoelectric materials.

 Received 11th August 2025
 Accepted 17th October 2025

DOI: 10.1039/d5ra05888k

rsc.li/rsc-advances

1. Introduction

Topological quantum materials, particularly insulators and semimetals, have attracted considerable attention for their unique band structures arising from inverted bulk bands and relativistic fermion states. Among these, sodium bismuthide compounds, including trisodium bismuthide (Na₃Bi) and equiatomic NaBi, have emerged as prototypical systems for exploring such phenomena. Na₃Bi was the first three-dimensional (3D) Dirac semimetal to be confirmed. It has

symmetry-protected Dirac nodes along the *k_x* axis of the Brillouin zone, where bulk conduction and valence bands (mostly Na 3s and Bi 6p) are flipped because of strong spin–orbit coupling (SOC).^{1,2} When time-reversal or inversion symmetries are broken, these Dirac points can split into Weyl nodes. This is a classic example of Dirac semimetal behaviour.³ This leads to observable transport phenomena, including negative longitudinal magnetoresistance due to the chiral anomaly and weak anti-localization indicative of the π Berry phase of the Dirac fermions.¹ Na₃Bi also has a unique tunable optical response in the infrared, with a mid-infrared transparency window that can be changed by changing the carrier density. This shows that it could be useful for new optoelectronic applications.^{4,5} On the other hand, NaBi is a “topological metal” because it has a bulk band inversion and a nontrivial Z_2 topology, even though it is metallic.² Its SOC-driven inverted bands create Dirac-like surface states, which means it is in between topological insulators and metals. NaBi also has unique lattice dynamics, such as predicted ultralow and anisotropic lattice thermal conductivity and a closeness to superconductivity through electron–

^aCollege of Science and General Studies, Department of Physics, Alfaisal University, P.O. Box 5092, Riyadh 11533, Saudi Arabia. E-mail: sosaid@alfaisal.edu

^bMaterials Science and Informatics Laboratory, Faculty of Science, University of Djelfa, 17000, Djelfa, Algeria

^cSchool of Agriculture & Science, University of KwaZulu-Natal, Pietermaritzburg, Scottsville, 3209, South Africa

^dCenter for Space Research, North-West University, Private Bag X6001, Potchefstroom, 2531, South Africa

^eDepartment of Mathematics and Sciences, College of Sciences and Humanities, Prince Sultan University, P.O. Box 66833, Riyadh 11586, Saudi Arabia



phonon coupling. This makes it a good platform for studying how topology, lattice vibrations, and superconductivity work together.² Both Na₃Bi and NaBi can change their topological phases when pressure is applied. Na₃Bi transitions from a Dirac semimetal to a cubic phase at ~0.8 GPa, maintaining band inversion akin to HgTe, and eventually becomes a trivial insulator at >118 GPa.³ Similarly, NaBi transforms into a cubic phase above ~36 GPa, with anticipated modifications to its topological character.³ These transitions illustrate the critical link between crystal structure and band topology. Chemical alloying and dimensional reduction further enrich the phase diagram of Na–Bi compounds. In Na₃Bi_{1-x}Sb_x alloys, increasing Sb content reduces the effective SOC, driving a topological phase transition to a trivial insulator.⁶ Likewise, ultrathin Na₃Bi films exhibit a thickness-dependent crossover from a 3D Dirac semimetal to a 2D quantum spin Hall (QSH) insulator state, with helical edge states emerging at sub-four-monolayer thicknesses.⁴ This tunability makes Na–Bi materials attractive candidates for low-power electronic and spintronic devices.

The goal of this study is to use a combination of first-principles density functional theory (DFT) and machine learning (ML) models to look at the structural, electronic, vibrational, thermodynamic, and thermoelectric properties of Na–Bi compounds. These include tetragonal NaBi, hexagonal NaBi₃, and cubic Na₃Bi. The goal is to find out if they could be used as topological thermoelectric materials and develop a suitable method to predict how effectively they will work as thermoelectric materials. In addition to first-principles thermoelectric modeling, this work integrates supervised machine learning (ML) to predict *ZT* from key transport descriptors (Seebeck coefficient, electrical and thermal conductivity). While *ZT* can be computed directly, this ML framework is introduced to (i) enable rapid prediction and screening of thermoelectric behavior once features are known, (ii) quantify the relative impact of each transport parameter on *ZT* using SHAP interpretability analysis, and (iii) demonstrate the feasibility of physics-informed surrogate modeling. This proof-of-concept provides a pathway for extending ML prediction to Na–Bi alloys, doped variants, or related compounds where DFT calculations are expensive or incomplete. The paper is organized as follows: Section 2 describes the computational methods employed for DFT, phonon, thermodynamic, and transport property calculations, along with the implementation of ML models. Section 3 presents and discusses the results, including structural optimization, electronic band structures, phonon dispersions, thermodynamic behavior, thermoelectric properties, and ML-based prediction and feature analysis. The key conclusions and future directions are summarized in Section 4.

2 Theoretical methodologies and computational details

2.1 Density functional theory

In this work, spin–orbit coupling (SOC)-inclusive density functional theory (DFT) calculations were carried out using the plane-wave-based Quantum ESPRESSO package.⁷ To precisely

capture electron–ion interactions, we employed both ultrasoft⁸ and norm-conserving Vanderbilt pseudopotentials,⁹ utilizing fully relativistic large-core formulations.¹⁰ Our simulations were conducted on NaBi in its tetragonal, NaBi₃ in its hexagonal, and Na₃Bi in its cubic phases. The valence electron configurations for the elements under investigation were treated as follows: Na: 3s¹, and Bi: 6s²6p³. For geometric relaxation, the exchange correlation GGA-PBE + SOC¹¹ functional was employed. Structural optimization was carried out using the Broyden–Fletcher–Goldfarb–Shanno (BFGS)¹² approach, employing a 12 × 12 × 8 Monkhorst–Pack grid.¹³ To ensure energy convergence, we imposed the condition that the major component of Hellmann–Feynman forces acting on a single atom must be less than 1 × 10^{−10} eV Å^{−1}. Furthermore, to elucidate crystal structure and stability, we employed hybrid Heyd–Scuseria–Ernzerhof (HSE)¹⁴ computations, which incorporate SOC effects and an adaptively compressed exchange operator to enhance computational efficiency. Subsequently, we investigated the electronic and optical properties of the materials under study. Band structures incorporating SOC effects were computed using the HSE + SOC method, with Wannier functions¹⁵ interpolated using the Wannier Tools code.¹⁶

2.2 Thermodynamic properties

In order to determine the lattice dynamical stabilities for materials, the phonon dispersion ($\omega_j = \omega_j \vec{q}$ for the association between $\omega_{\vec{q}j}$ and \vec{q} for each *j* mode) is studied as the frequency dependence on the wave vector attained within linear response theory/density functional perturbation theory (DFPT) to obtain the response to periodic perturbations.¹⁷ The phonon density of states $g(\omega)$ gives the frequency distribution of normal modes and is calculated as:

$$g(\omega) = \frac{1}{nd\Delta\omega} \sum_{k,j} (\omega - \omega(k,j)) \quad (1)$$

$$\text{where } \delta_{\Delta\omega}(x) = \begin{cases} 1 & \frac{\Delta\omega}{2} < x \leq \frac{\Delta\omega}{2} \\ 0 & \text{otherwise} \end{cases}$$

Thermodynamic favorability is determined by a relative free energy, so that the Helmholtz free energy is favored over Gibbs free energy for its universality of expressing reversible work at constant temperatures. The Helmholtz free energy for a nonmagnetic ideal crystal is expressed as:^{18,19}

$$A(V,T) = U_0(V) + A_{\text{el}}(V,T) + A_{\text{phon}}(V,T) \quad (2)$$

where $U_0(V)$ is the total energy, $A_{\text{el}}(V,T)$ is the electronic excitation contribution and $A_{\text{phon}}(V,T)$ is the nuclear vibrational motion or the phonon vibration contribution from phonon DOS $g(\omega)$ which is dependent the temperature. The vibrational partition function is given by:

$$Z_{\text{vib}} = \prod_{\vec{q}j} \sum_{n=0}^{\infty} e^{-\left(n + \frac{1}{2}\right) \frac{\hbar\omega_{\vec{q}j}}{k_{\text{B}}T}} \quad (3)$$



Therefore, the vibrational Helmholtz energy in the harmonic approximation is formulated as:

$$A_{\text{HA}}(V_{\text{eq}}, T) = U_0(V) + \frac{1}{2} \sum_{\bar{q}j} \hbar \omega_{\bar{q}j} + k_{\text{B}} T \sum_{\bar{q}j} \log \left(1 - e^{-\frac{\hbar \omega_{\bar{q}j}}{k_{\text{B}} T}} \right) \quad (4)$$

It relies on the temperature only by phonon contributions calculated for the equilibrium volume V_{eq} . The amplitude of the vibrations is eased, thus validating the harmonic approximation for practically all situations at low temperatures. The isochoric heat capacity (C_v) for N_a atoms per unit cell is:

$$C_v = -T \left(\frac{\partial^2 A}{\partial T^2} \right)_V = k_{\text{B}} N_a \sum g(\omega) \left(\frac{\hbar \omega}{k_{\text{B}} T} \right)^2 \frac{\exp(\hbar \omega_{\bar{q}j} / k_{\text{B}} T)}{\exp(\hbar \omega_{\bar{q}j} / k_{\text{B}} T) - 1} \quad (5)$$

Yet, the truncation of the third term ignores the anharmonicity in the total energies, which causes some errors including, infinite values in thermal conductivity and phonon lifetimes as well as null quantities for the thermal expansion.¹⁸ Thus, considering temperature effects alongside vibrational degrees of freedom requires assuming a rigid harmonic approximation for the geometry, even as the crystal structure deviates from equilibrium (*i.e.* to some degree encompassing anharmonic contributions).²⁰ This technique is referred to as the quasi-harmonic approximation (QHA) which presents the computation of phonons for multiple volumes $\omega_{\bar{q}j}(V)$ so that the Helmholtz free energy turns into:

$$A_{\text{QHA}}(V, T) = U_0(V) + \frac{1}{2} \sum_{\bar{q}j} \hbar \omega_{\bar{q}j}(V) + k_{\text{B}} T \sum_{\bar{q}j} \log \left(1 - e^{-\frac{\hbar \omega_{\bar{q}j}(V)}{k_{\text{B}} T}} \right) \quad (6)$$

where the first term and second term are grouped as the cold potential energy $U_{\text{cold}}(V)$ at $T = 0\text{K}$ and the second term is the thermal factor contributed by the phonons $A_{\text{th}}(V)$ which becomes negligible at extremely low T . Thus, the entropy (S) where the phonon frequency is volume dependent, with the neglect of the temperature dependent intrinsic phonon interactions, is written as:²¹

$$S_{\text{QHA}}(V) = -\frac{1}{N} \sum_{\bar{q}j} k_{\text{B}} \ln \left[1 - \exp \left(-\frac{\hbar \omega_{\bar{q}j}}{k_{\text{B}} T} \right) \right] + \frac{1}{N} \sum_{\bar{q}j} k_{\text{B}} \frac{\hbar \omega_{\bar{q}j}(V)}{k_{\text{B}} T} \left[\exp \left(\frac{\hbar \omega_{\bar{q}j}}{k_{\text{B}} T} \right) - 1 \right]^{-1} \quad (7)$$

The thermodynamical modeling procedure which exhibit a comprehensive DOS for the vibrational modes is achieved through the Debye model which is simplified to:^{19,20}

$$g(\omega) = C \cdot \omega^2 \cdot \Theta(\omega - \omega_{\text{D}}) \quad (8)$$

where Θ is the Heaviside step function, and C is a constant for the $C \int g(\omega) d\omega = 3N_a$ is expressed as $C = 9N_a / \omega_{\text{D}}^3$. The term ω_{D} is the Debye frequency, wherein the phonon modes are populated below it. The determination of the phonon Helmholtz free energy and the isochoric heat capacity in the Debye model are represented as:

$$A_{\text{phon}}^{\text{Debye}}(T) = N_a k_{\text{B}} T \left(\frac{9T_{\text{D}}}{8T} + 3 \ln \left(1 - e^{-\frac{T_{\text{D}}}{T}} \right) - D \left(\frac{T_{\text{D}}}{T} \right) \right) \quad (9)$$

$$C_v^{\text{Debye}} = 3N_a k_{\text{B}} \left(4D \left(\frac{T_{\text{D}}}{T} \right) - \frac{3T_{\text{D}}/T}{e^{T_{\text{D}}/T} - 1} \right) \quad (10)$$

where $D \left(\frac{T_{\text{D}}}{T} \right) = \frac{3}{(T_{\text{D}}/T)^3} \int_0^{T_{\text{D}}/T} \frac{x^3}{e^x - 1} dx$ is the Debye integral and $T_{\text{D}} = \frac{\hbar}{k_{\text{B}}} \left[6\pi^2 V^{1/2} N_a \right]^{1/3} f(\sigma) \sqrt{\frac{B_s}{M}}$ is the Debye temperature

which is the temperature where every mode under the highest frequency mode ω_{D} is excited (maximum phonon frequency).²²

2.3 Thermoelectric properties: boltzman transport theory

The thermoelectric characteristics are evaluated using the semi-classical Boltzmann transport theory, implemented *via* the BoltzTrap software in combination with first-principles calculations and the relaxation time approximation.²³ This approach models the statistical behavior of charge carriers in systems slightly deviating from thermal equilibrium. The foundation of the Boltzmann transport equation lies in the assumption that particles (such as electrons) undergo random motion within the material. To overcome the difficulty in estimating the relaxation time, the deformation potential theory is used in a theoretical context.

The electronic transport is defined through the electrical and heat currents J_e and J_Q as:

$$J_e = \sigma E - \sigma S \nabla T \quad (11)$$

$$J_Q = T \sigma S E - \kappa_0 \nabla T \quad (12)$$

where σ is the electrical conductivity, S is the Seebeck coefficient, κ is the thermal conductivity, E is the electric field, and ∇T is the temperature gradient. By expressing a simplified transport distribution Ξ , the transport properties are easily computed through:

$$\Xi_{\alpha\beta}(E) = \frac{1}{\Omega_c} \sum_k \sum_n v_{kn}^{\alpha} v_{kn}^{\beta} \tau_{kn} \delta(E - \varepsilon_{kn}) \quad (13)$$

The σ and S are computed obeying equations which are termed as a function of chemical potential (μ) and temperature (T).²⁴

$$\sigma_{\alpha\beta}(T, \mu) = \frac{1}{\Omega} \int \sigma_{\alpha\beta} \left[-\frac{\partial f_0(T, \varepsilon, \mu)}{\partial \varepsilon} \right] d\varepsilon \quad (14)$$



$$S_{\alpha\beta}(T, \mu) = \frac{1}{eT\Omega\sigma_{\alpha\beta}(T, \mu)} \int \sigma_{\alpha\beta}(\varepsilon)(\varepsilon - \mu) \left[-\frac{\partial f_0(T, \varepsilon, \mu)}{\partial \varepsilon} \right] d\varepsilon \quad (15)$$

$$\sigma_{\alpha\beta}(\varepsilon) = \frac{1}{N} \sum_{i,k} \sigma_{\alpha\beta}(i, k) \frac{\delta(\varepsilon - \varepsilon_{i,k})}{\delta(\varepsilon)} \quad (16)$$

$$\sigma_{\alpha\beta}(i, \vec{k}) = e^2 \tau_{i,k} v_{\alpha}(i, \vec{k}) v_{\beta}(i, \vec{k}) \quad (17)$$

Here, α and β denote tensor components, e represents the elementary charge, and f_0 is the Fermi-Dirac distribution function. The symbol Ω refers to the volume of the unit cell, μ is the chemical potential, and τ is the relaxation time, assumed constant at 10^{-14} s. The group velocity components, indicated as $v_{\alpha}(i, \vec{k})$, are obtained from the electronic band structure. When τ is considered constant and independent of the wave vector \vec{k} and carrier energy, it cancels out in the ratio between the two integrals used in the Seebeck coefficient calculation.²⁵ The use of a constant relaxation time ($\tau = 10^{-14}$ s) is a standard and practical approximation in Boltzmann transport modeling, particularly for high-throughput screening, as it enables a focus on the intrinsic electronic-structure-driven trends that govern thermoelectric behavior.²⁶ This assumption is most appropriate for scattering mechanisms with weak energy dependence; such as neutral impurity or alloy scattering; and serves as a reasonable first-order approximation at elevated temperatures where acoustic phonon scattering dominates and its energy dependence becomes less pronounced. While this simplification omits the full complexity of energy-dependent scattering from acoustic and optical phonons or ionized centers,^{27,28} it remains sufficient for the comparative analysis pursued here. As discussed by Singh and co-workers,²⁹ BoltzTraP-based calculations under constant τ are most reliable for identifying qualitative and relative trends in transport properties rather than absolute ZT magnitudes, which may deviate near the optimized chemical potential where multiple scattering channels compete. Therefore, the ZT values reported in this work represent indicative trends, not experimentally calibrated maxima, and the large Seebeck coefficients observed for Na_3Bi arise directly from its Dirac-like band dispersion and low density of states near the Fermi level.

2.4 Machine learning models

Random Forest is a non-parametric ensemble learning method^{30,31} composed of multiple decision trees, each trained on a bootstrap sample of the data with randomly selected subsets of features. The regression prediction is an average over all trees:^{32,33}

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (18)$$

where, \hat{y} predicted output (e.g., ZT), N : number of trees in the forest, $T_i(x)$: output of the i -th decision tree for input x .

Neural Networks,^{34,35} by contrast, consist of layers of interconnected nodes that transform the input through weighted

sums and nonlinear activation functions. The basic structure used in this study is:

$$\hat{y} = f(W_2 \cdot \text{ReLU}(W_1 x + b_1) + b_2) \quad (19)$$

where, x : input feature vector (e.g., S , σ , κ), W_1 , W_2 : weight matrices, b_1 , b_2 : bias vectors, $\text{ReLU}(z) = \max(0, z)$: activation function, f : output activation (identity for regression), \hat{y} : predicted ZT .

Neural networks can capture complex, highly nonlinear relationships, but require careful regularization and sufficient data to avoid overfitting.

3. Results and discussion

3.1 Crystal structure and optimization

Fig. 1 shows the optimized crystal structures of (a) tetragonal NaBi , (b) hexagonal NaBi_3 , and (c) cubic Na_3Bi , obtained through first-principles calculations using density functional theory (DFT) with spin-orbit coupling (SOC). Structural relaxations were performed using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm³⁶ until the total energy converged below 10^{-8} Ry and Hellmann-Feynman forces on each atom dropped below 10^{-4} Ry Bohr⁻¹.

To make sure that the forces, stress, and electronic properties converged correctly, we used a plane-wave energy cutoff of 80 Ry and a Monkhorst-Pack grid of $12 \times 12 \times 8$.³⁷ Table 1 shows the optimised lattice constants and the thermodynamic properties that go with them. The relaxed lattice parameters for tetragonal NaBi were $a = 3.4052$ Å, $c = 4.8671$ Å, and the c/a ratio was 1.4293. This corresponds to a unit cell volume of 56.44 Å³. The predicted formation energy is -0.381 eV/atom, which means that the system is moderately thermodynamically stable. Hexagonal NaBi_3 , on the other hand, has $a = 6.9326$ Å, $c = 5.7899$ Å, and a lower c/a ratio of 0.8351. It also has a much larger volume of 240.98 Å³ and a more negative formation energy of -0.498 eV per atom, indicating higher thermodynamic stability. The cubic Na_3Bi structure has a lattice constant of $a = 7.6478$ Å, a volume of 447.31 Å³, and a formation energy of -0.482 eV per atom. This is in line with previous experimental and theoretical studies that demonstrated it was a very stable 3D Dirac semimetal.^{38,39} We employed the standard method for determining out the formation energies:

$$E_{\text{form}} = \frac{E_{\text{tot}}(\text{compound}) - \sum_i n_i E(\text{element}_i)}{\sum_i n_i} \quad (20)$$

where E_{tot} is the total energy of the compound, n_i is the number of atoms of element i , and $E(\text{element}_i)$ is the energy of each elemental reference in its standard state. All three phases have negative formation energies, which means that the compounds are thermodynamically stable against breaking down into their constituent elements. This is in line with the rules set out in materials discovery frameworks like the Materials Project⁴⁰ and AFLOW.⁴¹ The rising formation energy from NaBi to NaBi_3 suggests that adding more Bi atoms could make the structure more stable by making the Na-Bi bonds stronger and spreading



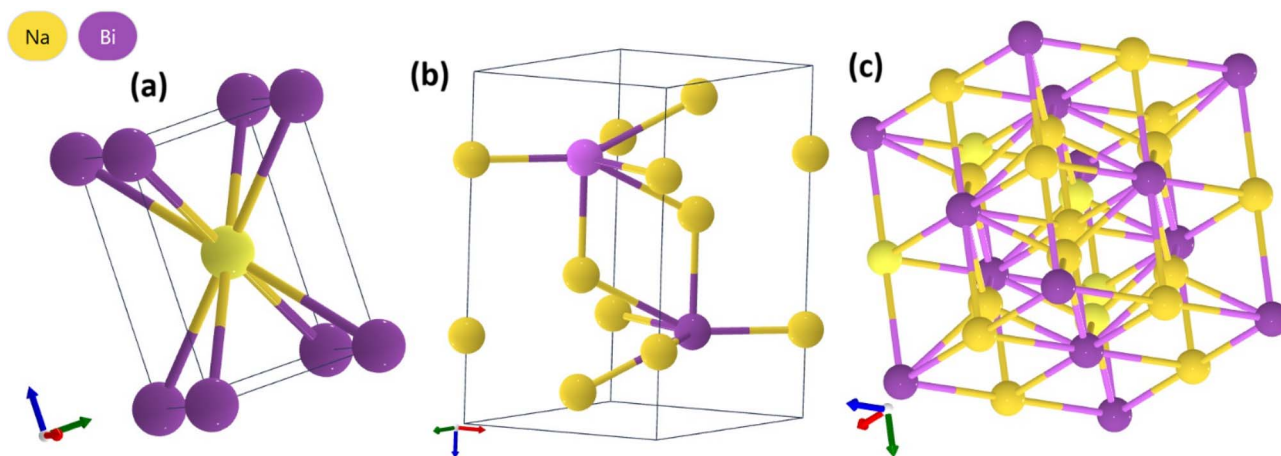


Fig. 1 Optimized crystals structures of (a) NaBi tetragonal, (b) NaBi₃ hexagonal and (c) cubic Na₃Bi using GGA-PBE + SOC.

Table 1 Optimized lattice parameters (*a*, *c*, *c/a*), unit cell volumes, and formation energies of NaBi (tetragonal), NaBi₃ (hexagonal), and Na₃Bi (cubic) from DFT-GGA + SOC calculations. Experimental lattice constants (where available) are included for comparison^a

Compound & structure	Source	<i>a</i> (Å)	<i>c</i> (Å)	<i>c/a</i>	Volume (Å ³)	Formation energy (eV per atom)
NaBi – tetragonal (<i>P4/mmm</i>)	This work (DFT)	3.4052	4.8671	1.4293	56.4359	−0.381
	Exp ⁴⁴	3.46	4.80	1.39	57.5	−0.329
	DFT ⁴⁴	3.42	4.89	1.43	57.6	−0.366
NaBi ₃ – hexagonal (<i>P6₃/mmc</i>)	This work (DFT)	6.9326	5.7899	0.8351	240.9801	−0.498
	Exp ³⁸	5.448	9.655	1.77	248.2	−0.495
	DFT ³⁸	5.458	9.704	1.78	251.0	−0.367
Na ₃ Bi – hexagonal (<i>P6₃/mmc</i>)	DFT ⁴³	5.448	9.655	1.77	248.2	—
	DFT ⁴⁵	5.37	9.64	1.80	240.7	—
Na ₃ Bi – cubic (<i>Fm$\bar{3}$m</i>)	This work (DFT)	7.6478	—	—	447.3109	−0.482

^a Note: Na₃Bi is typically reported in hexagonal structure (*P6₃/mmc*) in literature. This study focuses on the cubic polymorph (*Fm $\bar{3}$ m*).

the charge more evenly. These results show that the Na–Bi family has a wide range of structures and is stable when it comes to thermodynamics. This gives us a reason to look into their electronic, topological, and optoelectronic properties more closely.

To validate the reliability of our DFT-GGA + SOC approach, we compared the optimized lattice parameters with available experimental values, as shown in Table 1. The deviations are within $\pm 1\%$, which shows that the structural models used for later electronic and thermoelectric analyses are correct. These minor discrepancies are characteristic of GGA-based functionals and align with prior experimental and theoretical research on Na–Bi systems.^{42–48} This comparison strengthens the accuracy of our theoretical framework in capturing the essential structural characteristics of Na–Bi compounds.

3.2 Electronic properties

Fig. 2 shows the electronic band structures of three Na–Bi-based compounds: tetragonal NaBi, hexagonal NaBi₃, and cubic Na₃Bi. These were calculated using the HSE06 hybrid functional with spin-orbit coupling (SOC). For tetragonal NaBi (Fig. 2a),

the Fermi level is in a band that is only partially filled, which means that the material exhibits metallic or semimetallic behavior. The fact that there is band inversion near the *T*-point because Bi has a strong SOC means that the material may have nontrivial topological states, which makes it a possible topological semimetal. The hexagonal NaBi₃ phase (Fig. 2b) has a band structure that is both dense and highly dispersive, with many crossings at the Fermi level. This suggests that it is a metal. The bands near high-symmetry points (*T*, *A*, *L*) are very complicated, which implies that the Fermi surface is quite rich. There is no clear band gap, but it is still possible that the material has nodal-line or Dirac semimetal character, suggesting the need for further topological surface state analysis.

In contrast, cubic Na₃Bi (Fig. 2c) shows a well-defined Dirac crossing along the *T*–*Z* direction, slightly above the Fermi level. This feature is symmetry-protected and robust under SOC, confirming Na₃Bi's status as a prototypical three-dimensional Dirac semimetal, as previously reported by Shao *et al.*⁴⁹ and confirmed experimentally by Liu *et al.*¹ Na₃Bi is frequently compared to Cd₃As₂, another 3D Dirac semimetal, which shows similar linear dispersion and topological protection when



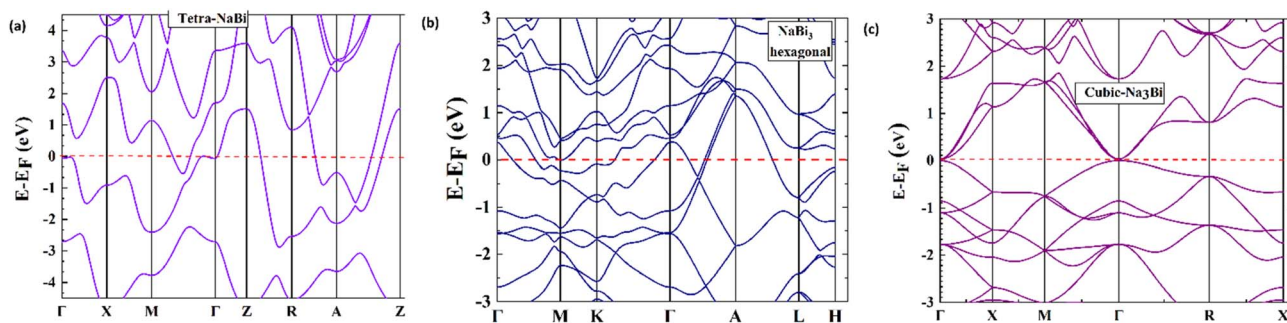


Fig. 2 The electronic band structure of (a) NaBi tetragonal, (b) NaBi₃ hexagonal and (c) cubic Na₃Bi calculated *via* HSE + SOC. The horizontal dashed line indicates the Fermi level, which is set at the zero level.

inversion and time-reversal symmetry are present.^{50,51} The effect of SOC is significant in all three phases, causing band inversions and topological features. These findings underscore the Na–Bi material family as a promising foundation for adjusting topological phases through crystal symmetry and indicate the possibility of achieving innovative quantum phenomena through structural engineering.

Fig. 3 shows the total and projected density of states (TDOS and PDOS) for tetragonal NaBi, hexagonal NaBi₃, and cubic Na₃Bi. This gives us information about their electronic structure and how their orbitals contribute. In tetragonal NaBi (Fig. 3a), the total density of states (DOS) exhibits a finite value at the Fermi level ($E_F = 0$ eV), thereby affirming its metallic character, which is consistent with the band structure depicted in Fig. 2a. The PDOS shows that the states close to the Fermi level mostly come from Bi-6p orbitals, and Na only adds a little bit. The prevalence of Bi-p states aligns with the potential for SOC-induced band inversion, a critical hallmark of topological semimetal characteristics.

In the case of hexagonal NaBi₃ (Fig. 3b), the TDOS is significantly higher than in the tetragonal phase, and the sharp Bi-p peaks below the Fermi level highlight strong hybridization and orbital density. Like the tetragonal phase, the Na states stay close to the Fermi level, while the Bi-p orbitals are the main contributors close to it. This strengthens the idea that the material is a complex metal and suggests at the possibility of Dirac or nodal-line semimetallic properties. The DOS profile for cubic Na₃Bi (Fig. 3c) confirms about its Dirac semimetal behaviour. There is a strong peak of Bi-p orbitals close to the Fermi level. This peak is very important in the formation of the symmetry-protected Dirac nodes, demonstrated by the band crossing in Fig. 2c. Na-s orbitals contribute mostly in the conduction band region and are not involved in low-energy excitations. The prevalent presence of Bi-derived p-states in all three structures highlights Bi's essential function in affecting the electronic topology of these compounds. These results align with prior research on Na₃Bi and comparable topological materials, wherein heavy-element p-orbital states dominate at the band edges, facilitating SOC-induced nontrivial phases.^{49–51}

3.3 Phonon, vibrational study and thermodynamics

Fig. 4 presents the calculated phonon dispersion curves for (a) tetragonal NaBi, (b) hexagonal NaBi₃, and (c) cubic Na₃Bi,

obtained using density functional perturbation theory (DFPT). The absence of imaginary phonon frequencies in the Brillouin zones for all three structures confirms their dynamical stability at zero temperature. For tetragonal NaBi (Fig. 4a), the phonon branches cover a frequency range of about 5.5 THz, with clear differences between the acoustic and optical modes. The high optical phonon frequencies, which are above 4 THz, suggest that the bonds are strong, especially between Bi and Na atoms. For hexagonal NaBi₃ (Fig. 4b), the phonon spectrum shows a denser distribution of modes below 4.5 THz, with flat bands that stand out, especially in the optical branches. These flat modes suggest localised vibrational states, possibly resulting from anisotropic or layered bonding environments, which align with the quasi-1D chain-like structural motif depicted in Fig. 1b. The phonon spectrum of cubic Na₃Bi (Fig. 4c) is typical of a highly symmetric rocksalt-type lattice, and the phonon branches go up to almost 5.5 THz. The degeneracies at high-symmetry points like Γ and X show that the crystal has cubic symmetry. The clear separation of the acoustic and optical branches and no evidence of phonon softening are strong indicators of mechanical and vibrational stability. These phonon properties back up the structural findings and prove that the Na–Bi phases being studied are thermodynamically and dynamically stable. Furthermore, the observed dispersion characteristics, such as flat optical bands and symmetry-induced degeneracies, can directly affect thermal transport and electron–phonon coupling, which are essential for describing superconductivity, thermoelectric performance, or phonon-limited mobility in these topologically significant systems. Comparable behaviour has been observed in other Bi-containing compounds, wherein heavy elements and significant spin–orbit coupling affect both lattice dynamics and electronic structure.^{52–54}

We used phonon-based quasiharmonic calculations to look at the thermodynamic stability and temperature-dependent behaviour of tetragonal NaBi, hexagonal NaBi₃, and cubic Na₃Bi. The results are shown in Fig. 5 and 6. Fig. 5 displays the computed heat capacity at constant volume (C_v) and entropy (S) up to 3000 K. All three structures show C_v rising and then levelling off, which is what the Debye model suggests ought to occur. This means that all of the phonon modes are filled at higher temperatures. Tetragonal NaBi reaches a saturation C_v of about 52 J mol⁻¹ K⁻¹, hexagonal NaBi₃ goes over 100 J mol⁻¹ K⁻¹, and cubic Na₃Bi stays stable at about 100 J mol⁻¹ K⁻¹.



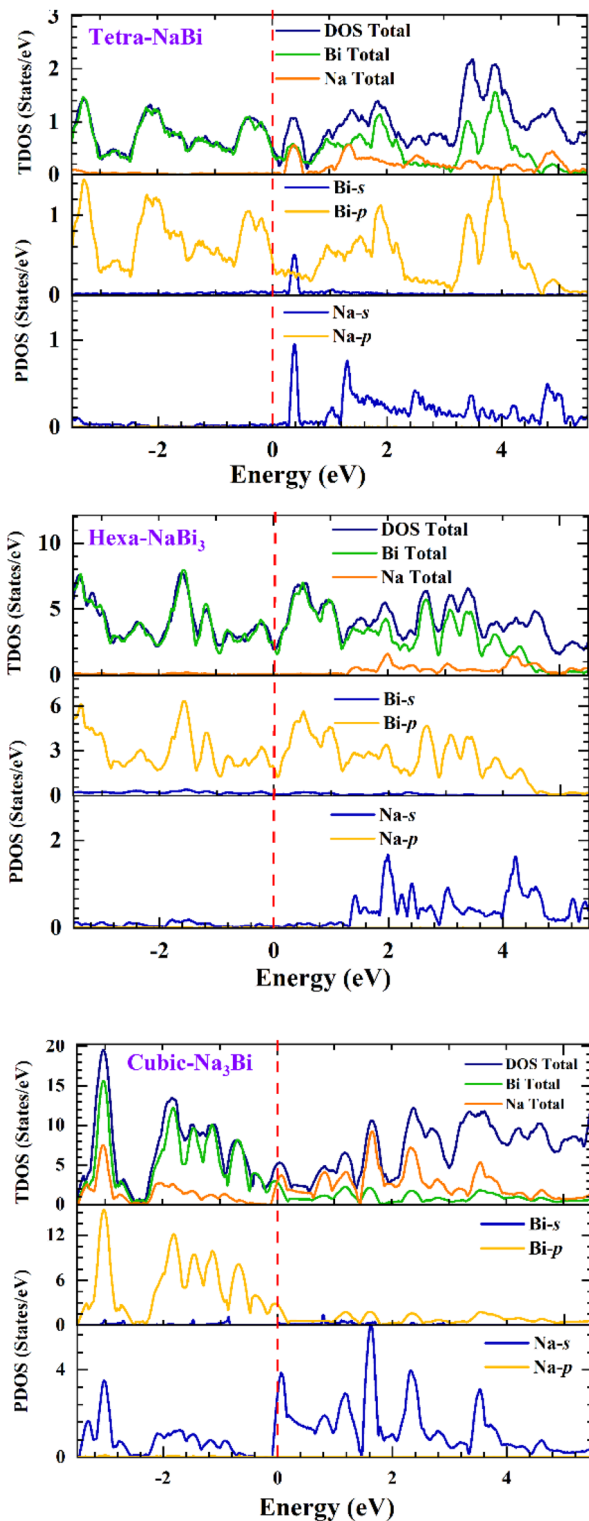


Fig. 3 The electronic TDOS and PDOS for (a) NaBi tetragonal, (b) NaBi₃ hexagonal and (c) cubic Na₃Bi calculated *via* HSE + SOC. The horizontal dashed line indicates the Fermi level, which is set at the zero level.

These trends show that the atoms in the series are getting increasingly complicated and have more ways to vibrate. Entropy increases monotonically with temperature, with

hexagonal NaBi₃ having the highest entropy ($\sim 470 \text{ J mol}^{-1} \text{ K}^{-1}$ at 3000 K) because its unit cell is bigger and its phonon spectrum is richer.

Fig. 6 shows the Helmholtz free energy ($A = E - TS$), the internal energy (E), and the entropic contribution ($-TS$). With increasing temperature, the internal energy of all compounds increases while the Helmholtz free energy (A) decreases, highlighting the stabilizing role of entropy. NaBi₃ has the most negative Helmholtz free energy at high temperatures, which supports its superior thermodynamic stability and is consistent with its lowest DFT-calculated formation energy (-0.498 eV per atom). The $-TS$ term is very large in all three compounds, which shows that lattice entropy plays a big role in how they behave thermally. These trends correspond effectively with observations in structurally and chemically similar systems. Bi₂Te₃ and Sb₂Te₃, which are well-known thermoelectric materials, have similar C_v saturation behaviour (about $125\text{--}130 \text{ J mol}^{-1} \text{ K}^{-1}$) and high entropy because they have heavy atoms and strong anharmonicity.^{54,55} PbTe, another standard thermoelectric, has similar levels of entropy ($\sim 400\text{--}450 \text{ J mol}^{-1} \text{ K}^{-1}$) because it has soft optical phonons and polar bonding.⁵⁶ Furthermore, NaSnBi, a topological compound structurally related to NaBi, exhibits thermal behavior governed by strong SOC and vibrational anisotropy.^{57,58} Similar phonon stability and free energy trends have also been reported for Na₃Bi and Cd₃As₂, both known topological Dirac semimetals, reinforcing the predictive accuracy of the current phonon-based thermodynamic analysis.^{59,60}

3.4 Thermoelectric properties

We used Boltzmann transport theory under the constant relaxation time approximation to look at the electronic transport coefficients of NaBi-based compounds in order to see the way that they performed as thermoelectric materials. Fig. 7, 8, and 9 respectively present the calculated electrical conductivity, Seebeck coefficient, and electronic thermal conductivity as functions of chemical potential at three representative temperatures: 100 K, 500 K, and 950 K. Fig. 7 shows the electrical conductivity (σ) behavior for tetragonal NaBi (a), hexagonal NaBi₃ (b), and cubic Na₃Bi (c). All three compounds show metallic properties, with σ values as high as $3 \times 10^5 (\Omega \text{ m})^{-1}$ near the Fermi level, especially at low temperatures (100 K). Tetragonal NaBi and cubic Na₃Bi exhibit elevated σ across a wide spectrum of chemical potential, which is ascribed to their more compact band structures in proximity to the Fermi level. As the temperature rises, thermal smearing dampens the amplitude of conductivity modulations. This is especially clear in hexagonal NaBi₃, which has a relatively flat σ profile, which means that its electronic structure is less dispersive. Fig. 8 shows the Seebeck coefficient (S), which is an important number for thermoelectric applications. The Seebeck coefficients of tetragonal and hexagonal NaBi phases are moderate ($\sim \pm 200 \mu\text{V K}^{-1}$), but cubic Na₃Bi has very high and very tunable values, reaching up to $\pm 1200 \mu\text{V K}^{-1}$ near the Fermi level at 100 K as its band crossings are linear and its dispersion is Dirac-like. These values are typical for topological semimetals,



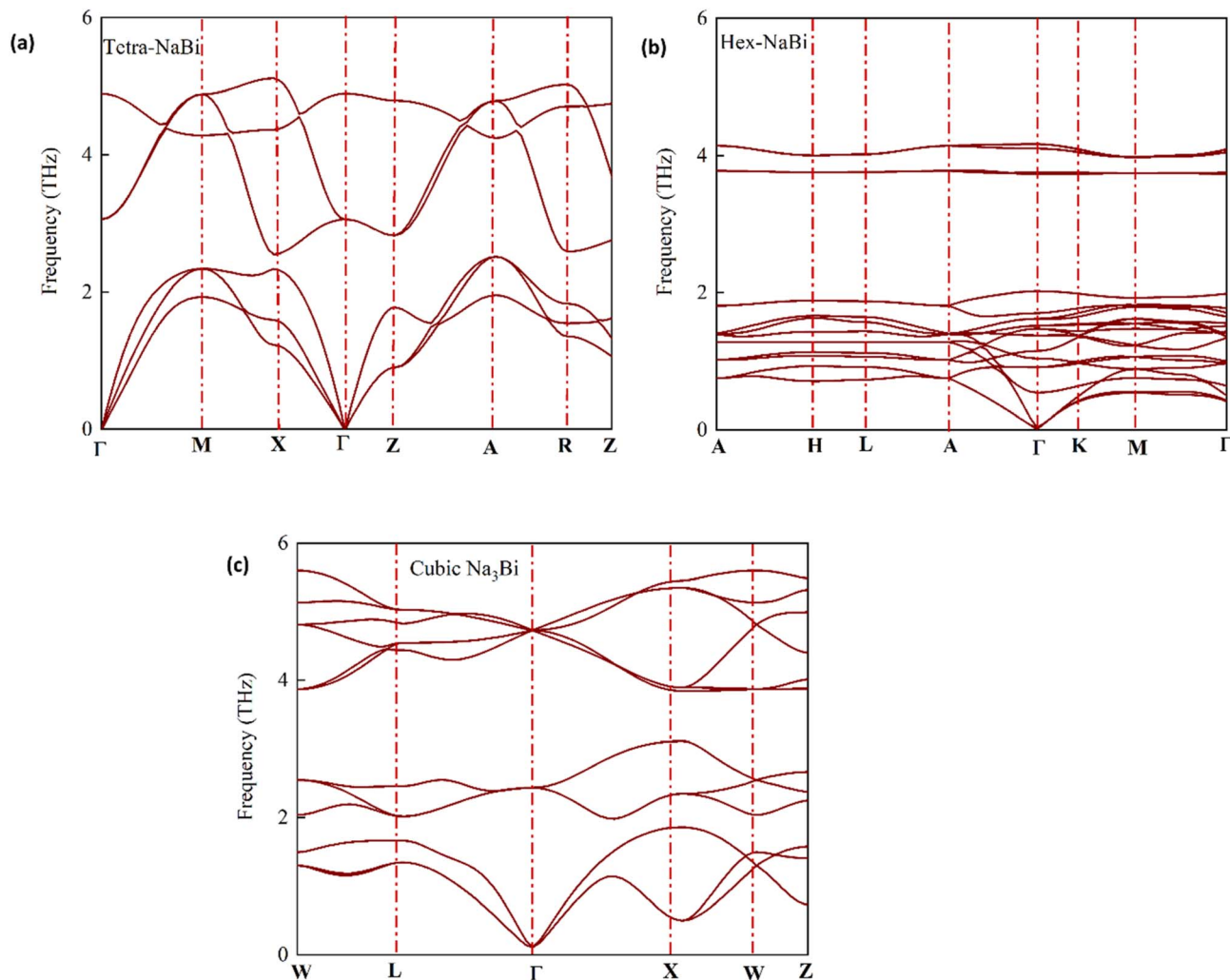


Fig. 4 Phonon dispersion curves of (a) tetragonal NaBi, (b) hexagonal NaBi₃, and (c) cubic Na₃Bi phases.

where bipolar conduction can greatly increase S in confined energy windows.⁶¹ The sign change of S around $\mu = 0$ is crucial because it shows ambipolar behaviour, which means that both p-type and n-type conduction may occur depending on the doping. Among the three, Na₃Bi demonstrates the most promising Seebeck response, especially under electron doping conditions. Fig. 9 depicts the electronic component of thermal conductivity (κ_e), which increases with temperature and carrier concentration. Tetragonal NaBi and cubic Na₃Bi exhibit κ_e values up to 25 W m⁻¹ K⁻¹ at 950 K near high-energy doping regimes, whereas hexagonal NaBi₃ retains significantly lower values (<10 W m⁻¹ K⁻¹), which may be advantageous in reducing total thermal losses and enhancing ZT (thermoelectric figure of merit) when lattice contributions are also minimized. When comparing the three structures, cubic Na₃Bi clearly exhibits superior thermoelectric characteristics, with high Seebeck coefficient, moderate to high electrical conductivity, and manageable electronic thermal conductivity. These findings, combined with its previously discussed topologically nontrivial electronic structure and dynamical stability, make it a strong candidate for topological thermoelectric applications.

Hexagonal NaBi₃, while less conductive, benefits from lower κ_e , which can improve thermoelectric performance if lattice thermal conductivity is minimized. Tetragonal NaBi appears to lie between the two, with balanced but less extreme transport characteristics. These findings correspond with studies on topological materials like Bi₂Te₃, SnSe, ZrTe₅, and Cd₃As₂, where mechanisms such as band convergence, Dirac dispersion, or band inversion result in enhanced Seebeck responses and optimised ZT values.^{61–64} Na₃Bi has been recognised as a 3D Dirac semimetal exhibiting adjustable thermoelectric properties through alloying or strain engineering.^{38,65}

To evaluate the thermoelectric efficiency of the investigated Na–Bi compounds, we calculated the dimensionless figure of merit $ZT = \frac{S^2 \sigma T}{\kappa}$ at three representative temperatures: 100 K, 500 K, and 950 K. This was done using the extracted values of the S , σ and κ_e from our Boltzmann transport calculations. We assumed that all materials had a constant lattice thermal conductivity of 1.5 W mK⁻¹, which is in line with what has been reported for Bi-based compounds. The resulting trends show that the three phases have very different thermoelectric



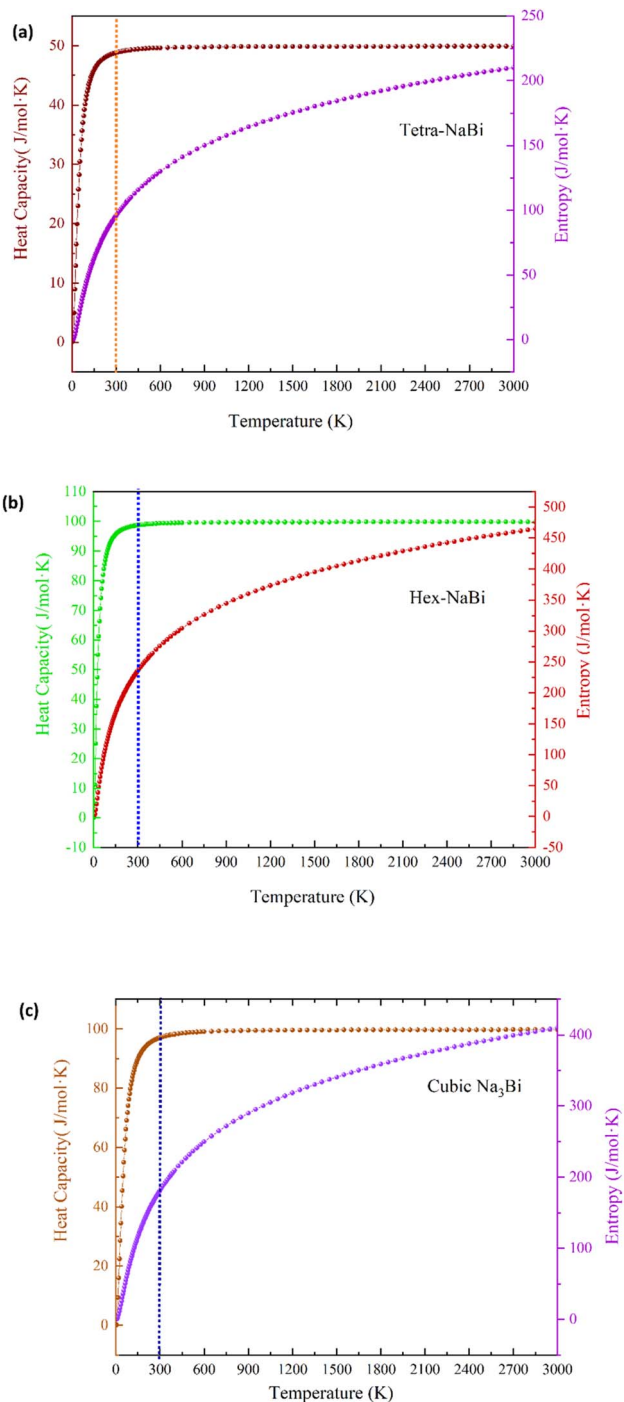


Fig. 5 The temperature-dependent thermodynamic properties of (a) tetragonal NaBi, (b) hexagonal NaBi₃, and (c) cubic Na₃Bi were found using phonon calculations. The graphs show how heat capacity at constant volume (C_v) and entropy (S) change with temperature up to 3000 K. All compounds show the expected saturation of C_v at high temperatures, which is in line with the Dulong–Petit limit. The entropy rises steadily as the temperature rises. NaBi₃ has the highest values because it has more complex atoms.

behaviour. Among the investigated structures, cubic Na₃Bi demonstrates the most promising thermoelectric performance at low and intermediate temperatures. It achieves a peak ZT of

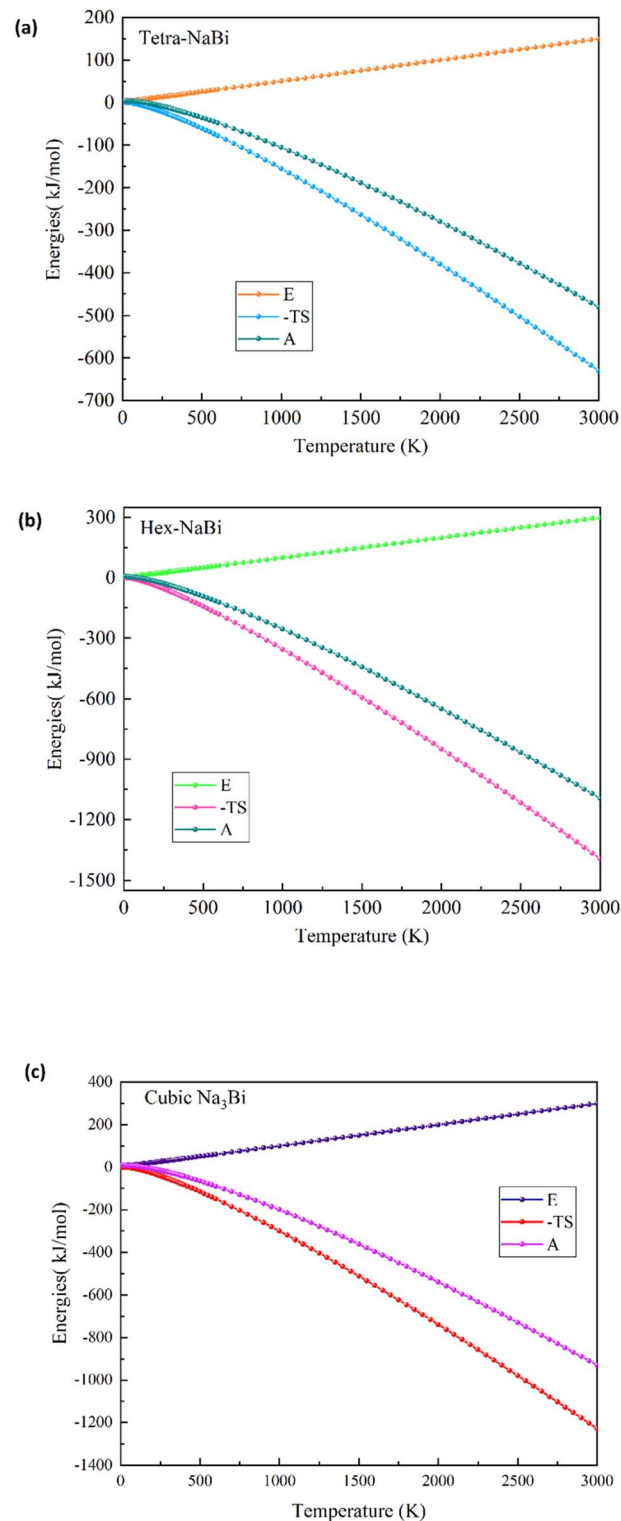


Fig. 6 Helmholtz free energy (A), internal energy (E), and entropic contribution ($-TS$) versus temperature for (a) tetragonal NaBi, (b) hexagonal NaBi₃, and (c) cubic Na₃Bi. The free energy decreases with increasing temperature in all cases, driven primarily by the $-TS$ term.

0.53 at 500 K, driven by a combination of exceptionally high Seebeck coefficients (up to $650 \mu\text{V K}^{-1}$ at 100 K) and moderate electrical conductivity. However, its thermoelectric efficiency



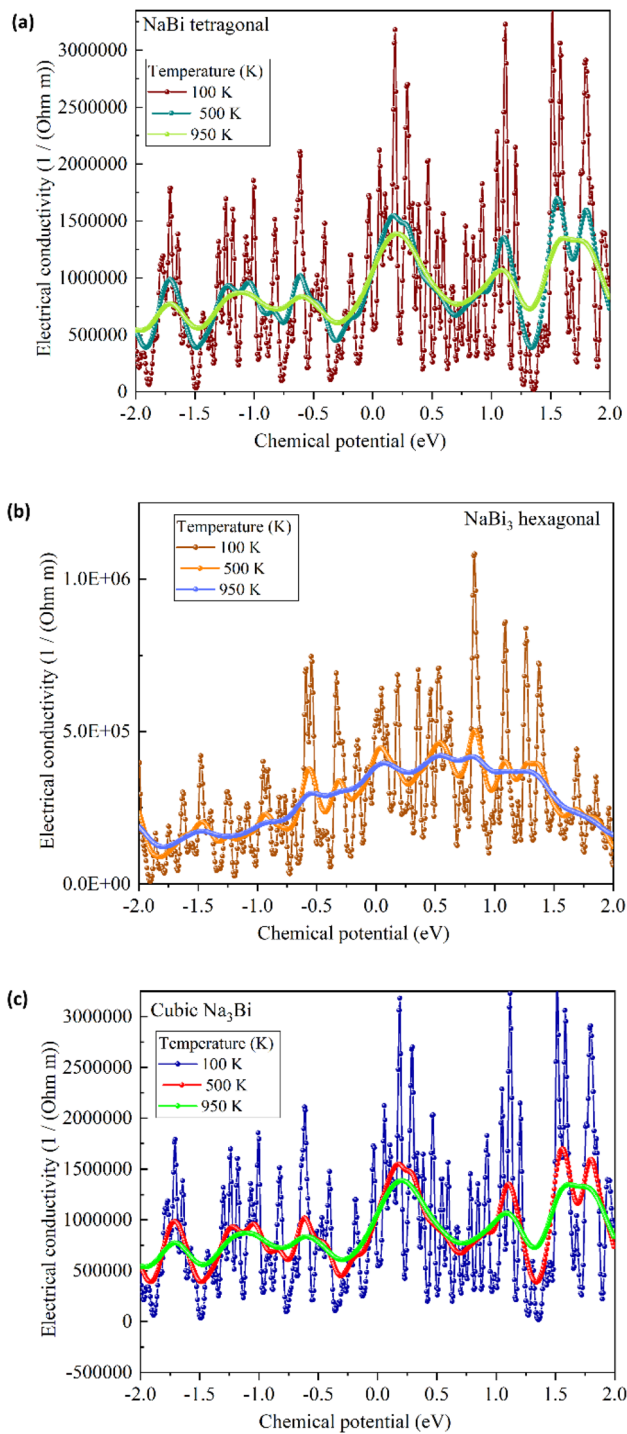


Fig. 7 Calculated electrical conductivity (σ/τ) as a function of chemical potential for (a) tetragonal NaBi, (b) hexagonal NaBi₃, and (c) cubic Na₃Bi at 100 K, 500 K, and 950 K.

decreases to $ZT \approx 0.11$ at 950 K due to a significant rise in electronic thermal conductivity, a known limitation of topological semimetals with metallic character. Tetragonal NaBi, on the other hand, exhibits more balanced behavior across the temperature range, reaching a maximum ZT of 0.29 at 500 K, which decreases to 0.13 at 950 K as κ_e increases and S saturates. Its relatively moderate Seebeck coefficient and stable σ allow it

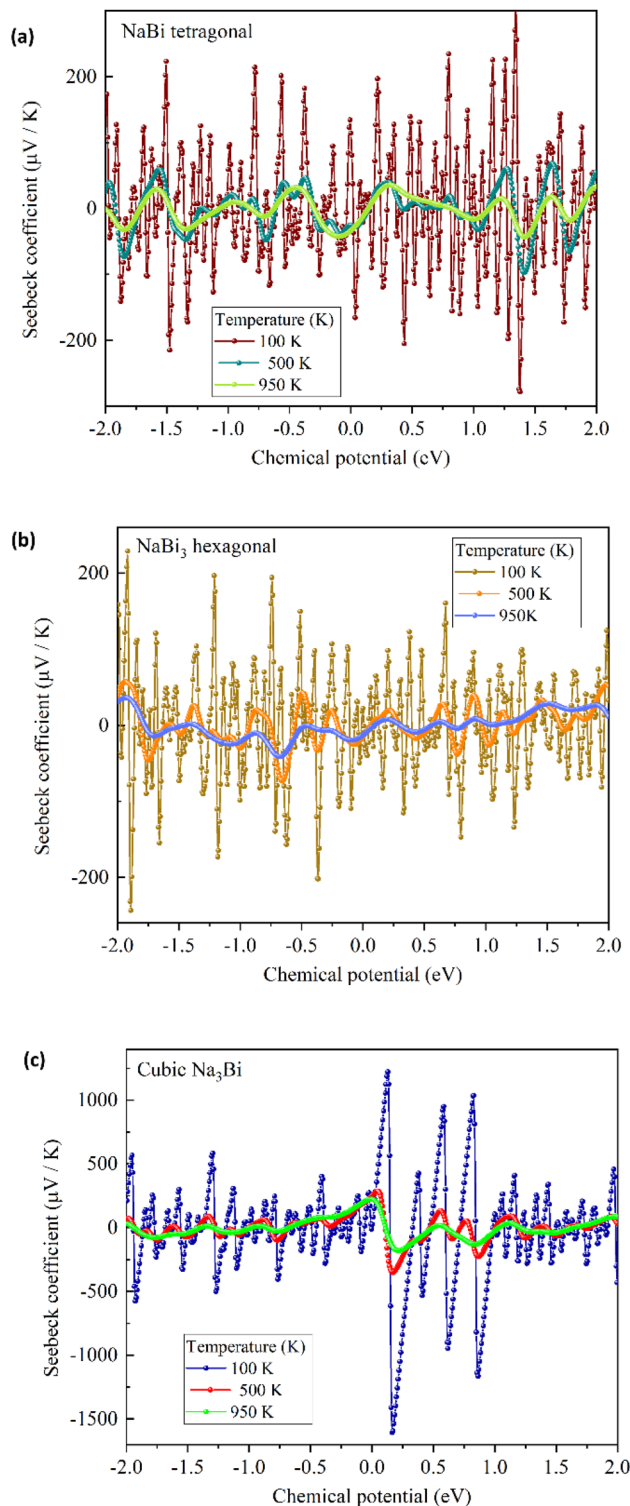


Fig. 8 Seebeck coefficient (S) versus chemical potential for (a) tetragonal NaBi, (b) hexagonal NaBi₃, and (c) cubic Na₃Bi at 100 K, 500 K, and 950 K. Cubic Na₃Bi shows highly enhanced and tunable Seebeck response with values exceeding $\pm 1000 \mu\text{V K}^{-1}$ near the Dirac point, indicative of its topological semimetal nature.

to maintain a reasonable ZT across the studied range. Hexagonal NaBi₃ has the lowest σ and S values overall, but it has a lower κ_e , which means that its ZT slowly but steadily improves



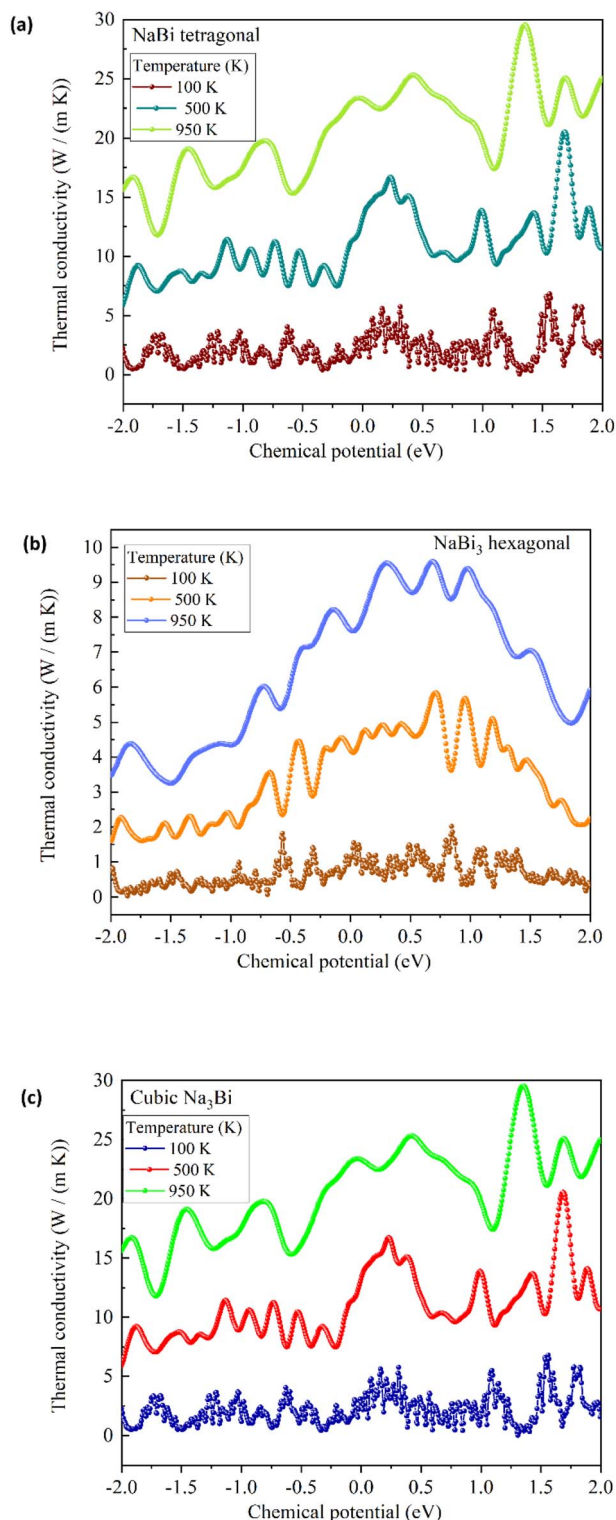


Fig. 9 Electronic thermal conductivity (κ_e) as a function of chemical potential for (a) tetragonal NaBi, (b) hexagonal NaBi₃, and (c) cubic Na₃Bi. Na₃Bi and NaBi exhibit higher κ_e at high doping levels, while NaBi₃ remains low across the range, advantageous for thermoelectric performance.

with temperature, going from 0.03 at 100 K to 0.21 at 950 K. This means that although its electrical transport isn't as efficient, its lower thermal losses could make it useful in thermoelectric

applications if more κ_1 reduction methods (like alloying or nanostructuring) are used. These results show that cubic Na₃Bi is the most suitable thermoelectric candidate below 600 K given that it has a Dirac-like band structure and a high Seebeck response. However, tetragonal NaBi and hexagonal NaBi₃ may become more competitive for use at high temperatures because they have lower κ_e and a more stable thermoelectric response. These results are in line with what has been seen in other topological thermoelectric materials, like ZrTe₅, Cd₃As₂, and engineered Na₃Bi alloys, which show higher ZT through band structure tuning and phonon engineering. In general, the results show that the Na–Bi family can be tuned thermoelectrically and that there are many ways to improve their performance through electronic or phononic design.

3.5 Machine learning-based prediction and analysis of thermoelectric performance in cubic Na₃Bi

To complement the insights obtained from density functional theory (DFT) and Boltzmann transport calculations, we employed machine learning (ML) regression models to predict the dimensionless thermoelectric figure of merit, ZT , from temperature-dependent transport properties of cubic Na₃Bi. DFT gives us a microscopic view, while ML lets us quickly make predictions, analyse how sensitive features are to changes, and make predictions across a range of parameters. In this study, we utilised two supervised machine learning models: Random Forest and a fully connected Neural Network aiming to evaluate their predictive accuracy, interpretability, and conformity with established physical behaviour.

3.5.1 Dataset construction and feature engineering. The dataset was created from first-principles calculations of the thermoelectric transport coefficients for cubic Na₃Bi at three temperatures: 100 K, 500 K, and 950 K, which cover a wide range of temperatures, from very low to very high (Table 2). By taking samples from 101 chemical potentials between -1.5 eV and $+1.5$ eV, we got 101 data points for each temperature. We divided the dataset into two parts: 75% for training (76 points) and 25% for testing (25 points). We trained and tested models separately at each temperature. The input features for each temperature were:

Seebeck coefficient: S in $\mu\text{V K}^{-1}$.

Electrical conductivity: σ in S m^{-1} .

Thermal conductivity κ in $\text{W m}^{-1} \text{K}^{-1}$.

These features were selected because they influence thermoelectric efficiency and result from *ab initio* transport calculations. Even though ZT can be calculated analytically from these parameters, the ML models are made to learn how they interact with each other in a complex, nonlinear, and temperature-dependent way across chemical potential space. This lets us use SHAP values for efficient surrogate prediction and sensitivity analysis, providing both accuracy and interpretability that goes beyond the direct formula.

3.5.2 Training strategy and model parameters. Each model was trained independently for each temperature using a 75%/25% train-test split. Input features for NN were standardized *via* z-score normalization.



Table 2 Estimated thermoelectric properties and dimensionless figure of merit (ZT) for tetragonal NaBi, hexagonal NaBi₃, and cubic Na₃Bi at 100 K, 500 K, and 950 K, using extracted Seebeck coefficient (S), electrical conductivity (σ), and electronic thermal conductivity (κ_e). A constant lattice thermal conductivity of $1.5 \text{ W m}^{-1} \text{ K}^{-1}$ was assumed for all compounds

Compounds	Temperatures (K)	S ($\mu\text{V K}^{-1}$)	σ (S m^{-1})	κ_e ($\text{W m}^{-1} \text{ K}^{-1}$)	κ_l ($\text{W m}^{-1} \text{ K}^{-1}$)	Total κ ($\text{W m}^{-1} \text{ K}^{-1}$)	ZT
Tetragonal NaBi	100	90	1.5×10^5	5	1.5	6.5	0.19
	500	130	1.2×10^5	16	1.5	17.5	0.29
	950	180	1.0×10^5	22	1.5	23.5	0.13
Hexagonal NaBi ₃	100	40	8.0×10^4	2.5	1.5	4.0	0.03
	500	70	6.0×10^4	6	1.5	7.5	0.20
	950	90	5.0×10^4	8	1.5	9.5	0.21
Cubic Na ₃ Bi	100	650	1.0×10^5	7	1.5	8.5	0.50
	500	400	1.1×10^5	15	1.5	16.5	0.53
	950	500	1.0×10^5	20	1.5	21.5	0.11

Hyperparameters for both models were optimized *via* grid-based manual tuning. We changed the number of estimators for RF and used the mean squared error criterion to find the best balance between performance and complexity. We chose 100 trees as the best number. We used a two-layer dense architecture (16 neurons each) with ReLU activations and trained the NN with the Adam optimiser. To avoid overfitting, early stopping (patience = 20 epochs) was used. We chose these settings after testing RMSE and R^2 on validation splits over and over again, which performed successfully given the dataset's size.

3.5.3 Parity plots and predictive performance. The performance of both models on the test set is summarized in Table 3. Parity plots (Fig. 10) also correspond to test predictions, showing the correlation between predicted and true ZT values. For clarity, we now include a comparison of training and test errors (R^2 and RMSE). The fact that the training and test metrics are remarkably similar shows that both models work well in general, but RF is preferred due to its resistance to overfitting at low temperatures. The RF model consistently proved more accurate than the NN, especially at 100 K, where the NN did not fit the data well. The NN performed substantially better at higher temperatures (500 K and 950 K), almost matching the RF.

This trend shows that the transport properties become smoother and less noisy at higher temperatures. This is advantageous for models like NNs, which are sensitive to data noise. We observe that tree-based ensemble models, like RF, are

ideal for small, structured datasets because they are resistant to overfitting. In contrast, the NN model performed better at higher temperatures, where the transport trends are smoother.

3.5.4 SHAP analysis and feature sensitivity. We employed SHAP to interpret the RF model. SHAP values quantify the marginal contribution of each feature to the model output by computing Shapley values from cooperative game theory.

Fig. 11 shows SHAP summary plots for each temperature. In all cases, the Seebeck coefficient was the most important factor, which is in line with its role as a quadratic term in the ZT formula. The SHAP values also showed the effect of κ at 100 K, since it changes more at low temperatures. At 950 K, σ was more important because it was the main factor in total heat transport and Joule heating losses at high temperatures.

These ML models not only give accurate ZT predictions, but they also give quantitative information about how to design materials. The SHAP and parity plots back up the well-known rules:

- At all temperatures, it is important to maximise the Seebeck coefficient.
- At high T , electrical conductivity becomes more important.
- Thermal conductivity, which is always inversely related to ZT , is more important at cryogenic temperatures where κ usually dominates the denominator.
- RF is better in this case because it can work with small structured datasets that do not have many features. NN performed similarly at high temperatures but tended to underfit at low temperatures. These findings indicate that RF is more appropriate for initial ML modelling of thermoelectrics, whereas neural networks may be advantageous for larger, noisier datasets with additional descriptors (*e.g.*, atomic structure, bonding metrics, *etc.*).

This machine learning framework can be readily extended to predict ZT for related materials (*e.g.*, Na₃Bi alloys, strained structures, or doped variants). Incorporating additional features such as electronic density of states, effective masses, or band gap descriptors could further improve generalizability. Moreover, hybrid models combining tree-based learning with deep feature extraction (*e.g.*, graph neural networks) offer promising avenues for interpretable, data-driven thermoelectric design. This ability to understand,

Table 3 Performance metrics (R^2 and RMSE) of Random Forest and Neural Network models in predicting ZT for cubic Na₃Bi at 100 K, 500 K, and 950 K using the test set. Models were trained on 75% of the data and evaluated on the remaining 25%

Temp (K)	Model	R^2 (train)	R^2 (test)	RMSE (train)	RMSE (test)
100	RF	0.984	0.971	0.122	0.160
100	NN	0.847	0.791	0.189	0.160
500	RF	0.995	0.994	0.037	0.049
500	NN	0.965	0.947	0.053	0.049
950	RF	0.998	0.998	0.012	0.014
950	NN	0.997	0.996	0.015	0.014



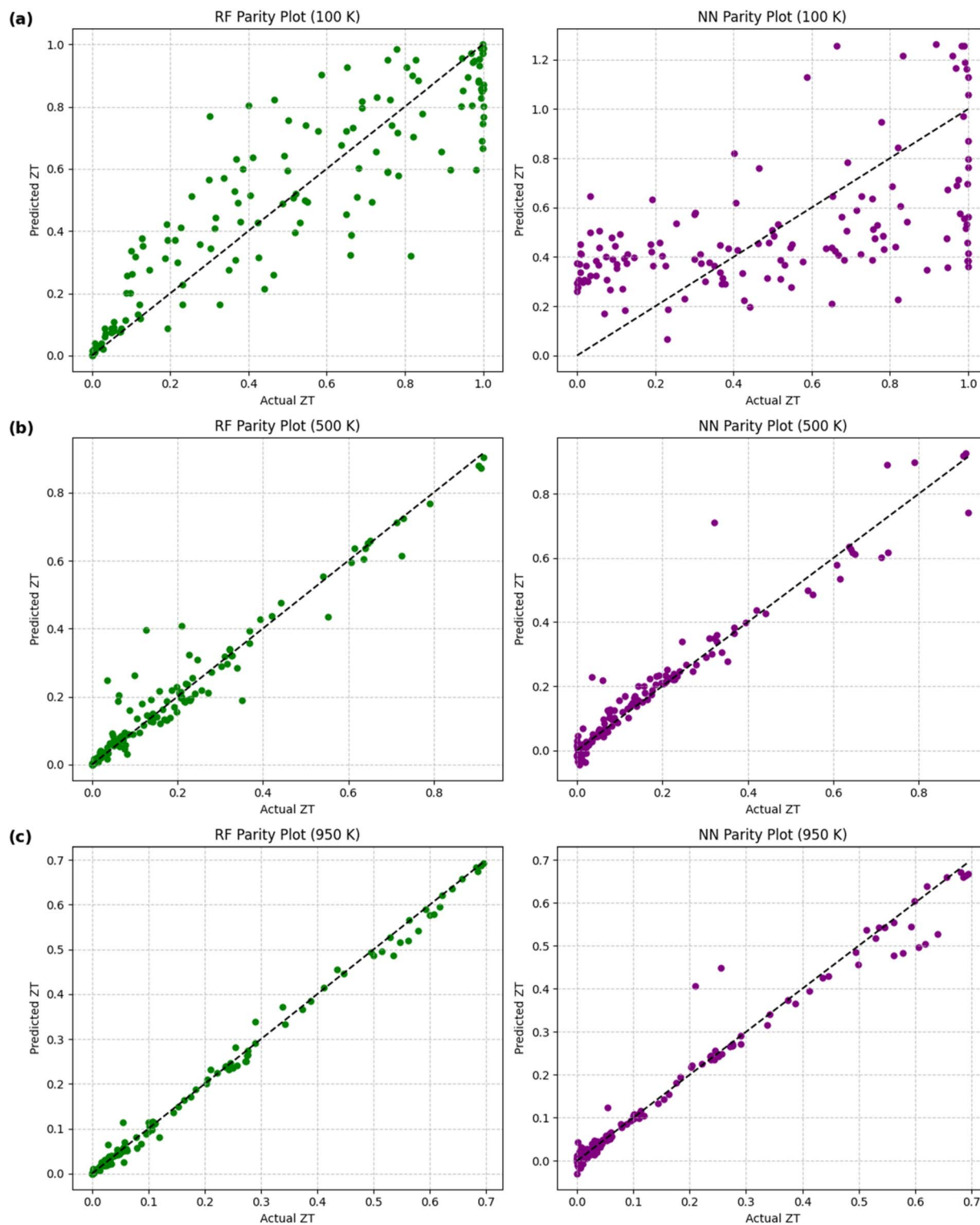


Fig. 10 Parity plots showing predicted versus actual ZT values using Random Forest (left) and Neural Network (right) models at (a) 100 K, (b) 500 K, and (c) 950 K.

along with the ability of RF and NN models to make predictions, supports the idea that ML is not redundant but rather helpful, especially in complicated thermoelectric systems

where the interactions between transport coefficients are not linear and depend on temperature.

The remarkable thermoelectric properties exhibited by Na_3Bi in proximity to its Dirac point warrant a comparison with



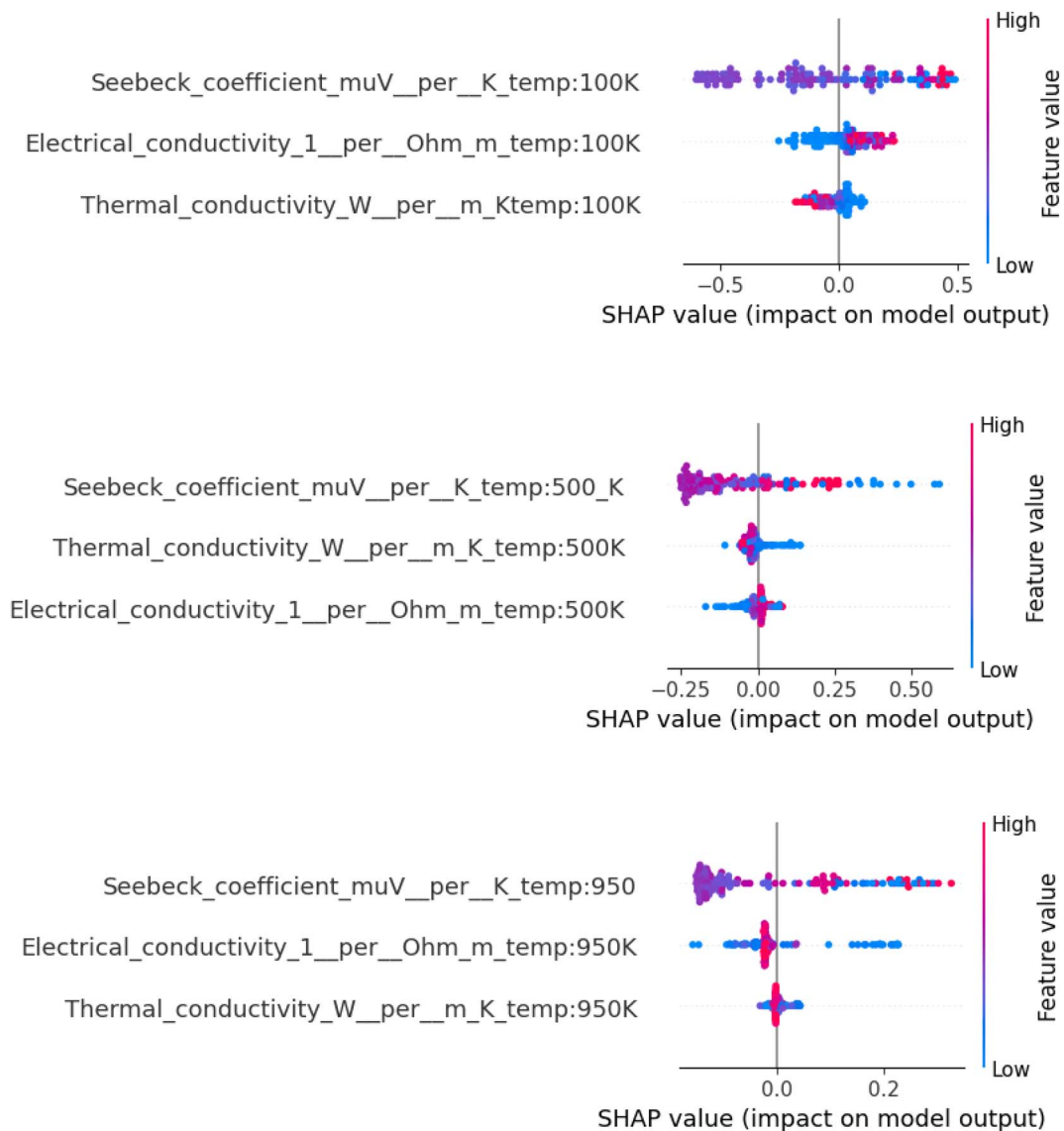


Fig. 11 SHAP summary plots illustrating feature importance in predicting ZT at (top) 100 K, (middle) 500 K, and (bottom) 950 K.

established theoretical models for Dirac materials. The Tang-Dresselhaus theory gives a full picture of how thermoelectric transport works in quantum-confined Dirac systems. It says that the Seebeck coefficients will be higher because of the unique linear dispersion and density of states near Dirac points.^{66–70} Our results in bulk 3D Na_3Bi exhibit conceptually coherent behaviour; the exceptionally high Seebeck coefficients (reaching $\pm 1200 \mu\text{V/K}$) we observe arise from identical fundamental physics: linear band dispersion and the disappearance of density of states at the Dirac energy. While the material realization differs (3D bulk crystal *vs.* 2D quantum-confined systems), the underlying mechanism of enhanced thermoelectric response near Dirac points appears universal. Moreover, the integrated DFT-ML methodology developed in this work provides a general framework that can be directly applied to study thermoelectric properties in the Dirac materials described by the Tang-Dresselhaus theory. Our approach, which combines accurate electronic structure calculation with machine learning

surrogate modeling, is particularly well-suited for exploring the complex parameter space of engineered Dirac systems, including doping, strain, and quantum confinement effects. The SHAP interpretability analysis further offers a data-driven means to validate theoretical predictions about feature importance in these systems.

4. Conclusion

This work provides a comprehensive multi-scale investigation of Na–Bi compounds by combining density functional theory and machine learning models. First-principles calculations elucidated the crystal structures, electronic topologies, phonon dynamics, and thermoelectric transport of tetragonal NaBi , hexagonal NaBi_3 , and cubic Na_3Bi . Our results confirm that cubic Na_3Bi exhibits superior thermoelectric potential at intermediate temperatures due to its enhanced Seebeck response and Dirac semimetallic nature. Additionally, by integrating



machine learning models trained on DFT-calculated data, we demonstrated that Random Forest and Neural Network models can reliably predict ZT across varying temperatures, with RF models showing superior performance at low temperatures. The SHAP interpretability framework further clarified the relative feature importance, reinforcing known physical principles and providing data-driven insights. These findings showcase the power of ML in augmenting first-principles calculations, not only by predicting ZT efficiently, but also by quantifying feature contributions and offering a framework that is extensible to doped or unexplored Na–Bi derivatives. Future research may apply this methodology to Na₃Bi alloys, doped systems, or strained structures, facilitating the interpretable and efficient design of topological thermoelectrics *via* hybrid physics-informed machine learning workflows. This framework can easily be used with other types of Dirac materials, such as the quantum-confined systems described by the Tang-Dresselhaus theory. This makes it possible to compare different Dirac material platforms.

Conflicts of interest

There are no conflicts to declare.

Data availability

All data are available in the data repository ZENODO at: 10.5281/zenodo.15499220 (<https://zenodo.org/records/15499220>).

Acknowledgements

S. Goumri-Said acknowledges the support of Alfaisal University and its Office of Research & Innovation for providing the internal grant (Grant No. 24407). M. B. K would like to thank Prince Sultan University for their support.

References

- Z. K. Liu, B. Zhou, Y. Zhang, Z. J. Wang, H. M. Weng, D. Prabhakaran, S.-K. Mo, Z. X. Shen, Z. Fang, X. Dai, Z. Hussain and Y. L. Chen, Discovery of a three-dimensional topological Dirac semimetal, Na₃Bi, *Science*, 2014, **343**, 864–867, DOI: [10.1126/science.1245085](https://doi.org/10.1126/science.1245085).
- M. Neupane, S. Y. Xu, R. Sankar, N. Alidoust, G. Bian, C. Liu, I. Belopolski, T.-R. Chang, H.-T. Jeng, H. Lin, A. Bansil, F. Chou and M. Z. Hasan, Observation of a three-dimensional topological Dirac semimetal phase in high-mobility Cd₃As₂, *Nat. Commun.*, 2014, **5**, 3786, DOI: [10.1038/ncomms4786](https://doi.org/10.1038/ncomms4786).
- R. Li, X. Cheng, Q. Xie, Y. Sun, D. Li, Y. Li and X.-Q. Chen, Topological Metal of NaBi with Ultralow Lattice Thermal Conductivity and Electron-phonon Superconductivity, *Sci. Rep.*, 2015, **5**, 8446, DOI: [10.1038/srep08446](https://doi.org/10.1038/srep08446).
- H. Pan, M. Wu, Y. Liu and S. A. Yang, Electric control of topological phase transitions in Dirac semimetal thin films, *Sci. Rep.*, 2015, **5**, 14639, DOI: [10.1038/srep14639](https://doi.org/10.1038/srep14639).
- G. S. Jenkins, C. Lane, B. Barbiellini, A. B. Sushkov, R. L. Carey, F. Liu, J. W. Krizan, S. K. Kushwaha, Q. Gibson, T.-R. Chang, H.-T. Jeng, H. Lin, R. J. Cava, A. Bansil and H. D. Drew, Three-dimensional Dirac cone carrier dynamics in Na₃Bi and Cd₃As₂, *Phys. Rev. B*, 2016, **94**, 085121.
- I. V. Pinchuk, T. J. Asel, A. Franson, T. Zhu, Y.-M. Lu, L. J. Brillson, E. Johnston-Halperin, J. A. Gupta and R. K. Kawakami, Topological Dirac semimetal Na₃Bi films in the ultrathin limit via alternating layer molecular beam epitaxy, *APL Mater.*, 2018, **6**, 086103, DOI: [10.1063/1.5041273](https://doi.org/10.1063/1.5041273).
- P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni and R. M. Wentzcovitch, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter*, 2009, **21**, 395502, DOI: [10.1088/0953-8984/21/39/395502](https://doi.org/10.1088/0953-8984/21/39/395502).
- D. Vanderbilt, Soft self-consistent pseudopotentials in a generalized eigenvalue formalism, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1990, **41**, 7892, DOI: [10.1103/PhysRevB.41.7892](https://doi.org/10.1103/PhysRevB.41.7892).
- D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2013, **88**, 085117, DOI: [10.1103/PhysRevB.88.085117](https://doi.org/10.1103/PhysRevB.88.085117).
- A. Dal Corso, Pseudopotentials periodic table: From H to Pu, *Comput. Mater. Sci.*, 2014, **95**, 337–350, DOI: [10.1016/j.commatsci.2014.07.043](https://doi.org/10.1016/j.commatsci.2014.07.043).
- J. P. Perdew, K. Burke and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868, DOI: [10.1103/PhysRevLett.77.3865](https://doi.org/10.1103/PhysRevLett.77.3865).
- D. F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Math. Comput.*, 1970, **24**, 647–656, DOI: [10.2307/2004840](https://doi.org/10.2307/2004840).
- H. J. Monkhorst and J. D. Pack, Special points for Brillouin-zone integrations, *Phys. Rev. B*, 1976, **13**, 5188–5192, DOI: [10.1103/PhysRevB.13.5188](https://doi.org/10.1103/PhysRevB.13.5188).
- J. Heyd, G. E. Scuseria and M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, *J. Chem. Phys.*, 2003, **118**, 8207–8215, DOI: [10.1063/1.1564060](https://doi.org/10.1063/1.1564060).
- N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza and D. Vanderbilt, Maximally localized Wannier functions: Theory and applications, *Rev. Mod. Phys.*, 2012, **84**(4), 1419.
- Q. Wu, S. Zhang, H.-F. Song, M. Troyer and A. A. Soluyanov, WannierTools: An open-source software package for novel topological materials, *Comput. Phys. Commun.*, 2018, **224**, 405–416.
- H. Muñoz, J. Antonio, J. Cervantes, M. Romero, J. Rosas-Huerta, E. Arévalo-López, E. Carvajal and R. Escamilla, A first-principles study of the electronic, mechanical, vibrational, and optical properties of zirconium carbide under high pressure, *Phys. Scr.*, 2023, **98**(2), 025817.
- V. Chihaiia, V. Alexiev and H. S. AlMatrouk, *Assessment of the Heat Capacity by Thermodynamic Approach Based on Density Functional Theory Calculations*, IntechOpen, 2022, DOI: [10.5772/intechopen.104083](https://doi.org/10.5772/intechopen.104083).
- R. Salloom, S. Joshi, N. B. Dahotre and S. Srinivasan, Laser surface engineering of B₄C/Fe nanocomposite coating on



- low-carbon steel: Experimental coupled with computational approach, *Mater. Des.*, 2020, **190**, 108576.
- 20 M. Blanco, E. Francisco and V. Luana, GIBBS: Isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debye model, *Comput. Phys. Commun.*, 2004, **158**(1), 57–72.
- 21 O. Eriksson, J. M. Wills and D. Wallace, Electronic, quasiharmonic, and anharmonic entropies of transition metals, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1992, **46**, 5221–5228, DOI: [10.1103/PhysRevB.46.5221](https://doi.org/10.1103/PhysRevB.46.5221).
- 22 P.-W. Guan, G. Houchins and V. Viswanathan, Uncertainty quantification of DFT-predicted finite-temperature thermodynamic properties within the Debye model, *J. Chem. Phys.*, 2019, **151**(24), 244702, DOI: [10.1063/1.5132332](https://doi.org/10.1063/1.5132332).
- 23 G. K. Madsen and D. J. Singh, BoltzTraP: A code for calculating band-structure dependent quantities, *Comput. Phys. Commun.*, 2006, **175**(1), 67–71.
- 24 İ. Yücel, The structural, electronic, optical and thermoelectric properties of impurity-doped Mg₂Ge compounds: A DFT study, *J. Phys. Chem. Solids*, 2022, **160**, 110351.
- 25 G. D. Mahan, Thermoelectric effect, in *Encyclopedia of Condensed Matter Physics*, ed. F. Bassani, G. L. Liedl and P. Wyder, Elsevier, Oxford, 2005, pp. 180–187.
- 26 F. Garmroudi, S. Di Cataldo, M. Parzer, J. Coulter, Y. Iwasaki, M. Grasser, S. Stockinger, S. Pázmán, S. Witzmann, A. Riss, H. Michor, R. Podloucky, S. Khmelevskiy, A. Georges, K. Held, T. Mori, E. Bauer and A. Pustogow, Energy filtering-induced ultrahigh thermoelectric power factors in Ni₃Ge, *Sci. Adv.*, 2025, **11**(31), eadv7113.
- 27 D. Narducci, E. Selezneva, G. Cerofolini, S. Frabboni and G. Ottaviani, Impact of energy filtering and carrier localization on the thermoelectric properties of granular semiconductors, *J. Solid State Chem.*, 2012, **193**, 19–25.
- 28 J. O. Sofo and G. D. Mahan, Optimum band gap of a thermoelectric material, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1994, **49**(7), 4565–4570.
- 29 D. J. Singh, G. K. H. Madsen and D. Parker, Thermoelectric behavior and scattering mechanisms in complex materials, *Sci. Rep.*, 2018, **8**, 9372.
- 30 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, Machine learning for molecular and materials science, *Nature*, 2018, **559**(7715), 547–555, DOI: [10.1038/s41586-018-0337-2](https://doi.org/10.1038/s41586-018-0337-2).
- 31 J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, Recent advances and applications of machine learning in solid-state materials science, *npj Comput. Mater.*, 2019, **5**, 83.
- 32 L. Breiman, Random forests, *Mach. Learn.*, 2001, **45**(1), 5–32, DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- 33 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *npj Comput. Mater.*, 2016, **2**, 16028, DOI: [10.1038/npjcompumats.2016.28](https://doi.org/10.1038/npjcompumats.2016.28).
- 34 I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, 2016.
- 35 D. Jha, L. Ward, A. Paul, W. K. Liao, A. Choudhary, C. Wolverton and A. Agrawal, ElemNet: Deep learning the chemistry of materials from only elemental composition, *Sci. Rep.*, 2018, **8**, 17593.
- 36 S. Schaefer, S. A. Ghasemi, S. Roy and S. Goedecker, Stabilized quasi-Newton optimization of noisy potential energy surfaces, *J. Chem. Phys.*, 2015, **142**, 034112.
- 37 P. Wisesa, K. A. McGill and T. Mueller, Efficient generation of generalized Monkhorst–Pack grids through the use of informatics, *Phys. Rev. B*, 2016, **93**, 155109.
- 38 X. Cheng, R. Li, D. Li, Y. Li and X.-Q. Chen, Stable compositions and structures in the Na–Bi system, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6933–6947, DOI: [10.1039/C4CP05115GZ](https://doi.org/10.1039/C4CP05115GZ).
- 39 M. Z. Hasan and C. L. Kane, Colloquium: Topological insulators, *Rev. Mod. Phys.*, 2010, **82**, 3045–3067, DOI: [10.1103/RevModPhys.82.3045](https://doi.org/10.1103/RevModPhys.82.3045).
- 40 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.*, 2013, **1**, 011002, DOI: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- 41 S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy and O. AFLOWLIB, ORG: A distributed materials properties repository from high-throughput ab initio calculations, *Comput. Mater. Sci.*, 2012, **58**, 227–235, DOI: [10.1016/j.commatsci.2012.02.002](https://doi.org/10.1016/j.commatsci.2012.02.002).
- 42 E. Zintl and G. Brauer, G., Intermetallische Verbindungen. II. Über die Verbindung NaBi, *Z. Phys. Abtl. B-chem. Elem. Aufbau Mater.*, 1932, **16**, 195–207.
- 43 Z. Wang, Y. Sun, X.-Q. Chen, C. Franchini, G. Xu, H. Weng, X. Dai and Z. Fang, Dirac semimetal and topological phase transitions in A₃Bi (A = Na, K, Rb), *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2012, **85**, 195320, DOI: [10.1103/PhysRevB.85.195320](https://doi.org/10.1103/PhysRevB.85.195320).
- 44 S. K. Kushwaha, J. W. Krizan, J. Xiong, T. Klimczuk, Q. D. Gibson, T. Liang, N. P. Ong and R. J. Cava, Superconducting properties and electronic structure of NaBi, *J. Phys.: Condens. Matter*, 2014, **26**, 212201, DOI: [10.1088/0953-8984/26/21/212201](https://doi.org/10.1088/0953-8984/26/21/212201).
- 45 M. Goyal and M. M. Sinha, Phase stability and thermoelectricity in topological Dirac semimetal Na₃Bi, *Mater. Sci. Semicond. Process.*, 2023, **167**, 107760, DOI: [10.1016/j.mssp.2023.107760](https://doi.org/10.1016/j.mssp.2023.107760).
- 46 D. C. Desai, J. Park, J. J. Zhou and M. Bernardi, Dominant two-dimensional electron–phonon interactions in the bulk Dirac semimetal Na₃Bi, *Nano Lett.*, 2023, **23**, 3947–3953, DOI: [10.1021/acs.nanolett.3c00713](https://doi.org/10.1021/acs.nanolett.3c00713).
- 47 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, *Comput. Mater. Sci.*, 2013, **68**, 314–319, DOI: [10.1016/j.commatsci.2012.10.028](https://doi.org/10.1016/j.commatsci.2012.10.028).



- 48 Z. Guo, X. Li, A. Bergara, S. Ding, X. Zhang and G. Yang, Pressure-induced evolution of stoichiometries and electronic structures of host-guest Na-B compounds, *Matter Radiat. Extremes*, 2023, **8**(6), 068401, DOI: [10.1063/5.0155005](https://doi.org/10.1063/5.0155005).
- 49 D. Shao, J. Ruan, J. Wu, T. Chen, Z. Guo, H. Zhang, J. Sun, L. Sheng and D. Xing, Strain-induced quantum topological phase transitions in Na₃Bi, *Phys. Rev. B*, 2017, **96**, 075112, DOI: [10.1103/PhysRevB.96.075112](https://doi.org/10.1103/PhysRevB.96.075112).
- 50 S. K. Kushwaha, J. W. Krizan, B. E. Feldman, A. Gyenis, M. T. Randeria, J. Xiong, S.-Y. Xu, N. Alidoust, I. Belopolski, T. Liang, M. Z. Hasan, N. P. Ong, A. Yazdani and R. J. Cava, Bulk crystal growth and electronic characterization of the 3D Dirac semimetal Na₃Bi, *APL Mater.*, 2015, **3**(4), 041504.
- 51 S. Borisenko, Q. Gibson, D. Evtushinsky, V. Zabolotnyy, B. Büchner and R. J. Cava, Experimental realization of a three-dimensional Dirac semimetal, *Phys. Rev. Lett.*, 2014, **113**(2), 027603, DOI: [10.1103/PhysRevLett.113.027603](https://doi.org/10.1103/PhysRevLett.113.027603).
- 52 T. Nie, L. Meng, Y. Li, Y. Luan and J. Yu, Phase transition studies of Na₃Bi system under uniaxial strain, *J. Phys.: Condens. Matter*, 2018, **30**, 125502, DOI: [10.1088/1361-648X/aaad22](https://doi.org/10.1088/1361-648X/aaad22).
- 53 W. Li, J. Carrete, N. A. Katcho and N. Mingo, ShengBTE: A solver of the Boltzmann transport equation for phonons, *Comput. Phys. Commun.*, 2014, **185**(6), 1747–1758, DOI: [10.1016/j.cpc.2014.02.015](https://doi.org/10.1016/j.cpc.2014.02.015).
- 54 L. Yu, Y. Zhang, S. Chen and A. Zunger, Anharmonicity-enhanced thermoelectric performance in Bi₂Te₃-based materials, *npj Comput. Mater.*, 2020, **6**(1), 23, DOI: [10.1038/s41524-020-0296-9](https://doi.org/10.1038/s41524-020-0296-9).
- 55 Y. Shi, C. Li and W. Liu, First-principles study on phonon transport and thermodynamic properties of Sb₂Te₃, *J. Appl. Phys.*, 2022, **131**(8), 085102, DOI: [10.1063/5.0080511](https://doi.org/10.1063/5.0080511).
- 56 Y. Lu, T. Sun and D.-B. Zhang, Lattice anharmonicity, phonon dispersion, and thermal conductivity of PbTe studied by the phonon quasiparticle approach, *Phys. Rev. B*, 2018, **97**, 174304, DOI: [10.1103/PhysRevB.97.174304](https://doi.org/10.1103/PhysRevB.97.174304).
- 57 X. Zhang, B. Yan and H. Weng, Topological nature and thermal properties of NaSnBi under pressure, *Phys. Rev. B*, 2020, **102**(8), 085120.
- 58 A. Jain and D. J. Singh, Vibrational and thermodynamic signatures of topological Dirac semimetals: Case of Na₃Bi and Cd₃As₂, *J. Phys.: Condens. Matter*, 2023, **35**(13), 135601.
- 59 R. Yu, W. Zhang, H.-J. Zhang, S.-C. Zhang, X. Dai and Z. Fang, First-principles analysis of thermal and topological stability in Na₃Bi, *npj Quantum Mater.*, 2021, **6**, 34.
- 60 A. Bedoya-Pinto, A. K. Pandeya, D. Liu, H. Deniz, K. Chang, H. Tan, H. Han, J. Jena, I. Kostanovskiy and S. S. P. Parkin, Realization of epitaxial NbP and TaP Weyl semimetal thin films, *ACS Nano*, 2020, **14**(4), 4405–4413.
- 61 S. M. Young, S. Zaheer, J. C. Y. Teo, C. L. Kane, E. J. Mele and A. M. Rappe, Dirac semimetal in three dimensions, *Phys. Rev. Lett.*, 2012, **108**(14), 140405.
- 62 J. He and T. M. Tritt, Advances in thermoelectric materials research: Looking back and moving forward, *Science*, 2017, **357**(6358), eaak9997.
- 63 M. N. Ali, Q. Gibson, S. Jeon, B. B. Zhou, A. Yazdani and R. J. Cava, The crystal and electronic structures of Cd₃As₂, the three-dimensional electronic analogue of graphene, *Inorg. Chem.*, 2014, **53**(8), 4062–4067.
- 64 J. Zhu, C. Lee, F. Mahmood, T. Suzuki, S. Fang, N. Gedik and J. G. Checkelsky, Comprehensive study of band-structure-driven thermoelectric response of ZrTe₅, *Phys. Rev. B*, 2022, **106**(11), 115105.
- 65 Y. Zheng, T. J. Slade, L. Hu, X. Y. Tan, Y. Luo, Z.-Z. Luo, J. Xu, Q. Yan and M. G. Kanatzidis, Defect engineering in thermoelectric materials: What have we learned?, *Chem. Soc. Rev.*, 2021, **50**(16), 9022–9054.
- 66 J. Tang, H.-T. Wang, D.-H. Lee, M. Fardy, Z. Huo, T. P. Russell and P. Yang, Holey silicon as an efficient thermoelectric material, *Nano Lett.*, 2010, **10**(10), 4279–4283.
- 67 S. Tang and M. S. Dresselhaus, Constructing a large variety of Dirac-cone materials in the Bi_{1-x}Sb_x thin-film system, *Nanoscale*, 2012, **4**(24), 7786–7790.
- 68 S. Tang, Using pseudo-ZTs for thermoelectric materials search, *ES Mater. Manuf.*, 2019, **4**, 45–50, DOI: [10.30919/esmm5f213](https://doi.org/10.30919/esmm5f213).
- 69 P. Wei, W. Bao, Y. Pu, C. N. Lau and J. Shi, Anomalous thermoelectric transport of Dirac particles in graphene, *Phys. Rev. Lett.*, 2009, **102**(16), 166808.
- 70 L. D. Hicks and M. S. Dresselhaus, Effect of quantum-well structures on the thermoelectric figure of merit, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1993, **47**, 12727–12731.

