Supplementary Information (SI) for Journal of Materials Chemistry C.

Data-Driven Exploration of $NaXTe_2$ (X=Al, Ga, In): From High-Throughput Screening to Tailored Optoelectronic Functionalities

Changcheng Chen[a*], Wenkang Yu[a], Xinhui Zhang[a], Chunling Zhang[a], Yue Cheng[a], Weijun Wang[a], Chunlian Xiong[a], Xunzhe Zhang[a], Jinkang Yu[a], Zhengjun Wang[a], Gang Liu[b], Jiangzhou Xie[c], Yaxin Xu[d], Xiaoning Guan[b*], Pengfei Lu[b,e*]

a. School of Science, Xi'an University of Architecture and Technology, Xi'an, 710055, China.

b. State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

c. School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney, New South Wales 2052, Australia.

d. MOE Key Laboratory for Non-Equilibrium Synthesis and Modulation of Condensed Matter, Key Laboratory for Advanced Materials and Mesoscopic Physics of Shaanxi Province, School of Physics, Xi'an Jiaotong University, Xi'an 710049, China.

e. Ningbo Weiyuan Optoelectronic Research Institute Co.,Ltd.,Ningbo 315200, China.

Email: chenchangcheng@xauat.edu.cn(C. Chen), guanxn@bupt.edu.cn(X. Guan), photon.bupt@gmail.com (P. Lu).

Section S1: Machine Learning Model Details

Data preprocessing

All descriptor values were extracted from the Materials Project database. The initial dataset contained 723 compounds, some with missing values. After removing entries with incomplete information, 439 valid samples were retained for model training. All features were standardized using StandardScaler from the scikit-learn library to have zero mean and unit variance prior to model training.

Hyperparameter settings

All models were implemented using scikit-learn (except XGBoost, which used the xgboost library) and optimized via grid search with 5-fold cross-validation on the training set. The optimal hyperparameters for each model are summarized in Table S1. Table S2 lists the complete set of nine descriptors used in this work, including their categories, sources, and descriptions.

Table S1 Optimal hyperparameters for all regression models.

Model	Hyperparameters
SVR	kernel='rbf', C=15, gamma=0.1, epsilon=0.2
Random Forest (RF)	n_estimators=300, max_depth=15, max_features='sqrt', min_samples_leaf=2, min_samples_split=3, random_state=42
Gradient Boosting (GBR)	n_estimators=300, learning_rate=0.05, max_depth=4, max_features='sqrt', min_samples_leaf=5, min_samples_split=10, subsample=0.8, random_state=42
XGBoost	n_estimators=400, learning_rate=0.1, max_depth=4, min_child_weight=3, subsample=0.8, colsample_bytree=0.6, gamma=0.1, reg_alpha=0.1, reg_lambda=1, random_state=42
Decision Tree (DT)	max_depth=15, min_samples_leaf=3, min_samples_split=3, random_state=42

Table S2 Complete list of descriptors used in this study.

Descriptor	Category	Source	Description
CBM	Electronic structure	Materials Project	Conduction band minimum (eV)
VBM	Electronic structure	Materials Project	Valence band maximum (eV)
FermiEnergy	Electronic structure	Materials Project	Fermi energy (eV)
FormationEnergy	Thermodynamic	Materials	Formation energy per atom (eV/atom)

EnergyAboveHull	Thermodynamic	Project Materials	Energy above convex hull (eV/atom)
CellVolume	Structural	Project Materials	Unit cell volume ($Å^3$)
MassDensity	Structural	Project Materials	Mass density (g/cm^3)
DirectGap	State indicator	Project Materials	Whether the band gap is direct (1) or indirect (0)
IsMagnetic	State indicator	Project Materials	Whether the material is magnetic (1) or non-magnetic (0)

Table S3: Comparison of band gaps (in eV) calculated with different methods.

Compound	PBE	PBE+SOC	HSE06
$NaAlTe_2$	1.29	1.28	1.96
$NaGaTe_2$	0.31	0.21	0.82
$NaInTe_2$	0.66	0.62	1.31

Table S4: Comparison of elastic constants (in GPa) for $NaInTe_2$ calculated with different van der Waals functionals.

Functional	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	C_{66}	C_{11}/C_{12}	$\lvert C_{11}-C_{12}\rvert$
optB88-vdW	18.68	19.11	5.04	56.64	6.78	8.93	0.98	0.43 GPa
optB86b-vdW	21.34	24.94	8.30	58.69	12.04	8.14	0.86	3.60 GPa
rev-vdW-DF2	22.99	29.72	10.84	62.21	12.47	9.61	0.77	6.73 GPa

Table S5: Comparison of possible approaches for handling marginal Born violations in elastic constants.

Procedure	Description	Limitations
Raw DFT reporting	Report original values and note marginal Born violation	Precludes polycrystalline averaging; limits quantitative comparison with other compounds
Full symmetrization	Force $C_{11} = C_{22} = (C_{11}+C_{12})/2$	Artificially suppresses intrinsic anisotropy; would alter the relative ranking of materials
Structural re-optimization	Re-optimize structure with stricter convergence criteria	Computationally expensive; unlikely to change a 0.44 GPa difference given typical DFT precision
Error sensitivity analysis	Use C_{11} and C_{12} within their uncertainty range	Useful for validation but does not provide a single consistent tensor for analysis
Our approach: minimal	Swap C_{11} and C_{12} while preserving their average	Preserves all physical trends and relative rankings; enables consistent polycrystalline averaging;

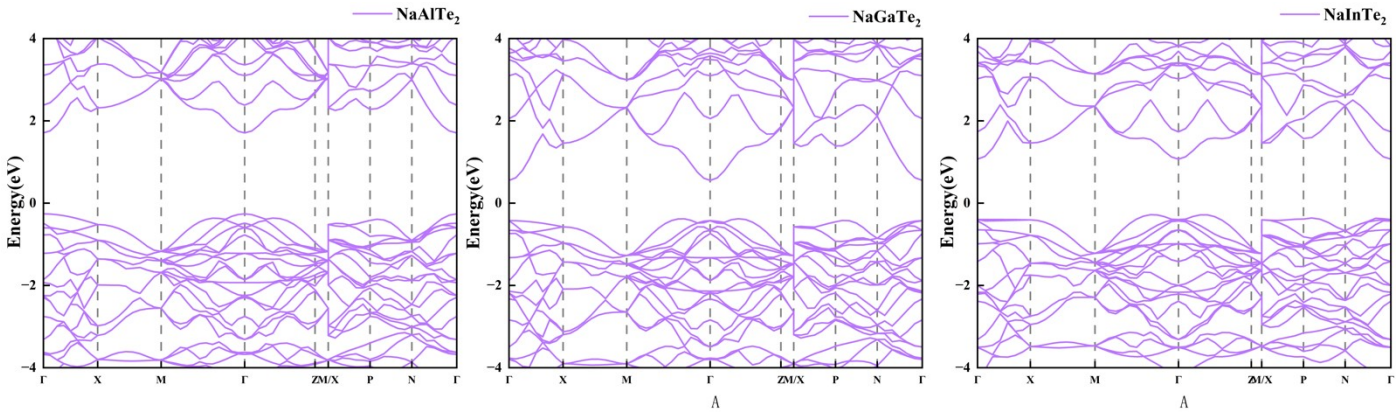

Fig. S1: HSE06 band structures of $NaXTe_2$ (X = Al, Ga, In) ◦

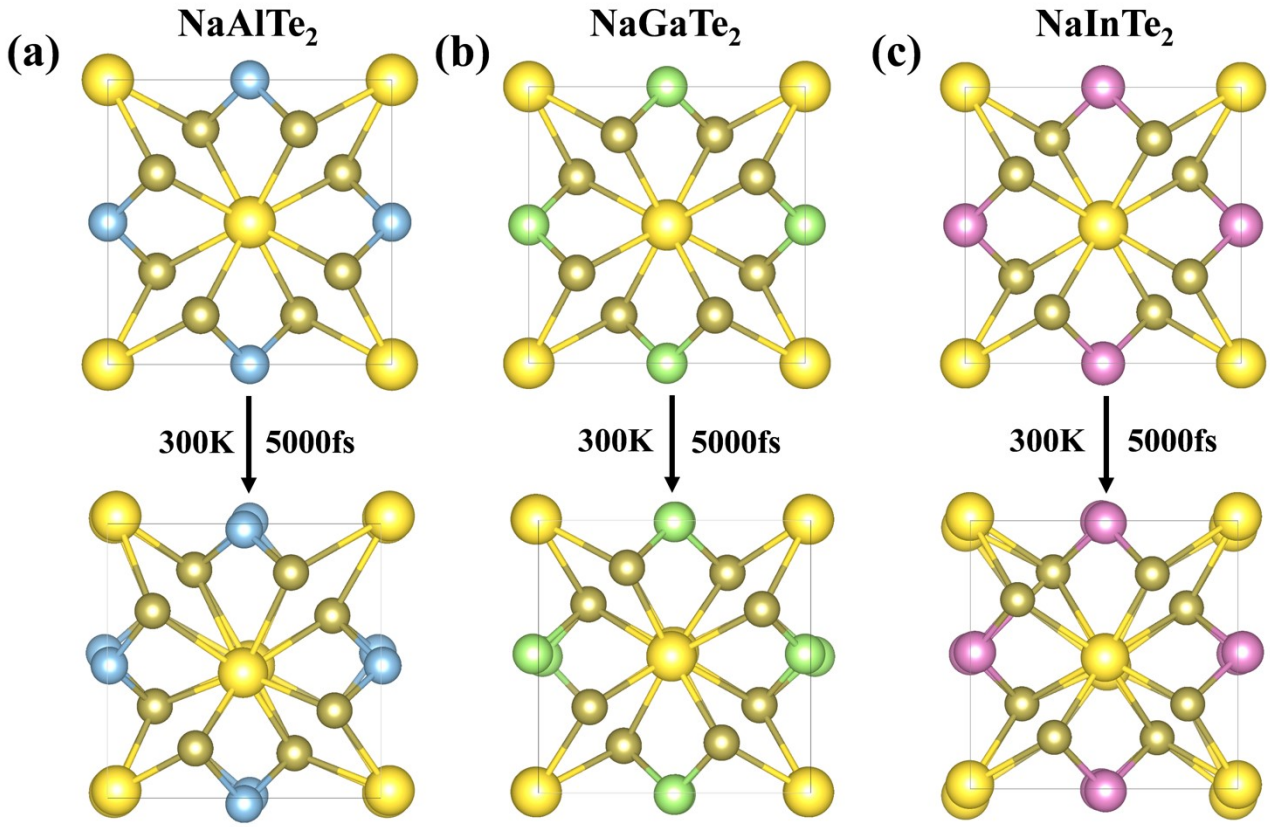

Fig. S2 Final snapshots of AIMD simulations at 300 K after 5 ps for (a) $NaAlTe_2$, (b) $NaGaTe_2$, and (c) $NaInTe_2$. The structures maintain their integrity with no bond breaking or reconstruction, confirming thermal stability.

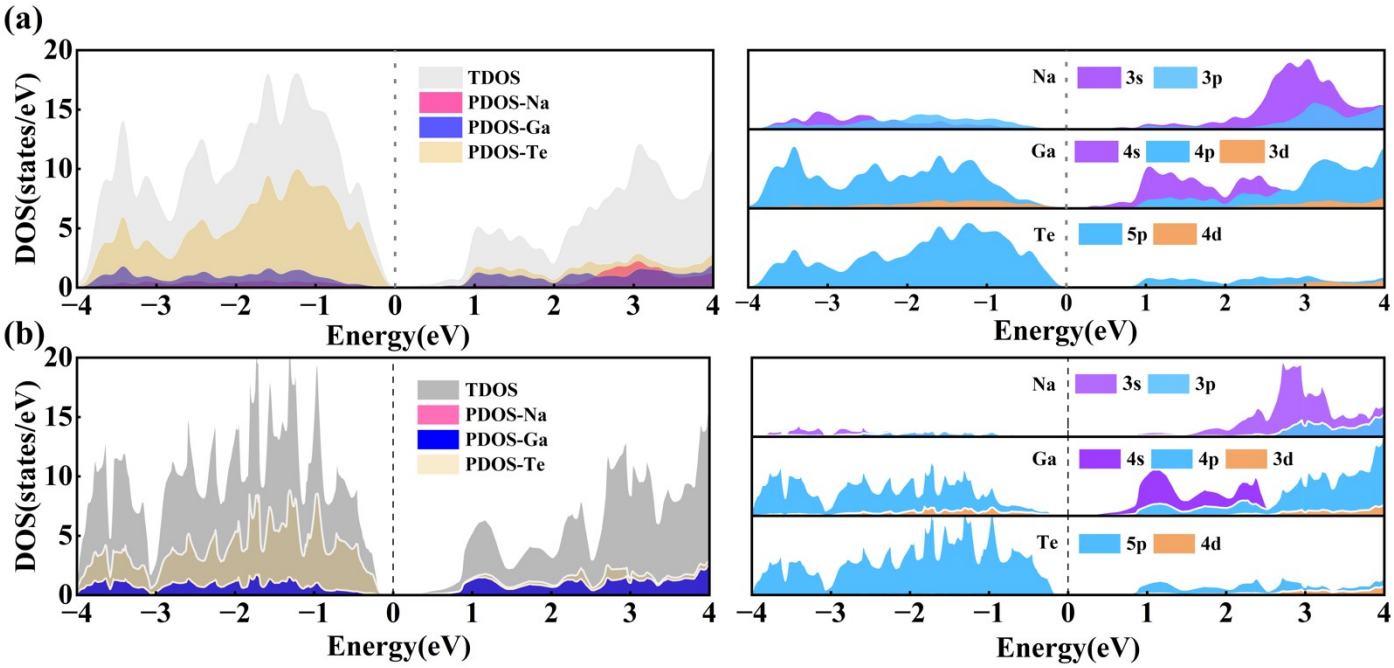

Fig. S3: Projected density of states (PDOS) of $NaGaTe_2$ (a) without and (b) with spin-orbit coupling.

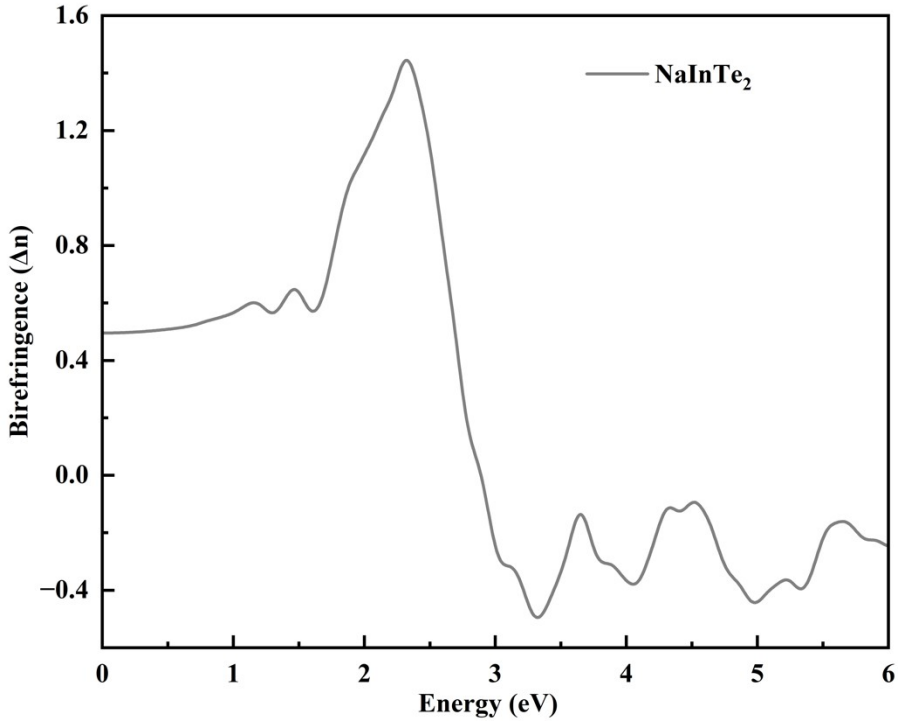

Fig. S4: Birefringence of $NaInTe_2$ as a function of photon energy