Electronic Supplementary Material (ESI) for Nanoscale.
This journal is © The Royal Society of Chemistry 2021

Supplementary Information

High-Efficient *ab initio* Bayesian Active Learning Method and Applications in Prediction of Two-dimensional Functional Materials

Xing-Yu Ma,[1†] Hou-Yi Lyu,[1, 2†] Kuan-Rong Hao,[1] Zhen-Gang Zhu,[3] Qing-Bo Yan,[2*] Gang Su[4, 1, 2*]

[1]School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.

[2]Center of Materials Science and Optoelectronics Engineering, College of Materials Science and Optoelectronic Technology, University of Chinese Academy of Sciences, Beijing 100049, China.

[3]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China.

[4]Kavli Institute for Theoretical Sciences, and CAS Center of Excellence in Topological Quantum Computation, University of Chinese Academy of Sciences, Beijing 100190, China.

*Correspondence authors. Email: yan@ucas.ac.cn; gsu@ucas.ac.cn

Table of contents

1. **How to generate 3,119 2D hexagonal binary compounds (MX and MX_2)**.......3
2. **The first-principles calculations and relevant physical properties**...................5
3. **Definition of 22 initial features**...8
4. ***Ab initio* Bayesian active learning details**..9
5. **Supervised machine learning details** 15
6. **The prediction of 2D ferroelectric HBCs** 23
7. **The energy versus polarization and Monte Carlo simulation**..........................26
8. **2D ferroelectric photovoltaic materials BiAs and BiSb**...................................29
9. **The stability of the predicted 2D ferroelectric HBCs** 31
10. **Supplementary References**..33

1. How to generate 3,119 2D hexagonal binary compounds (MX and MX_2)

Table S1. 45 elements for M and X sites of MX

Site	Elements
M, X	Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd, Ag, Cd, In, Sn, Sb, La, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg, Tl, Pb, Bi, Al, Si, P, B, C, N

Table S2. 50 elements for M and X sites of MX_2

Site	Elements
M, X	Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd, Ag, Cd, In, Sn, Sb, La, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg, Tl, Pb, Bi, Al, Si, P, B, C, N, Be, Ma, Ca, Sr, Ba

Different 2D MX and MX_2 materials are generated by replacing M and X atoms by 45 and 50 elements (Table S1 and S2) across the periodic table, respectively. Note that MX_2 and XM_2 structures are stoichiometrically inequivalent, and MX and XM structures are stoichiometrically equivalent. There are total 2450 MX_2 and 990 MX structures. In addition, when MX structures contain alkaline earth metals, the vertical displacement between M and X atoms (d) is about 3.30 Å, such as 3.30 Å for BaSr and 3.25 Å for LaBa (see Fig. S1). Such displacements are far longer than those of stable MX structures (see d in Table S8), thus those structures containing alkaline earth metals could be unstable and we preclude alkaline earth metals in MX. Before executing the machine learning training, the data cleaning method is performed to ensure the consistency of data by removing the abnormal and unnecessary data. We performed full relaxed geometric structure optimizations and discovered that two X atoms of 315 structures in 2450 MX_2 are not sitting on the same plane (see Fig. S1), and they could not form a hexagonal lattice. The remaining 2135 MX_2 structures are considered to guarantee the consistency of data. We also found that the optimization of geometric structures of 6 MX structures cannot be finished because their self-consistent calculations of electrons do not converge, leaving us 984 MX structures. To the end, the total 3,119 2D hexagonal binary compounds (HBCs) are considered as the total data set.

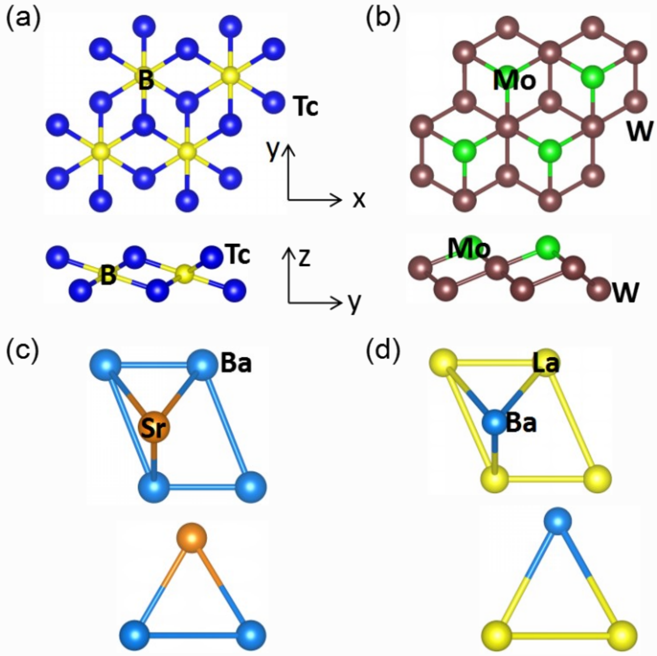

Figure S1. Geometric structures (top and side view) of (a) BTc_2, (b) MoW_2, (c) BaSr and (d) LaBa.

2. The first-principles calculations and relevant physical properties

The first-principles calculations are performed based on the density functional theory (DFT) implemented in the Vienna *Ab initio* Simulation Package (VASP) [1, 2] within projected augmented wave (PAW) method.[3,4] Perdew-Burke-Ernzerhof generalized gradient approximation (PBE-GGA) [5] for exchange-correlation potential was adopted. A cutoff of 450 eV was chosen, and the Brillouin zone was sampled using Γ-centered Monkhorst–Pack k-point grids of 12×12×1. Optimized atomic structures were achieved when the force on all atoms and the energy were converged to 0.01 eV/Å and 10^{-6} eV, respectively. All electronic structures are calculated using the PBE-GGA + U method. We apply Hubbard U corrections to account for Coulomb correlated potential in 3d, 4d and 5d transition metals, which were chosen as U = 4 eV, 2.5 eV and 0.5 eV, respectively.[6, 7] In addition, the electronic structures of 24 stable ferroelectric materials were calculated by the hybrid functional of Hyed-Scuseria-Ernzerhof (HSE).[8] The phonon calculations were calculated using the finite displacement method with a 6×6×1 supercell.[9, 10] The phase transition between ferroelectric and paraelectric phases was investigated by calculating the minimum energy pathway for polarization switching using the climbing image nudged elastic band (CI-NEB) method.[11] We performed *ab initio* molecular dynamics (AIMD) simulations at 300 K by using canonical ensemble (NVT) with Nosé-Hoover thermostat.[12, 13] The total simulation time is 12 ps with a time step of 3 fs. Here, the heat of formation (E_f) is calculated by the expression $E_f = \frac{E(MX) - E(M) - E(X)}{2}$, where E(MX), E(M) and E(X) represent the energy of MX per unit cell, the average energy (per atom) of simple substance M and simple substance X, respectively.

The total out-of-plane polarization has the form of $P = P_{\mathrm{ion}} + P_{\mathrm{el}}$, where P_{ion} and P_{el} represents the out-of-plane ion polarization and out-of-plane electron polarization, respectively. We used the point charge model to calculate the out-of-plane ion polarization. The equation was frequently adopted to obtain the out-of-plane electron polarization in previous works [14, 15] in the form of $P_{\mathrm{el}} = \iiint z\rho(\vec{r})\mathrm{d}\vec{r}$, where $\rho(\vec{r})$ is the electron density and *z* is the spatial position along the *z* direction. The conversion equations for different types of units of polarization have the following three forms $P(\mathrm{pC/m}) = P(e\text{Å})/S, P(\mathrm{e/cm^2}) = P(e\text{Å})/V$, where $P(e\text{Å})$, $P(\mathrm{pC/m})$, and P(e/cm^2) are the polarization with units of eÅ/unit cell, pC/m, and e/cm^2, respectively, S and V are the area and volume of the unit cell of specific materials, respectively. Our results show that the out-of-plane polarization of WTe$_2$ bilayer is 2.35×10^{11} e/cm^2, which is nearly consistent with the value of 2×10^{11} e/cm^2 experimentally measured at 20K.[16]

We calculated the carrier mobility (μ) using the deformation potential theory.[17] The carrier mobility (μ) of 2D materials can be calculated by $\mu_{2D} = \frac{e\hbar^3 C_{2D}}{k_B T m^* m_d {E_i}^2}$, here, $\hbar$, k_B and T represent the reduced Planck constant, Boltzmann constant and temperature (here 300 K in this work), respectively. The elastic modulus (C_{2D}) is given by $C_{2D} = \frac{1}{S_0} \frac{\partial^2 E}{\partial (l/l_0)^2}|_{l=l_0}$, where E represents the total energy of the supercell and S_0 represents the area of the optimized supercell, $m^* = \hbar^2 (\frac{\partial^2 E}{\partial k^2})^{-1}$ represents the carrier effective mass along x and y directions, and $m_d = \sqrt{{m^*}_x {m^*}_y}$ represents the reduced effective mass of carriers. E_i is the deformation potential constant of the conducting band minima (CBM) (electron) and valence band maxima (VBM) (hole) along the transport direction, obtained by $E_i = \Delta E/(\Delta l/l_0)$, where ΔE represents the energy change of the CBM or VBM under a proper cell compression and dilatation, l_0 and Δl represents the lattice constant along the transport direction and the length of deformation on l_0 ($\Delta l/l_0$ is set to be 0.1% in this work), respectively.

The quasiparticle (QP) energies were calculated via the perturbative solution to the Dyson equation based on the GW approximation [18, 19] in the form of $[-\frac{\hbar^2}{2m_e}\nabla^2 + V_{ion} + V_H + \sum E_{nk}^{QP}]\psi_{nk}^{QP} = E_{nk}^{QP}\psi_{nk}^{QP}$, where m_e represents the mass of electron, $\hbar$ represents the reduced Planck constant, V_{ion} represents the electrostatic potential contributed by ions, V_H represents the Hartree potential, and E_{nk}^{QP} represents the quasi-particle energy. The optical excitation energies and exciton wave functions are calculated by the Bethe-Salpeter equation (BSE),[20-22] $\left(E_{ck}^{QP}-E_{vk}^{QP}\right)A_{vck}^{S}+\sum_{v'c'k'}\langle vck\,|\,K^{eh}\,|\,v'c'k'\rangle A_{v'c'k'}^{S}=\Omega^{S}A_{vck}^{S}$, where E_{ck}^{QP} and E_{vk}^{QP} represent the quasi-particle energies for conduction and valence bands, respectively, K^{eh} and Ω^{S} represent the electron-hole interaction kernel and excitation energy, respectively. The imaginary part is calculated through the excitation energies and exciton wave functions,[22] given by $\varepsilon(\omega)=\frac{16\pi e^{2}}{\omega^{2}}\sum_{S}|\vec{\lambda}\langle 0\,|\,\vec{v}\,|\,S\rangle|^{2}\delta(\omega-\Omega^{S}$), where $\vec{\lambda}$ and $\vec{v}$ represent the polarization vector of the incident light and velocity operator, respectively, and $\langle 0\,|\,\vec{v}\,|\,S\rangle$ represent the transition matrix element. The absorption coefficient $\alpha(\omega)$ can be obtained by the following equation,[23] $\alpha(\omega)=\sqrt{2}\omega[\sqrt{\varepsilon_{1}^{2}(\omega)+\varepsilon_{2}^{2}(\omega)}-\varepsilon_{1}(\omega)]^{1/2}$, where $\varepsilon_{1}(\omega)$ and $\varepsilon_{2}(\omega)$ represent the real and imaginary parts of the dielectric function, respectively. The kinetic energy cutoff for GW and BSE is taken as 300 eV, and we use 10 occupied and 10 unoccupied orbitals to determine the electron-hole interaction kernel.

3. Definition of 22 initial features

Table S3. 22 Initial features with definition

Features	Definition
$MDEDM$	Manhattan distance of eigenvalue of the distance matrix, named as $MDEDM = \frac{1}{2}\sum_{i,j}\left\|\lambda_i - \lambda_j\right\|$, where λ_i and λ_j are the i-th and j-th eigenvalues of the distance matrix, respectively.
d	The distance between M and X atoms along the plane normal
r_M, r_X	Ion radii of M and X-site atoms
IE_M, IE_X	Ionization energy of M and X-site atoms
P_M, P_X	Ionic polarizability of M and X-site atoms
E_M, E_X	Martynov-Batsanov electronegativity of M and X-site atoms
IC_M, IC_X	Ionic charge of M and X-site atoms
n_M, n_X	Number of valence electrons of M and X-site atoms
$n_{M(in)}$, $n_{M(out)}$, $n_{X(in)}$, $n_{X(out)}$	Number of valence electrons in the inner/outer shell of M and X-site atoms
$r_{M(in)}$, $r_{M(out)}$, $r_{X(in)}$, $r_{X(out)}$	Inner/outer valence electrons orbital radius of M and X-site atoms

The distance matrix can be defined as **D** (d_{kl}), where d_{kl} represents the distance between k-th atom and l-th atom, which is an $n \times n$ square matrix, where n is the number of atoms in unit cell. We use the data of element-related property, such as the orbital radius of valence electron from the python Mendeleev package 0.4.1[24] and Ref. [25], respectively. Here, we mean the 'outer valence electrons' and 'inner valence electrons' by taking Fe atom as an example. For the valence electron configuration of Fe atom with $3d^6 4s^2$, the inner and outer valence electron configuration represents $3d^6$ and $4s^2$ electrons, respectively.

4. *Ab initio* Bayesian active learning details

4.1 Distributions of polarization and band gap of all 2D HBCs

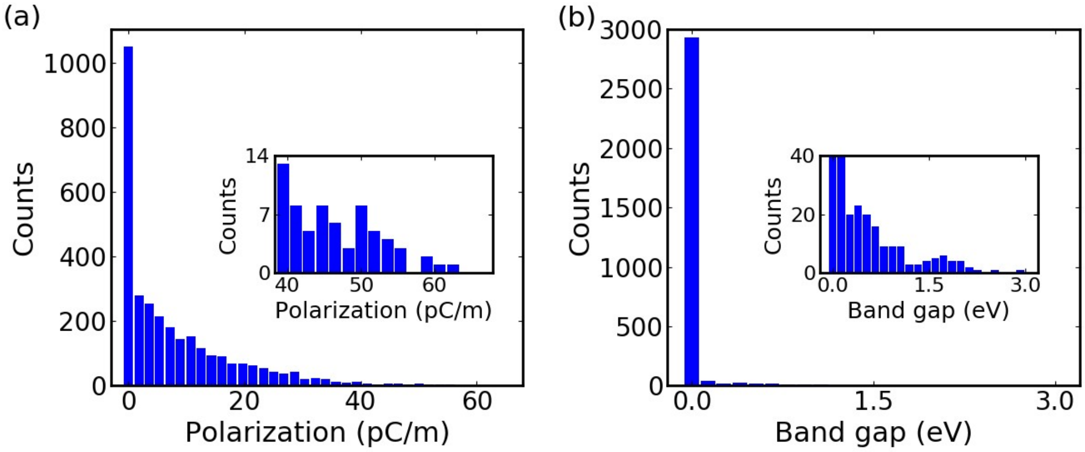

Figure S2. Distributions of (a) polarization and (b) band gap obtained from calculations of all 2D HBCs by DFT.

As shown in Fig. S2 (a), the distributions of the polarizations and band gaps of all 2D HBCs obtained by calculating all the candidates via DFT confirm that the counts significantly decrease as polarization and band gap values increase, respectively. Fig. S2(b) shows the extremely unbalanced distributions of band gaps for all 2D HBCs, where most of the 2D HBCs are metal and the nonmetallic candidates are extremely rare. The maximum polarization is 64.359 pC/m, corresponding with $BaPt_2$. The band gap of $GeBi_2$ is 1.490 eV, which is the closest to 1.5 eV.

4.2 Gaussian process regression

The Gaussian process regression (GPR) [26] is adopted as the surrogate model in active learning, which is in the open-source scikit-learn package.[27] Gaussian process (GP) is a collection of random variables satisfying a joint Gaussian distribution. The targeted functions $f(x)$ can be described by using the Gaussian process, which has the form $f(x) \sim GP(m(x), k(x,x^{'}))$ where x and $x^{'}$ are input feature vectors of two different data, m(x) and $k(x,x^{'})$ are their mean function and kernel function, respectively. GPR is a non-parametric regression technique, which builds a distribution of functions that are accord with the training data set. An n-dimensional squared exponential kernel is adopted as the kernel function, which has the form $k(x,x^{'}) = exp\left\{ - \frac{\sum_{i=1}^{n} |x_i - x_i^{'}|^2}{\sigma_i^2} \right\}$, where x and $x^{'}$ are input features vectors of two different data, x_i is the i-th feature of n-dimensional input feature vectors for each data, n is the dimensional of input feature vectors of each data, and σ_i^2 is a hyper-parameter obtained using the maximum likelihood estimate.

4.3 The number of added candidates at each iteration

Table S4. Dependence of the optimization performance for searching structures with maximal polarization on the number of calculated structures added at each iteration.

Number of calculated structures added at each iteration	1	**5**	10	15	20	25	30	35	40	45	50
Time spent for training GPR model (s)	47	**9**	9	11	7	8	8	8	10	13	10
The total number of calculated structures	43	**40**	60	75	66	72	84	84	101	116	109

Table S5. Dependence of the optimization performance for searching structures with band gap the closest 1.50 eV on the number of calculated structures added at each iteration.

Number of calculated structures added at each iteration	1	**5**	10	15	20	25	30	35	40	45	50
Time spent for training GPR model (s)	22	**6**	16	14	12	10	11	12	11	22	18
The total number of calculated structures	34	**35**	67	64	71	75	84	95	90	119	119

In the *ab initio* Bayesian active learning, the number of candidates added at each iteration is important as it influences the optimization speed. Table S4 and S5 show the testing runs for different choice of numbers. If the number is small, the fitting to the prediction model needs to be performed too often, which results in a long training time. If it is large, the total number of calculations needed to obtain the final optimal result becomes large, resulting in the huge DFT calculations. By balancing these two factors, we chose the number to be 5. Note that 15 independent optimization runs are performed for searching for the materials with maximum polarization and band gap the closest 1.50 eV from randomly sampling 400 candidates in the total data set, respectively. Time spent for training GPR model and the total number of structures that need to be calculated for all iterations are the average of these 15 independent runs. Due to the small training data set, the time spent of training GPR model is much shorter than that of DFT calculations. However, for a larger data set, the time spent of training GPR model becomes very long and could not be ignored (see Fig. S7 (c)(d)).

4.4 Comparison of Optimizations via the *ab initio* Bayesian active learning and random search

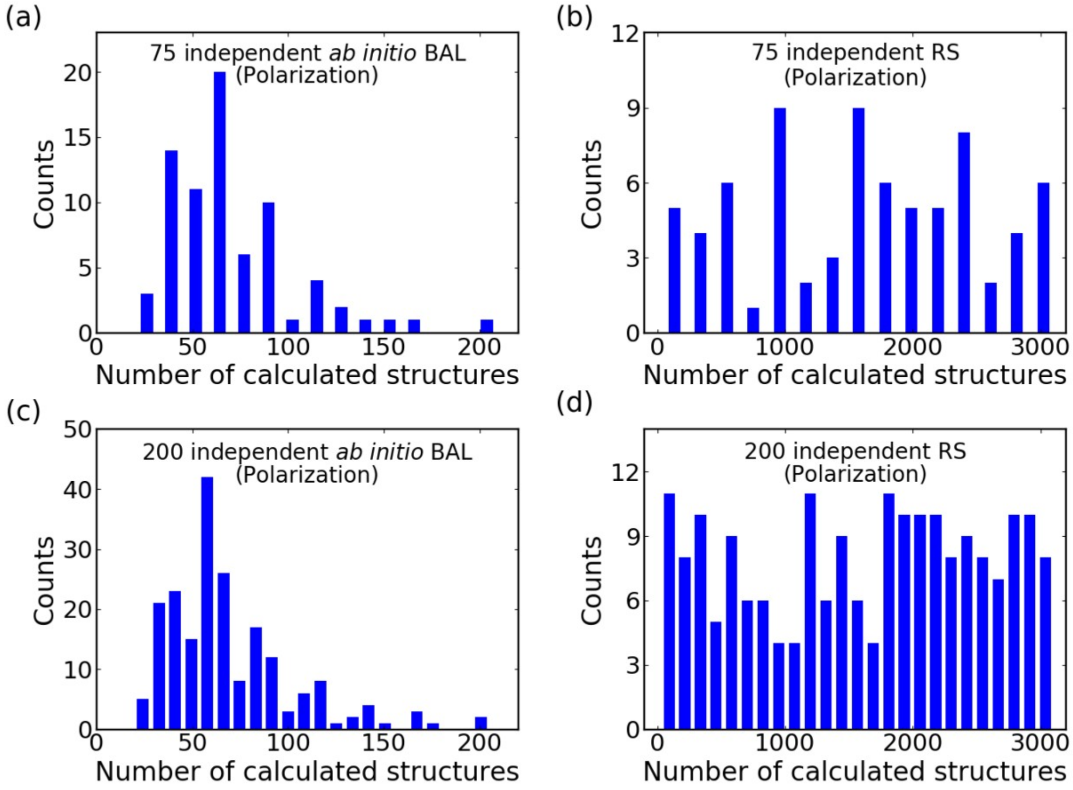

Figure S3. Histogram of the number of calculated structures required to discover the structures with maximum electric polarization in 75 and 200 independent optimization runs. (a)(c) The 75 and 200 optimization runs via the *ab initio* Bayesian active learning (BAL), respectively. (b)(d) The 75 and 200 optimization runs via random search (RS), respectively.

The 75 and 200 independent optimization runs are conducted with different initial 15 structures. Distributions of how many calculations it took to find the structures with maximum polarization (optimal structure) in each case are shown in Fig. S3. The optimal structure is found within about 65 calculated structures in the most of 75 and 200 optimization runs via the *ab initio* Bayesian active learning, respectively (Figs. S3(a) and (c)). For random search optimization, the distributions for random search are very random (Figs. S3(b) and (d)). Moreover, for 75 optimization runs, the average number of calculated structures of the above two methods is 71 and 1620, respectively, which are consistent with the cases of 200 optimization runs (the numbers are 69 and 1618).

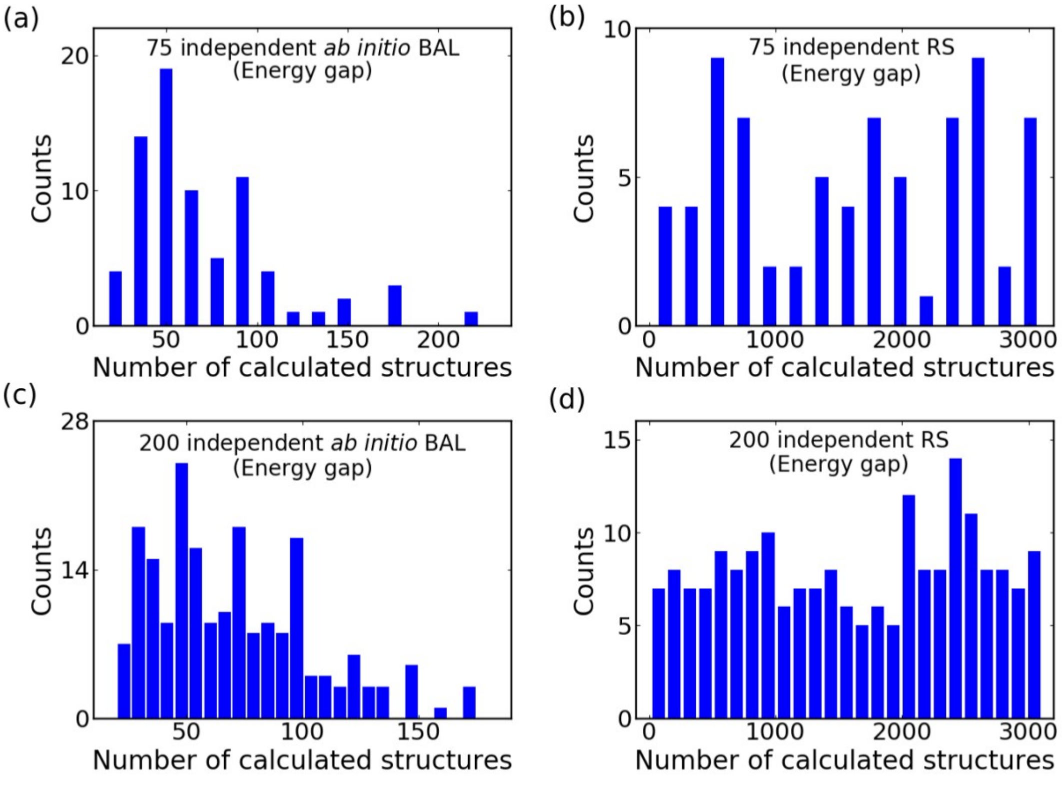

Figure S4. Histogram of the number of calculated structures required to discover the structures with band gap the closest 1.5 eV in 75 and 200 independent optimization runs. (a)(c) The 75 and 200 optimization runs via the *ab initio* Bayesian active learning (BAL), respectively. (b)(d) The 75 and 200 optimization runs via random search (RS), respectively.

We performed the 75 and 200 independent optimization runs beginning with different initial 15 structures for finding the structures with band gap the closest 1.5 eV (optimal structure). The optimal structure is discovered within about 50 calculated structures in the most of 75 and 200 optimization runs by the *ab initio* Bayesian active learning, respectively (Figs. S4(a) and (c)). The distributions for random search are very random (Figs. S4(b) and (d)). The average number of calculated structures for 75 optimization runs of the above two methods is 70 and 1612, respectively, which are accord with the cases of 200 optimization runs (the numbers are 70 and 1614).

4.5 Comparison of Optimizations via the *ab initio* Bayesian active learning and greedy method

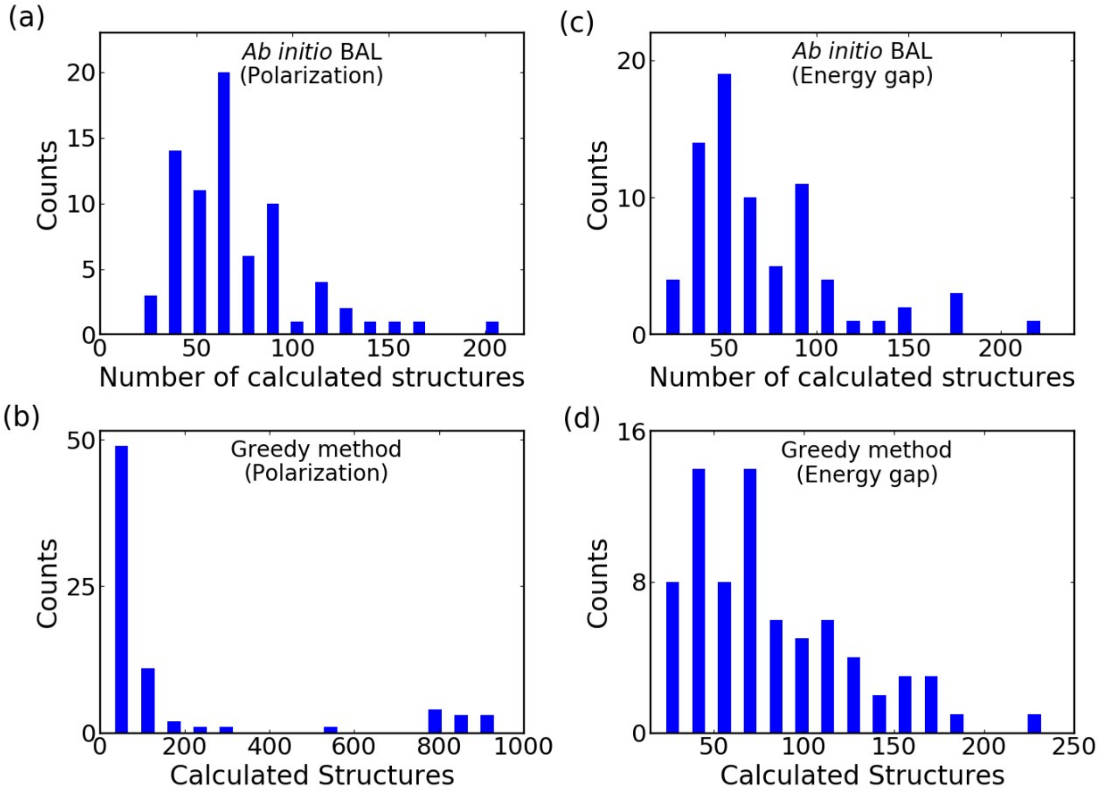

Figure S5. Histogram of the number of calculated structures required to discover the structures with maximum polarization and band gap the closest 1.5 eV in 75 independent optimization runs, respectively. (a)(c) The 75 optimization runs via the *ab initio* Bayesian active learning (BAL). (b)(d) The 75 optimization runs via the greedy method, respectively.

We compare the *ab initio* Bayesian active learning with a greedy method, and latter selected the top five materials with highest prediction of polarization or bandgap at each iteration. The 75 independent optimization runs are conducted with different initial 15 structures. For finding the structure with maximum polarization, the average number of calculated structures for 75 optimization runs of the above two methods is 71 and 183, respectively. For discovering the structure with band gap the closest 1.5 eV, the average values for 75 optimization runs of the above two methods is 70 and 81, respectively. It is dictated that the calculated candidates of the *ab initio* BAL are less than that of the greedy method. Because the greedy method tends to optimize data locally and it may need a lot of data to reach the global optimal value.

5. Supervised machine learning details

5.1 Gradient boosting regression (GBR)

The Gradient boosted regression (GBR) [28] is a flexible non-parametric supervised machine leaning algorithm in the open-source scikit-learn package.[27] The algorithm was a highly accurate algorithms combining several custom base weakly accurate algorithms (e.g., the decision tree algorithm), each of which contributes to the performance of the final model. The principle to improve the accuracy of the results of final model is to minimize the loss function on the training data set by iteratively integrating many weakly accurate algorithms. The final highly accurate algorithm is the weighted sum of several weak accurate algorithms at each iteration in the following form

$$F_M(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{S1}$$

where $h_m(x)$ are the weak algorithms, x is the input features of each data, γ_m is the weight of each weak algorithm, and M is the number of the weak algorithms.

The weight $\gamma_m$of next weak algorithms at each iteration, is calculated by minimizing the loss function (f_L) for all data points (count N), in the form of

$$\gamma_m = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{n} f_L(y_i, \mathrm{F}_{m-1}(x_i) + \gamma_m h(x_i)) \tag{S2}$$

The minimization problem was solved numerically by the steepest descent. The steepest descent direction is the negative gradient of the loss function evaluated at the current mode $\mathrm{F}_{m-1}(x_i)$.

Selecting a set of optimal hyper-parameters is necessary for improving the performance of the model. We selected the optimal five hyper-parameters in the GBR model by the grid searching method: loss function (least squares), maximum depth of the individual regression estimators (15), maximum leaf nodes of the individual regression estimators (15) and the number of learning algorithms (300).

5.2 Algorithm evaluation criteria

The coefficient of determination (R^2) is adopted to evaluate the deviation between the predicted values and actual values, which has the form

$$R^2 = 1 - \frac{\sum_i (y_i^{True} - y_i^{Pred})^2}{\sum_i (y_i^{True} - y_i^{True})^2} \tag{S3}$$

where y_i^{True} is the actual value, y_i^{Pred} is the predicted value, N is the number of data. R^2 equals to 1, implying the ideal fitting results.

The mean absolute error (MAE) is used to evaluate the overall error between the predicted values and actual values, in the following form

$$MAE = \frac{1}{N}\sum_{i}^{N} |y_i^{True} - y_i^{Pred}| \tag{S4}$$

The correlation between predictive values and actual values can be described by Pearson coefficient (r), which has the form:

$$r^2 = \frac{\sum_{i=1}^{N} (y_i^{\mathrm{True}} - \overline{y_i^{\mathrm{True}}})(y_i^{\mathrm{Pred}} - \overline{y_i^{\mathrm{Pred}}})}{\sqrt{\sum_{i=1}^{N} (y_i^{\mathrm{True}} - \overline{y_i^{\mathrm{True}}})^2 (y_i^{\mathrm{Pred}} - \overline{y_i^{\mathrm{Pred}}})^2}} \quad \text{(S5)}$$

The value of r is from -1 to +1. There is positive correlation between two variables when r is larger than zero. Otherwise, there is negative correlation between two variables. Moreover, the larger the absolute value of r, the stronger the correlation.

5.3 Comparison of various feature combinations

Table S6. Determination coefficient (R^2) and mean absolute errors (MAE) of polarization on the total data set for various feature combinations with the GBR algorithm

Table S2. 35 Initial features with definition

Feature combinations	R^2	MAE (pC/m)
MDEDM+element-related properties+*d*	**0.871**	**2.620**
element-related properties+*d*	0.802	3.086
element-related properties	0.542	5.028

Table S7. Determination coefficient (R^2) and mean absolute errors (MAE) of band gap on the total data set for various feature combinations with the GBR algorithm

Feature combinations	R^2	MAE (eV)
MDEDM+element-related properties+*d*	**0.230**	**0.055**
element-related properties+*d*	0.069	0.072
element-related properties	-0.401	0.081

To verify that MDEDM can improve the performance of model for polarization and band gap, we compare three different feature vectors (features of element-related properties; features of element-related properties plus *d*; features of element-related properties plus *d* and MDEDM). The latter gives a better performance regardless of polarization and band gap (see Table S6 and S7), indicating that MDEDM and *d* can indeed improve the performance of model for polarization and band gap. This is because polarization and band gap depend on the electron density of the materials, and the electron density is related to the atomic positions, and the geometric descriptor *d* and MDEDM describe geometrical information about atomic positions.

5.4 Comparison between different algorithms

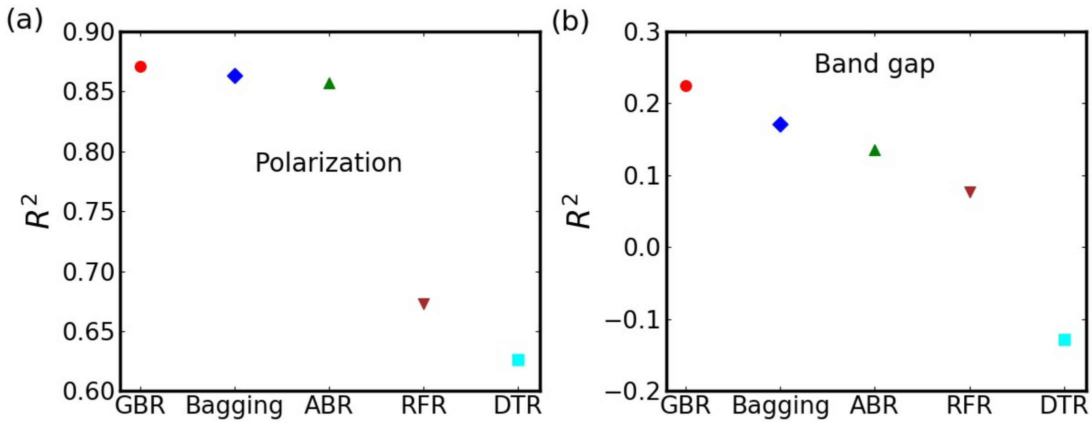

Figure S6. Determination coefficients (R^2) of five various supervised machine learning algorithms.

We tested five various supervised machine learning algorithms such as the gradient boosting regression (GBR), [28] bagging, [29] adaboost regression (ABR), [30] random forest regression (RFR),[31] and decision tree regression.[32] The results demonstrate that GBR model outperforms other algorithms regardless of polarization and band gap.

5.5 Comparison of the gradient boosting regression with Gaussian process regression algorithm

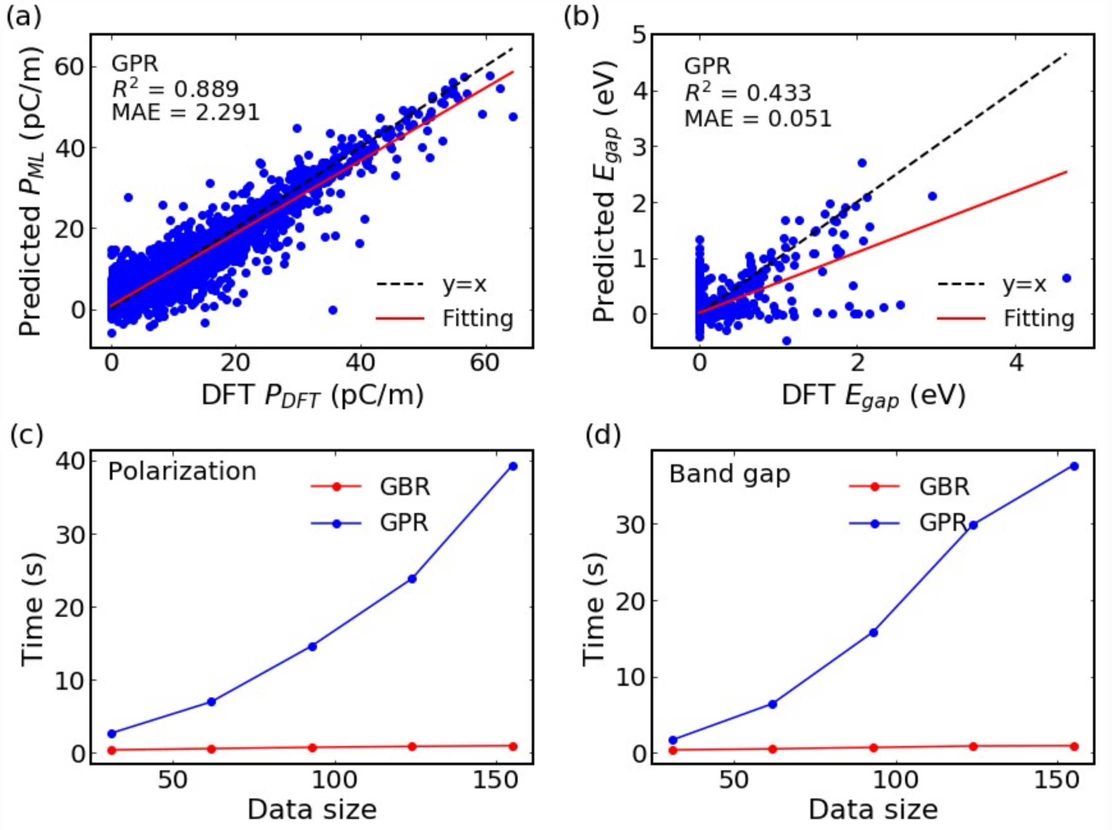

Figure S7. The five-fold cross-validation predicted values via Gaussian process regression (GPR) algorithm versus DFT-calculated values for (a) polarization and (b) band gap. The spent time of training gradient boosting regression (GBR) model and Gaussian process regression (GPR) model with different size of data for (c) polarization and (d) band gap.

As shown in Fig. S7(c)(d), the training time of the GPR model is significantly increasing with the data set size increasing, and it is several orders of magnitudes larger than the time needed for training GBR model on same data set. Meanwhile, the accuracy of GPR model is nearly consistent with that of GBR model. Thus, we adopted GBR model.

5.6 Feature reduction of supervised machine learning

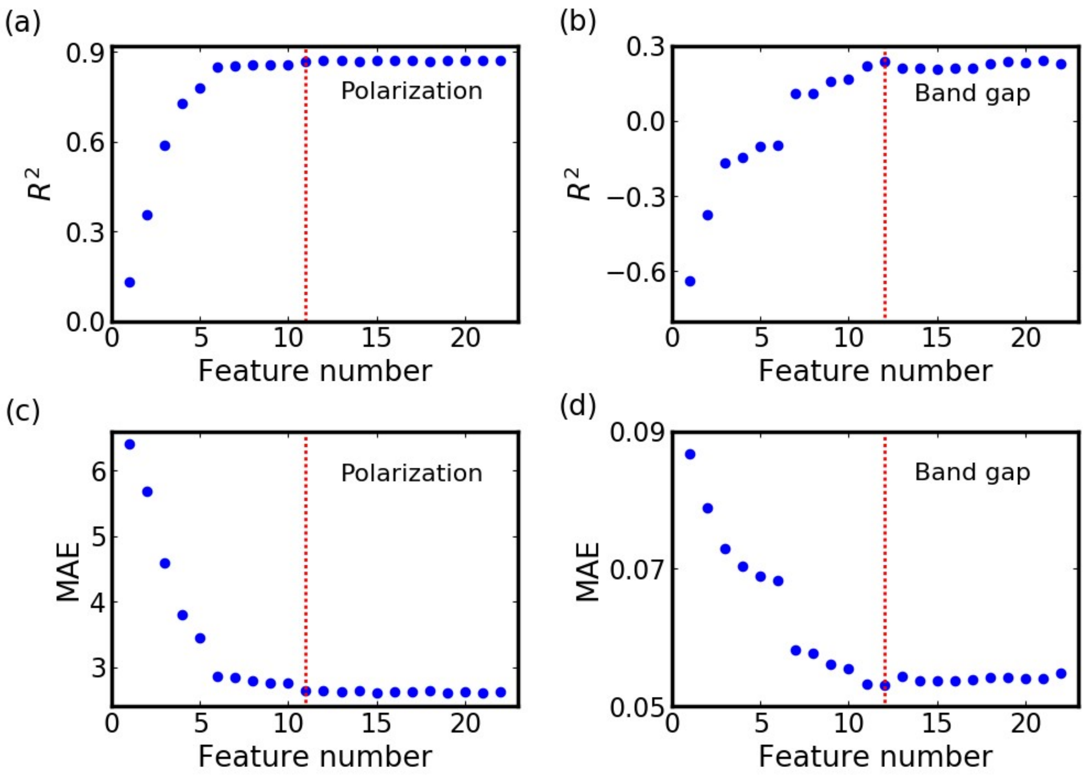

Figure S8. Determination coefficient (R^2) and mean absolute error (MAE) of GBR model with increasing the feature numbers. The positions of the red lines are the high convergence value of R^2 and the low convergence value of MAE, corresponding to 11 and 12 feature numbers for the models of polarization and band gap, respectively.

The 22 (MDEDM and *d* plus 20 element-related properties) features were chosen as the initial feature set for the models of polarization and band gap (see Table S3). We utilized GBR algorithm to efficiently exclude the features that have less impact on the polarization until R^2 and MAE achieve the high value (0.869) and the low convergence value (2.645), respectively, while only the 11 features construct an optimal features space. By utilizing GBR algorithm, we exclude the features that have less impact on the band gap until R^2 and MAE achieve maximum value (0.236) and minimum value (0.053), respectively, and an optimal feature space is constructed by the 12 features.

5.7 The results of supervised machine learning model

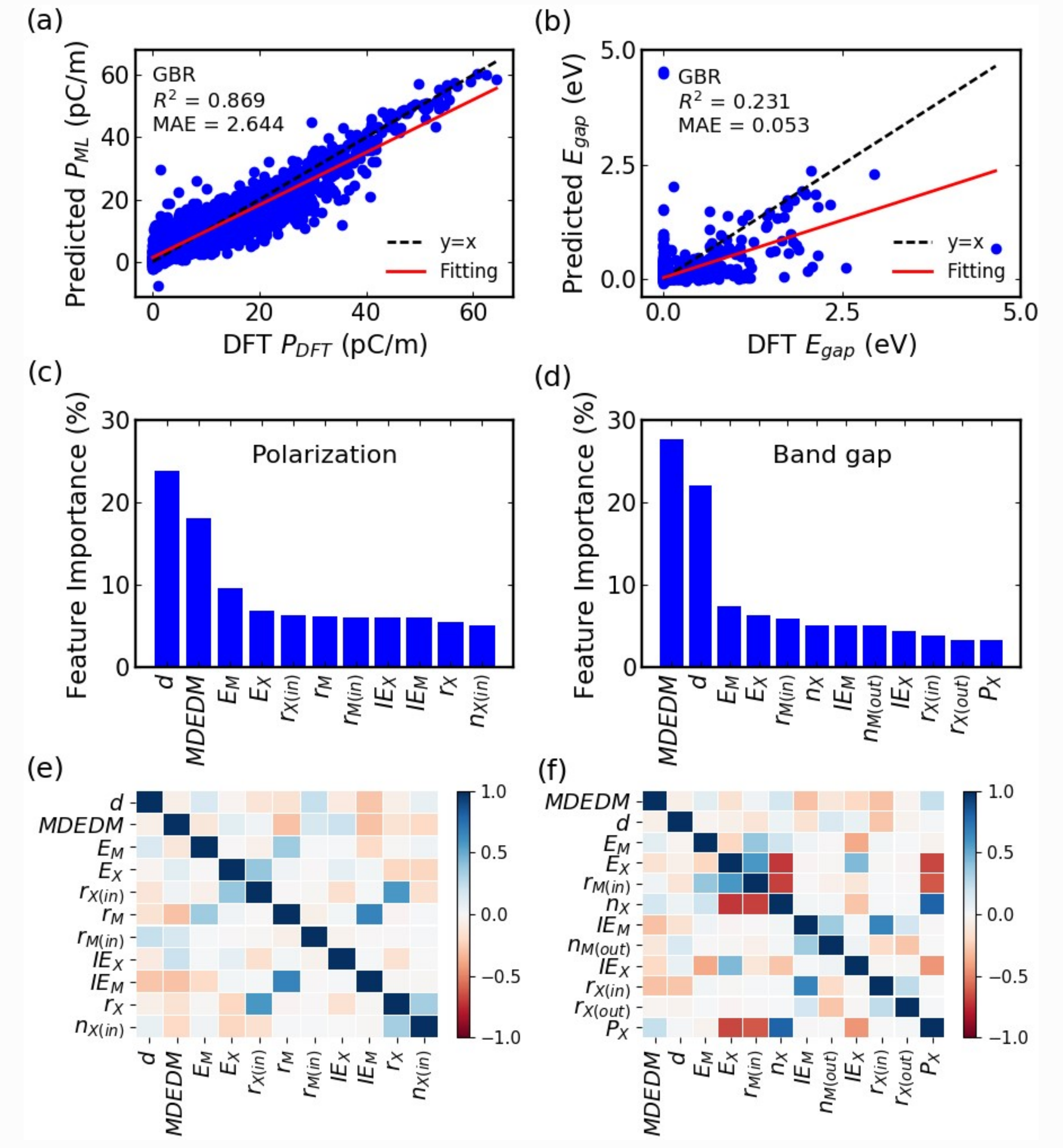

Figure S9. The predicted results, importance and correlation of the optimal features in the supervised machine learning model. (a)(b) The ten-fold cross-validation predicted values via gradient boosting regression (GBR) algorithm versus DFT-calculated values for polarization and band gap. (c)(d) The features are ranked with their importance by GBR algorithm. (e)(f) The Pearson correlation coefficient heat map of the optimal feature sets for (e) polarization and (f) band gap. The coefficient of determination (R^2) and the mean absolute error (MAE) evaluate the performance of the models. The red solid line and black dash line represent the actual fitting line and the ideal fitting line whose prediction error is zero, respectively.

Here, we build a supervised machine learning model for electric polarization and band gap. To avoid a complex supervised machine learning model, we perform a feature reduction by utilizing the GBR algorithm and exclude the feature of less impact (see Fig. S8). As a result, the two optimal feature spaces for polarization and band gap are composed of two different sets of features, which have 11 and 12 features, respectively (Table S3 and Figs. S9(c) and (d)). Figures S9(c) and (d) illustrate the importance of selected features obtained by GBR algorithm, *d* and MDEDM that play the most vital role in the prediction of polarization and band gap, respectively. As shown in Figs. S9(e) and (f), the weak correlation coefficients for most of features reveal that the redundant and irrelevant features are removed. When 3,119 HBC structures are all chosen as training data set in the supervised machine learning models of polarization and band gap, it can only give two accuracies with the mean absolute error (MAE) of 2.644 pC/m and 0.053 eV for polarization and band gap, respectively (Figs. S9(a) and (b)).

Due to extremely unbalanced distributions of band gaps, when all 3,119 HBC structures are selected as the training data set, the R^2 for the supervised machine learning model of band gap is very low (0.231) (Fig. S9(b)). Therefore, the unbalanced distribution of target property could result in the unreasonable accuracy of the supervised learning model, and it is very difficult to accurately discover materials with desired property.

6. The prediction of 2D ferroelectric HBCs

6.1 All stable ferroelectric materials MX

Table S8. Formula, space group, band gap at HSE level, energy barrier, polarization, the displacement (*d*) and the heat of formation (E_f) for 24 stable 2D ferroelectrics. Bold represents unreported materials.

Formula	Space group	Gap at HSE (eV)	Direct	Barrier (eV/unit cell)	Polarization (eÅ/ unit cell)/(pC/m)	d (Å)	E_f (eV/atom)
BiSb	P3m1	1.345	True	4.999	0.085 (8.80)	1.69	0.109
SbAs	P3m1	1.969	False	4.921	0.033 (4.07)	1.52	0.035
SbP	P3m1	2.281	False	4.747	0.060 (7.98)	1.44	0.153
BiAs	P3m1	1.479	True	4.719	0.126 (14.60)	1.56	0.099
BiP	P3m1	1.931	True	4.547	0.158 (19.44)	1.48	0.259
AsP	P3m1	2.430	False	4.331	0.035 (5.42)	1.32	0.066
WC	P3m1	1.716	False	3.238	0.041 (8.43)	0.94	1.707
AsN	P3m1	2.912	False	3.737	0.036 (7.65)	0.96	0.424
PN	P3m1	2.767	False	3.729	0.050 (12.50)	0.86	0.094
SbN	P3m1	2.476	False	3.246	0.058 (10.09)	1.02	0.496
MoC	P3m1	1.281	False	2.860	0.059 (12.00)	0.93	2.556
NbN	P3m1	0.791	False	1.058	0.037 (6.67)	0.79	0.928
BiGa	P3m1	0.432	True	0.339	0.026 (2.37)	0.79	0.395
SnGe	P3m1	0.566	True	0.322	0.031 (2.97)	0.77	0.625
SbIn	P3m1	1.188	True	0.211	0.096 (8.09)	0.77	0.352
SbGa	P3m1	1.439	True	0.200	0.063 (6.09)	0.70	0.276
SnSi	P3m1	0.491	True	0.196	0.047 (4.70)	0.69	0.770
InAs	P3m1	1.334	True	0.168	0.111 (10.72)	0.67	0.263
GeSi	P3m1	0.567	True	0.146	0.012 (1.40)	0.59	0.659
GaAs	P3m1	1.854	False	0.143	0.078 (8.75)	0.59	0.109
SbAl	P3m1	2.053	False	0.097	0.077 (7.33)	0.63	0.304
BiB	P3m1	1.090	True	0.083	0.056 (6.86)	0.51	1.243

InP	P3m1	1.811	False	0.069	0.116 (11.84)	0.53	0.354
GaP	P3m1	2.538	False	0.042	0.078 (9.35)	0.42	0.111

The E_f of cubic-diamond is greater than zero (0.136 eV/atom), but it has been synthesized experimentally.[33] Similarly, the E_f of T-carbon is far greater than zero (1.318 eV/atom),[34] but it is also synthesized experimentally.[35,36] Although the E_f of these materials is positive, they may still be synthesized experimentally.

6.2 The ferroelectric phase transition barrier

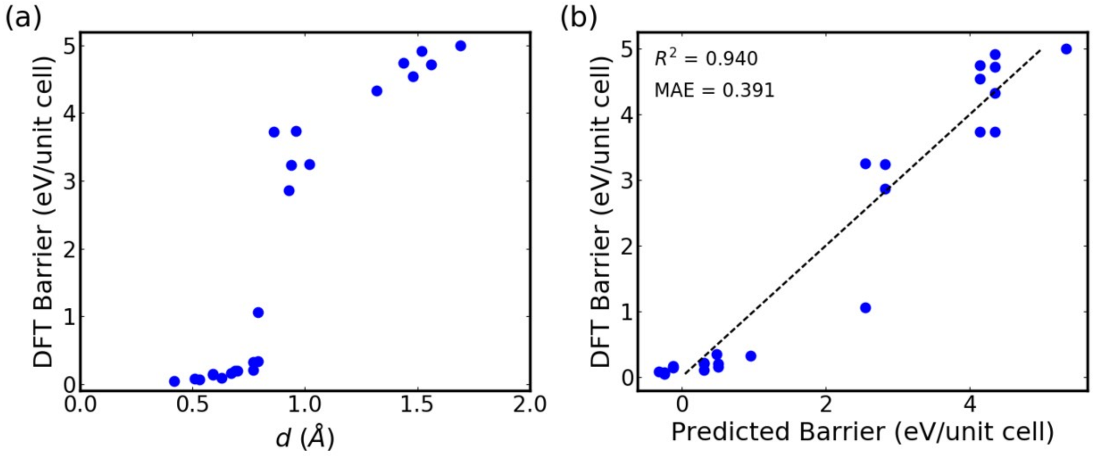

Figure S10. (a) The energy barrier for the ferroelectric-paraelectric phase transition versus the displacement (*d*). (b) Comparison of energy barriers between the predicted values obtained by the model $Barrier = \alpha \times n_X \times n_M \times {r_{M(in)}}^{\frac{1}{3}} + \beta$ and the DFT-calculated values. The dash line presents the ideal fitting line whose

We observed that the energy barriers of 24 stable 2D ferroelectrics generally increase with the increasing vertical displacements (*d*) (see Fig. S10 (a)). The sure independence screening and sparsifying operator (SISSO) algorithm[37] was performed to obtain a feature for the energy barriers (Fig. S10 (b)), which has the form $Barrier = \alpha \times n_X \times n_M \times {r_{M(in)}}^{\frac{1}{3}} + \beta$, where $r_{M(in)}$, n_X and n_M are the radius of inner valence electrons of M, the number of valence electrons of M and X atoms, respectively, and the coefficients α and β are 0.095 $eV \cdot (pm)^{-\frac{1}{3}}$ and -6.815 eV, respectively. Thus the increasing number of valence electrons of constituent atoms could result in the larger energy barrier in these ferroelectrics. The high coefficient of determination value (R^2=0.940) and the low mean absolute error (MAE=0.391) indicate the good fitting. Besides, the material with the largest error is NbN.

7. The energy versus polarization and Monte Carlo simulation

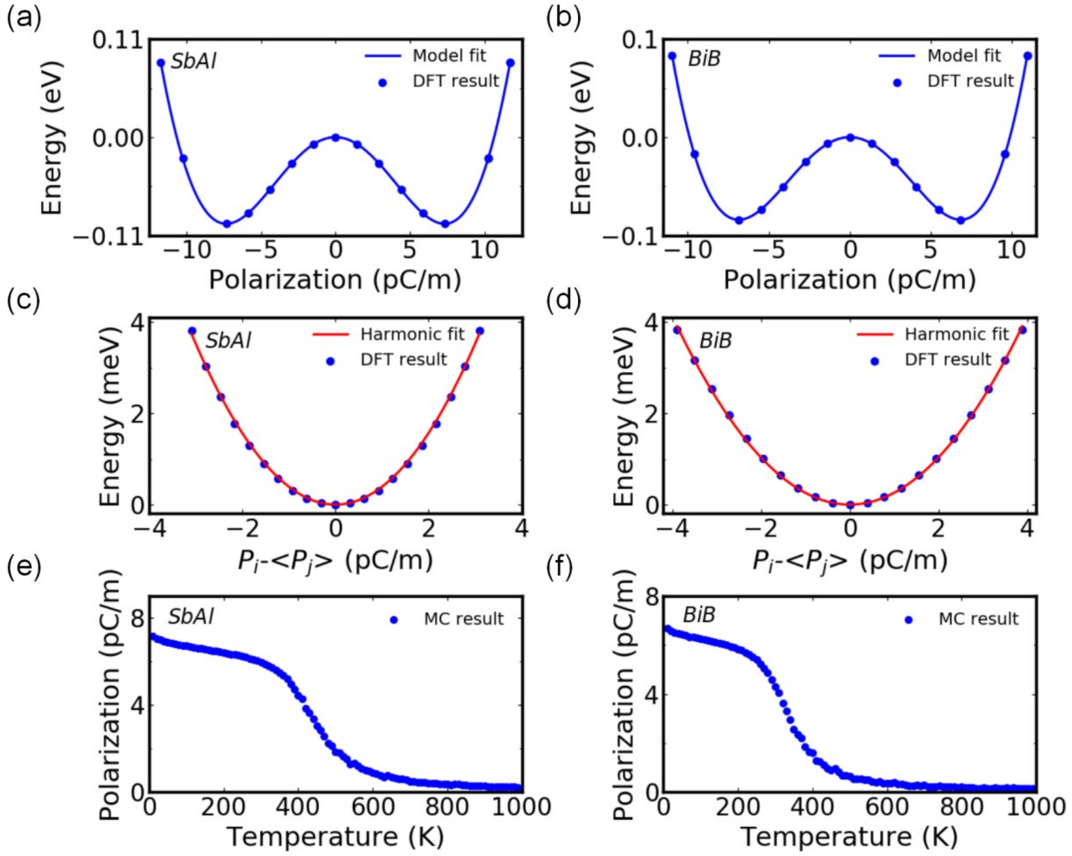

Figure S11. (a)(c)(e) SbAl, (b)(d)(f) BiB. (a)(b) The double-well energy versus polarization. The blue points are obtained by DFT. The blue line is the fitted curve by Landau-Ginzburg model. (c)(d) The dipole-dipole interactions. The blue points are the DFT-calculated total energy for different P_i-<P_j>. The red line is the fitting curve within harmonic approximation. (e)(f) The temperature dependence of polarization obtained by Monte Carlo (MC) simulations.

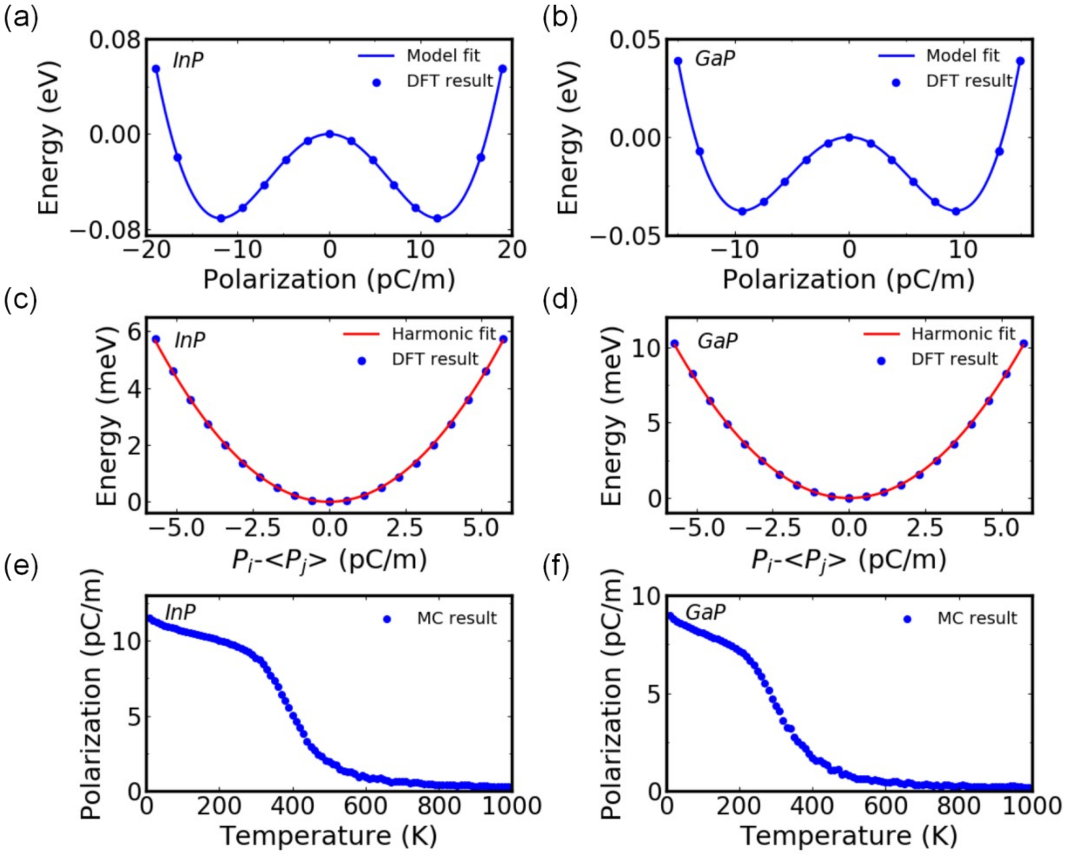

Figure S12. (a)(c)(e) InP, (b)(d)(f) GaP. (a)(b) The energy versus polarization. The blue points are results of DFT. The blue lines are the fitting curves by Landau-Ginzburg model. (c)(d) The dipole-dipole interactions. The blue points are the DFT-calculated total energy for different P_i-<P_j>. The red lines are the fitting curves within harmonic approximation. (e)(f) The temperature dependence of polarization obtained by Monte Carlo (MC) simulations.

Table S9. The energy barrier, the spontaneous polarization, the fitted parameters A, B, C and D in Eq. (S6), the estimated Curie temperature T_c (K).

Formula	Barrier (meV/unit cell)	Polarization (pC/m)	A	B	C	D	T_c (K)
SbAl	97	7.300	-7.587	1.602×10^{-1}	-3.489×10^{-4}	0.784	800
BiB	85	6.861	-7.444	1.753×10^{-1}	-3.527×10^{-4}	0.517	700
InP	71	11.835	-2.121	1.727×10^{-2}	-1.542×10^{-5}	0.349	800
GaP	38	9.353	-1.784	2.240×10^{-2}	-2.220×10^{-5}	0.622	700

The Landau-Ginzburg expansion of the energy in terms of the polarization P_i as the order parameter was widely adopted to investigate phase transitions, [38, 39] in the form of

$$E = \sum_i \left[\frac{A}{2}P_i^2 + \frac{B}{4}P_i^4 + \frac{C}{6}P_i^6\right] + \frac{D}{2}\sum_{<\mathrm{i,j}>}(P_i - P_j)^2 \qquad \text{(S6)}$$

where P_i is the polarization of the i-th unit cell, <i,j> indicates the nearest neighbors, and *A, B, C, D* are fitting parameters. The double-well potential energy versus polarization curves are fitted using the first three terms of Eq. (S6) (Figs. S11(a)(b) and S12(a)(b)). The dipole-dipole interaction between the nearest neighboring unit cells was fitted by the last term of Eq. (S6) (Figs. S11(c)(d) and S12(c)(d)). The above parameters are obtained by fitting the DFT results, as listed in Table S9. With the effective model, the temperature dependence of polarization was calculated to identify the ferroelectric Curie temperatures T_c using Monte Carlo simulations. The results reveal that T_c of SbAl, BiB, InP and GaP can be about 800K, 700K, 800K and 700K, respectively (Figs. S11(e) (f) and S12(e) (f)), indicating that the ferroelectricity in the these 2D materials are very robust against thermal perturbations.

8. 2D ferroelectric photovoltaic materials BiAs and BiSb

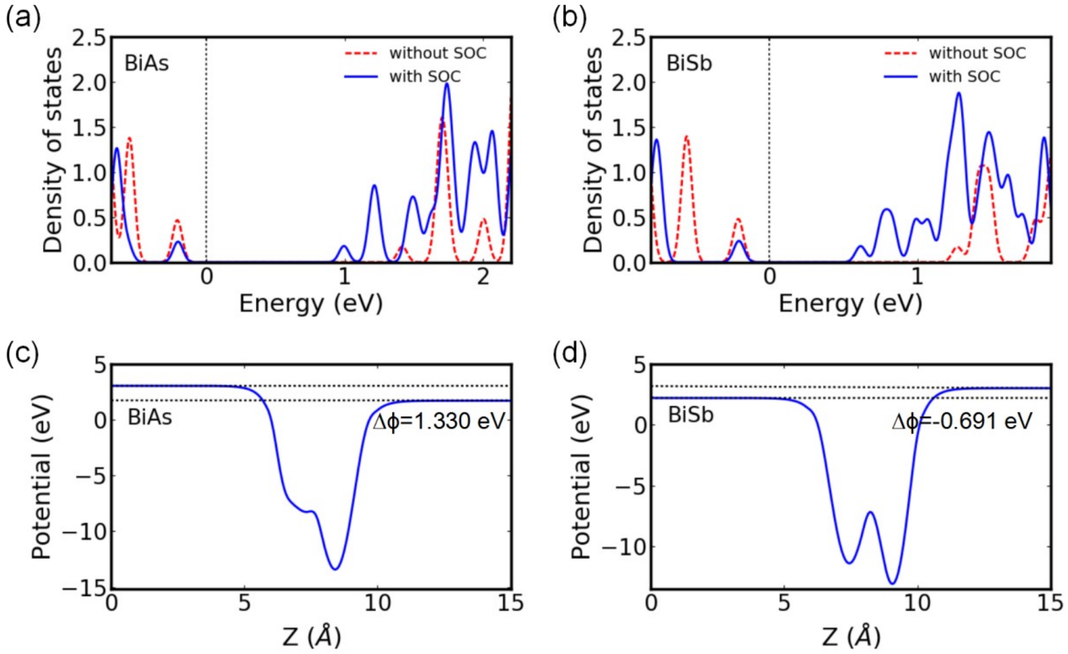

Figure S13. Density of states and the electrostatic potentials averaged in xy-plane along the vertical z-direction of (a)(c) BiAs and (b)(d) BiSb. Blue solid lines and red dash lines of figure (a) and (b) indicate density of states with/without SOC effect, respectively.

Comparing the density of states of the cases with and without SOC effect, we discovered that the SOC effect obviously increase the density of states near conduction bands of BiAs and BiSb (Fig. S13(a)(b)). The built-in fields existing in ferroelectric materials BiAs and BiSb generate two potential differences of 1.330 and 0.691 eV between the top and bottom surfaces (Figs. S13(c) (d)), respectively, which are on the same order of magnitude as that of In_2Se_3 (1.37 eV), [14] suggesting that the strong built-in field could be favorable for decreasing the carrier recombination probability.

Table S10. Calculated elastic modulus (C_{2D}), effective mass (m/m_0), deformation potential constant (E_i), and carrier mobility (μ) in the zigzag and armchair directions for 2D BiAs and BiSb at 300 K.

Structure	Carrier	direction	C_{2D} (N/m)	m/m_0	E_i (eV)	μ ($cm^2V^{-1}s^{-1}$)
BiAs	electron	zigzag	42.77	0.081	-5.086	**5335.68**
		armchair	42.71	0.081	-5.086	**5327.41**
	hole	zigzag	42.77	-0.643	-5.137	20.66
		armchair	42.71	-1.625	-5.097	52.95
BiSb	electron	zigzag	36.09	0.072	-4.745	**6544.82**
		armchair	36.10	0.072	-4.745	**6547.90**
	hole	zigzag	36.09	-0.545	-5.270	28.52
		armchair	36.10	-1.195	-5.180	64.76

The carrier mobility at room temperature was calculated using the deformation potential theory. The results reveal that the electron mobility is much larger than that of the holes regardless of BiAs and BiSb (Table S10), indicating an electron-hole asymmetry, which is very beneficial to the separation of carriers.

9. The stability of the predicted 2D ferroelectric HBCs

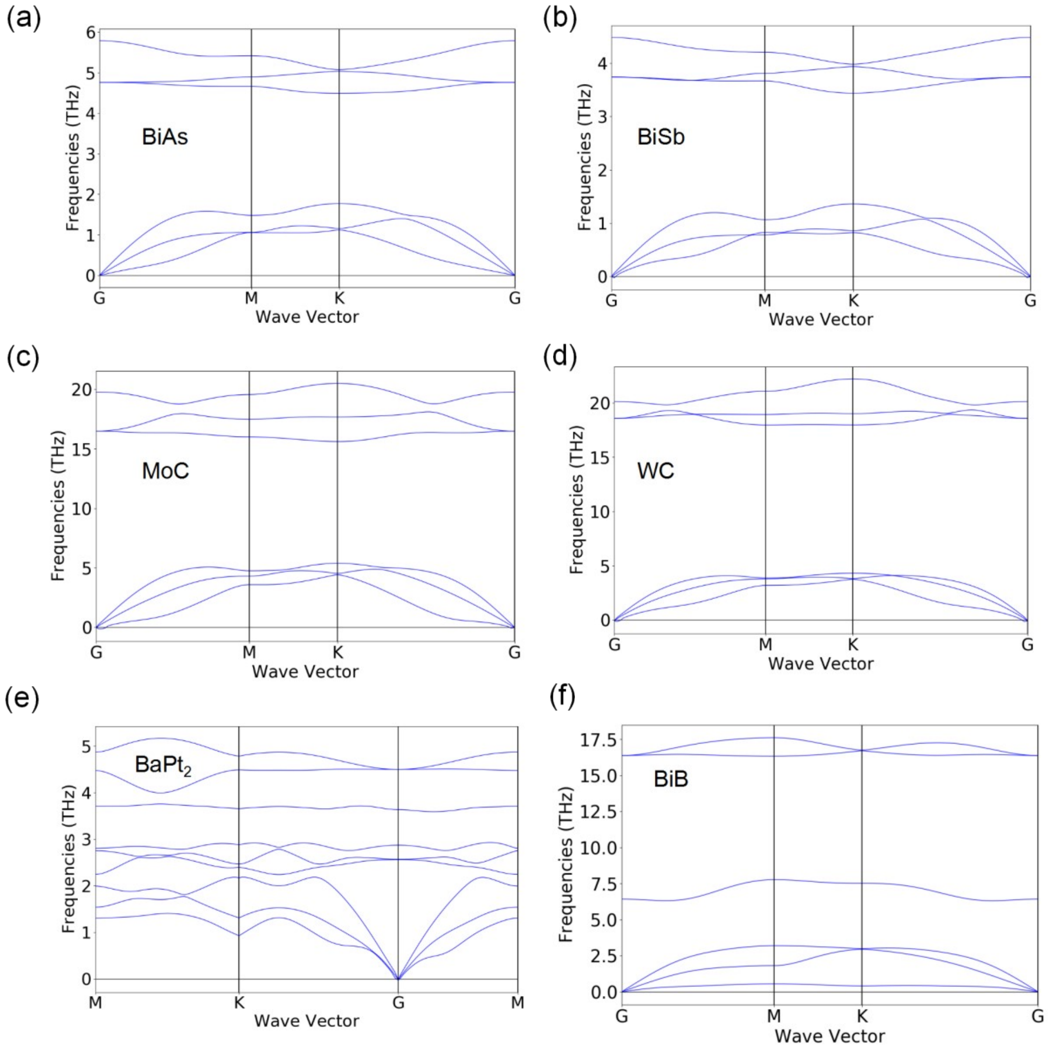

Figure S14. The phonon spectra of (a) BiAs, (b) BiSb, (c) MoC, (d) WC, (e) BaPt$_2$ and (f) BiB.

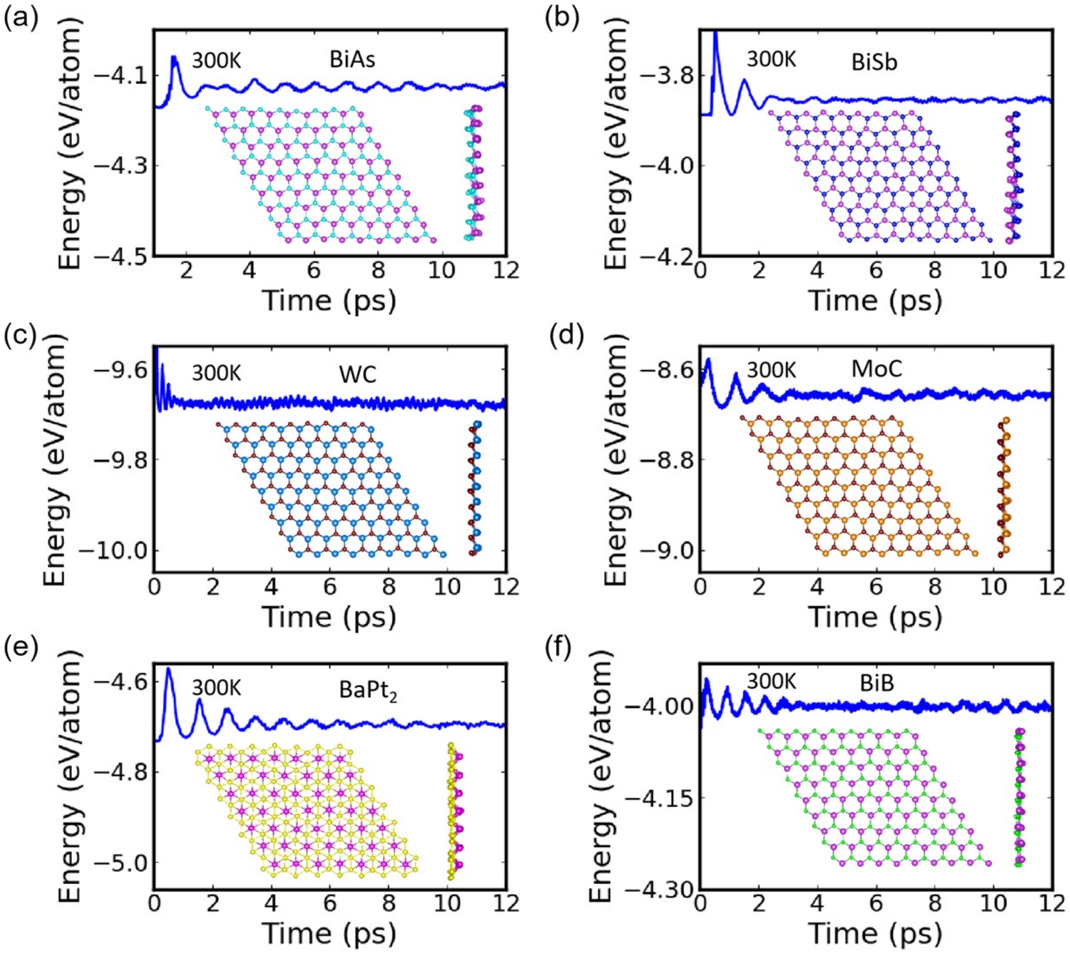

Figure S15. AIMD evolutions of average energy per atom for (a) BiAs, (b) BiSb, (c) WC, (d) MoC (e) $BaPt_2$ and (f) BiB. The insets show the snapshots of these structures at 300 K after 12 ps AIMD simulations.

Figure S14 shows the phonon spectra of six unreported ferroelectric materials, and they are dynamically stable with no observed imaginary modes in phonon spectra.

We performed *ab initio* molecular dynamics simulations at room temperature. The results revealed that the time-dependent variations of energy per atom are oscillating within a very narrow range, showing these structures could maintain their original configurations at room temperature (see in Fig. S15), demonstrating that they possess good thermodynamic stabilities.

10. Supplementary Reference

(1) G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15−50.

(2) G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169−11186.

(3) P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953−17979.

(4) G. Kresse and D. Joubert, *Phys. Rev. B*, 1999, **59**, 1758−1775.

(5) J. P. Perdew and K. Burke, *Phys. Rev. Lett.*, 1996, **77**, 3865−3868.

(6) S. Maekawa, T. Tohyama, S. E. Barnes, S. Ishihara, W. Koshibae and G. Khaliullin, Physics of Transition Metal Oxides, Series in Solid State Sciences *Springer, Berlin Heidelberg*, 2004, *Vol. 144*.

(7) J. Y. You, Z. Zhang, B. Gu and G. Su, *Phys. Rev. Appl.*, 2019, **12**, 024063.

(8) J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207-8215.

(9) G. Kresse, J. Furthmüller, and J. Hafner, *Europhys. Lett.*, 1995, **32**, 729.

(10) K. Parlinski, Z. Q. Li and Y. Kawazoe, *Phys. Rev. Lett.*, 1997, **78**, 4063–4066.

(11) G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901-9904.

(12) S. Nosé, *J. Chem. Phys*., 1984, **81**, 511-519.

(13) W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.

(14) W. Ding, J. Zhu, Z. Wang, Y. Gao, D. Xiao, Y. Gu, Z. Zhang and W. Zhu, *Nat. Commun.*, 2017, **8**, 14956.

(15) Q. Yang, M. Wu and J. Li, *J. Phys. Chem. Lett.*, 2018, **9**, 7160-7164.

(16) Z. Fei, W. Zhao, T. A. Palomaki, B. Sun, M. K. Miller, Z. Zhao, J. Yan, X. Xu and D. H. Cobden, *Nature*, 2018, **560**, 336.

(17) C. Herring and E. Vogt, *Phys. Rev.*, 1956, **101**, 944.

(18) L. Hedin, *Phys. Rev.*, 1965, **139**, 796−823.

(19) L. Hedin, *J. Phys.: Condens. Matter*, 1999, **11**, 489−528.

(20) E. E. Salpeter and H. A. Bethe, *Phys. Rev.*, 1951, **84**, 1232−1242.

(21) G. Strinati, *Phys. Rev. B*, 1984, **29**, 5718−5726.

(22) M. Rohlfing and S. G. Louie, *Phys. Rev. B*, 2000, **62**, 4927−4944.

(23) S. Saha, T. P. Sinha and A. Mookerjee, *Phys. Rev. B*, 2000, **62**, 8828–8834.

(24) L. Mentel, Mendeleev – a python resource for properties of chemical elements, ions and isotopes, 0.4.1 2014.

(25) J. B. Mann, Atomic Structure Calculations II. Hartree-Fock wave functions and radial expectation values: hydrogen to lawrencium, LA-3691, Los Alamos Scientific Laboratory, USA, 1968.

(26) C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*; Massachusetts Institute of Technology Press, 2006.

(27) F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, 12, 2825 – 2830.

(28) J. H. Friedman, *Ann. Stat.*, 2001, **29**, 1189–1232.

(29) L. Breiman, *Mach. Learn.*, 1996, **24**, 123-140.

(30) Y. Freung and R. Shapire, *J. Comput. Syst. Sci*, 1997, **55**, 119-139.

(31) A. Liaw and W. Matthew, *R news*, 2002, **2**, 18-22.

(32) T. Hastie, R. Tibshirani and J. Friedman, *Elements of Statistical Learning*; Springer Press, 2009.

(33) R. A. Andrievski, *Int. J. Refract. Met. Hard Mater.*, 2001, **19**, 447.

(34) X. L. Sheng, Q. B. Yan, F. Ye, Q. R. Zheng and G. Su, *Phys. Rev. Lett.*, 2011, **106**, 155703.

(35) J. Zhang, R. Wang, X. Zhu, A. Pan, C. Han, X. Li, D. Zhao, C. Ma, W. Wang, H. Su and C. Niu, *Nat. Commun.*, 2017, **8**, 683.

(36) K. Xu, H. Liu, Y. C. Shi, J. Y. You, X. Y. Ma, H. J. Cui, Q. B. Yan, G. C. Chen and G. Su, *Carbon*, 2020, **157**, 270-276.

(37) R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler and L. M. Ghiringhelli, *Phys. Rev. Mater.*, 2018, **2**, 083802.

(38) R. Fei, W. Kang and L. Yang, *Phys. Rev. Lett.*, 2016, **117**, 097601.

(39) S. Guan, C. Liu, Y. Lu, Y. Yao and S. A. Yang, *Phys. Rev. B*, 2018, **97**, 144104.