Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.

Supplementary Materials

Machine learning-assisted design of AlN-based high-performance piezoelectric materials

Huirong Jing,[a] Chaohong Guan,[a] Yu Yang,[a] Hong Zhu [a,*]

[a] University of Michigan− Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China

[*] Authors to whom correspondence should be addressed: hong.zhu@sjtu.edu.cn

Feature construction and selection

Table S1. Original features in the features pool.

Types	Features abbreviation	Descriptions
Electronic features	*FED*	Average formation energy of decomposition products (e.g. AlN, ScN, Be_3N_2…)
	BGD	Average band gap of decomposition products
	FIE	Average first ionization energy
	Nd	Average number of *d* orbital electron
	TF	Tetrahedral factor, average value of cation radius divided by anions radius
	V	Average DFT volume per atom of cations (T=0K ground state)
	CR	Average covalent radius of each cations
	P	Average polarizabilities of cations
	EN	Average electronegativity difference between cations and nitrogen
	AR	Average atomic radius of each cations
	WF	Average work function of cations
Structural features	$N_{i/o}^{jth}(A-A)$, $N_{i/o}^{jth}(B-B)$, $N_{i/o}^{jth}(AB-AB)$	The number of *j*th (*jth*=1*st*, 2*nd*, 3*rd*, 4*th*) in-plane/out-of-plane *A*-*A*, *B*-*B*, *A*-*B* bonds
	A_{xy}, A_z, B_{xy}, B_z	Uniformity of the distribution of doped cations along *xy* and*d z* direction (A, B represent different cation element)

Table S1 shows the original 11 features extracted from the properties of the elements or decomposition products of $A_xB_yAl_{1-x-y}N$ and 19 features come from crystal structure. We describe the average as a weighted fraction of the elements or decomposition products, and define the materials features as following,

$$X = \sum_i f_i x_i \tag{1}$$

where f_i is the mole fraction and x_i correspond the property of an element or decomposition products. The decomposition products are obtained according to phase diagram. For instance, the decomposition products of $Al_xSc_{1-x}N$ are AlN with x and ScN with 1-x. The *FED* (*BGD*) of $Al_xSc_{1-x}N$ is calculated by eq. (1), where x_i is the formation energy (band gap) of AlN and ScN, respectively, f_i is x for AlN and 1-x for ScN.

The tetrahedral factor (*TF*) is define by

$$TF = \frac{R_{cation}}{R_N}, \tag{2}$$

where R_{cation} is the mole average of the Shannon's (1976) ionic radii of cations in $A_xB_yAl_{1-x-y}N$ at four coordination and R_N is the radius of nitrogen at 4 coordination.

The electronegativity difference (*EN*) is mainly used to compare the difference between cations and anions, so we define it by

$$EN = 3.04 - EN_{cations}, \tag{3}$$

3.04 is the electronegativity of nitrogen and $EN_{cations}$ is the electronegativity of cations, then using the equation (1) to make a further calculation.

We plot the Pearson correlation matrix to group features highly correlated with each other (|p| > 0.9). Within each group, we choose a descriptor that is physically related to the target property and representative of other features in the group. For example, tetrahedral factor, volume, covalent radius, polarizability, electronegativity difference, atomic radius and work function are highly correlated. we select the tetrahedral factor (*TF*) as it reflects the size and change of polarization simultaneously.

Distinct cation bonds are distinguished by their relative positions. In the 2 × 2 × 2 supercell, five cation pairs are ordered according to the distance between two atoms. The type of different bond is plotted in Fig. S1. Fig. S1(a) show the *jth* (*jth=1st, 2nd, 3rd, 4th*) in-plane/out-of-plane bond with the same type elements (*A-A* or *B-B*), Fig. S1(b) show the *jth* (*jth=1st, 2nd, 3rd, 4th*) in-plane/out-of-plane bond with the different type elements (*AB*). The definition of A_{xy} (B_{xy}) and A_z (B_z) is as following,

$$\bar{a} = \left(\frac{a1}{n}, \frac{a2}{n}, \frac{a3}{n}, \frac{a4}{n}\right)/4, \tag{4}$$

$$\bar{b} = \left(\frac{b1}{n}, \frac{b2}{n}, \frac{b3}{n}, \frac{b4}{n}\right)/4, \tag{5}$$

$$\bar{c} = (\frac{c1}{n}, \frac{c2}{n}, \frac{c3}{n}, \frac{c4}{n})/4 , \quad (6)$$

$$A_x = \frac{\sum_{i=1}^{4} |ai - \bar{a}|}{4} , \quad (7)$$

$$A_y = \frac{\sum_{i=1}^{4} |bi - \bar{b}|}{4} , \quad (8)$$

$$A_z = \frac{\sum_{i=1}^{4} |ci - \bar{c}|}{4} , \quad (9)$$

$$A_{xy} = \frac{1}{2}(A_x + A_y) , \quad (10)$$

where $a1 \sim a4$, $b1 \sim b4$, $c1 \sim c4$ are the number of $A(B)$ atoms per cation layer along a-, b- and c-directions, respectively, n is the total number of $A(B)$ atoms in the $2 \times 2 \times 2$ supercell.

Structural features of the AlN-based supercell

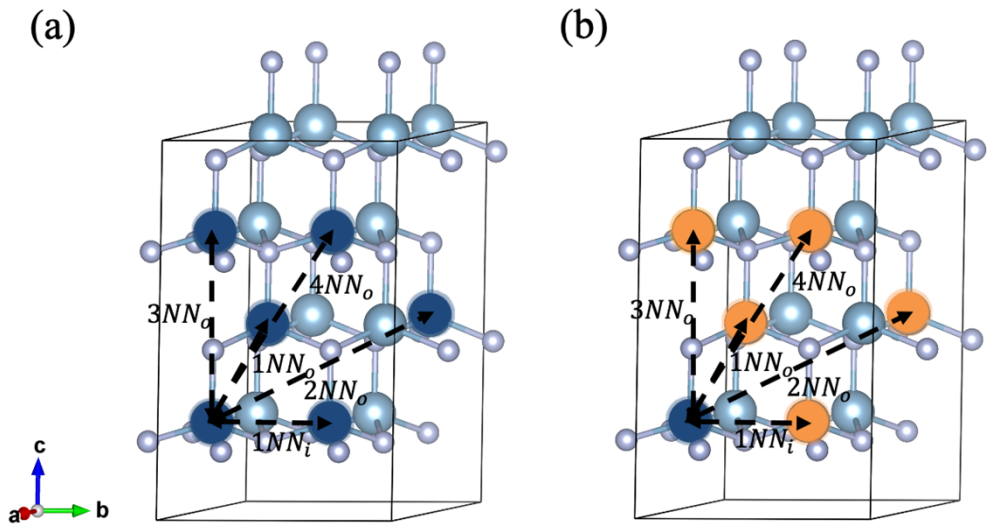

Fig. S1 Structural features. Schematic diagrams of the *jth* (*jth=1st, 2nd, 3rd, 4th*) in-plane/out-of-plane bond with the same type elements (*A-A* or *B-B*), (b) with the different type elements (*A-B*).

Pearson correlation matrix among 30 original features

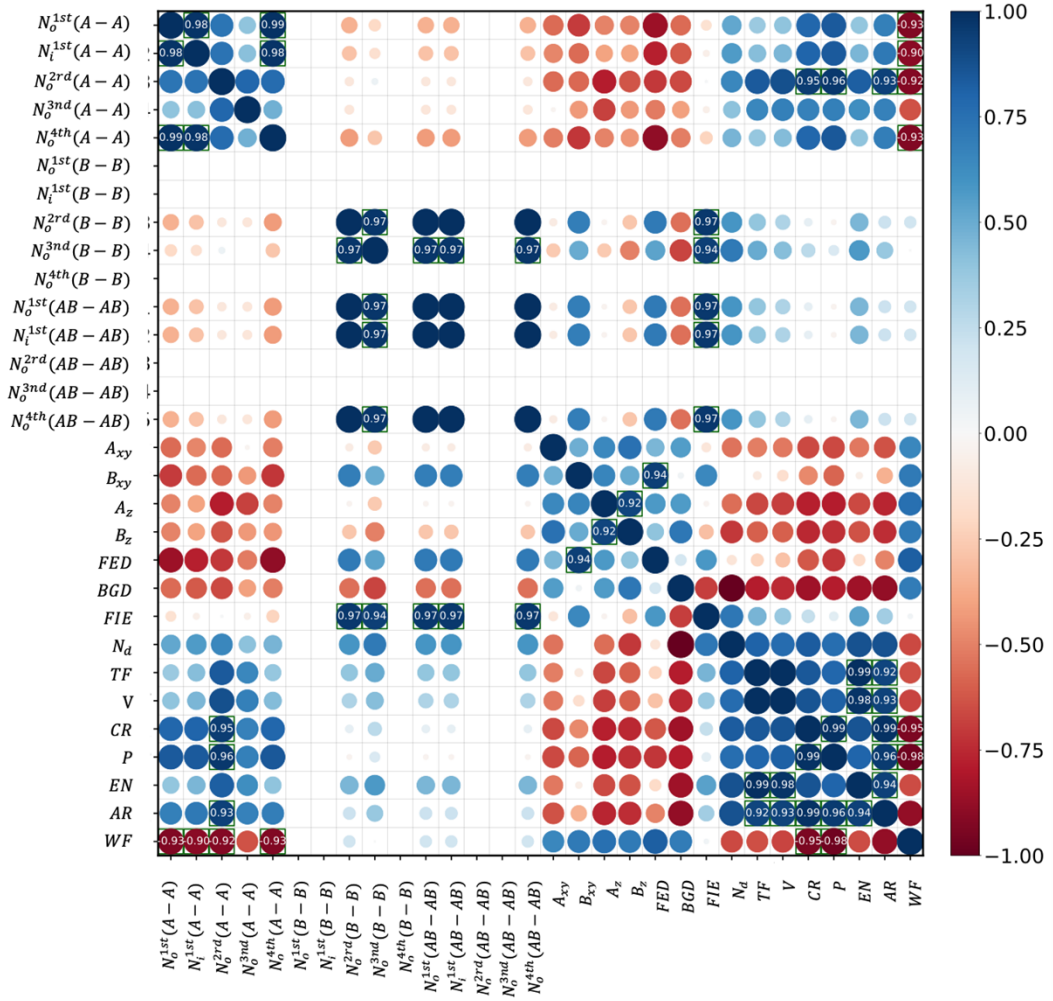

Fig. S2 Pearson correlation coefficient (PCC) correlation map of 30 features. The blue and red colors represent positive and negative correlations, respectively. Darker color and bigger circles indicate stronger correlation. Two features highly correlated with each other (|p| > 0.9) are shown in green squares.

Application of GBRT model in larger supercell

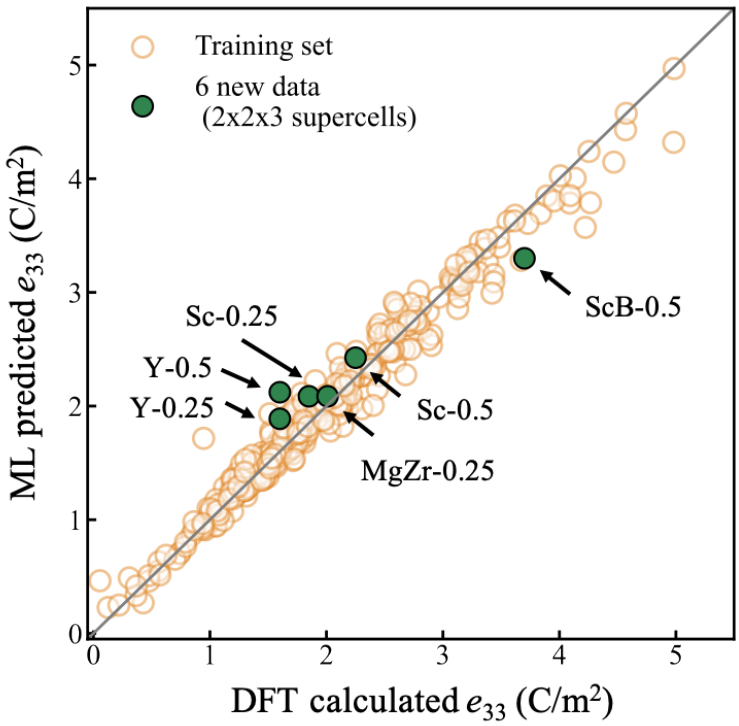

Fig. S3 Parity plots comparing DFT-computed e_{33} against ML-predicted e_{33} of 6 new structures with 2 × 2 × 3 supercell not included in the training dataset.