Supporting information for

Inverse Design for Materials Discovery from the Multidimensional Electronic Density of States

Kihoon Bang,[1,†] Jeongrae Kim, [1,2,†] Doosun Hong, [1,†] Donghun Kim, [1,*] and Sang Soo Han[1,*]

[1]Computational Science Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea

[2]Department of Artificial Intelligence Software Convergence, Korea Polytechnics Chuncheon Campus, Gangwon-do, 24409, Republic of Korea

[†]These authors contributed equally.

*Corresponding authors: sangsoo@kist.re.kr (SSH); donghun@kist.re.kr (DK)

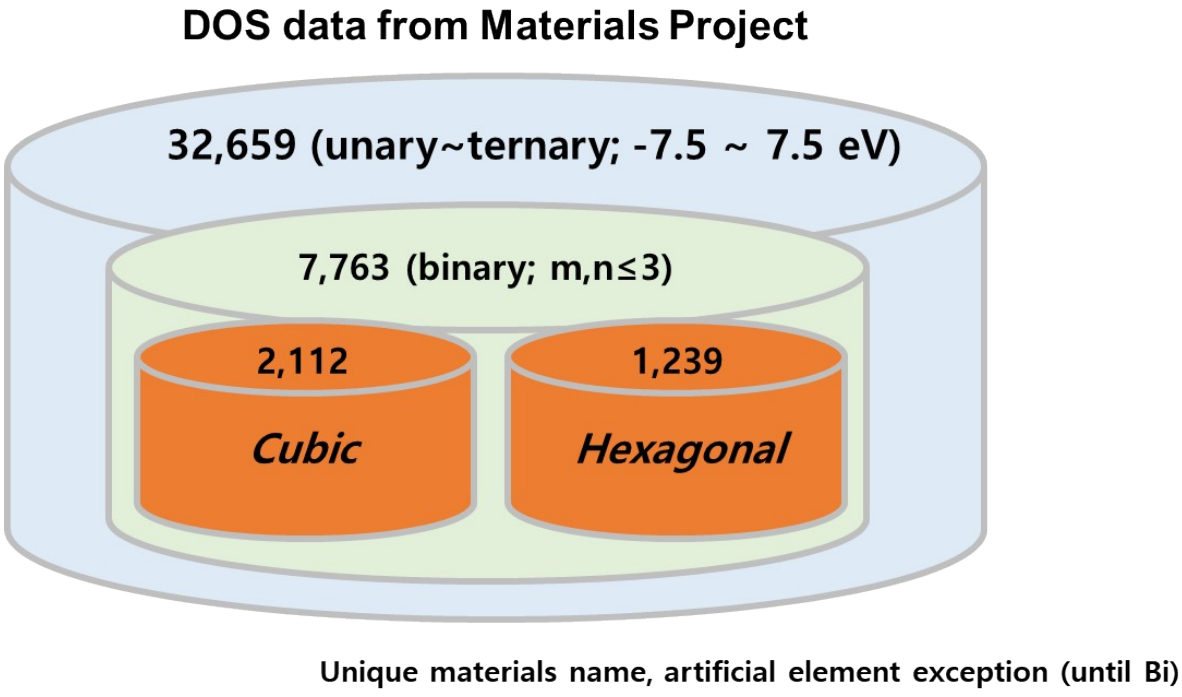


Figure S1 DOS database used in this study, obtained from the Materials Project library (next-gen.materialsproject.org). A total of 32,659 DOS patterns from unary to ternary compositions were collected and preprocessed to energy levels from -7.5 to 7.5 eV. There are 7,763 DOS patterns with the binary composition of A_mB_n, where $m,n \leq 3$. Among them, 2,112 compositions have a cubic Bravais lattice, and 1,239 compositions have a hexagonal lattice.

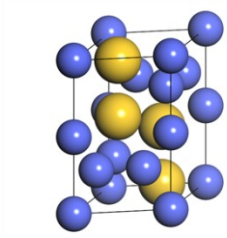
Validation of the composition vector (CV) performance by bandgap energy classification

We evaluated the CV performance by training an artificial neural network (ANN) model. This model was employed to classify the bandgap of input compositions, allowing us to compare the accuracy between the DOS-based CVs in our work and the CVs created with the element embedding with normalized composition matrix (EENCM) method. The bandgap was classified into three categories: small gap ($E_g < 0.2$ eV), medium gap (0.2 eV $< E_g \leq 3.6$ eV), and large gap ($E_g > 3.6$ eV). As a result, the ANN model had three output nodes. The input was either the DOS-based CV or the EENCM CV, which represented the composition of the material with a size of 60. The ANN model was composed of 3 hidden layers with 100, 50, and 25 nodes. To gather data for training and testing, we collected the material composition and bandgap data from the Materials Project database. We sampled 1,998 examples with a cubic structure and a binary composition. Among these samples, 90% (1,799) were utilized as the training set, while 10% (199) were designated as the test set. The results obtained from the test set are presented in Table S1.

Table S1 Accuracy and F1-score of the bandgap classification models with different CVs

Type of CV	Accuracy	F1-score
DOS-based CV (our work)	0.92 ± 0.01	0.54 ± 0.14
EENCM CV	0.93 ± 0.01	0.58 ± 0.12

a



b

Rank	Material	Cosine similarity	Formation energy (eV/atom)
1	$AuCo_2$	**0.91**	0.03
2	$PtMn_2$	**0.89**	-0.09
3	$PtFe_2$	**0.87**	-0.06

c

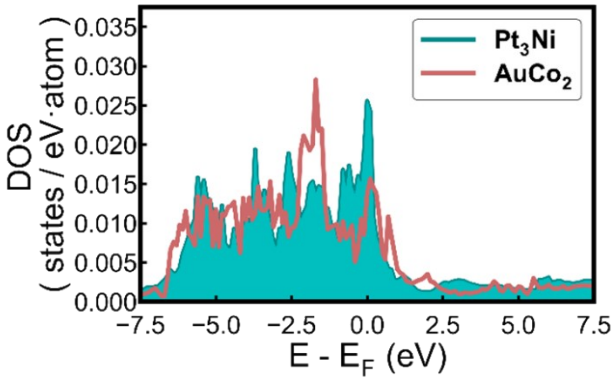
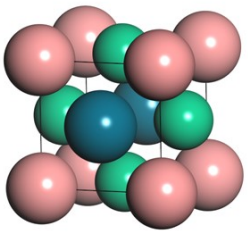


Figure S2. Candidate materials with a hexagonal structure ($P6_3/mmc$) exhibiting the DOS pattern comparable to that of Pt_3Ni. a Atomic structure of the candidate materials. **b** A list of the candidate materials, including their DOS similarity values and DFT formation energies. **c** The calculated DOSs of the candidate material (AuCo2) and target material (Pt_3Ni).

Figure S3. Candidate materials with a tetragonal structure (P4/mmm) exhibiting the DOS pattern comparable to that of Pt3Ni. a Atomic structure of the candidate materials. b A list of the candidate materials, including their DOS similarity values and DFT formation energies. c The calculated DOSs of the candidate material ($CoRh_2Pd$) and the target material (Pt_3Ni).

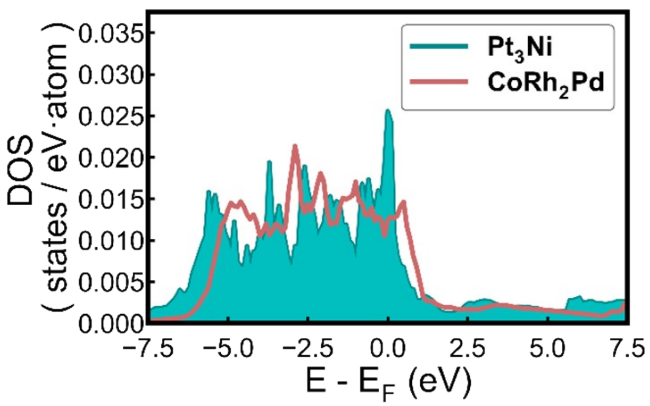
a



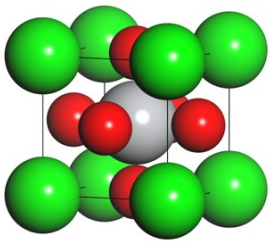
b

Rank	Material	Cosine similarity	Formation energy (eV/atom)
1	$CoRh_2Pd$	**0.90**	0.04
2	$ZrRh_2Ir$	**0.80**	-0.71
3	$ScRh_2Ir$	**0.80**	-0.60

c



a



b

Rank	Material	Cosine similarity	Formation energy (eV/atom)
1	$SrTiO_3$	**0.85**	-3.18
2	$BaMnO_3$	**0.82**	-2.15
3	$BaZrO_3$	**0.77**	-3.24

c

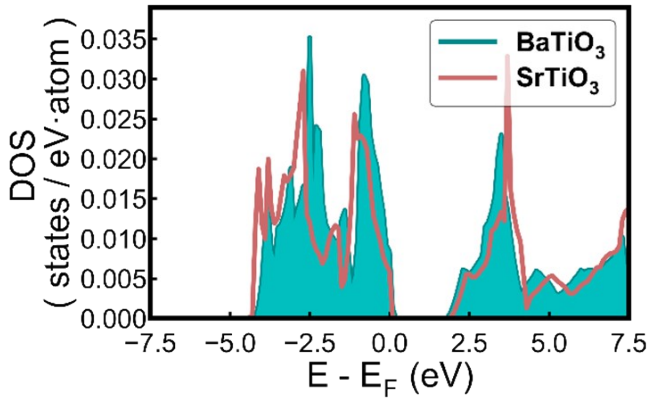


Figure S4. Candidate materials of the ternary oxide system (space group: Pm-3m) exhibiting the DOS pattern comparable to that of $BaTiO_3$. a Atomic structure of the candidate materials. **b** A list of the candidate materials, including their DOS similarity values and DFT formation energies. **c** The calculated DOSs of the candidate material ($SrTiO_3$) and the target material ($BaTiO_3$).