

Combination of recommender system and single-particle diagnosis for accelerated discovery of novel nitrides

Yukinori Koyama,^{1,a)} Atsuto Seko,² Isao Tanaka,^{2, 3} Shiro Funahashi⁴ and Naoto Hirosaki⁴

¹*Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan*

²*Department of Materials Science and Engineering, Kyoto University, Kyoto, Kyoto 606-8501, Japan*

³*Nanostructures Research Laboratory, Japan Fine Ceramics Center, Nagoya, Aichi 456-8587, Japan*

⁴*Research Center for Functional Materials, National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan*

^{a)}Corresponding author: KOYAMA.Yukinori@nims.go.jp

ABSTRACT

Discovery of new compounds from wide chemical space is attractive for materials researchers. However, theoretical prediction and validation experiments have not been systematically integrated. Here, we demonstrate that a new combined approach is powerful to accelerate the discovery rate of new compounds significantly, which should be useful for exploration of wide chemical space in general. A recommender system for chemically relevant composition is constructed by machine learning of Inorganic Crystal Structure Database (ICSD) using chemical compositional descriptors. Synthesis and identification experiments are made at the chemical compositions with high recommendation scores by the single-particle diagnosis method. Two new compounds,

$\text{La}_4\text{Si}_3\text{AlN}_9$ and $\text{La}_{26}\text{Si}_{41}\text{N}_{80}\text{O}$, and two new variants (isomorphic substitutions) of known compounds, $\text{La}_7\text{Si}_6\text{N}_{15}$ and $\text{La}_4\text{Si}_5\text{N}_{10}\text{O}$, are successfully discovered. Finally, density functional theory calculations are conducted for $\text{La}_4\text{Si}_3\text{AlN}_9$ to confirm the energetic and dynamical stability and to reveal its atomic arrangement.

I. INTRODUCTION

Discovery of new high-performance materials often triggers new technology, which has been one of the central topics of materials research. Although real materials usually consist of multiple phases and synergistic effects with point defects, dopants, surface, grain boundary, etc., are often crucial to key properties, we focus on bulk of single-phase compounds in this study as the first step. There are at least three types of “new compounds”. First type of new compounds are variants of known compounds in a specific field. Isomorphic substitution of elements is a common strategy for this type. Second type of new compounds are already known compounds in another field, which have not been investigated in the field of concern. A typical strategy for this type is a high-throughput screening of compounds in databases, such as Inorganic Crystal Structure Database (ICSD),¹ International Centre for Diffraction Data Powder Diffraction File (ICDD-PDF),² SpringerMaterials,³ Pearson’s Crystal Data,⁴ Crystallography Open Data (COD),⁵ and AtomWork-Adv.⁶ The third type is as-yet-unknown compounds having new crystal structures and/or new chemical compositions. Improvement of performance may be achieved by searching the first type of new compounds, whereas breakthroughs are expected by discovery of the second and third types of new compounds. Especially, the discovery of the third type of new compounds from wide chemical space is quite attractive for materials researchers. However, no systematic approach has thus far been established.

Density functional theory (DFT) calculation is one of the most powerful approaches to predict energetics and physical properties. Some databases of computed information are available, such as the Materials Project,⁷ AFLOW,⁸ OQMD,⁹ and NOMAD.¹⁰ However, it should be emphasized that the majority of these databases consist of known compounds. DFT calculations can be routinely made for new compounds of the first and second types. For the third type of new compounds, chemical compositions and crystal structures should be determined before DFT calculations, and this requires great challenges because of two reasons. Firstly, chemically relevant compositions are unknown *a priori*. Considering vast chemical space, a huge number of compounds are uninvestigated. However, stable compounds are very sparse in the chemical space. According to Villars and Iwata, average number of compounds per chemical system is only 3.2 for ternary and 2 for quaternary systems.¹¹ Searching new compounds is like looking for a needle in a haystack. The second challenge is to search the stable crystal structures for given chemical compositions. Although a few methods have been used for searching stable structures,¹²⁻¹⁶ it is still a long time-consuming task.

Machine learning approaches have been proposed to predict probable substitution elements for known compounds,^{17,18} formation energies of arbitrary compositions,¹⁹ and chemically relevant compositions.^{20,21} These approaches enable systematic prediction of new compounds. Suzuki et al.²² succeeded in experimental synthesis of new lithium-ion conducting oxides based on machine learning prediction.²¹ They adopted conventional synthesis method and powder x-ray diffraction (XRD) analysis to identify the new compounds, and successfully found two new compounds. As the composition of one of the two new compounds was deviated from the exact value by the machine learning prediction, a systematic series of synthesis experiments were needed. In the conventional

approach, the synthesized sample needs to be mostly composed of a single phase to identify composition and crystal structure of the new compound.

In this paper, we adopt single-particle diagnosis approach²³ for efficient identification of new crystal structures. In the single-particle diagnosis approach, starting powders are mixed in compositions of machine learning prediction and fired under predetermined conditions. Then, well grown particles of about 10 μm in size are hand-picked in an optical microscope from the fired samples. As well grown particles are usually single crystals, it is not necessary to have the whole sample be a single phase at this stage. The picked particles are then characterized one by one using scanning electron microscopy energy dispersive x-ray spectroscopy (SEM-EDX) to evaluate ratios of cations, followed by a single-crystal XRD analysis to reveal the crystal structures. The ability to identify phases and crystal structures of single particles even from mixed-phase samples is a great strength of the single-particle diagnosis approach. Hirosaki et al. successfully discovered new phosphor materials using this approach.²³ We adopt this approach to LaN-Si₃N₄-AlN pseudo-ternary system to search new nitrides.

II. Methods

A. Machine learning

A machine learning model of a compositional descriptor-based recommender system was built to evaluate chemical relevance of compositions.²¹ In this study, target compounds were restricted to ionic ones with normal oxidation states. Training dataset consisted of compositions registered in ICSD as positive cases and compositions not registered in ICSD as negative cases. Compounds having partial occupancy, unusual oxidation states, and more than 15 atoms in chemical formula for any constituent element were excluded from the training dataset. The number of positive cases was 33,367. The negative cases

were hypothetical compositions composed of 1.3 million pseudo-binary compounds and 2.6 million pseudo-ternary oxides, nitrides, and sulfides. A set of descriptors composed of means, standard deviations, and covariances of 22 types of elemental representations were used to describe the compositions. Random forest classifier was used. The size of the ensemble of the random forest classifier was increased to get convergence of prediction on a validation dataset, which consisted of pseudo-binary and pseudo-ternary oxides registered in ICDD but not registered in ICSD. Finally, the size of the ensemble was 10,000. The expectant probabilities as the positive cases were used as recommendation scores of the compositions. Other details can be found in Ref. 21.

B. Experiments

Samples in the LaN-Si₃N₄-AlN pseudo-ternary system were prepared by firing mixture of appropriate amounts of LaN (Kojundo Chemical Laboratory, 3N), Si₃N₄ (Ube Industries, SN-E10) and AlN (Tokuyama, Type-E) with additional trace amount of EuN (MATERION, -60 mesh typically 99.9% pure) at 1900 °C for 2 hours in 1.0 MPa nitrogen atmosphere using a gas-pressure sintering furnace (Fujidempa High Multi 5000). EuN was added to distinguish particles of different phases by color,²³ and the amount of Eu was fixed at 0.1 at% of the compositions by the machine learning prediction. Nominal contents of oxygen impurity in starting powders were <2.0 wt% and 0.8 wt% for Si₃N₄ and AlN, respectively.

The products were characterized by powder XRD using SmartLab (Rigaku) with Cu-K α ₁ radiation operated at 45 kV and 200 mA. The diffraction patterns were collected at room temperature in a 2θ range of 5-95° with a step width of 0.05°. The obtained XRD patterns were analyzed using PDXL software (Rigaku) to match phases against known compounds and to refine lattice constants. The XRD patterns were first matched against

known compounds in the La-Si-Al-N-O system. Oxygen is a probable impurity in real experiments introduced from the starting powders or during the synthesis process. If notable amount of unidentified diffraction peaks remained, the patterns were then matched against all compounds in the pattern library to find variants (isomorphic substitutions) of known compounds.

Single-crystal XRD data of single particles were collected on a SMART APEX II Ultra diffractometer (Bruker) with Mo-K α radiation and multilayer mirrors as a monochromator, operated at 50 kV and 50 mA. Applied absorption corrections were done using the multi-scan procedure SADABS. The structures were solved by direct methods implemented in SHELX. Refinement of crystal structures was conducted with anisotropic displacement parameters for all atoms by full-matrix least-squares calculation. The elemental analysis was carried out using a SEM (Hitachi High-Technologies SU1510) equipped with an energy dispersive spectroscopy (Bruker XFlash SDD) operated at 10 kV.

C. DFT calculations

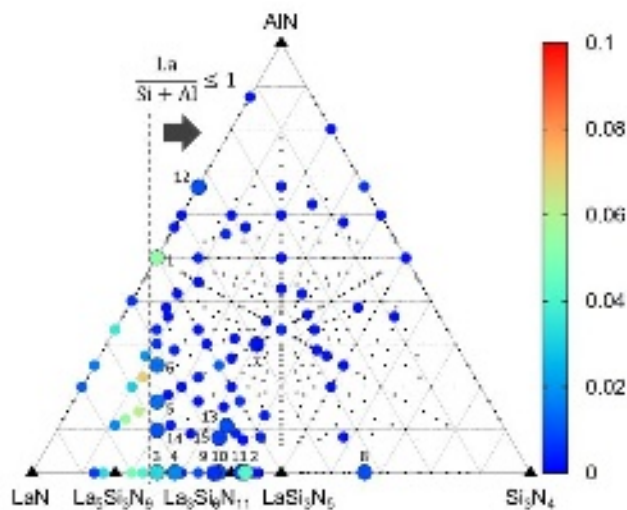
DFT calculations were conducted to investigate discovered compounds in detail using the plane-wave basis projector augmented wave (PAW) method as implemented in the Vienna ab-initio simulation package (VASP) 6.1.^{24,25} The Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional²⁶ was used. Cut-off energy was set at 520 eV. Number of k -points was determined so that the k -point density in the reciprocal space was more than 100 \AA^3 . Total energy converged in 0.1 meV/atom. Atomic positions and lattice constants were optimized until total energy converged in 1 meV/atom. Dynamical stability of crystals was examined by harmonic phonon calculations using phonopy code²⁷ with total energy convergence in 10^{-10} eV/atom and force convergence in 0.01 eV/ \AA .

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50049981

III. RESULTS and DISCUSSION

A. Chemical relevance of compositions

Figure 1 illustrates recommendation scores of compositions in the LaN-Si₃N₄-AlN pseudo-ternary system. In this pseudo-ternary system, LaN, Si₃N₄, AlN, LaSi₃N₅, La₃Si₆N₁₁, La₅Si₃N₉, and La₁₇Si₉Al₄N₃₃ are registered in ICSD. La₁₇Si₉Al₄N₃₃ is registered but not used to build the machine learning model because of its complex composition. The machine learning model suggested various compositions having relatively high recommendation scores, especially for La-rich compositions. However, to our experience, La-rich nitrides are often unstable in the ambient atmosphere, and it is difficult to characterize samples by the single-particle diagnosis. Therefore, we set our experimental target to compositions in which La atoms are fewer than or equal to sum of Si and Al atoms, $La/(Si+Al) \leq 1$. Top 15 recommended compositions are summarized in Table 1. They are seven LaN-Si₃N₄ pseudo-binary, two LaN-AlN pseudo-binary, and six LaN-Si₃N₄-AlN pseudo-ternary compositions.



This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50049981

FIG. 1. Chemically relevant compositions in the LaN-Si₃N₄-AlN pseudo-ternary system. Closed circles correspond to compositions with non-zero recommendation scores as indicated by colors. Large circles indicate top 15 recommended compositions with notations of their ranks (see also Table 1). Small dots are compositions whose scores are zero. Black triangles are compositions that are registered in ICSD and used to build the machine learning model.

TABLE 1. Top-15 recommended compositions, corresponding recommendation scores, and identified candidate phases by powder XRD analysis.

Rank and Sample Number	Composition	Recommendation Score	Identified Candidate Phases
1	La1 Al1 N2	0.054	LaN, AlN
2	La4 Si9 N16	0.042	LaSi ₃ N ₅ , La ₃ BaSi ₅ N ₉ O ₂ , La ₃ Si ₆ N ₁₁
3	La3 Si3 N7	0.032	La ₁₁ Si ₁₃ N _{27.64} O _{1.04} , Ce ₇ Si ₆ N ₁₅
4	La5 Si6 N13	0.018	La ₁₁ Si ₁₃ N _{27.64} O _{1.04}
5	La4 Si3 Al1 N9	0.014	No identified phase
6	La5 Si3 Al2 N11	0.014	No identified phase
7	La2 Si3 Al1 N7	0.012	La ₃ BaSi ₅ N ₉ O ₂ , La ₃ Si ₆ N ₁₁ , La _{26-x} Sr _x Si ₄₁ O _{x+1} N _{80-x}
8	La1 Si6 N9	0.010	LaSi ₃ N ₅ , Si ₃ N ₄
9	La2 Si3 N6	0.010	La ₃ Si ₆ N ₁₁ , La ₁₁ Si ₁₃ N _{27.64} O _{1.04} , LaSiNO ₂
10	La5 Si9 N17	0.010	La ₃ Si ₆ N ₁₁ , La ₃ BaSi ₅ N ₉ O ₂
11	La7 Si15 N27	0.010	LaSi ₃ N ₅ , La ₃ Si ₆ N ₁₁ , LaN
12	La1 Al2 N3	0.010	AlN, LaN
13	La5 Si9 Al1 N18	0.010	La ₃ Si ₆ N ₁₁ , La ₃ BaSi ₅ N ₉ O ₂
14	La7 Si6 Al1 N16	0.010	La ₁₁ Si ₁₃ N _{27.64} O _{1.04} , Ce ₇ Si ₆ N ₁₅
15	La7 Si12 Al1 N24	0.008	La ₃ Si ₆ N ₁₁ , LaSi ₃ N ₅

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50049981

B. Powder XRD analysis

Table 1 summarizes sample compositions and identified candidate phases by powder XRD analysis. Most of samples exhibited peaks that could not be identified, apart from the amount. Those minor unidentified diffraction peaks are not shown in this table. Most of the identified candidate phases are already known compounds in the LaN-Si₃N₄-AlN pseudo-ternary system, namely LaN, Si₃N₄, AlN, LaSi₃N₅, and La₃Si₆N₁₁. In addition to these nitrides, two known oxynitrides, LaSiNO₂ and La₁₁Si₁₃N_{27.64}O_{1.04} were identified. Oxygen impurity must be introduced from the starting powders or during the synthesis process. Seven samples, namely #1, #4, #8, #9, #11, #12, and #15, were shown as mixtures of these known compounds. La/(Si+Al) ratios in the products of the seven samples were estimated from the powder XRD analysis, which were found to be comparable to those in the starting powders. The identified candidate phases were confirmed to be existent. Sample #15 was found as a mixture of La₃Si₆N₁₁ and LaSi₃N₅, whereas the starting powders contained Al. It is well known that silicon (oxy)nitrides often form solid solutions, as exemplified in SiAlON, in which Al and O atoms respectively substitute for Si and N atoms.^{28,29} It is likely that one or both of the identified phases form solid solutions. The products of samples #7, #13, and #14 may also contain solid solutions with Al and O.

By the powder XRD analysis, three other candidate phases were identified: La₃BaSi₅N₉O₂ in samples #2, #7, #10, and #13, Ce₇Si₆N₁₅ in samples #3 and #14, La_{26-x}Sr_xSi₄₁O_{x+1}N_{80-x} in sample #7. As Ba, Ce and Sr are unlikely introduced as impurities, the products are most probably new variants of the candidate phases, namely La₄Si₅N₁₀O, La₇Si₆N₁₅, La₂₆Si₄₁N₈₀O, and their solid solutions with Al and O.

The first candidate, $\text{La}_4\text{Si}_5\text{N}_{10}\text{O}$, had an orthorhombic structure of space group $\text{Pmn}2_1$ (No. 31). Lattice constants of $\text{La}_4\text{Si}_5\text{N}_{10}\text{O}$ in sample #10 were refined as $a = 9.4772(14) \text{ \AA}$, $b = 19.116(3) \text{ \AA}$, $c = 12.0804(16) \text{ \AA}$, and $V = 2188.5(6) \text{ \AA}^3$. The lattice constants a and c are slightly (less than 1%) smaller than those of $\text{La}_3\text{BaSi}_5\text{N}_9\text{O}_2$,³⁰ whereas the lattice constant b is slightly larger. The cell volume is slightly smaller than that of $\text{La}_3\text{BaSi}_5\text{N}_9\text{O}_2$.

The second candidate, $\text{La}_7\text{Si}_6\text{N}_{15}$, had a high recommendation score of 0.040. As a matter of fact, $A_7\text{Si}_6\text{N}_{15}$ ($A = \text{La}, \text{Ce}, \text{ and Pr}$) have been reported by Schmolke et al. in 2009.³¹ $\text{La}_7\text{Si}_6\text{N}_{15}$ is not registered in ICSD, but $\text{Ce}_7\text{Si}_6\text{N}_{15}$ and $\text{Pr}_7\text{Si}_6\text{N}_{15}$ are registered and used to build the machine learning model. This composition was, however, excluded from the experimental target because of the larger La/Si ratio than 1.

The third candidate, $\text{La}_{26}\text{Si}_{41}\text{N}_{80}\text{O}$, was a valiant of $\text{La}_{26-x}\text{Sr}_x\text{Si}_{41}\text{O}_{x+1}\text{N}_{80-x}$ ($x = 12.72\text{-}12.90$) reported by Zhang et al. in 2020.³² The original phase, $\text{La}_{26-x}\text{Sr}_x\text{Si}_{41}\text{O}_{x+1}\text{N}_{80-x}$, is not yet registered in ICSD, but it has been registered in our in-house pattern library. Therefore, it was identified by the powder XRD analysis. It should be noted that the Sr/La ratio in Ref. 32 is 0.958-0.985 and is quite different from zero in this study. Its structure is hexagonal of space group $P\bar{6}$ (No. 174). The lattice constants of this phase in sample #7 were refined as $a = 17.3738(18) \text{ \AA}$, $c = 22.500(3) \text{ \AA}$, and $V = 5881.6(12) \text{ \AA}^3$. The lattice constant a of this phase is almost the same as that of $\text{La}_{13.1}\text{Sr}_{12.9}\text{Si}_{41}\text{O}_{13.9}\text{N}_{67.1}$, whereas the lattice constant c and the cell volume are slightly larger. Solubility of Al and O was not analyzed, but this phase would be a solid solution with Al and O.

C. Single-particle diagnosis

The powder XRD analysis identified no known phases in samples #5 and #6. Therefore, single-particle diagnosis was adopted to these samples. The diagnosed particle of sample

#5 was $8\ \mu\text{m} \times 10\ \mu\text{m} \times 10\ \mu\text{m}$ in size as shown in Fig. 2a. The measured ratio of cations by SEM-EDX analysis was La:Si:Al = 44.3:42.5:13.2 (at%). The amount of Eu was below the detection limit. The measured Si/Al ratio, 3.22, is in good agreement with the ratio of the starting powders, whereas the measured La/(Si+Al) ratio, 0.797, was somewhat smaller than the ratio of the starting powders. This may be due to the large energy difference between the La-L line and the Si- and Al-K lines used in the SEM-EDX analysis. The chemical composition may not have been calibrated correctly. The Si/Al ratio of this new phase will be hereafter assumed as the ratio of the starting powders, 3. Based on the single-crystal XRD analysis, the particle was identified to be $\text{La}_4\text{Si}_3\text{AlN}_9$ having a new monoclinic structure of space group $P2_1/c$ (No. 14) with $a = 6.7353(5)\ \text{\AA}$, $b = 5.5648(4)\ \text{\AA}$, $c = 12.8828(9)\ \text{\AA}$, $\beta = 105.4423(13)^\circ$, $V = 465.42(6)\ \text{\AA}^3$, and $Z = 2$ with $R = 2.2\%$, $wR = 4.3\%$, and $S = 1.03$. The crystal structure data is summarized in Table 2 and Supplementary Material S1 and S2. A schematic view of the crystal structure is illustrated in Fig. 2b. Here, Si and Al atoms were assumed to share the same sites with random distribution. This structure accommodates two types of La sites, two types of Si/Al sites, and five types of N sites. Both of the two symmetrically inequivalent La sites are coordinated to seven N atoms. Si/Al sites are coordinated to four N atoms, and (Si/Al) N_4 tetrahedra are linked via common corners. The powder XRD pattern of sample #5 can be fitted well using this structure only.

The powder XRD pattern of sample #6 can be fitted using the structure of $\text{La}_4\text{Si}_3\text{AlN}_9$. The refined lattice constants are $a = 6.7293(5)\ \text{\AA}$, $b = 5.5659(4)\ \text{\AA}$, $c = 12.8598(9)\ \text{\AA}$, $\beta = 104.731(3)^\circ$, and $V = 465.82(6)\ \text{\AA}^3$. As the Si/Al ratio of the starting powders is different between samples #5 and #6, this phase might form a solid solution with Al and O. But the lattice constants are almost the same, and there are minor unidentified diffraction peaks remained in the powder XRD pattern of sample #6. Further

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50049981

study is necessary to identify the origin of the unidentified peaks and to evaluate the solubility of Al and O in $\text{La}_4\text{Si}_3\text{AlN}_9$. $\text{La}_4\text{Si}_3\text{AlN}_9$ was also identified in the powder XRD pattern of sample #14.

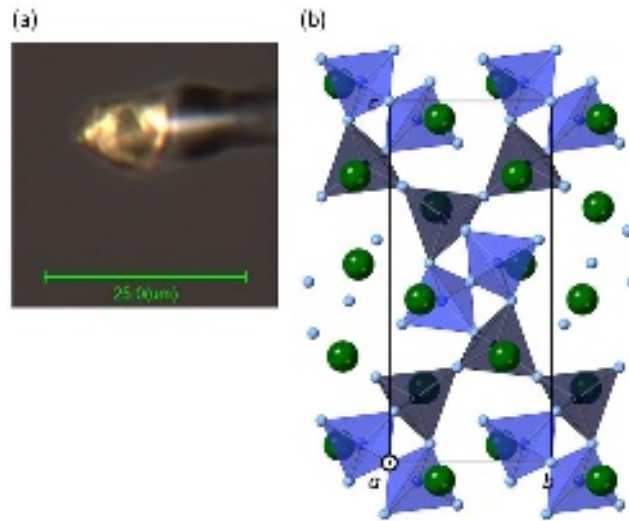


FIG. 2. (a) The diagnosed particle taken from sample #5 supported on a needle, and (b) a schematic view of the crystal structure of $\text{La}_4\text{Si}_3\text{AlN}_9$ from the a -axis direction. The b -axis is horizontal, and the reciprocal lattice vector of c^* is vertical. Light and dark blue tetrahedra represent $(\text{Si}/\text{Al})\text{N}_4$ at the Si/Al1 and Si/Al2 sites, respectively. Large green spheres and small light-blue spheres represent La and N atoms, respectively.

TABLE 2. Crystal structure data of $\text{La}_4\text{Si}_3\text{AlN}_9$. Space group $P2_1/c$ (No. 14), $a = 6.7353(5) \text{ \AA}$, $b = 5.5648(4) \text{ \AA}$, $c = 12.8828(9) \text{ \AA}$, $\beta = 105.4423(13)^\circ$, $V = 465.42(6) \text{ \AA}^3$, and $Z = 2$.

Site	Wyc koff	x	y	z	$U_{\text{iso}} (\text{\AA}^2)$	Occupa ncy (<1)
La1	4 e	0.30554(3)	0.69104(3)	0.04856(2)	0.00869(4)	
La2	4 e	0.12157(3)	0.21149(3)	0.20612(2)	0.00651(4)	
Si1	4 e	0.16588(14)	0.32634(16)	0.44798(7)	0.00457(15)	0.75
Al1	4 e	0.16588(14)	0.32634(16)	0.44798(7)	0.00457(15)	0.25
Si2	4 e	0.55124(15)	0.15765(17)	0.17746(8)	0.00725(17)	0.75
Al2	4 e	0.55124(15)	0.15765(17)	0.17746(8)	0.00725(17)	0.25
N1	4 e	0.3746(5)	0.2512(5)	0.0521(2)	0.0101(5)	
N2	4 e	0.2323(4)	0.5286(5)	0.3582(2)	0.0088(5)	
N3	4 e	0.0413(5)	0.0790(5)	0.3835(2)	0.0105(5)	
N4	4 e	0.5717(5)	0.4176(6)	0.2635(3)	0.0133(6)	
N5	2 a	0	0	0	0.0092(7)	

D. DFT calculations on $\text{La}_4\text{Si}_3\text{AlN}_9$

In the single-crystal XRD analysis on $\text{La}_4\text{Si}_3\text{AlN}_9$, Si and Al atoms were assumed to share the same sites with random distribution, since it is difficult to distinguish these two elements by XRD. However, the two sites have considerably different distances to the coordinating N atoms: 1.71-1.76 \AA of the Si/Al1 site, whereas 1.79-1.84 \AA of the Si/Al2 site. This implies that the Si and Al distribution is not random and the larger Al ions

preferentially occupy the Si/Al2 site. Therefore, DFT calculations on $\text{La}_4\text{Si}_3\text{AlN}_9$ were conducted to analyze the distribution of Si and Al atoms.

There are 10 symmetrically distinct configurations of Si and Al atoms within the unit cell of $\text{La}_4\text{Si}_3\text{AlN}_9$. Figure 3a illustrates the DFT total energies of these configurations with respect to the number of Al atoms at the Si/Al1 site. As more Al atoms occupy the Si/Al1 site, the energy becomes higher. Therefore, Al atoms likely occupy the Si/Al2 site only. This is consistent with the distances to the coordinating N atoms as mentioned above.

Among the three configurations in which all Al atoms occupy the Si/Al2 site, two of them have almost the same energy, whereas the other has considerably high energy. Hence the Si and Al configurations at the Si/Al2 site was additionally investigated using $2\times 1\times 1$ and $1\times 2\times 1$ supercells. There are 11 and 10 symmetrically distinct irreducible configurations for the $2\times 1\times 1$ and $1\times 2\times 1$ supercells, respectively, assuming all Al atoms occupy the Si/Al2 site. $(\text{Si}/\text{Al})\text{N}_4$ tetrahedra at the Si/Al2 site are one-dimensionally linked to each other via common corners of the N4 site along the *b*-axis direction (see Fig. 2b). Figure 3b illustrates the DFT total energies of all the configurations with respect to the number of AlN_4 - AlN_4 links at the Si/Al2 site, which is equal to the number of SiN_4 - SiN_4 links at the Si/Al2 site. Configurations with no AlN_4 - AlN_4 link are in the lowest energy group and have almost the same energies. As the number of AlN_4 - AlN_4 links increases, the energy becomes higher. These calculation results suggest that Si and Al atoms alternately occupy the Si/Al2 site on the one-dimensional links along the *b*-axis direction. In contrast to this one-dimensional ordering, configuration of Si and Al atoms on different links is probably random.

Harmonic phonon calculations were conducted on $\text{La}_4\text{Si}_3\text{AlN}_9$ with the lowest energy configuration of Si and Al atoms within the unit cell, which is also lower in energy

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50049981

than the other configurations within the $2\times 1\times 1$ and $1\times 2\times 1$ supercells. The structure of $\text{La}_4\text{Si}_3\text{AlN}_9$ optimized by the DFT calculation with the lowest energy configuration of Si and Al atoms is supplied as a CIF file of Supplementary Material S3. The phonon dispersion curves exhibit no phonon mode with imaginary frequency as shown in Fig. S1 in Supplementary Material, indicating that $\text{La}_4\text{Si}_3\text{AlN}_9$ is dynamically stable.

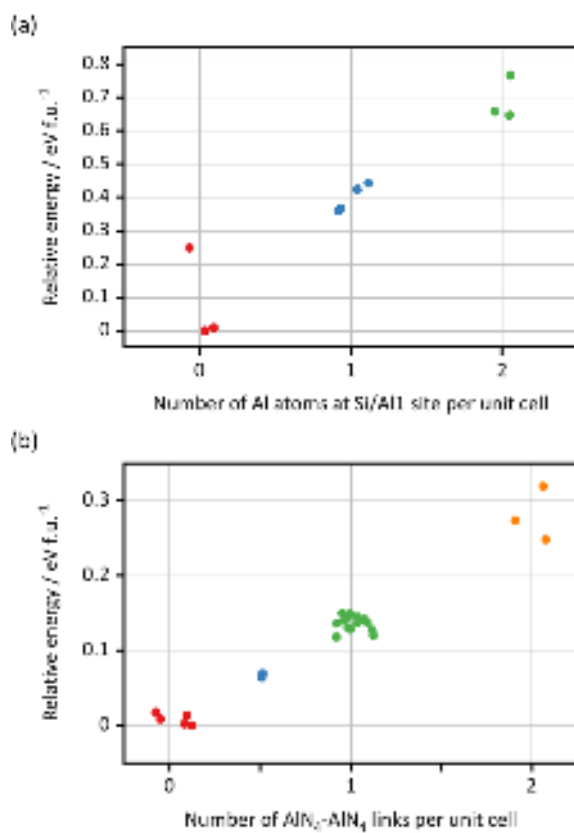


FIG. 3. (a) DFT total energy of $\text{La}_4\text{Si}_3\text{AlN}_9$ with a variety of configurations of Si and Al atoms within the unit cell with respect to the number of Al atoms at the Si/Al1 site, and (b) DFT total energy within the unit cell, $2\times 1\times 1$ supercell, and $1\times 2\times 1$ supercell with respect to the number of AlN_4 - AlN_4 links at the Si/Al2 site per unit cell assuming all Al atoms occupy the Si/Al2 site. The marks are jittered in the horizontal axes to avoid overlaps. Energies are relative to the lowest energy configuration.

E. Discussion

15 compositions in the LaN-Si₃N₄-AlN pseudo-ternary system by the machine learning prediction were experimentally validated. As the results, four new compounds, La₄Si₃AlN₉, La₇Si₆N₁₅, La₄Si₅N₁₀O, and La₂₆Si₄₁N₈₀O have been discovered. They might form solid solutions, in which Al and O atoms respectively substitute for Si and N atoms, although the solubility has not been analyzed in detail in this study. La₇Si₆N₁₅ and La₄Si₅N₁₀O are variants (isostructural substitutions) of already registered compounds in ICSD, whereas La₄Si₃AlN₉ and La₂₆Si₄₁N₈₀O have no similar-structural compounds in ICSD. The yield to discover new compounds is notably high. The success of this study demonstrates powerfulness of the proposed approach.

The new compound La₄Si₃AlN₉ discovered by the experimental validation has the exact composition by the machine learning prediction. The chemical compositions of AB₃C₄X₉ are quite rare in ICSD. They can be found only for borates such as NaCa₄B₃O₉, silicates and germanates such as K₄CaSi₃O₉, copper oxides such as YBa₄Cu₃O₉. It is therefore hard to imagine the unusual composition of La₄Si₃AlN₉ to be chemically relevant without the machine learning prediction.

In the experimental validation, the poor fitting of the powder XRD patterns only with known compounds in a pattern library is a promising indication of the presence of new compounds, as in samples #5 and #6. Fortunately, we have obtained almost single-phase products in these samples. It is, however, hard to know in general whether the sample is a single phase or not, unless the composition and crystal structure of the new compound is known. This has been a dilemma in the search of new compounds. The single-particle diagnosis is extremely helpful in such situation, since this approach does not require single-phase products in contrast to the conventional powder XRD analysis.

From the DFT total energies of $\text{La}_4\text{Si}_3\text{AlN}_9$ and other nitrides in the $\text{LaN-Si}_3\text{N}_4\text{-AlN}$ pseudo-ternary system, the phase diagram of this pseudo-ternary system is illustrated in Fig. 4. Computed formation energies of $\text{La}_4\text{Si}_3\text{AlN}_9$ and the known nitrides are listed on Table S1 in Supplementary Material. $\text{La}_4\text{Si}_3\text{AlN}_9$ is located on the energy convex hull, indicating that it is energetically stable even without consideration of configurational entropy associated with the mixture of Si and Al atoms.

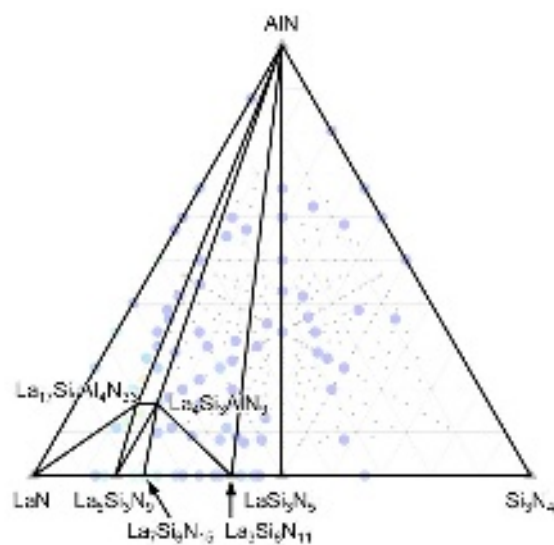


FIG. 4. Phase diagram of the $\text{LaN-Si}_3\text{N}_4\text{-AlN}$ pseudo-ternary system based on the DFT total energies, which is overlaid on the chemically relevant composition map shown in Fig. 1.

IV. CONCLUSION

We have proposed the new approach to search new compounds, in which chemical relevance of compositions is predicted by the machine learning and the prediction is validated by experiments using the single-particle diagnosis. The proposed approach has been adopted to the $\text{LaN-Si}_3\text{N}_4\text{-AlN}$ pseudo-ternary system, and four new

compounds, $\text{La}_4\text{Si}_3\text{AlN}_9$, $\text{La}_7\text{Si}_6\text{N}_{15}$, $\text{La}_4\text{Si}_5\text{N}_{10}\text{O}$, and $\text{La}_{26}\text{Si}_{41}\text{N}_{80}\text{O}$ have been discovered. They might form solid solutions, in which Al and O atoms respectively substitute for Si and N atoms. The yield to discover new compounds in this study is notably high. The success of this study demonstrates powerfulness of the proposed approach to search new compounds.

The new compound, $\text{La}_4\text{Si}_3\text{AlN}_9$, was found to be energetically and dynamically stable by the DFT calculations. It was suggested that Al atoms preferentially occupy the Si/Al₂ site and that Si and Al atoms are arranged alternately.

SUPPLEMENTARY MATERIAL

See Supplementary Material for the detailed crystallographic data and two CIF files for $\text{La}_4\text{Si}_3\text{AlN}_9$ that are determined by the experimental analysis and optimized by the DFT calculation with the lowest energy configuration of Si and Al atoms within the unit cell. Phonon dispersion curves of $\text{La}_4\text{Si}_3\text{AlN}_9$ with the lowest energy configuration by the DFT calculation, and formation energies of $\text{La}_4\text{Si}_3\text{AlN}_9$ and the known nitrides in the LaN-Si₃N₄-AlN pseudo-ternary system by the DFT calculation are also included in the Supplementary Material.

ACKNOWLEDGEMENTS

This work was supported in part by the Japan Science and Technology Agency (JST), CREST Grant Number JPMJCR19J2, and the “Materials Research by Information Integration” Initiative (MI²I) of the Support Program for Starting Up Innovation Hub. Y.K. Acknowledges financial support by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP18K04716. The DFT calculations in this study were performed on Numerical Materials Simulator at National Institute for Materials Science.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

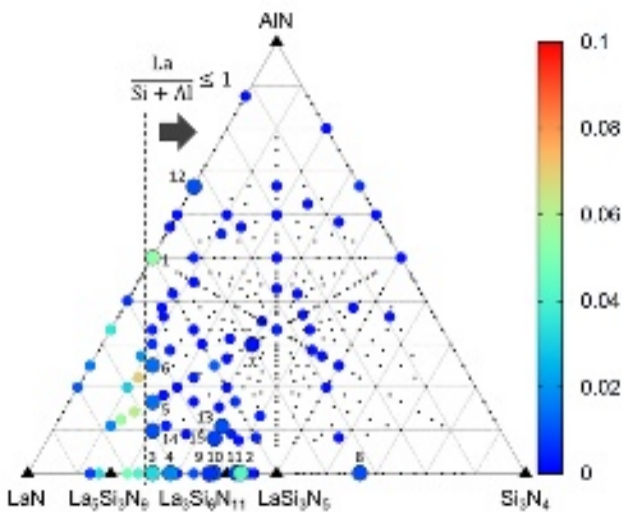
REFERENCES

1. Inorganic Crystal Structure Database (ICSD), FIZ Karlsruhe GmbH, Germany.
2. International Centre for Diffraction Data Powder Diffraction File (ICDD-PDF), JCPDS-International Centre for Diffraction Data, USA.
3. SpringerMaterials, Springer Nature Switzerland AG, Switzerland.
4. Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds, ASM International, USA.
5. S. Grazulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quiros, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, *J. Appl. Crystallogr.* 42, 726 (2009).
6. AtomWork-Adv, National Institute for Materials Science, Japan.
7. 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, *APL Mater.* 1, 11 (2013).
8. S. Curtarolo, W. Setyawan, S. D. Wang, J. K. Xue, K. S. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, *Comp. Mater. Sci.* 58, 227 (2012).
9. J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, *JOM* 65, 1501 (2013).
10. C. Draxl and M. Scheffler, *MRS Bull.* 43, 676 (2018).
11. P. Villars and S. Iwata, *Chem. Met. Alloys* 6, 81 (2013).
12. A. R. Oganov and C. W. Glass, *J. Chem. Phys.* 124, 15 (2006).
13. Y. C. Wang, J. A. Lv, L. Zhu, and Y. M. Ma, *Phys. Rev. B* 82, 8 (2010).
14. M. Amsler and S. Goedecker, *J. Chem. Phys.* 133, 8 (2010).
15. C. J. Pickard and R. J. Needs, *J. Phys.-Condens. Mat.* 23, 053201 (2011).
16. K. Terayama, T. Yamashita, T. Oguchi, and K. Tsuda, *npj Comput. Mater.* 4, 8 (2018).
17. G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, *Chem. Mater.* 22, 3762 (2010).

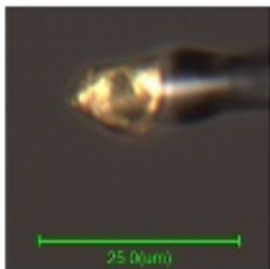
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50049981

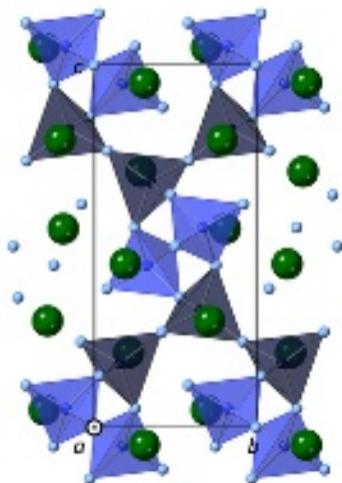
18. G. Hautier, C. Fischer, V. Ehrlacher, A. Jain, and G. Ceder, *Inorg. Chem.* 50, 656 (2011).
19. B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, *Phys. Rev. B* 89, 7 (2014).
20. A. Seko, H. Hayashi, H. Kashima, and I. Tanaka, *Phys. Rev. Mater.* 2, 8 (2018).
21. A. Seko, H. Hayashi, and I. Tanaka, *J. Chem. Phys.* 148, 7 (2018).
22. K. Suzuki, K. Ohura, A. Seko, Y. Iwamizu, G. W. Zhao, M. Hirayama, I. Tanaka, and R. Kanno, *J. Mater. Chem. A* 8, 11582 (2020).
23. N. Hirosaki, T. Takeda, S. Funahashi, and R. J. Xie, *Chem. Mater.* 26, 4280 (2014).
24. G. Kresse and J. Furthmüller, *Phys. Rev. B* 54, 11169 (1996).
25. G. Kresse and D. Joubert, *Phys. Rev. B* 59, 1758 (1999).
26. J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* 77, 3865 (1996).
27. A. Togo and I. Tanaka, *Scr. Mater.* 108, 1 (2015).
28. R. J. Xie and N. Hirosaki, *Sci. Technol. Adv. Mater.* 8, 588 (2007).
29. G. G. Li, Y. Tian, Y. Zhao, and J. Lin, *Chem. Soc. Rev.* 44, 8688 (2015).
30. D. Durach, L. Neudert, P. J. Schmidt, O. Oeckler, and W. Schnick, *Chem. Mater.* 27, 4832 (2015).
31. C. Schmolke, O. Oeckler, D. Bichler, D. Johrendt, and W. Schnick, *Chem.-Eur. J.* 15, 9215 (2009).
32. Y. X. Zhang, S. X. Li, T. Takeda, S. Funahashi, N. Hirosaki, and R. J. Xie, *J. Mater. Chem. C* 8, 13458 (2020).



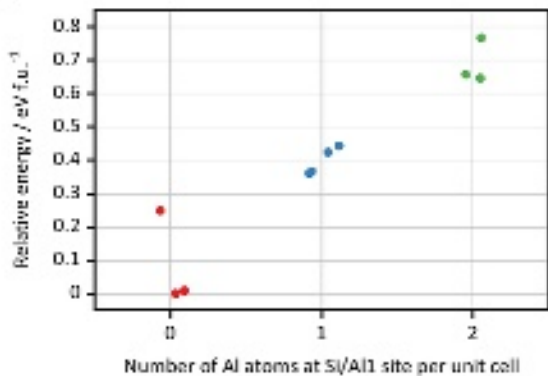
(a)



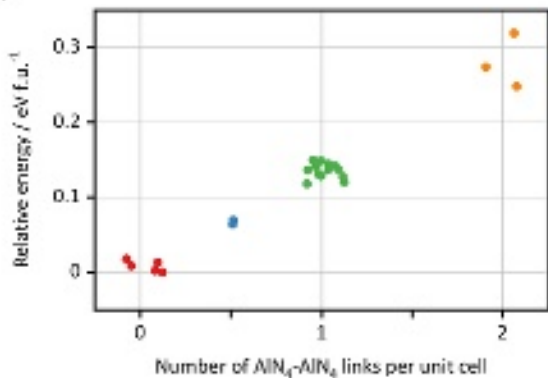
(b)



(a)



(b)



AlN

