Machine learning guided discovery of ternary compounds containing La, P and group 14 elements

Huaijun Sun[1,3], Chao Zhang[2], Weiyi Xia[3,4], Ling Tang[5,3], Renhai Wang[6], Georgiy Akopov[3,7], Nethmi W. Hewage,[3,7] Kai-Ming Ho[3,4], Kirill Kovnir[3,7], Cai-Zhuang Wang[3,4,*]

[1]Jiyang College of Zhejiang Agriculture and Forestry University, Zhuji, 311800, China

[2]Department of Physics, Yantai University, Yantai 264005, China

[3]Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States;

[4]Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, United States

[5]Department of Applied Physics, College of Science, Zhejiang University of Technology, Hangzhou, 310023, China

[6]School of Physics and Optoelectronic Engineering, Guangdong University of Technology, Guangzhou 510006, China

[7]Department of Chemistry, Iowa State University, Ames, IA 50011, United States.

*Corresponding authors: wangcz@ameslab.gov

ABSTRACT

We integrate a deep machine learning (ML) method with first-principles calculations to efficiently search for the energetically favourable ternary compounds. Using La-Si-P as a prototype system, we demonstrate that ML-guided first-principles calculations can efficiently explore crystal structures and their relative energetic stabilities, thus greatly accelerate the pace of materials discovery. A number of new La-Si-P ternary compounds with formation energies less than 30 meV/atom above the known ternary convex hull are discovered. Among them, the formation energies of a La_5SiP_3 and a La_2SiP phases are only 2 meV/atom and 10 meV/atom respectively above the convex hull. These two compounds are dynamically stable with no imaginary phonon modes. Moreover, by replacing Si with heavier group 14 elements in the eight lowest-energy La-Si-P structures from our ML-guided predictions, a number of low-energy La-*X*-P phases (*X* = Ge, Sn, Pb) are predicted.

1. Introduction

Ternary compounds formed by La, P, and group 14 element *X* (*X* = Si, Ge, Sn, and Pb) are relatively unexplored but promising for novel quantum materials discovery. Many reported silicon phosphides of transition and/or rare-earth metals crystalize in the non-centrosymmetric (NCS) structures. The absence of inversion symmetry leads to antisymmetric spin-orbit coupling which splits the Fermi surface and removes the spin degeneracy of electrons. Resulting NCS intermetallics may exhibit significant hybridization of *d*-, *f*-, and *p*-orbitals resulting in a plethora of emergent properties, such as unconventional superconductivity, topologically non-trivial quantum properties over large energy windows, and quasiparticle behavior [1-4]. In contrast to the abundance of binary *X*-P and molecular compounds with *X*-P covalent bonds [5-14], much less is known regarding ternary systems containing La, P, and group 14 elements *X*. A limited number of reported compounds [15-17] with composition ratio (*X*+P)/La > 1 have an extended covalent *X*-P framework resulting in unique properties. For example, a fraction of La-Si-P compounds are non-centrosymmetric semiconductors with promising non-linear optical applications [18]. One of the interesting questions regarding to the La-*X*-P ternary compounds is "*can any stable La-X-P ternary compounds be formed when the composition (X+P)/La is less than or equal to 1*?"

A successful experimental search for such ternaries using traditional "trial-and-error" methods would require an enormous dose of good fortune [19]. Computationally, one can in principle search for ternary La-*X*-P structures with favorable formation enthalpy using the state-of-the art crystal structure prediction algorithms such as genetic algorithm (GA) [20-23], particle swarm optimization (PSO) [24, 25], and adaptive genetic algorithm (AGA) [26-28]. These algorithms and methods have been demonstrated to be very useful for finding low-energy crystal structures with given chemical compositions [20-28]. However, even for a given 3 elements, the number of possible combinations of different stoichiometry is enormous. Using these structure search algorithms directly to examine all possible compositions will be very time consuming. Efficient computational approaches to accelerate the discovery new materials remain a great challenge.

In this paper, we use a machine learning (ML) approach to efficiently address this problem. We use the La-Si-P ternary system as a prototype La-*X*-P system for the proof of principle of the ML approach. **Fig. 1** shows the currently known ternary convex hull of La-Si-P [16]. Two experimental

observed ternary compounds (i.e, $La(SiP_3)_2$ and $LaSiP_3$) have been reported in Refs 16 and 17, and another compound (La_2SiP_4) has also been recently experimentally synthesized [17, 29]. These three ternary phases satisfy the composition condition (Si+P)/La > 1. However, no ternary compounds with compositions (Si+P)/La $\leqslant$ 1 have been reported. This composition domain covers one quarter of the whole ternary phase diagram as shown by the shaded area in **Fig. 1**. We will show that our proposed ML approach leads to discovery of a number of low-energy metastable compounds in the composition region (Si+P)/La $\leqslant$ 1. In particular, a La_5SiP_3 and a La_2SiP phases are predicted to be only 2 meV/atom and 10 meV/atom respectively above the convex hull. These two compounds are dynamically stable with no imaginary phonon modes from first-principles density functional theory (DFT) calculations. Moreover, by replacing Si with other group 14 elements in the 8 lowest-energy La-Si-P structures from our ML-guided predictions, a number of low-energy La-*X*-P (*X* = Ge, Sn, Pb) are predicted.

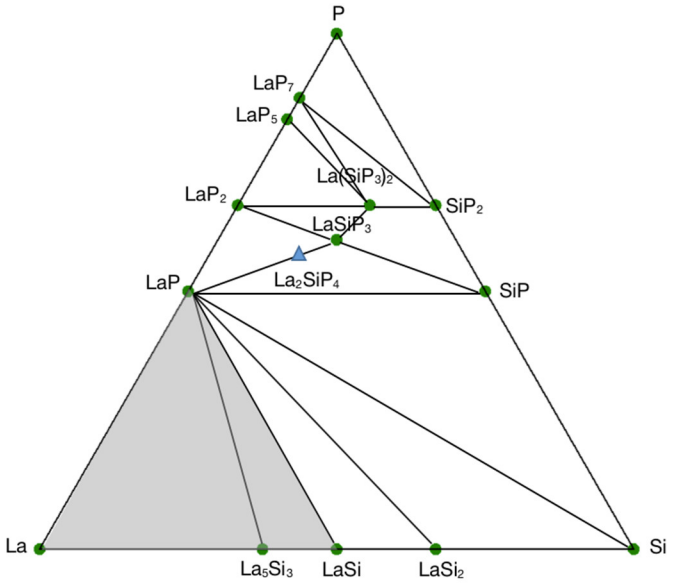


Fig. 1. The currently known convex hull of ternary La-Si-P system. The La_2SiP_4 compound synthesized recently by experiment is found to be 20 meV/atom above the convex hull and is shown by the blue triangle in the figure.

2. Computational details

For given three elements (La, Si, and P in the present study), we employ a crystal graph convolutional neural network (CGCNN) ML model to efficiently guide the selection of promising stoichiometry and structures for further optimization by first-principles DFT calculations. In CGCNN, a crystal structure is represented by a crystal graph which encodes both atomic information and bonding interactions between atoms. A convolutional neural network is then built on top of the graph to construct the proper descriptors that are optimal for predicting target properties. The CGCNN model used in the present study was developed by Xie and Grossman [30]. This model was trained using the structures and energies of 28,046 binary and ternary compounds from DFT calculations in the Materials Project (MP) database [15]. The nodes of the crystal graph were represented by 9 atomic properties of the elements including group number, period number, electronegativity, covalent radius, valence electrons, first ionization energy, electron affinity,

block and atomic volume. The graph edges were characterized by neighboring bonds for each atom/node. The dataset was divided into training set (80%), validation set (10%) and test set (10%), respectively. The mean absolute error for the CGCNN model in comparison with the DFT calculated formation energy is 0.039 eV/atom [30], indicating that the trained CGCNN model has sufficient accuracy for fast screening of formation energies of the ternary compounds. We adopt this CGCNN formation energy model to perform fast predictions for the relationship among the crystal structures, chemical compositions, and formation energies.

Initial structures of the ternary La-Si-P compounds are generated by replacing the three elements on the lattice sites of 28,469 ternary structures extracted from the MP database. For each ternary structure from the MP database, five hypothetical lattices are generated by uniformly scaling the bond lengths of the structure so that the volume of the unit cell changes from the original volume by a factor of 0.96 to 1.04 of in an increment of 0.02. Since CGCNN model does not have the interatomic forces to relax the bond lengths in the structures, the use of scaling factor for the crystal unit cell volume helps the CGCNN model to differentiate the energetic stability of the same structure with different bond lengths. There are also six ways to shuffle the three elements on a given template ternary structure. Therefore, by multiplying 6 combinations of elements and 5 volume steps, 854,070 ternary La-Si-P hypothetical structures are generated from the 28,469 ternary structures extracted from the MP database. Although the binary structures in the MP database are not used in the generation of the hypothetical structures of ternary La-Si-P compounds in our scheme, they are included in training the CGCNN model for the formation energy predictions. We also noted that the generated hypothetical La-Si-P compounds are based on the known structure types in the MP database. Different structure types in other structural database and unknown structure types are not included in our structure pool. The formation energy distribution from the CGCNN prediction for this set of structures generated by our scheme is shown in **Fig. 2.**

From the CGCNN ML model prediction shown in **Fig. 2**, we can see that there are substantial fraction of the La-Si-P ternary compounds have negative formation energy E_f. Here E_f is defined with respect to the elemental ground-state bulk phases of the constituent elements La, Si, and P, i.e.,

$$E_f = \frac{E(La_mSi_nP_p) - mE(La) - nE(Si) - pE(P)}{m+n+p}, \qquad (1)$$

where $E(La_mSi_nP_p)$ is the total energy of the $La_mSi_nP_p$ compound, and $E(La)$, $E(Si)$ and $E(P)$ are the per-atom energy of the ground state of La, Si and P crystals, respectively. Negative formation energy indicates that the formation of the ternary compounds is energetically favourable with respect to the three elemental phases.

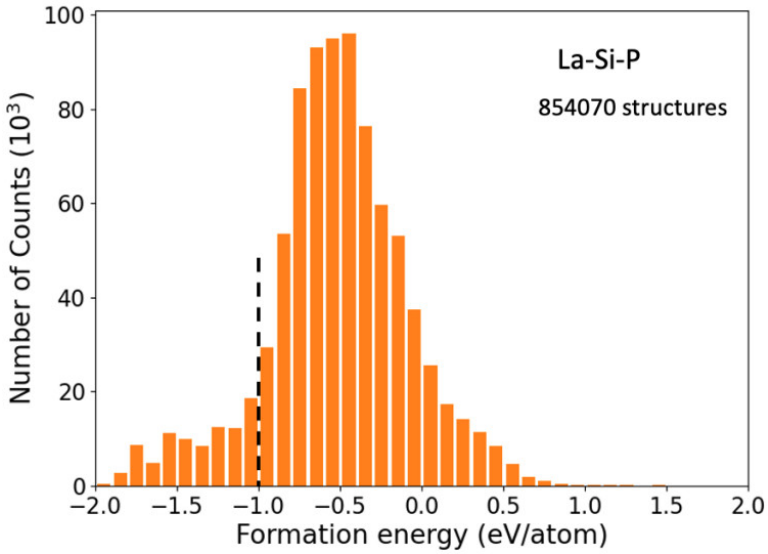


Fig. 2. The formation energies (E_f) of the 854070 ternary La-Si-P compounds predicted by CGCNN ML model. 1466 distinct structures with E_f less than -1.0 eV/atom (left of black dashed line) and the composition ratio of (Si+P)/La $\leqslant$ 1 are selected for further optimization by first-principles DFT calculations.

Among the structures predicted by CGCNN that have negative formation energies, 1466 distinct structures with E_f less than -1.0 eV/atom and the composition ratio of (Si+P)/La $\leqslant$ 1 are founded. These 1466 structures from the original 854070 structures are therefore selected for further optimization by first-principles DFT calculations.

To further investigate the energetic stability of the structures at different compositions, the formation energy with respect to the known convex hull of the ternary system (denoted as E_d) are also calculated using the energy data from the DFT calculations. More details of the DFT calculations are shown in the Supporting Information. E_d of any given phase on the ternary convex hull can be calculated by comparing its formation energy with respect to the nearby three known phases on the convex hull. These three known phases can be ternary, binary, or elementary crystalline phases, and the chemical compositions of these three phases are located at the vertexes of a triangle (called a Gibbs triangle) that encloses the composition of the phase for which E_d is calculated. Therefore, E_d determines the thermodynamic stability of the given phase against decomposition into the nearby three known phases.

3. Results and discussion

Our ML-guided DFT calculations reveal 130 metastable La-Si-P ternary compounds with the formation energy within 100 meV/atom from the convex hull in the composition range (Si+P)/La $\leqslant$ 1. In particular, we find that 4 La-Si-P ternary compounds with formation energy less than 15 meV/atom above the convex hull, 4 compounds with formation energy between 15-30 meV/atom above the convex hull, and another 17 compounds with formation energy between 30 and 50 meV/atom from the convex hull. These low-energy structures could be synthesizable by experiment at finite temperatures. More discussions about 4 lowest-energy structures from our ML-guided ab initio predictions are given in the following subsections. The structures and

crystallographic information of the other 4 low-energy compounds, and the energy information of all the 130 metastable ternary compounds are given as **Fig. S1, Table S2** and **Table S3** in Supporting Information.

3.1 Structures of the lowest-energy metastable phases

The structures of the 4 lowest-energy metastable compounds (with formation energies less than 15 meV/atom above the convex hull) predicted by our ML-guided DFT calculations are plotted in **Fig. 3.** The crystallographic information of these 4 compounds is given as **Table S1** in Supporting Information. Both of the La_5SiP_3 and La_2SiP compounds are *C*-centered orthorhombic structures with *Cmcm* (No. 63) symmetry, whose conventional cells contains four formula units ($Z = 4$). They share one important structural feature, that is, they both contain P-centered La_6 octahedron and Si-centered La_6 prism. The P-centered octahedra share edges with each other forming a layer in the *a-b* plane for the La_5SiP_3 and in the *a-c* plane for the La_2SiP. The orientation of P-centered octahedron in the two compounds is different, as shown in Figs. 3(a) and 3(b). The layers of P-centered octahedra sandwich Si-centered prism forming a three-dimensional structure. The Si-centered prisms in the *Cmcm*-La_5SiP_3 structure share trigonal faces along the *a* axis, whereas in the *Cmcm*-La_2SiP, the Si-centered prisms form a layer. The P-centered octahedron and Si-centered prism have similar volume in the two compounds, i.e., 34.9 $Å^3$ for P-centered octahedron and 29.5 $Å^3$ for Si-centered prism. The average bond length of La-P and La-Si in the two compounds are 2.97 Å and 3.17 Å, respectively. The arrangement of the Si atoms in the *Cmcm*-La_2SiP compound can also be view as zigzag chain with Si-Si bond length of 2.542 Å. The latter is 0.2 Å longer than the sum of Si covalent radii, 2.34 Å. We performed the Wyckoff sequence search by comparing to those of the known ternary structures in MP database. We found the structure type for the newly predicted La_5SiP_3 (*Cmcm*) and La_2SiP (*Cmcm*) phases are the same as that of $SrTi_3O_5$ and Mo_2BC respectively in MP database. The Mo_2BC structure has been observed experimentally, while the $SrTi_3O_5$ is only computationally predicted structure.

Both $La_{12}SiP_8$ and La_8SiP_5 crystallizes are noncentrosymmetric body-centered tetragonal structures. The symmetry of $La_{12}SiP_8$ is $I\bar{4}2d$ (No. 122) with four formula units (i.e., Z=4) in the conventional cell, whereas the symmetry of La_8SiP_5 is $I\bar{4}$ (No. 82) with 2 formula units (i.e., Z=2). The $I\bar{4}2d$-$La_{12}SiP_8$ and $I\bar{4}$-La_8SiP_5 are composed of P- and Si-centered distorted snub disphenoids or Siamese dodecahedra. These dodecahedra share faces and edges forming a three-dimensional network. The average volume of P-centered dodecahedra is 54.0 $Å^3$ and 54.2 $Å^3$ for the $I\bar{4}2d$-$La_{12}SiP_8$ and $I\bar{4}$-La_8SiP_5, respectively. The La-P distance is in the range of 2.96-3.32 Å. The average volume of Si-centered dodecahedra of the $I\bar{4}2d$-$La_{12}SiP_8$ (55.9 $Å^3$) is slightly smaller than that of the $I\bar{4}$-La_8SiP_5 (56.3 $Å^3$). Two types of Si-La distances in both of the $I\bar{4}2d$-$La_{12}SiP_8$ and $I\bar{4}$-La_8SiP_5 are 3.01 Å and 3.31 Å. An interesting observation is that half of the four most stable predicted structures shown in **Fig. 3** are non-centrosymmetric which is in line with reported La-poor La-Si-P phases, where $LaSi_2P_6$ and two polymorphs of $LaSiP_3$ are non-centrosymmetric, while another polymorph of $LaSiP_3$ and La_2SiP_4 are centrosymmetric.

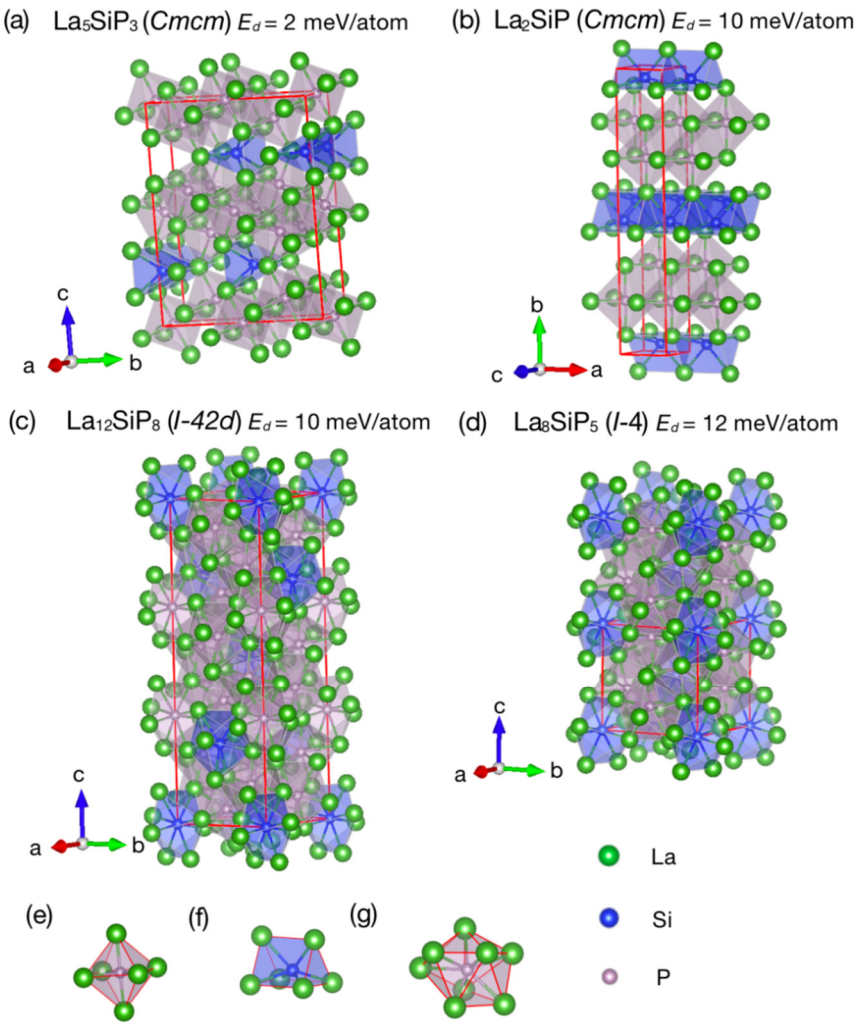


Fig. 3. The structures of the 4 lowest-energy La-Si-P compounds are obtained from our ML-guided ab initio calculation prediction [(a)-(d)]. The formation energies of the predicted La_5SiP_3, La_2SiP, $La_{12}SiP_8$, and La_8SiP_5 are 2 meV/atom, 10 meV/atom, 10 meV/atom, and 12 meV/atom above the convex hull, respectively. P-centered La_6 octahedron (e) and Si-centered La_6 prism (f) are contained in both La_5SiP_3 and La_2SiP, whereas the $La_{12}SiP_8$ and La_8SiP_5 compounds are composed of P- and Si-centered Siamese dodecahedra (g).

3.2 Dynamical stability of the La_5SiP_3 and La_2SiP compounds

To further investigate the dynamical stability, we calculated the phonon dispersion for La_5SiP_3 and La_2SiP using the DFT via Phonopy code [31]. The phonon dispersion curves of the La_5SiP_3 and La_2SiP compounds are shown in **Fig. 4**. The absence of imaginary frequencies confirms dynamical stability of the La_5SiP_3 and La_2SiP phases.

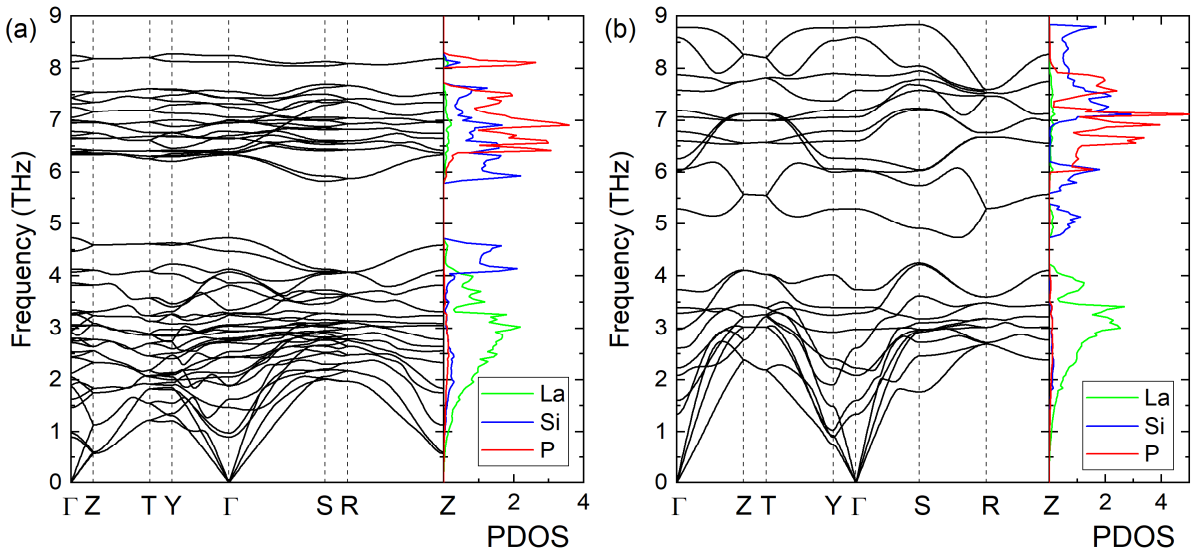


Fig. 4. The phonon dispersion and phonon density of states of the predicted (a) La_5SiP_3 and (b) La_2SiP compounds. No imaginary vibration frequencies are observed, indicating these structures are dynamically stable.

There is a gap (~1.1 THz in La_5SiP_3 and ~0.5 THz in La_2SiP) in the phonon dispersions of the two compounds, which divides the phonon dispersion into two groups. For the *Cmcm*-La_5SiP_3, the La atoms dominate the low-frequency range below 4 THz, whereas the Si atom dominate the vibration modes between 4 and 6.3 THz. Both Si and P atoms contribute strongly to the vibration modes above 6.3 THz. The phonon dispersion of the *Cmcm*-La_2SiP exhibits some difference from that of *Cmcm*-La_5SiP_3. While the low frequency range below 4 THz is still dominated by the vibration of La atoms as seen in the *Cmcm*-La_5SiP_3 phase, the maximum vibrational frequency of *Cmcm*-La_2SiP is larger than that of *Cmcm*-La_5SiP_3. For the high-frequency range above 4.7 THz in the *Cmcm*-La_2SiP compound, the vibration modes of P atom distribute between 6 THz and 8 THz, whereas the vibration modes related to Si atom split into two ranges, i.e., 4.7 – 6.2 THz and 6.8 – 8.9 THz. The difference between the vibration properties of the Si and P atoms can be attributed to the different interaction strength among the La, Si, and P atoms in the two structures.

We also performed molecular dynamics (MD) simulations to assess the stability of the La_5SiP_3 and La_2SiP crystalline phases at finite temperature. The MD simulations are performed using a deep machine leaning interatomic potential recently developed by some of the authors and the LAMMPS package [32]. The initial MD simulation box is a 12×3×8 and 12×4×12 crystalline supercell containing 5184 and 4608 atoms for La_5SiP_3 and La_2SiP, respectively. An isothermal-isobaric (NPT) ensemble and a Nose-Hoover thermostat are used in the simulations [33, 34]. The MD time step for the simulation is 2.5 fs. The La_5SiP_3 and La_2SiP supercells are first heated to 1200 K for 50 ps and then continuously heated up to 2000 K and 2400 K respectively at rates of 10^{12} K/s. **Fig. 5 (a)** and **(b)** show the evolution of instantaneous potential energy E-$3k_BT$ as function of temperature for the La_5SiP_3 and La_2SiP samples respectively. We can see a clear jump in the potential energy at T≈1600 K and T≈2100 K for La_5SiP_3 and La_2SiP phases, respectively, indicating occurrence of crystal-liquid transitions around these temperatures. Snapshot structures

of the simulation cell before and after the melting are plotted in the insets of **Fig. 5** to give a visual view of the changes in the atomic structures upon the crystal melting. We note the melting temperatures from these simulations would be overestimated due to the finite system size and simulation time used in the MD simulations. Nevertheless, it is reasonable to believe that these two crystalline structures are stable at high temperature at least above 1000 K. We have also performed ab initio MD (AIMD) simulations at 1300 K using supercells of 216 and 256 atoms for the La_5SiP_3 and La_2SiP phases respectively. The crystal structures remain stable at 1300 K during the AIMD simulation time of 20 ps. More detail information obtained from the AIMD simulations are shown in **Fig. S2** and **Fig. S3** in the Supporting Information.

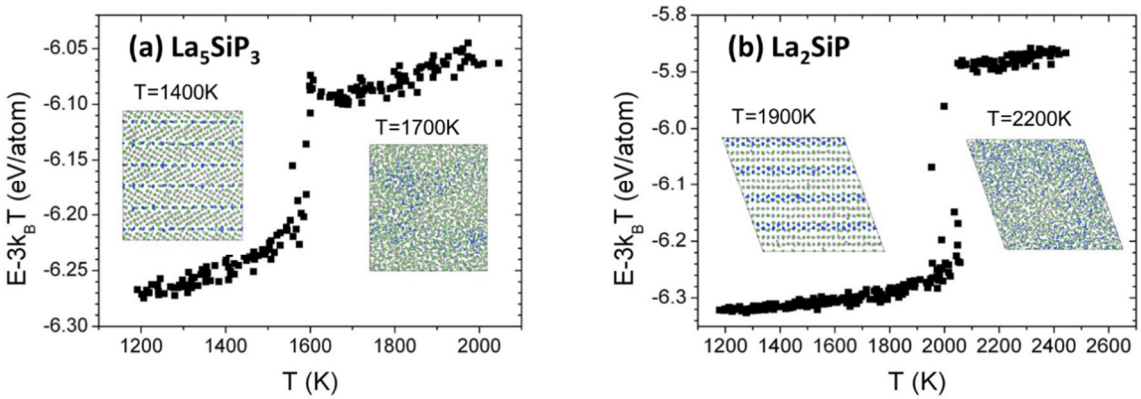


Fig. 5. The instantaneous potential energy E-3k_BT as function of temperature for (a) La_5SiP_3 and (b) La_2SiP phase, respectively. The insets are the snapshot atomistic structures before and after melting from our MD simulations.

3.3 Gibbs free energy calculation

In order to gain insights into the effects of temperature on the thermodynamics stability of the La_5SiP_3 and La_2SiP compounds, we evaluated the Gibbs free energy as a function of temperature for the La_5SiP_3, La_2SiP and relevant binary and ternary phases in the La-Si-P system by including the vibrational contributions.

Under zero pressure and at given temperature T and volume V, the Gibbs free energy G of a crystalline compound is given by

$$G = E_0(V) + \frac{1}{2}\sum_{q,\nu}\hbar\omega_{q,\nu} + k_B T \sum_{q,\nu} ln\left[1 - exp\left(-\hbar\omega_{q,\nu}/k_B T\right)\right], \qquad (2)$$

where $E_0(V)$ is the energy at T = 0 K, k_B and $\hbar$ are Boltzmann constant and the reduced Planck constant, respectively, $\omega_{q,\nu}$ is the phonon frequency at q and υ, where q and υ are the wave vector and band index, respectively [35].

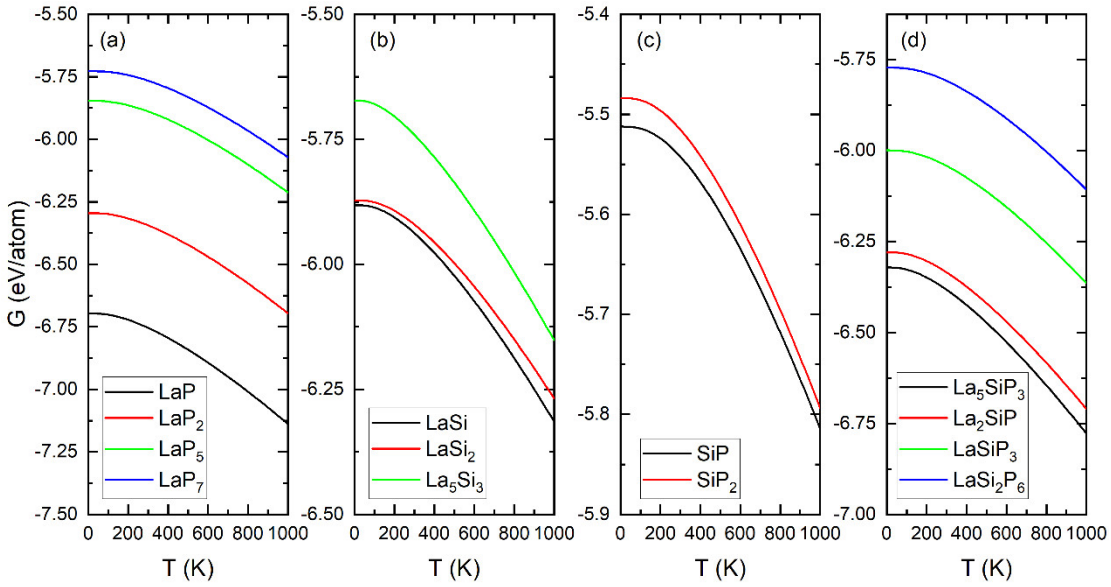


Fig. 6. Gibbs free energy as a function of temperature for (a) La-P binary system, (b) La-Si binary system, (c) Si-P binary system, and (d) La-Si-P ternary system.

The results of the Gibbs free energy as the function of temperature for the relevant phases from our calculations are shown in **Fig. 6**. To characterize the thermodynamic stability of the ternary La_5SiP_3 and La_2SiP structures at different temperature, the formation Gibbs energies ΔG of each ternary phase with respect to the nearby competing binary and ternary phases in the ternary convex hull at each T are calculated. The results are plotted in **Fig. 7**. We can see that formation Gibbs free-energies of the La_5SiP_3 and La_2SiP phases with respect to the ternary convex hull increase monotonically from 0 K to 1000 K. However, the changes are very moderate from 3.6 to 8.4 meV/atom for the La_5SiP_3 phase and 9.6 to 16.2 meV/atom for the La_2SiP phase. These results suggest the La_5SiP_3 and La_2SiP are still thermodynamically competitive metastable phases even at temperatures as high as 1000 K.

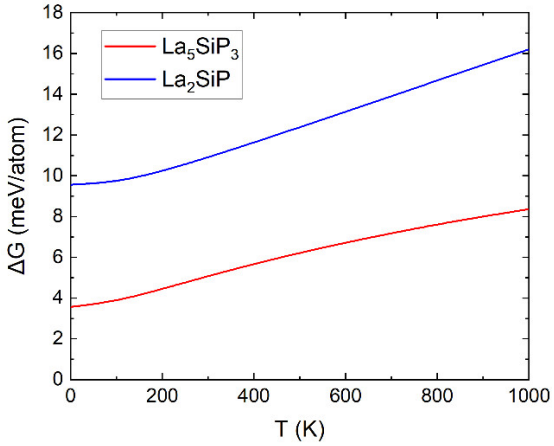


Fig. 7. Temperature-dependent formation Gibbs free-energies ΔG of the La_5SiP_3 and La_2SiP phases with respect to the La-Si-P ternary convex hull.

3.4 Synthetic attempts

First of all, a safety warning with phosphide synthesis should be noted. Phosphorus vapor pressure produced at high temperatures may be sufficient to compromise the reaction ampoule and may lead to ampoule shattering. For related reactions, the phosphorus amounts should be kept to a minimum. For the same reason, arc-melting of samples with phosphorus is highly discouraged – a significant depletion of sample with phosphorus will occur simultaneously with contamination of arc-melted with very reactive phosphorus vapor and deposits of pyrophoric white phosphorus.

Using the expertise developed in the synthesis of $LaSiP_3$ and La_2SiP_4 phases we have attempted syntheses of the predicted La-rich compounds, namely La_2SiP and La_5SiP_3. Various attempts were undertaken utilizing elements as stating materials as well as La_5Si_3 arc-melted precursor, reaction temperature was varied from 500 to 1000 ^oC. Synthesis of La_2SiP was attempted using a nominal composition of $La_5Si_3P_3$=$La_{1.67}SiP$, using a La_5Si_3 precursor and elemental P in a 1:3.1 ratio. Synthesis of La_5SiP_3 was attempted using elemental La, Si and P in a 5:3:3.1 ratio. More details about the synthesis process are given in the Supporting Information.

All the efforts resulted in the formation of stable binary intermediate LaP and unreacted Si. Synthesis of La-rich La-Si-P phases remains a challenge. Previous success in synthesis of (Si+P)-rich phases was due to chemical potential of P which was present in all reactions in molar excess, note that both $LaSiP_3$ and La_2SiP_4 have more than 50 at-% of P. Reducing P content to 25% resulted in only partial reaction of predominantly La with P resulting in a formation of LaP, a kinetic dead-end of the reaction. Novel synthetic methods to explore metal-rich parts of phase diagrams containing highly volatile components, like P, need to be developed. This is a focus of undergoing efforts.

3.5 Substitution of group 14 elements for Si

By substituting the Si atoms in the eight lowest-energy La-Si-P compounds with heavier group 14 elements *X* (*X* = Ge, Sn, Pb), we found that some of these *X*-substituted compounds have the formations either below or very close to the currently known convex hulls. The formation energy relative to the convex hull of these *X*-substituted compounds after first-principles DFT optimization are shown in **Table 1**, in comparison with those of the corresponding La-Si-P compounds. La_5SnP_3 and La_5PbP_3 are predicted to be stable compounds with their formation energies below the currently known convex hulls for these systems. The crystal structures after the substitution and DFT relaxation are given in **Table S4, Table S5 and Table S6** in Supporting Information. Some of these La-*X*-P ternary compounds would be synthesizable by experiment.

Table 1. Comparison of the formation energy above the convex hull for the 8 lowest-energy La-Si-P compounds with Si being substituted by Ge, Sn, and Pb respectively. The energies are in the unit of eV/atom.

Composition	X = Si	X = Ge	X = Sn	X = Pb
La_5XP_3	0.002	0.058	-0.016	-0.019
La_2XP	0.010	-0.001	0.037	0.023
$La_{12}XP_8$	0.010	0.006	0.029	0.041
La_8XP_5	0.012	0.004	0.034	0.052
La_4XP_2	0.019	0.003	0.063	0.077
$La_{13}X_2P_8$	0.026	0.010	0.006	0.004
$La_{24}X_3P_{14}$	0.027	0.020	0.053	0.072
$La_{10}XP_6$	0.028	0.022	0.048	0.055

4. Conclusions

In summary, we searched for possible thermodynamically stable ternary La-Si-P compounds in composition range (Si+P)/La $\leqslant$ 1 using a CGCNN ML method and first-principles DFT calculations. The CGCNN ML screening enable to quickly identify a short list of 1446 candidate structures over many possible combinations of compositions and crystal structures for DFT calculations. This small fraction of structures can be computed in days by standard DFT calculation methods on commonly used cluster computers. Our DFT explorations quickly predict 130 ternary compounds in the composition domain of (Si+P)/La $\leqslant$ 1 which have formation energy less than 100 meV/atom above the convex hull. In particular, La_5SiP_3 and La_2SiP compounds have formation energies only 2 meV/atom and 10 meV/atom, respectively, above the convex hull. It should be noted that partial occupancies and atomic mixings between Si and P are not considered in our predictions because it is challenging for DFT to deal with such situations. Nevertheless, majority of studied M-Si-P phases show no such mixing [36-38]. Initial assignment of Si/P is done based on single crystal X-ray diffraction. High quality datasets allow to refine P/Si. We have utilized ^{31}P and ^{29}Si solid state NMR spectroscopy to verify Si/P assignments [36-38]. We hypothesized that the ML-guided DFT predictions for ordered La-Si-P system are reliable.

Phonon calculations and MD simulations show that the La_5SiP_3 and La_2SiP compounds obtained from our prediction are dynamically stable with no imaginary vibration frequencies. Moreover, we calculated the Gibbs free energy including the effects of vibration entropy for the competing phases. The calculation results show that formation Gibbs free-energies of the La_5SiP_3 and La_2SiP phases with respect to the ternary convex hull increase monotonically from 0 K to 1000 K. However, the changes are very moderate from 3.6 to 8.4 meV/atom for the La_5SiP_3 phase and 9.6

to 16.2 meV/atom for the La_2SiP phase. These results suggest the La_5SiP_3 and La_2SiP are still thermodynamically competitive metastable phases even at as high as 1000 K. Our present manuscript is not dealing with predicting synthesizability, but further quantitative analysis of kinetics of the phase nucleation and growth are important for understanding the synthesizability which will be the subject of consequence studies. Our synthetic attempts also show that new methods of synthesis to overcome kinetic stability of LaP intermediate need to be developed.

By substituting Si with heavier group 14 elements *X* (*X* = Ge, Sn, Pb) in the 8 lowest-energy La-Si-P structures obtained from our ML-guided DFT calculations, we also predict a number of La-*X*-P compounds with formation energies either below or only slightly above the currently known convex hulls. Our studies demonstrated that combination ML with first-principles calculations is a powerful approach to accelerate the exploration of crystal structures relative energetic stabilities for complex compounds. Such a ML-guided and data science-based approach represents a promising paradigm for efficient design and discovery of novel materials in the digital era.

It is worthwhile to note that while CGCNN model trained using a general dataset from commonly available database (e.g., MP database) would be a reasonably good start for the structure prediction as shown in the present paper, the accuracy of the general CGCNN model can be adaptively improved for specific systems of interest once more structures for the specific systems are available to use in further training the CGCNN model. Such an adaptive improvement of the accuracy of the CGCNN model would be an interesting topic of further studies.

ASSOCIATED CONTENT

Supporting Information Available:

Supporting Information include more detail descriptions of First-principles density functional theory (DFT) calculation and Experimental synthesis process. The structures of the other 4 low-energy La-Si-P compounds obtained from our ML-guided ab initio calculation prediction are shown in **Fig. S1**. The partial pair correlation functions and the time evolution of potential energies of La_5SiP_3 and La_2SiP compounds at T=1300K, as well as the snapshot structures of La_5SiP_3 (4×1×3 supercell, 216 atoms) and La_2SiP (4×2×4 supercell, 256 atoms) at simulation time of t=20ps from AIMD simulations are shown in **Fig. S2** and **Fig. S3** respectively. The lattice parameters and atomic positions of the 8 lowest-energy structures of La-Si-P ternary compounds obtained from our ML-guided ab initio predictions are shown in **Table S1 and Table S2**. The cohesive energies (E_0), formation energies (E_f) and the formation energies above the convex hull (E_d) of the 130 metastable La-Si-P ternary compounds obtained from our ML-guided ab initio calculation predictions are shown in **Table S3**. The lattice parameters and atomic positions of La-*X*-P (*X* = Ge, Sn, Pb) ternary compounds by substituting the Si atoms in the 8 lowest-energy La-Si-P compounds with *X* element are shown in **Tables S4-S6** respectively..

This information is available free of charge at the website: http://pubs.acs.org/

AUTHOR INFORMATION

Corresponding Author

Cai-Zhuang Wang － Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States;
Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, United States
Email: wangcz@ameslab.gov

Authors

Huaijun Sun － Jiyang College of Zhejiang Agriculture and Forestry University, Zhuji, 311800, China;
Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States

Chao Zhang － Department of Physics, Yantai University, Yantai 264005, China

Weiyi Xia －Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States;
Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, United States

Ling Tang －Department of Applied Physics, College of Science, Zhejiang University of Technology, Hangzhou, 310023, China;
Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States

Renhai Wang － School of Physics and Optoelectronic Engineering, Guangdong University of Technology, Guangzhou 510006, China

Georgiy Akopov － Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States;
Department of Chemistry, Iowa State University, Ames, IA 50011, United States

Nethmi W. Hewage － Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States;
Department of Chemistry, Iowa State University, Ames, IA 50011, United States

Kai-Ming Ho － Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States;
Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, United States

Kirill Kovnir － Ames Laboratory, U.S. Department of Energy, Ames, IA 50011, United States;
Department of Chemistry, Iowa State University, Ames, IA 50011, United States

Author Contributions

H.J. Sun, C. Zhang, W.Y. Xia, R.H. Wang and C.Z. Wang performed machine learning predictions and first-principles calculations. L. Tang and C.Z. Wang performed the molecular dynamics simulations to study the stability of the compounds at finite temperature. G. Akopov, N.W. Hewage, and K. Kovnir performed the experimental synthesis and characterization. C.Z. Wang, K.M. Ho, and K. Kovnir conceptualized and planned the research and C.Z. Wang coordinated the research. All authors contributed to the discussions, analyses of the data, and writing of the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Prof. V. Pecharsky for access to the arc-melting setup for our sample preparation. Work at Ames Laboratory was supported by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Science and Engineering Division including a grant of computer time at the National Energy Research Scientific Computing Centre (NERSC) in Berkeley, CA. Ames Laboratory is operated for the U.S. DOE by Iowa State University under Contract No. DE-AC02-07CH11358. G.A. is grateful to the Ames Laboratory Spedding Postdoctoral Fellowship for financial support. K. Kovnir was also supported by the Ames Laboratory's Laboratory Directed Research and Development (LDRD) program. C. Zhang acknowledges financial support from the National Natural Science Foundation of China (Grant No. 11874318).

REFERENCES

(1) Bauer E.; Sigrist, M. (Eds.) Non-centrosymmetric superconductors: introduction and overview (Springer, Berlin, 2012).

(2) Carnicom E. M.; Xie W.; Klimczuk T.; Lin J.; Górnicka K.; Sobczak Z.; Ong N. P.; Cava R. J. $TaRh_2B_2$ and $NbRh_2B_2$: Superconductors with a chiral noncentrosymmetric crystal structure. *Sci. Adv.* **2018***,* 4(5), eaar7969.

(3) Smidman M.; Salamon M. B.; Yuan H. Q.; Agterberg D. F. Superconductivity and spin–orbit coupling in non-centrosymmetric materials: a review. *Rep. Prog. Phys.* **2017**, 80, 036501.

(4) Kneidinger F.; Bauer E.; Zeiringer L.; Rogl P.; Blaas-Schenner C.; Reith D.; Podloucky R. Superconductivity in non-centrosymmetric materials. *Phys. C*, **2015**, 514, 388-398.

(5) Lee, K.; Synnestvedt, S.; Bellard, M.; Kovnir, K. GeP and $(Ge_{1-x}Sn_x)(P_{1-y}Ge_y)$ (x=0.12, y=0.05): Synthesis, structure, and properties of two-dimensional layered tetrel phosphides. *J. Solid State Chem.* **2015,** 224, 62-70.

(6) Wadsten, T.; Jerslev, B.; Cyvin, S.J.; Brunvoll, J.; Dumanovi, J. Synthesis of a Pyrite-Type Modification of SiP_2. *Acta Chem. Scand.* **1967,** 21, 1374-1376.

(7) Mentzen, B.F.; Hillel, R.; Michaelides, A.; Tranquard, A.; Bouix, J. Study on the structure of cadmium arsenide. *C. R. Acad. Sci.* **1981,** 293, 965-967.

(8) Wadsten, T.; Vikan, M.; Krohn, C.; Nilsson, K.; Dumanovi, J. The crystal structures of SiP_2, $SiAs_2$, and GeP. *Acta Chem. Scand.* **1967,** 21, 593-594.

(9) Groom, C.R.; Bruno, I.J.; Lightfoota, M.P.; Warda, S.C. The Cambridge Structural Database. *Acta Crystallogr B*. **2016,** 72(2), 171-179.

(10) Mark, J.; McBride, B.C.; Lee, S.; Yox, P.; Kovnir, K. Synthesis, crystal growth, and transport properties of van-der-Waals tetrel pnictide GeAs2. *ACS Appl. Energy Mater.* **2020,** 3, 4168-4172.

(11) Lee, S.; Owens-Baird, B.; Kovnir, K. Aliovalent substitutions of the 2D layered semiconductor GeAs. *J. Solid State Chem.* **2019,** 276, 361-367.

(12) Guo, J.; Liu, Y.; Ma, Y.; Zhu, E.; Lee, S.; Lu, Z.; Zhao, Z.; Xu, C.; Lee, S.J.; Wu, H.; Kovnir, K.; Huang,Y.; Duan, X. Few-Layer GeAs Field-Effect Transistors and Infrared Photodetectors. *Adv. Mater.* **2018,** 30(21), 1705934.

(13) Lee, K.; Kamali, S.; Ericsson, T.; Bellard, M.; Kovnir, K. GeAs: Highly anisotropic van der Waals thermoelectric material. *Chem. Mater.* **2016,** 28, 2776-2785.

(14) Koster, K.G.; Wang, Y.; Scudder, M.R.; Moore, C.E.; Windl, W.; Goldberger, J.E. Synthesis and characterization of a new family of layered PbxSn4-xAs3 alloys. *Mater. Chem. C*, 9, **2021,** 6477-6483.

(15) Jain, A.; Ong, S.P.; Hautier, G. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013,** 1(1), 011002. Materials Project https://materialsproject.org/ .

(16) Kaiser, P.; Jeitschko, W. The rare earth silicon phosphides $LnSi_2P_6$ (Ln = La, Ce, Pr, and Nd). *J. Solid State Chem.* **1996,** 124(2), 346-352.

(17) Akopov, G.; Mark, J.; Viswanathan, G.; Lee, S.J.; McBride, B.C.; Won, J.; Perras, F.A.; Paterson, A.L.; Yuan, B.; Sen, S.; Adeyemi, A.N.; Zhang, F.; Wang, C.Z.; Ho, K.M.; Miller, G.J.; Kovnir, K. Third time's the charm: intricate non-centrosymmetric polymorphism in LnSiP(3) (Ln = La and Ce) induced by distortions of phosphorus square layers. *Dalton T.* **2021,** 50(19), 6463-6476.

(18) Sun, Y.S.; Chen, J.D.; Yang, S.D.; Li, B.X.; Chai, G.L.; Lin, C.S.; Luo, M.; Ye, N. $LaSiP_3$ and $LaSi_2P_6$: Two Excellent Rare-Earth Pnictides with Strong SHG Responses as Mid- and Far-Infrared Nonlinear Optical Crystals. *Adv. Opt. Mater.* **2021,** 9(10), 2002176.

(19) Kovnir, K. Predictive Synthesis. *Chem. Mater.* **2021,** 33(13), 4835-4841.

(20) Oganov, A.R.; Glass, C.W. Evolutionary crystal structure prediction as a tool in materials design. *J. Phys.: Condens. Matter* **2008,** 20(6), 064210.

(21) Lyakhov, A.O.; Oganov, A.R.; Stokes, H.; Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comp. Phys. Comm.* **2013,** 184, 1172-1182.

(22) Ji, M.; Umemoto, K.; Wang, C.Z.; Ho, K.M.; Wentzcovitch, R.M. Ultrahigh-pressure phases of H_2O ice predicted using an adaptive genetic algorithm. *Phys. Rev. B* **2011,** 84, 220105.

(23) Wu, S.Q.; Umemoto, K.; Ji, M.; Wang, C.Z.; Ho, K.M.; Wentzcovitch, R.M. Identification of post-pyrite phase transitions in SiO_2 by a genetic algorithm. *Phys. Rev. B* **2011,** 83, 184102.

(24) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B* **2010,** 82, 094116.

(25) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. CALYPSO: a method for crystal structure prediction. *Comp. Phys. Commun.* **2012,** 183(10), 2063-2070.

(26) Zhao, X.; Nguyen, M.C.; Zhang, W.Y.; Wang, C.Z.; Kramer, M.J.; Sellmyer, D.J.; Li, X.Z.; Zhang, F.; Ke, L.Q.; Antropov, V.P.; Ho, K.M. Exploring the structural complexity of intermetallic compounds by an adaptive genetic algorithm. *Phys. Rev. Lett.* **2014,** 112, 045502.

(27) Wu, S.Q.; Ji, M.; Wang, C.Z.; Nguyen, M.C.; Zhao, X.; Umemoto, K.; Wentzcovitch, R.M.; Ho, K.M. Adaptive Genetic Algorithm for Crystal Structure Prediction. *J. Phys.: Condens. Matter* 26, **2014,** 035402.

(28) Zhao, X.; Shu, Q.; Nguyen, M.C.; Wang, Y.; Ji, M.; Xiang, H.; Ho, K.M.; Gong, X.; Wang, C.Z. Interface structure prediction from first-principles. *J. Phys. Chem. C* **2014,** 118, 9524-9530.

(29) Akopov, G.; Viswanathan, G.; Kovnir, K. Synthesis, Crystal and Electronic Structure of La_2SiP_4. *ZAAC* **2021,** 647(2-3), 91-97.

(30) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018,** 120, 145301.

(31) Togo, A.; Tanaka, I. First principles phonon calculations in materials science. *Scripta Mater.* **2015,** 108, 1-5.

(32) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995,** 117(1), 1-19.

(33) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984,** 81, 511-519.

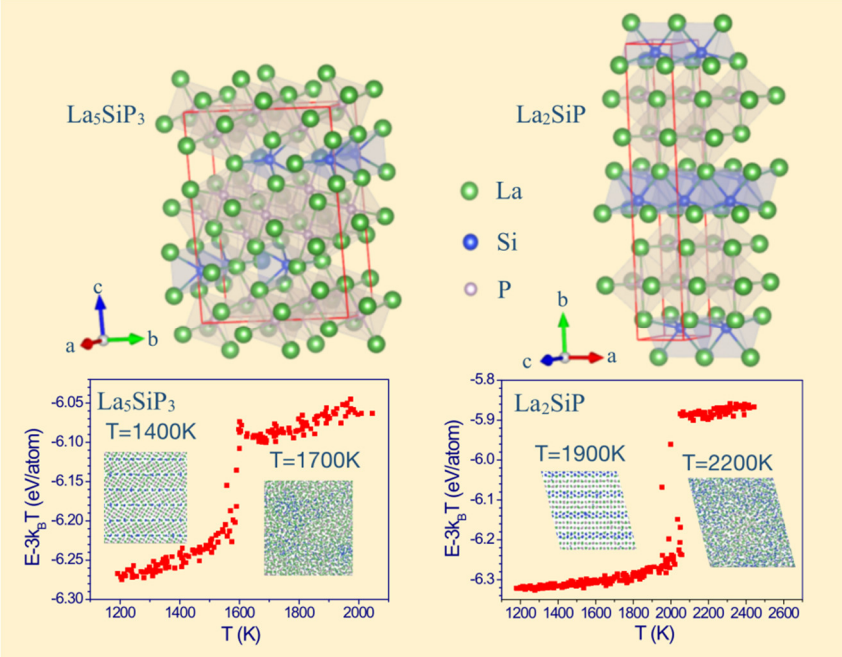
(34) Hoover, W.G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985,** 31(3), 1695-1697.

(35) Togo, A.; Tanaka, I. First Principles Phonon Calculations in Materials Science. *Scripta Mater.* **2015**, 108, 1-5.

(36) Yox, P.; Porter, A.P.; Dorn, R.W.; Kyveryga, V.; Rossini, A.; Kovnir, K. Semiconducting silicon-phosphorous frameworks for caging exotic polycations. *Chem. Commun.* **2022**, 58, 7622-7625.

(37) Lee, S.J.; Viswanathan, G.;Carnahan, S.C.; Harmer, C.P.; Akopov, G.; Rossini, A.J.; Miller, G.J.; Kovnir, K. Add a Pinch of Tetrel: The Transformation of a Centrosymmetric Metal into a Nonsymmorphic and Chiral Semiconductor. *Chem. Europ. J.* **2022**, 28, e202104319.

(38) Lee, S.; Carnahan, S.; Akopov, G.; Yox, P.; Wang, L.-L.; Rossini, A.; Wu, K.; Kovnir, K. Noncentrosymmetric Tetrel Pnictides $RuSi_4P_4$ and $IrSi_3P_3$: Nonlinear Optical Materials with Outstanding Laser Damage Threshold. *Adv. Funct. Mater.* **2021**, 2010293.



Synopsis

We integrate machine-learning with first-principles calculations to efficiently search for novel La-Si-P ternary compounds over a wide composition range. 130 La-rich compounds within 100 meV/atom from the convex hull are discovered. In particular, La_5SiP_3 and La_2SiP phases are only 2 and 10 meV/atom respectively above the convex hull. These two structures are dynamically stable at least up to 1000 K and would be synthesizable by experiment.