THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL Letter

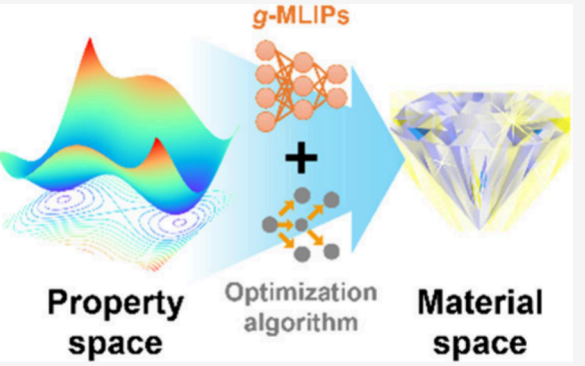
De Novo Inverse Design Superhard C−N Compounds via Global Machine Learning Interatomic Potentials and Multiobjective Optimization Algorithm

Guanjian Cheng and Wan-Jian Yin*

Cite This: J. Phys. Chem. Lett. 2025, 16, 4392−4400 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A major challenge in the field of superhard materials is the identification of compounds with a hardness exceeding that of diamond. In this study, we developed a variable-composition inverse material design (VC-IMD) approach for designing C−N superhard materials. In this approach, an improved multiobjective optimization algorithm is introduced, utilizing structure similarity constraint to prevent convergence toward local minima. Combined with active learning, it trains global machine learning interatomic potentials (*g*-MLIPs) while exploring target materials. By comparing several *g*-MLIPs and selecting the best, the resulting *g*-MLIPs achieved reasonable precision within three iterations. Through multiple searches, 38 novel and stable C−N superhard materials not present in major computational materials databases were identified. Notably, the material $C_3(P6_422)$ with a hardness of 97.4 GPa was discovered, potentially exceeding that of diamond (94.0 GPa). This approach provided a new pathway for materials design with target properties.



A holy grail in the field of superhard materials is the search for compounds that exceed the hardness of diamond.[1,2] Superhard materials with excellent Vickers hardness (over 40 GPa)[3] are excellent in abrasion resistance and ability to withstand extreme conditions. Diamond, one of the hardest known materials, has a Vickers hardness of about 100 GPa and is therefore widely used in industrial operations.[4−8] However, diamond is unstable at moderate temperatures and in strong oxidizing environments, as well as reacting readily with iron-containing materials,[9] prompting scientists to search actively for other superhard materials with excellent mechanical properties and chemical stability. In recent years, C−N compounds have become one of the popular research areas due to their potential high hardness and favorable chemical stability.[9−15] Exploring new C−N superhard compounds would be expected to overcome the limitations of diamond and develop wider applications.

In the design of C−N superhard materials, the variability of composition offers opportunities to explore new material families with superior properties. In the early stages, traditional methods for addressing variable-composition problems typically relied on element substitution based on prototype structures[15,16] or crystal structure prediction (CSP).[9,12,17] The former is inefficient and heavily dependent on the experience and intuition of scientists, while the latter involves searching for materials by repeatedly altering compositions, making it challenging to predict complex compositions. With the development of computer technology and computational methods, de novo inverse materials design has become feasible. However, de novo inverse design of C−N superhard compounds requires addressing two key problems: (i) the variable-composition properties, i.e., the variable number of C and N atoms; and (ii) the multiple target properties, such as bulk modulus, shear modulus, and so on. The variable-composition problem is a challenge for optimization algorithms (OA), because it involves continuous-discrete combination optimization and variable-dimension parameter space. Therefore, for this problem, materials are usually designed by changing the composition several times in combination with CSP methods, i.e., variable-composition CSP[11,18−20] However, as the number of atoms increases, the number of searches and computational complexity will increase exponentially. Moreover, these methods are usually based on density functional theory (DFT) calculations.

Fortunately, recent advances and breakthroughs in machine learning interatomic potentials (MLIPs) have significantly impacted in various computational materials fields.[21−25] Using

Received: January 19, 2025
Revised: April 19, 2025
Accepted: April 22, 2025
Published: April 24, 2025

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400

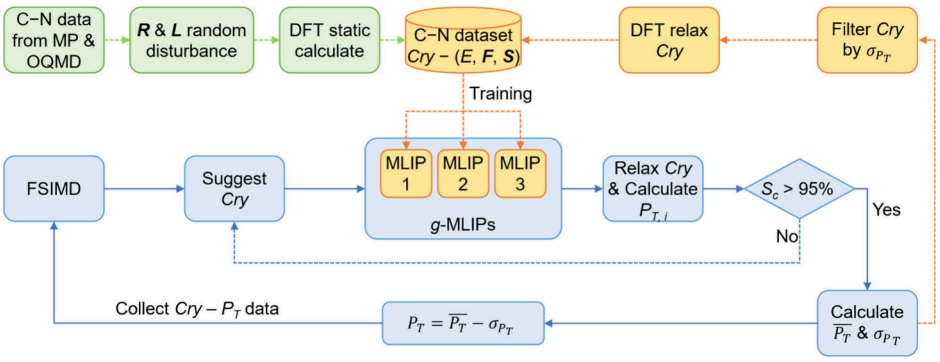


Figure 1. The VC-IMD workflow. Green blocks are for initial dataset generation flow, blue blocks are for VC-IMD material search, and yellow blocks are for *g*-MLIPs update iteration. Here, $\boldsymbol{R}$ represents atomic coordinates, $\boldsymbol{L}$ represents lattice parameters, $P_{T,i}$ is the target property predicted by the *i*th *g*-MLIP, $\overline{P}_T$ and σ_{P_T} are the mean and standard deviation of the $P_{T,i}$, S_c is the structural similarity.

MLIPs, it is possible to quickly relax crystal structure and calculate the stiffness tensors, thus allowing rapid prediction of various elastic constants of the material. Notably, training MLIPs is generally much less costly than training property prediction models, as a large training dataset can be obtained through DFT relaxations or a short time for molecular dynamics simulations. The state-of-the-art MLIPs can achieve near-DFT accuracy (errors within a few meV).[22,26−28] However, most of the current studies of MLIPs are based on specific prototype structures or several phases of specific compositions,[29,30] which here is called local MLIPs (l-MLIPs), and these methods are difficult to use for variable-composition materials design. For the problem of variable-composition materials design under specific elements, we need a global MLIPs (g-MLIPs),[31] that is applicable to the whole composition space. Of course, there are many pre-trained universal MLIPs (u-MLIPs) have been developed, e.g., M3GNet,[32] CHGNet,[33] ALIGNN-FF,[34] DPA-2,[35] etc. However, these methods typically exhibit lower accuracy in specific domains; they are difficult to be used directly, and they require fine-tuning through transfer learning methods before they can be used. They need to be fine-tuned using the transfer learning method before being used, and it is important to note that the DFT data used for fine-tuning should be generated with the same exchange-correlation functional as the original training data.

In this study, we develop a general variable-composition inverse material design (VC-IMD) method, which is based on our previous development of the full space inverse material design (FSIMD) method.[36] The flow of VC-IMD is shown in Figure 1, which consists of three main steps: (i) initial dataset generation (green-block flow), (ii) VC-IMD material search (blue-block flow), and (iii) *g*-MLIPs update iteration (yellow-block flow). First, following the green block flow, the crystal structures of C−N system compounds were collected from the existing databases. And after perturbing these structures, an initial dataset was generated by DFT static calculations, which was used to train three initial *g*-MLIPs of the same neural network architecture. Next, these three *g*-MLIPs are employed as calculators for crystal structure optimization and stiffness tensor calculations, using a customed multiobjective optimization algorithm to search the target materials via the VC-IMD framework. After one round of searching is completed, the crystal structures with large errors are filtered out based on the uncertainty ranking of the predicted target properties by *g*-MLIPs. Finally, these crystal structures undergo structure optimization by DFT calculations, and the data from the optimization trajectories are added to the dataset for a new round of *g*-MLIPs training, and next followed by repeating last two steps.

VC-IMD is developed based on the FSIMD approach. In FSIMD, a crystal structure (Cry) with N_{comp} compositions is represented as Cry = ($\boldsymbol{a}$, $\boldsymbol{c}$, $\boldsymbol{s}$), where $\boldsymbol{a}$ denotes the atomic type, $\boldsymbol{a}$ = (a_1,...,$a_{N_{\text{comp}}}$) $\in \mathbb{A}^{N_{\text{comp}}}$ (where $\mathbb{A}$ denotes the set of all chemical elements); $\boldsymbol{c}$ denotes the chemical stoichiometry (the number of atoms for each a_i), $\boldsymbol{c}$ = (c_1,...,$c_{N_{\text{comp}}}$) $\in \mathbb{N}^{N_{\text{comp}}}$; and $\boldsymbol{s}$ refers to the structural information, $\boldsymbol{s}$ = ($\boldsymbol{R} \in \mathbb{R}^{3\times\sum_i a_i c_i}$, $\boldsymbol{L} \in \mathbb{R}^{3\times3}$) (where $\boldsymbol{R}$ denotes atomic coordinates and $\boldsymbol{L}$ represents lattice parameters). The optimization algorithm searches for the Cry that yields the target property (P_T) by optimizing ($\boldsymbol{a}$, $\boldsymbol{c}$, $\boldsymbol{s}$), i.e.,

$$\text{Cry} = \arg\max_{\text{Cry}} P_T(\text{Cry}) = \arg\max_{\boldsymbol{a},\boldsymbol{c},\boldsymbol{s}} P_T(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{s}) \quad (1)$$

In this work, the goal is to design superhard materials in the C−N system. Therefore, $\boldsymbol{a}$ is a constant value, i.e., $\boldsymbol{a}$ = (C, N); the parameters $\boldsymbol{c}$ and $\boldsymbol{s}$ are then optimized by maximizing the target P_T, i.e.,

$$\text{Cry} = \arg\max_{\text{Cry}} P_T(\text{Cry}|\boldsymbol{a}) = \arg\max_{\boldsymbol{c},\boldsymbol{s}} P_T(\boldsymbol{c}, \boldsymbol{s}) \quad (2)$$

However, there are two main challenges in this optimization problem. First, given that $\boldsymbol{c}$ is a discrete variable and $\boldsymbol{s}$ is a continuous variable, it makes the optimization become a difficult problem, called combinatorial optimization.[37] The second challenge is due to the dynamic change in the dimension of the parameter space for $\boldsymbol{s}$, which is caused by variations in $\boldsymbol{c}$ and, consequently, the total number of atoms $N = \sum_i c_i$, making it difficult for traditional optimization algorithms to handle. In our method, for the combinatorial optimization challenge, we convert the discrete variable $\boldsymbol{c}$ into a continuous variable before putting it into the OA. After the optimization process, the continuous variable is converted back into a discrete value, thus resolving the combinatorial optimization challenge. To address the problem of variable-dimension parameter space, we impose a maximum atom number limit, N_{max}, ensuring that the sum of atomic counts $N = \sum_i c_i \le N_{\text{max}}$ for any $\boldsymbol{c}$ in its variation range. The optimization algorithm then operates on the N_{max} atomic coordinates and selects the top N of them to construct the crystal structure. These ways effectively resolve the issue of combinatorial optimization and variable-dimension parameter space, enabling the OA to be used for inverse materials design with variable composition.

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400

To further enhance the efficiency of the OA, we introduced symmetry constraint[17,38] to optimize the structure generation process. This constraint ensures that the generated crystal structures conform to the physical symmetry requirements, thereby reducing the production of invalid structures and improving search efficiency. Specifically, we incorporated space group $G \in \{1, 2, ..., 230\}$ and Wyckoff position $\boldsymbol{W} = (\{w_1, .., w_i\}|G)$ variables into the structure $\boldsymbol{s}$, represented as $\boldsymbol{s} = (\boldsymbol{R}|G;\boldsymbol{W}, \boldsymbol{L}|G)$, and optimized G and $\boldsymbol{W}$ using the OA (as shown in eq 3).

$$\mathrm{Cry} = \underset{\boldsymbol{c},\boldsymbol{s}}{\operatorname{argmax}}\, P_{\mathrm{T}}(\boldsymbol{c}, \boldsymbol{s}) = \underset{\boldsymbol{c},\boldsymbol{R},G,\boldsymbol{W}}{\operatorname{argmax}}\, P_{\mathrm{T}}(\boldsymbol{c}, \boldsymbol{R}|G; \boldsymbol{W}, \boldsymbol{L}|G) \tag{3}$$

Meanwhile, the MLIPs learn interatomic interactions by training models, which significantly reduces the computational cost while ensuring high accuracy.[39−41] Compared to DFT calculations, MLIPs greatly improve the speed of structure optimization and property prediction and thus have been widely applied in materials science in recent years. In this study, *g*-MLIPs are the core component of the VC-IMD framework. Not only do they need to achieve an accuracy comparable to *l*-MLIPs, but they must also cover the full potential energy surface of the system, similar to *u*-MLIPs, but *u*-MLIPs are typically applicable to a broader range of elements that are less accurate. The goal of MLIPs is to learn the energy $E = \mathrm{MLIP}(\mathrm{Cry})$ and use the autodifferentiation capabilities of neural networks to compute the forces $\boldsymbol{F} = -\partial E/\partial \mathbf{R}$ and stresses $\boldsymbol{S} = V^{-1}\partial E/\partial \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is the lattice strain tensor. Different MLIPs differ in the way they characterize the crystal structure and the neural network frameworks that they employ. Some MLIPs may focus on a detailed description of interatomic distances or local environments,[21,42] while others may use more complex graph neural networks to capture the overall features of the crystal structure.[27,43] Furthermore, the neural framework chosen also influences how MLIPs perform in different systems. For example, some frameworks are better suited to simpler binary or ternary systems,[44] while others can handle more complex multielement systems.[32] These differences mean that each MLIP has its own strengths and weaknesses, making it suitable for different material design tasks.

To further explore the performance of different MLIP frameworks in the design of C−N system superhard materials, we selected several representative MLIPs, including NEP,[21] M3GNet,[32] CHGNet,[33] and MACE.[45] NEP utilizes simple local descriptors, while M3GNet and CHGNet employ graph neural networks to capture many-body interactions. In contrast, MACE incorporates equivariant message-passing[46] and higher-order interactions, further enhancing model performance. Moreover, it is worth mentioning that, as shown in Figure 1, there are two reasons for training the same three *g*-MLIPs simultaneously: first, by assessing the uncertainty in the prediction differences between multiple *g*-MLIPs for a structure, it can be determined whether the structure should be sampled for DFT calculations and used in the next round of *g*-MLIPs training; second, to judge the confidence of the prediction properties by the multiple prediction results. By using the three *g*-MLIPs, three predicted properties $P_{\mathrm{T},i}$ are obtained, from which the mean $(\overline{P}_{\mathrm{T}})$ and standard deviation $(\sigma_{P_{\mathrm{T}}})$ are calculated. The lower confidence bound of one standard deviation is then used as the final target property P_{T} (i.e., $P_{\mathrm{T}} = \overline{P}_{\mathrm{T}} - \sigma_{P_{\mathrm{T}}}$)) for OA, which helps to ensure the reliability of the prediction results.

Mechanical properties are the key to determining whether a material is a superhard material or not. By calculating the energy−volume (*E*-*V*) curve of a material and fitting its equation of state (EOS),[47] the bulk modulus *B* can be quickly estimated. However, the bulk modulus is only one component of hardness; material hardness typically requires the support of other mechanical properties, such as the shear modulus. To calculate these mechanical properties, it is necessary to use more advanced calculation methods. The energy−strain (*E*−*S*) relationship[48] is a common method for calculating the elastic constants of materials. By applying a series of small strains to a material and calculating the total energy under each strain, the *E*−*S* relationship can be constructed, from which the stiffness tensor is obtained. The components of the stiffness tensor correspond to the material's elastic response under different strain states, allowing for the calculation of elastic constants, including bulk modulus (*B*), shear modulus (*G*), Young's modulus (*Y*), and Poisson's ratio (*υ*). These mechanical properties are fundamental for calculating the Vickers hardness. In this study, the empirical formula[49] used for calculating Vickers hardness is as follows:

$$H_{\mathrm{v}} = 2(k^2G)^{0.585} - 3 \tag{4}$$

where *k* is Pugh's ratio ($k = G/B$). These methods for calculating mechanical properties are available via MLIPs calculations in this work, and the pre-processing and post-processing of the calculations are realized via VASPKIT.[48]

In superhard materials design with light elements (such as B, C, N, O), there is a certain correlation between a material's density and its hardness; in general, higher density corresponds to greater hardness.[50] Accordingly, before calculating the hardness of a material, the density is a crucial indicator. In this study, we used density along with a mechanical property P_m ($P_{\mathrm{m}} \in \{B,G,k,Y,v,H_v\}$) as the target property in the MOO algorithm for material search, i.e., $P_{\mathrm{T}} = (\rho, P_m)$.

Subsequently, following the workflow in Figure 1, the first step involves training three identical MLIPs. For training the MLIPs, a high-quality dataset is crucial in the training of MLIPs.[22] Since of the scarcity of C−N system data in the existing databases, the available data cannot satisfy the large amount of data required for MLIPs training, we have to expand the dataset. Therefore, we needed to expand the dataset to ensure effective MLIPs training. Here, we employed the *ab initio* approach, generating an initial dataset covering a wide range of C−N configurations based on DFT calculations.

We extracted known crystal structures containing carbon and nitrogen elements from the Materials Project (MP)[51] and the Open Quantum Materials Database (OQMD),[52] including 208 structures with elementary substances and binary compounds. Due to the limited amount of directly available data, we expanded these structures to generate more diverse structures based on the following three ways.

(i) *Atomic coordinate perturbation*. For each structure, all atoms were randomly perturbed in three directions with a perturbation amplitude of −0.5 to 0.5 Å and an interval of 0.1 Å. Ten new structures can be generated for each structure, for a total of 2080 new structures.

(ii) *Lattice parameters perturbation*. For each structure, the lattices were scaled by random perturbation with a perturbation amplitude of −10% to 10% and interval of 2%. Ten new structures can be generated for each structure, for a total of 2080 new structures.

(iii) *Random generation structures*. Randomly generating the number of atoms, atomic coordinates, and lattices, and

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400

ensuring that the atomic distance $d \geq 1.0$ Å and the cell volume $V \leq 5 \times \sum_i V_a^i$ (where V_a^i is the ith atomic volume in a structure), generating 500 structures with 100 carbon elementary substances, 100 nitrogen elementary substances, and 300 C_xN_y binary compounds.

Through these expansion methods, 4868 C−N crystal structures were obtained. Following DFT static calculations on these structures, successfully yielding energies (E), forces ($\boldsymbol{F}$), and stresses ($\boldsymbol{S}$) data for 4560 crystal structures Cry. To ensure data quality, we filtered the data with energy $E < 0$ eV and forces -5 eV/Å $\leq \boldsymbol{F} \leq 5$ eV/Å, which finally generated a dataset containing 2449 Cry $-$ (E, $\boldsymbol{F}$, $\boldsymbol{S}$) data items, and the distribution of the dataset is shown in Figure S1. These high-quality data as the initial dataset were divided into training and validation sets in the ratio of 90% and 10%, which were used in the subsequent training of the g-MLIPs. It is important to note that the dataset, although expanded, is still relatively poor for training g-MLIPs. To ensure the generalization of the model, the test dataset is not divided here, as the test dataset will be generated from new structures identified during the material design process by the VC-IMD framework.

Then, we trained these MLIPs on the initial dataset and evaluated their performance on out-of-distribution (OOD) data. The OOD data refer to samples outside the training set, and testing on such data effectively examines the generalization ability of the model. Here, we utilized the initial dataset to train NEP, M3GNet, CHGNet and MACE models, with three identical models trained for each MLIP. Subsequently, using each of the four models, a single round of superhard material search was conducted with the target properties $PT = (\rho, H_v)$, resulting in the construction of an OOD dataset. As shown in Figure 2a, the mean absolute errors (MAEs) for energy and forces are as follows: NEP (MAE_E = 324.6 meV/atom, MAE_F = 1117.7 meV/Å), M3GNet (MAE_E = 2745.9 meV/atom, MAE_F = 683.7 meV/Å), CHGNet (MAE_E = 277.5 meV/atom, MAE_F = 1249.3 meV/Å), and MACE (MAE_E = 212.6 meV/atom, MAE_F = 754.6 meV/Å). The experimental results showed that while NEP, M3GNet and CHGNet performed well on the training set, their accuracy decreases on the OOD data. In contrast, MACE showed better performance across all test data, with smaller errors and greater stability. MACE's high accuracy and strong generalization performance make it an ideal tool for dealing with complex material design problems. Based on these test results, we finally selected MACE as the g-MLIP for the VC-IMD framework in the design of variable-composition C−N system superhard materials.

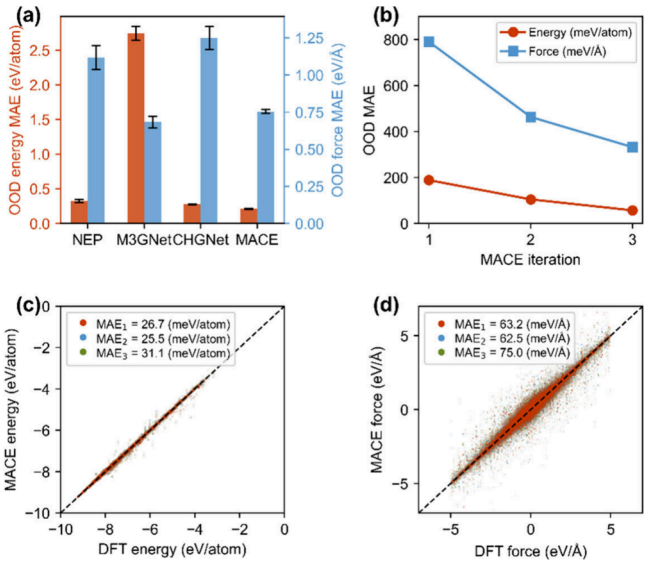


Figure 2. The MLIPs training results. (a) Comparison of errors in the energy and force of OOD dataset for initial trained MLIPs. (b) Energy and force errors on OOD dataset for three iterations of MACE-based MLIPs. (c) Energy errors of MACE-based MLIPs on all datasets. (d) Force errors of MACE-based MLIPs on all datasets.

Meanwhile, during the first round of exploration, it was observed that the optimization algorithm frequently became trapped in local minima, limiting the efficiency and effectiveness of the search process. Therefore, we introduced structural similarity analysis to identify and bypass similar structures that appear during the OA's sampling process. With this approach, the OA can be effectively prevented from trapping in local minima, thus enhancing the global search capability and increasing the diversity of the discovered compounds. The structural similarity is based on X-ray diffraction (XRD) patterns of the crystal structures and is calculated using cosine similarity[53] (eq 5).

$$S_c(\boldsymbol{A}, \boldsymbol{B}) := \cos(\theta) = \frac{\boldsymbol{A}\cdot\boldsymbol{B}}{\|\boldsymbol{A}\|\|\boldsymbol{B}\|} = \frac{\sum_i A_i B_i}{\sqrt{\sum_i A_i^2}\cdot\sqrt{\sum_i B_i^2}} \tag{5}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are the XRD intensity values of any two crystal structures, θ is the angle between $\boldsymbol{A}$ and $\boldsymbol{B}$, and A_i and B_i are the ith values of $\boldsymbol{A}$ and $\boldsymbol{B}$. Since $A_iB_i \geq 0$, $S_c \in [0, 1]$. The closer the S_c value is to 1, the more similar the two structures are. For example, as shown in Figure 3a, the XRD spectra of diamond, cubic C_3N_4 (c-C_3N_4) and β-C_3N_4 are presented, and Figure 3b shows their similarity matrix. We determine whether a sampling instance is accepted by the OA by setting a maximum number of samplings (N_{samp}) for the same structure. If a structure has been sampled N_{samp} times and is sampled again, the OA will skip that sampling. To evaluate the performance of the OA, we run the material searches with and without structural similarity constraints, and the results are shown in Figures 3c and 3d. Without the structural similarity constraint, the OA falls into a local optimum at ∼500 steps, with nearly 40% of the structures being similar within 1000 steps. In contrast, when the structural similarity constraint was applied with setting N_{samp} = 3, the ratio of duplicate structures is significantly reduced to 2%, which shows that the similarity constraint is effective in helping the algorithm to avoid local minima and improve the global search capability.

To improve the prediction accuracy of the g-MLIPs, we have updated the MACE-based g-MLIPs in several iterations. In this study, we conducted three updates of the g-MLIPs from the initial dataset trained g-MLIPs. The datasets for each update were obtained through material searches using the VC-IMD in 1500 steps, with the target property $P_T = (\rho, H_v)$. After each search had been completed, we removed duplicate structures and extracted new structures. These structures were then ranked based on the σ_{P_T}, and those with $\sigma_{P_T} > 5$ GPa were selected for DFT structure optimization. From the structure optimization trajectories, if the number of converged ionic steps $N_{traj} \leq 10$, all were selected; if $N_{traj} > 10$, 10 configurations were uniformly sampled, and then these configurations were added into the training dataset. To accelerate the g-MLIPs training, each update of the models is based on the previous trained models in full,

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400

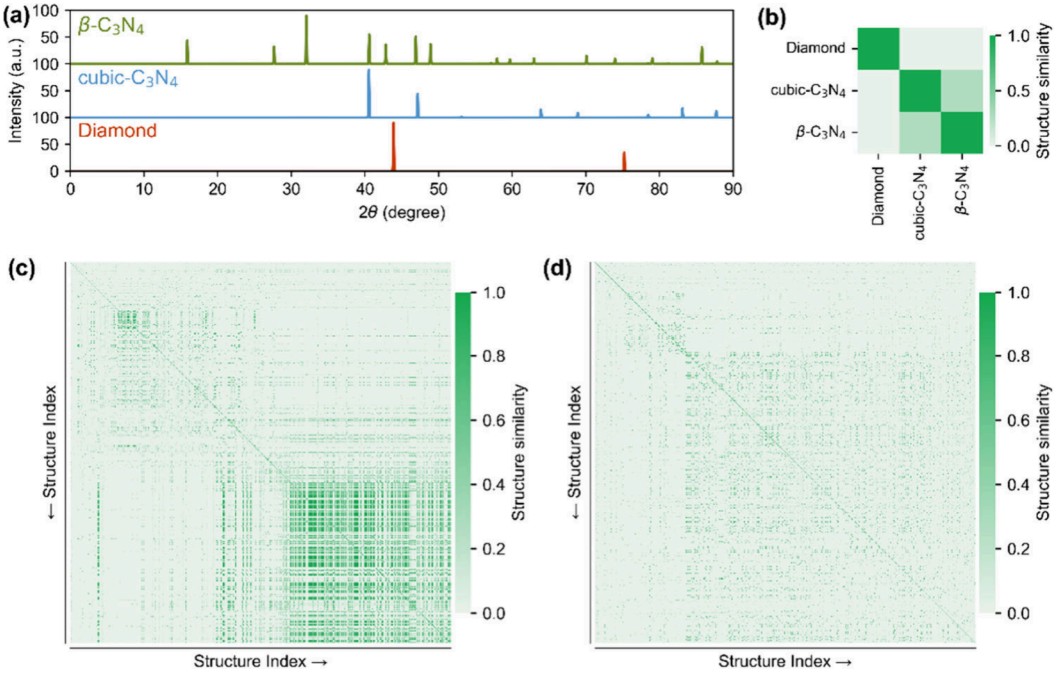


Figure 3. Performance of multiobjective optimization algorithms for structure similarity constraint. (a) XRD spectra of diamond, c-C_3N_4, and β-C_3N_4. (b) The structure similarity matrix of diamond, cubic-C_3N_4, and β-C_3N_4. (c) The structure similarity matrix of all structures in the material search without using structure similarity constraint. (d) The structure similarity matrix of all structures in the material search with using structure similarity constraint.

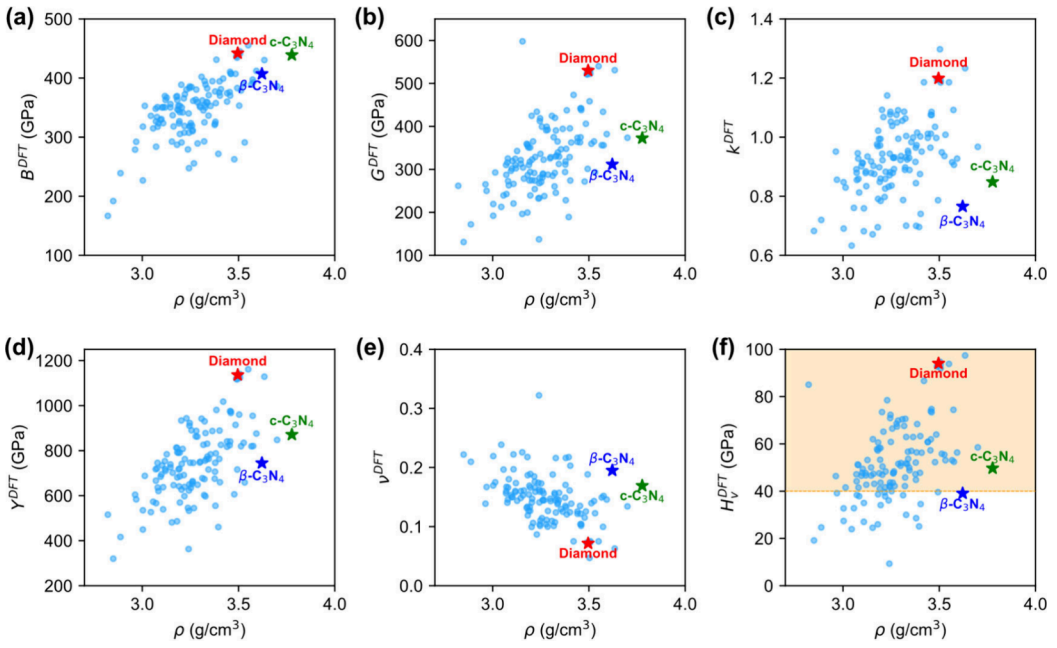


Figure 4. The results of DFT calculations for the candidates: plots of (a) bulk modulus vs density, (b) shear modulus vs density, (c) Pugh's modulus ratio vs density, (d) Young's modulus vs density, (e) Poisson's ratio vs density and (f) Vickers hardness vs density, respectively. The areas with yellow regions are the candidates that were $H_v \geq 40$ GPa in panel (f).

instead of training from scratch. In the three rounds of searches, 514, 608, and 117 new structures with $\sigma_{H_v} > 10$ GPa were generated, respectively. After DFT optimization, 3409, 3937, and 1108 new configurations were obtained. From these calculation results, Cry − (E, $\boldsymbol{F}$, $\boldsymbol{S}$) data were extracted, with their distribution shown in Figures S2−S4. After one initial training and three update trainings, the g-MLIPs underwent four generations of iterative optimization. During these iterations, the prediction error of the g-MLIPs gradually decreased, and the overall performance was continuously optimized.

As shown in Figure 2b, the MAEs of the out-of-distribution predictions for energy in the first three generations were 188.6 104.9, and 57.2 meV/atom, respectively, and the MAEs for force were 790.6 463.4, and 332.2 meV/Å, all showing a decreasing trend. This indicates that the prediction accuracy of the model improved with each update. Finally, in the fourth-generation

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400

Table 1. Top 20 Stable Candidates and Their Mechanical Properties[a])

#	compound	space group	ρ (g/cm^3)	B (GPa)	G (GPa)	k (GPa)	Y (GPa)	ν	H_v (GPa)
–	diamond	$Fd\bar{3}m$	3.50	442.09	530.05	1.20	1136.10	0.07	94.04
–	c-C_3N_4	$I\bar{4}3d$	3.78	439.07	372.58	0.85	871.29	0.17	49.69
–	β-C_3N_4	$P6_3/m$	3.62	407.11	311.71	0.77	744.99	0.20	39.09
1	C_3	$P6_422$	3.63	430.44	**530.89**	**1.23**	1128.66	0.06	**97.40**
2	C_2	$I4_1/amd$	3.55	**455.52**	**540.08**	1.19	**1161.28**	0.08	93.84
3	C_6	$R\bar{3}m$	3.49	434.51	520.76	1.20	1116.32	0.07	93.00
4	C_4	$P6_3/mmc$	3.49	436.95	522.01	1.19	1120.01	0.07	92.77
5	C_2	$Fd\bar{3}m$	3.51	441.04	522.62	1.18	1123.92	0.08	91.93
6	C_6	$P6_522$	3.42	399.16	473.29	1.19	1017.65	0.08	86.65
7	C_6	$Cmcm$	3.23	380.52	434.14	1.14	943.57	0.09	78.48
8	C_{12}	$Ia\bar{3}d$	3.46	410.60	441.73	1.08	975.40	0.10	73.84
9	C_6	$P6_122$	3.20	388.72	424.68	1.09	933.92	0.10	73.45
10	C_6	$Fmmm$	3.25	385.74	418.82	1.09	922.56	0.10	72.29
11	C_6	$C2/m$	3.24	383.08	412.51	1.08	910.66	0.10	70.90
12	C_8	$Cmce$	3.31	402.15	424.48	1.06	942.00	0.11	70.42
13	C_{15}	$P1$	3.29	359.29	392.07	1.09	862.49	0.10	69.86
14	C_4	$I4/mmm$	3.38	431.02	437.52	1.02	980.71	0.12	68.39
15	C_8	$Imma$	3.26	384.83	399.44	1.04	890.28	0.11	66.47
16	C_7N_4	$P\bar{4}$	3.44	360.19	364.28	1.01	817.31	0.12	60.86
17	C_5	$C2$	3.25	319.27	336.03	1.05	746.28	0.11	60.82
18	C_8N_4	$P\bar{1}$	3.44	374.39	373.48	1.00	840.85	0.13	60.77
19	C_5	$C2$	3.18	346.72	354.26	1.02	792.77	0.12	60.58
20	C_3N_4	$P\bar{4}3m$	3.70	386.30	373.67	0.97	847.68	0.13	58.52

[a]For the mechanical properties, bold indicates that the property of the carbon material exceeds that of diamond, while underlined indicates that the mechanical property of the C_xN_y material exceeds that of c-C_3N_4.

model, the errors in the whole dataset for energy and force were reduced to 27.8 meV/atom and 66.9 meV/Å, respectively, as shown in Figures 2c and 2d. These results demonstrate that, through multiple iterative updates, the performance of the g-MLIPs was significantly improved across the entire dataset, with marked enhancements in prediction accuracy and reliability. Meanwhile, as shown in Figure S5, some candidate materials with relatively high bulk moduli were identified during these material searches, indicating that some of these materials may be potential superhard materials.

Through the VC-IMD framework, combined with g-MLIPs and the MOO, we can efficiently work on inverse material design with variable composition. To explore superhard materials in the C−N system, we set multiple target properties $P_T = (\rho, P_m)$ for several rounds of searches. In this study, the choice of P_m is crucial, we selected four mechanical properties closely related to material hardness: B, G, k, and H_v. And these mechanical properties were all calculated by using the E-S method. There were two search flows for each P_m, and all search flows were run for 1000 steps. For selecting excellent candidates, we consolidated all search results, including the data used to train the g-MLIPs, yielding a total of 9944 materials. Six mechanical properties of these searched structures were calculated by g-MLIPs via the E-S method: B, G, k, Y, ν, H_v. For further accurate screening, we focused on structures with a density (ρ) within the range of 3−5 g/cm^3, as there is often a correlation between density and hardness in the design of light-element superhard materials. Based on this density criterion, 634 materials were selected from an initial pool of 9944 materials. The relationship between these mechanical properties and ρ is shown in Figure S6 for assessing the potential properties of the materials more visually. Given that materials with $k > 0.9$ or $H_v > 40$ GPa are widely considered as potential superhard materials,[9,54] to maximize the discovery of candidate materials, we further expanded the screening criteria to $k > 0.8$ or $H_v > 30$ GPa. This extended screening approach increases the number of candidates and prevents more potential candidates are missed. As a result, 154 materials were ultimately selected for further validation.

The screened materials were further validated by DFT calculations with 130 materials successfully passing the validation. The DFT-calculated mechanical properties are presented in Figure 4 and Table S1. These results were compared with those of common materials, such as diamond, c-C_3N_4, and β-C_3N_4, to better assess the potential of the identified materials. Notably, we observed that several materials exceeded the properties of diamond in each of the mechanical property categories (as shown in Figure 4). Among the 130 materials, 100 exhibited $H_v \geq 40$ GPa, classifying them as potential superhard materials. To validate the stability of these superhard materials, we further computed their phonon spectra. This led to the identification of 38 new stable superhard materials. Additionally, we compared the elastic properties (B, G, Y, and H_v) calculated by g-MLIPs with the DFT results for these materials, as shown in Figures S7−S10, further demonstrating the reliability of the g-MLIPs. The mechanical properties of these materials, alongside three common materials (diamond, c-C_3N_4, and β-C_3N_4), are summarized in Table 1, with the complete results available in Table S2. Interestingly, several C_xN_y materials demonstrated superior properties compared to c-C_3N_4. For example, the H_v values of $C_7N_4(P\bar{4})$, $C_8N_4(P\bar{1})$, $C_3N_4(P\bar{4}3m)$ were 60.86, 60.77, and 58.52 GPa, respectively, all of which exceeded that of c-C_3N_4 (49.69 GPa). Figures 5a−c showcases the structures and phonon spectra of three C_xN_y materials with hardnesses exceeding that of c-C_3N_4. These C_xN_y materials, with their higher hardness, demonstrate significant potential as superhard materials. Moreover, a substantial number of carbon materials were found to exhibit superhard characteristics. Several carbon

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400

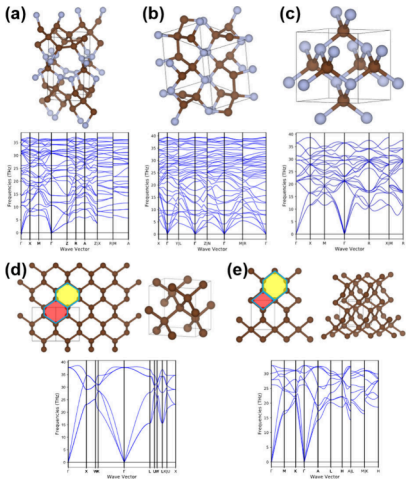


Figure 5. Structures and phonon spectra for the top three C−N compounds, diamond and $C_3(P6_422)$. The structure and phonon spectra of (a) C_7N_4 ($P\bar{4}$, 60.86 GPa), (b) C_8N_4 ($P\bar{1}$, 60.77 GPa), (c) C_3N_4 ($P\bar{4}3m$, 58.52 GPa), (d) diamond ($Fd\bar{3}m$, 94.04 GPa), and (e) C_3 ($P6_422$, 97.40 GPa), respectively.

materials displayed mechanical properties comparable to or even surpassing those of diamond. Among these, $C_3(P6_422)$ stood out as particularly intriguing, because its hardness exceeds that of diamond. Figures 5d and 5e display the structures and phonon spectra of both diamond and $C_3(P6_422)$ for comparison. Our analysis revealed that the arrangement of atoms in $C_3(P6_422)$ is notably more compact, with a density of 3.63 g/cm^3, which could explain its superior H_v over diamond. This discovery highlights the potential of $C_3(P6_422)$ as an extraordinary material, warranting further investigation into its structural and mechanical properties. This discovery highlights the potential of $C_3(P6_422)$ as an extraordinary superhard material, warranting further investigation into its structural and mechanical properties.

In this work, we developed a framework combining MLIPs and optimization algorithms for variable-composition inverse material design. The MACE model, selected for its superior performance among NEP, M3GNet, and CHGNet, was iteratively improved using an active learning strategy, enhancing the predictive accuracy and generalization. And we introduced an improved multiobjective optimization algorithm to optimize properties like density and hardness, incorporating structural similarity constraints to reduce redundancy and improve global search efficiency. Using this framework, we explored the C−N chemical space, identifying 38 novel superhard materials absent from the major computational materials databases (including MP, OQMD, AFlow,[55] and Alexandria[56]) after DFT and phonon spectrum validations. This method provides a versatile approach for variable-composition inverse material design, extendable to other systems and properties, such as bandgap or adsorption energies, offering significant potential for functional material discovery.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.5c00181.

DFT calculation details, training data-set distributions, *g*-MLIP vs DFT comparisons, complete list of discovered superhard materials (PDF)

Data for 38 superhard materials are given in Table S2 (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Wan-Jian Yin − *College of Energy, Soochow Institute for Energy and Materials InnovationS (SIEMIS), and Jiangsu Provincial Key Laboratory for Advanced Carbon Materials and Wearable Energy Technologies, Soochow University, Suzhou 215006, China; Hefei National Laboratory, Hefei 230088, China;* orcid.org/0000-0003-0932-2789; Email: wjyin@suda.edu.cn

Author

Guanjian Cheng − *College of Energy, Soochow Institute for Energy and Materials InnovationS (SIEMIS), and Jiangsu Provincial Key Laboratory for Advanced Carbon Materials and Wearable Energy Technologies, Soochow University, Suzhou 215006, China; Key Laboratory for Computational Physical Sciences (MOE), State Key Laboratory of Surface Physics, Department of Physics, Fudan University, Shanghai 200433, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.5c00181

Author Contributions

Wan-Jian Yin conceived the idea. Guanjian Cheng wrote the code and conducted the calculations. Guanjian Cheng and Wan-Jian Yin discussed the results and wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledged funding support by the National Key Research and Development Program of China (No. 2020YFB1506400), National Natural Science Foundation of China (Nos. 11974257 and 12188101), Jiangsu Distinguished Young Talent Funding (No. BK20200003), Innovation Program for Quantum Science and Technology (No. 2024ZD0300102), Soochow Municipal Laboratory for low carbon technologies and industries. DFT calculations were carried out at the National Supercomputer Center in Tianjin [TianHe-1(A)].

■ REFERENCES

(1) Zhang, S.; He, J.; Zhao, Z.; Yu, D.; Tian, Y. Discovery of Superhard Materials via CALYPSO Methodology. *Chin. Phys. B* **2019**, *28* (10), No. 106104.

(2) Zhukov, V. V.; Zheng, Z.; Sorokin, P. B. Toward Superior Stiffness in Carbon: The Interplay between Bond Strength and Anisotropic Response. *J. Phys. Chem. Lett.* **2024**, *15* (40), 10198−10203.

(3) Badzian, A. R. Superhard Material Comparable in Hardness to Diamond. *Appl. Phys. Lett.* **1988**, *53* (25), 2495−2497.

(4) Monteiro, S. N.; Skury, A. L. D.; de Azevedo, M. G.; Bobrovnitchii, G. S. Cubic Boron Nitride Competing with Diamond as a Superhard Engineering Material − an Overview. *J. Mater. Res. Technol.* **2013**, *2* (1), 68−74.

(5) Wort, C. J. H. Applications for Superhard and Ultra-Hard Materials. In *Microstructure-Property Correlations for Hard, Superhard, and Ultrahard Materials*; Kanyanta, V., Ed.; Springer International Publishing: Cham, Switzerland, 2016; pp 25−74, DOI: 10.1007/978-3-319-29291-5_2.

(6) Kanyanta, V. Hard, Superhard and Ultrahard Materials: An Overview. In *Microstructure-Property Correlations for Hard, Superhard,*

and Ultrahard Materials; Kanyanta, V., Ed.; Springer International Publishing: Cham, Switzerland, 2016; pp 1−23, DOI: 10.1007/978-3-319-29291-5_1.
(7) Filonenko, V. P.; Zibrov, I. P.; Anokhin, A. S.; Kukueva, E. V. Diamond-Based Superhard Composites: New Synthesis Approaches and Application Prospects. *Phys.-Uspekhi* **2019**, *62* (2), 207.
(8) Mazhnik, E.; Oganov, A. R. Application of Machine Learning Methods for Predicting New Superhard Materials. *J. Appl. Phys.* **2020**, *128* (7), No. 075102.
(9) Hao, J.; Liu, H.; Lei, W.; Tang, X.; Lu, J.; Liu, D.; Li, Y. Prediction of a Superhard Carbon-Rich C−N Compound Comparable to Diamond. *J. Phys. Chem. C* **2015**, *119* (51), 28614−28619.
(10) Tang, X.; Hao, J.; Li, Y. A First-Principles Study of Orthorhombic CN as a Potential Superhard Material. *Phys. Chem. Chem. Phys.* **2015**, *17* (41), 27821−27825.
(11) Dong, H.; Oganov, A. R.; Zhu, Q.; Qian, G.-R. The Phase Diagram and Hardness of Carbon Nitrides. *Sci. Rep.* **2015**, *5* (1), 9870.
(12) Wang, Y.; Ding, C.; Yuan, J.; Sun, J. Novel Superhard Structures of High-Pressure C−N Compounds. *arXiv Preprints* **2022**, arXiv.2212.03009 DOI: 10.48550/arXiv.2212.03009 (accessed July 25, 2024).
(13) Tian, F.; Wang, J.; He, Z.; Ma, Y.; Wang, L.; Cui, T.; Chen, C.; Liu, B.; Zou, G. Superhard Semiconducting C_3N_2 Compounds Predicted via First-Principles Calculations. *Phys. Rev. B* **2008**, *78* (23), No. 235431.
(14) Zhang, S.; He, J.; Zhao, Z.; Yu, D.; Tian, Y. Discovery of Superhard Materials via CALYPSO Methodology. *Chin. Phys. B* **2019**, *28* (10), No. 106104.
(15) Laniel, D.; Trybel, F.; Aslandukov, A.; Khandarkhaeva, S.; Fedotenko, T.; Yin, Y.; Miyajima, N.; Tasnádi, F.; Ponomareva, A. V.; Jena, N.; Akbar, F. I.; Winkler, B.; Néri, A.; Chariton, S.; Prakapenka, V.; Milman, V.; Schnick, W.; Rudenko, A. N.; Katsnelson, M. I.; Abrikosov, I. A.; Dubrovinsky, L.; Dubrovinskaia, N. Synthesis of Ultra-Incompressible and Recoverable Carbon Nitrides Featuring CN4 Tetrahedra. *Adv. Mater.* **2024**, *36* (3), No. 2308030.
(16) Gou, H.; Hou, L.; Zhang, J.; Wang, Z.; Gao, L.; Gao, F. Cubic γ-Be3N2: A Superhard Semiconductor Predicted from First Principles. *Appl. Phys. Lett.* **2007**, *90* (19), No. 191905.
(17) Cheng, G.; Gong, X.-G.; Yin, W.-J. Crystal Structure Prediction by Combining Graph Network and Optimization Algorithm. *Nat. Commun.* **2022**, *13*, 1492.
(18) Oganov, A. R.; Lyakhov, A. O.; Valle, M. How Evolutionary Crystal Structure Prediction Works—and Why. *Acc. Chem. Res.* **2011**, *44* (3), 227−237.
(19) Wang, Z.; Wang, X.; Luo, X.; Gao, P.; Sun, Y.; Lv, J.; Wang, H.; Wang, Y.; Ma, Y. Concurrent Learning Scheme for Crystal Structure Prediction. *Phys. Rev. B* **2024**, *109* (9), No. 094117.
(20) Xue, H.; Cheng, G.; Yin, W. Computational Design of Energy-related Materials: From First-principles Calculations to Machine Learning. *WIREs Comput. Mol. Sci.* **2024**, *14* (5), No. e1732.
(21) Fan, Z.; Zeng, Z.; Zhang, C.; Wang, Y.; Song, K.; Dong, H.; Chen, Y.; Ala-Nissila, T. Neuroevolution Machine Learning Potentials: Combining High Accuracy and Low Cost in Atomistic Simulations and Application to Heat Transport. *Phys. Rev. B* **2021**, *104* (10), No. 104309.
(22) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), No. 143001.
(23) Roy, S.; Dürholt, J. P.; Asche, T. S.; Zipoli, F.; Gómez-Bombarelli, R. Learning a Reactive Potential for Silica-Water through Uncertainty Attribution. *Nat. Commun.* **2024**, *15* (1), 6030.
(24) Wang, J.; Wang, Y.; Zhang, H.; Yang, Z.; Liang, Z.; Shi, J.; Wang, H.-T.; Xing, D.; Sun, J. E(n)-Equivariant Cartesian Tensor Message Passing Interatomic Potential. *Nat. Commun.* **2024**, *15* (1), 7607.
(25) Xie, F.; Lu, T.; Meng, S.; Liu, M. GPTFF: A High-Accuracy out-of-the-Box Universal AI Force Field for Arbitrary Inorganic Materials. *Sci. Bull.* **2024**, *69* (22), 3525−3532.
(26) Song, K.; Zhao, R.; Liu, J.; Wang, Y.; Lindgren, E.; Wang, Y.; Chen, S.; Xu, K.; Liang, T.; Ying, P.; Xu, N.; Zhao, Z.; Shi, J.; Wang, J.; Lyu, S.; Zeng, Z.; Liang, S.; Dong, H.; Sun, L.; Chen, Y.; Zhang, Z.; Guo, W.; Qian, P.; Sun, J.; Erhart, P.; Ala-Nissila, T.; Su, Y.; Fan, Z. General-Purpose Machine-Learned Potential for 16 Elemental Metals and Their Alloys. *Nat. Commun.* **2024**, *15* (1), No. 10208.
(27) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat. Commun.* **2022**, *13* (1), 2453.
(28) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. *Nat. Commun.* **2023**, *14* (1), 579.
(29) Zhao, J.; Byggmästar, J.; He, H.; Nordlund, K.; Djurabekova, F.; Hua, M. Complex Ga2O3 Polymorphs Explored by Accurate and General-Purpose Machine-Learning Interatomic Potentials. *Npj Comput. Mater.* **2023**, *9* (1), 159.
(30) Chen, T.; Yuan, F.; Liu, J.; Geng, H.; Zhang, L.; Wang, H.; Chen, M. Modeling the High-Pressure Solid and Liquid Phases of Tin from Deep Potentials with Ab Initio Accuracy. *Phys. Rev. Mater.* **2023**, *7* (5), No. 053603.
(31) Huang, S.-D.; Shang, C.; Kang, P.-L.; Zhang, X.-J.; Liu, Z.-P. LASP: Fast Global Potential Energy Surface Exploration. *WIREs Comput. Mol. Sci.* **2019**, *9* (6), No. e1415.
(32) Chen, C.; Ong, S. P. A Universal Graph Deep Learning Interatomic Potential for the Periodic Table. *Nat. Comput. Sci.* **2022**, *2*, 718−728.
(33) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a Pretrained Universal Neural Network Potential for Charge-Informed Atomistic Modelling. *Nat. Mach. Intell.* **2023**, *5* (9), 1031−1041.
(34) Choudhary, K.; DeCost, B.; Major, L.; Butler, K.; Thiyagalingam, J.; Tavazza, F. Unified Graph Neural Network Force-Field for the Periodic Table: Solid State Applications. *Digit. Discovery* **2023**, *2* (2), 346−355.
(35) Zhang, D.; Liu, X.; Zhang, X.; Zhang, C.; Cai, C.; Bi, H.; Du, Y.; Qin, X.; Peng, A.; Huang, J.; Li, B.; Shan, Y.; Zeng, J.; Zhang, Y.; Liu, S.; Li, Y.; Chang, J.; Wang, X.; Zhou, S.; Liu, J.; Luo, X.; Wang, Z.; Jiang, W.; Wu, J.; Yang, Y.; Yang, J.; Yang, M.; Gong, F.-Q.; Zhang, L.; Shi, M.; Dai, F.-Z.; York, D. M.; Liu, S.; Zhu, T.; Zhong, Z.; Lv, J.; Cheng, J.; Jia, W.; Chen, M.; Ke, G.; E, W.; Zhang, L.; Wang, H. DPA-2: A Large Atomic Model as a Multi-Task Learner. *Npj Comput. Mater.* **2024**, *10* (1), 1−15.
(36) Cheng, G.; Gong, X.-G.; Yin, W.-J. An approach for full space inverse materials design by combining universal machine learning potential, universal property model, and optimization algorithm. *Sci. Bull.* **2024**, *69* (19), 3066−3074.
(37) Lee, J. *A First Course in Combinatorial Optimization*; Cambridge Texts in Applied Mathematics; Cambridge University Press: Cambridge, 2004, DOI: 10.1017/CBO9780511616655.
(38) Breuck, P.-P. D.; Piracha, H. A.; Rignanese, G.-M.; Marques, M. A. L. A Generative Material Transformer Using Wyckoff Representation. *arXiv Preprints* **2025**, arXiv:2501.16051, DOI: 10.48550/arXiv.2501.16051 (accessed April 19, 2025) .
(39) Mortazavi, B.; Zhuang, X.; Rabczuk, T.; Shapeev, V. A. Atomistic Modeling of the Mechanical Properties: The Rise of Machine Learning Interatomic Potentials. *Mater. Horiz.* **2023**, *10* (6), 1956−1968.
(40) Liu, Y.; Meng, H.; Zhu, Z.; Yu, H.; Zhuang, L.; Chu, Y. Predicting Mechanical and Thermal Properties of High-Entropy Ceramics via Transferable Machine-Learning-Potential-Based Molecular Dynamics. *Adv. Funct. Mater.* **2025**, *35*, 2418802.
(41) Tong, Q.; Gao, P.; Liu, H.; Xie, Y.; Lv, J.; Wang, Y.; Zhao, J. Combining Machine Learning Potential and Structure Prediction for Accelerated Materials Design and Discovery. *J. Phys. Chem. Lett.* **2020**, *11* (20), 8710−8720.
(42) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-Kit: A Deep Learning Package for Many-Body Potential Energy Representation and Molecular Dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178−184.

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400

(43) Liao, Y.-L.; Smidt, T. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. *arXiv Preprints*, **2023**, arXiv:2206.11990 (accessed July 15, 2023).
(44) Fan, Z.; Chen, W.; Vierimaa, V.; Harju, A. Efficient Molecular Dynamics Simulations with Many-Body Potentials on Graphics Processing Units. *Comput. Phys. Commun.* **2017**, *218*, 10−16.
(45) Batatia, I.; Kovacs, D. P.; Simm, G. N. C.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 11423−11436.
(46) Batatia, I.; Batzner, S.; Kovács, D. P.; Musaelian, A.; Simm, G. N. C.; Drautz, R.; Ortner, C.; Kozinsky, B.; Csányi, G. The Design Space of E(3)-Equivariant Atom-Centred Interatomic Potentials. *Nat. Mach. Intell.* **2025**, *7*, 56−67.
(47) Vinet, P.; Ferrante, J.; Rose, J. H.; Smith, J. R. Compressibility of Solids. *J. Geophys. Res. Solid Earth* **1987**, *92* (B9), 9319−9325.
(48) Wang, V.; Xu, N.; Liu, J.-C.; Tang, G.; Geng, W.-T. VASPKIT: A User-Friendly Interface Facilitating High-Throughput Computing and Analysis Using VASP Code. *Comput. Phys. Commun.* **2021**, *267*, No. 108033.
(49) Chen, X.-Q.; Niu, H.; Li, D.; Li, Y. Modeling Hardness of Polycrystalline Materials and Bulk Metallic Glasses. *Intermetallics* **2011**, *19* (9), 1275−1281.
(50) Blatov, V. A.; Yang, C.; Tang, D.; Zeng, Q.; Golov, A. A.; Kabanov, A. A. High-Throughput Systematic Topological Generation of Low-Energy Carbon Allotropes. *Npj Comput. Mater.* **2021**, *7* (1), 15.
(51) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), No. 011002.
(52) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies. *Npj Comput. Mater.* **2015**, *1* (1), No. 15010.
(53) Nguyen, H. V.; Bai, L. Cosine Similarity Metric Learning for Face Verification. In *Computer Vision−ACCV 2010*; Kimmel, R., Klette, R., Sugimoto, A., Eds.; Springer: Berlin, Heidelberg, 2011; pp 709−720.
(54) Zhang, L.; Wang, Y.; Lv, J.; Ma, Y. Materials Discovery at High Pressures. *Nat. Rev. Mater.* **2017**, *2* (4), No. 17005.
(55) Curtarolo, S.; Setyawan, W.; Hart, G. L. W.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O.; Mehl, M. J.; Stokes, H. T.; Demchenko, D. O.; Morgan, D. AFLOW: An Automatic Framework for High-Throughput Materials Discovery. *Comput. Mater. Sci.* **2012**, *58*, 218−226.
(56) Schmidt, J.; Cerqueira, T. F. T.; Romero, A. H.; Loew, A.; Jäger, F.; Wang, H.-C.; Botti, S.; Marques, M. A. L. Improving Machine-Learning Models in Materials Science through Large Datasets. *Mater. Today Phys.* **2024**, *48*, No. 101560.

https://doi.org/10.1021/acs.jpclett.5c00181
J. Phys. Chem. Lett. 2025, 16, 4392−4400