Deep learning interatomic potential for Ca-O system at high pressure

Fulun Wu,[1,*] Feng Zheng,[1,*] Weibo He,[1] Xinrui Cao,[1,2] Tie-Yu Lü,[1] Zi-Zhong Zhu,[1,2] and Shunqing Wu[1,†]

[1]*Department of Physics, OSED, Key Laboratory of Low Dimensional Condensed Matter Physics (Department of Education of Fujian Province), Jiujiang Research Institute, Xiamen University, Xiamen 361005, China*

[2]*Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Xiamen University, Xiamen 361005, China*

(Received 10 June 2022; accepted 12 October 2022; published 28 October 2022)

Calcium-containing oxides are fundamental components of the Earth's crust and mantle. Analysis of their structural behavior contributes to our understanding of the Earth's interior, which needs a reliable interatomic potential. Here, we present an interatomic potential for the Ca-O system using the deep neural network method. The initial training data set for the deep interatomic potential (DP) consists of snapshots of prototype binary structures from materials projects and crystal-derived structures based on Ca-O binary compounds, as well as their molecular dynamic trajectories. The accuracy of the DP is evaluated by predicting forces and energies in comparison with those from *ab initio* calculations. We demonstrate that the vibrational and thermodynamic properties based on DP calculations are in excellent agreement with those from *ab initio* calculations. Besides, we also construct a temperature-pressure phase diagram of Ca-O compounds with DP at a lower cost compared to *ab initio* methods. Finally, we use the DP to explore Ca-O structures by combining it with the genetic algorithm, the accuracy of which is validated by a principal component analysis for the local atomic environments. The DP training procedure used in this work is equally applicable to other systems for accurate atomistic simulations as an effective method.

DOI: 10.1103/PhysRevMaterials.6.103802

I. INTRODUCTION

Calcium and oxygen are well known as two of the most abundant elements of the Earth's crust and mantle, respectively [1]. Studies on the structures and fundamental properties in the Ca-O system under high pressure can help us understand Earth's interior. It has been shown that the typical oxide formed by calcium metal is CaO, with the space group $Fm\bar{3}m$ under ambient conditions, which will transform into $Pm\bar{3}m$ at high pressure (∼60 GPa) [2]. Recently, using the *ab initio* random structure searching (AIRSS) method, Joseph *et al.* discovered several stable calcium peroxides (CaO_2) at high temperatures and high pressures, emphasizing the potential role for CaO_2 in Earth's mantle [3]. Also noteworthy are the recent theoretical studies that led to the discovery of CaO_3 ($P\bar{4}2_1m$), which holds unusual divalent ozone anions, providing new information to understand the cycles of oxygen within our planet [4]. This successful discovery was confirmed by experimental synthesis [4]. These results provide a structural database of Ca-O system for modeling the interior of Earth.

However, to fully understand the Earth's interior, a more comprehensive investigation of structural behavior of Ca-O compounds with variable stoichiometry under high-pressure conditions is still necessary. In doing so, a reliable interatomic potential is needed to describe complicated atomic interactions with high accuracy. While density functional theory (DFT) can offer high accuracy in interatomic interactions [5,6], such *ab initio* calculations require expensive computational costs as a cubic function of the number of electrons. Alternatively, empirical interatomic potentials are computationally efficient, but their accuracy and transferability are questionable due to their limited algebraic form that is specifically designed for a few known structures. A legitimate question is are there any tradeoffs between computational cost and accuracy to investigate crystal structures?

Recently advances in machine learning potential enables the evaluation of the free energy of *ab initio* quality at the cost of empirical force fields [7–16]. This method can construct a direct mapping of structures to forces and energies by using a large amount of experimental or computational data, saving a significant amount of computational time required for first-principles calculations. The machine learning potential was initially proposed by Behler and Parrinello in 2007 as the neural network potential (NNP) [9]. In the last few years, various types of machine learning potentials have been proposed: Gaussian approximation potential (GAP) [12], the gradient-domain machine learning (GDML) [15], the moment tensor potential (MTP) [17], the spectral neighbor analysis potential (SNAP) [7], and the deep potential for molecular dynamics (DEEPMD) [18]. In particular, the "deep potential" and DEEPMD are garnering wide interest recently with applications to challenging simulations such as the 1D cooperative diffusion of Ca [19], the phase diagram of water [20], silicon liquid structure and crystal nucleation [21], proton transfer at the water-TiO_2 interface [22], accurate irradiation damage simulations [23], nonadiabatic excited-state dynamics [24], potential for metalloid-containing Pd-Si compounds [25], Al-Tb alloys [26], etc.

*These authors contributed equally to this work.
†wsq@xmu.edu.cn

In this paper, we developed a general deep interatomic potential (DP) for the Ca-O system and demonstrated how it can be used to predict forces, energies, and vibrational and thermodynamic properties. The computational efficiency of DP makes it possible to construct phase diagrams and perform a crystal-structure search at a lower cost.

II. METHODS

A. First-principles calculations

Our first-principles calculations were performed by using the projector-augmented wave (PAW) [27,28] representation with density functional theory as implemented in the Vienna *ab initio* simulation package (VASP) [5,29]. The exchange and correlation energy was treated within the spin-polarized generalized gradient approximation (GGA) and parametrized by the Perdew-Burke-Ernzerhof (PBE) [30] formula. We used pseudopotential descriptions with valence electronic configurations $3s^2 3p^6 4s^2$ and $2s^2 2p^4$ for Ca and O, respectively. A kinetic-energy cutoff of 500 eV was used, and Brillouin-zone integrations were performed using the Monkhorst-Pack scheme with a k-point grid resolution of $2\pi \times 0.033$ Å^{-1}. The convergence test results on the wave function energy cutoff and the k-point density are shown in Tables S3 and S4 of the Supplemental Material [31].

B. Adaptive genetic algorithm

To collect enough structures as an initial training set, Ca_xO_y (x, $y = 1$, 2, 3, 4) was investigated by using the adaptive genetic algorithm (AGA) crystal structure prediction method [32], which is a method that combines fast structure exploration by auxiliary classical potentials and the accurate *ab initio* calculations adaptively and iteratively. In the genetic algorithm (GA) loop, the Ca and O atoms' initial atomic positions were randomly generated without assuming the Bravais lattice type, symmetry, atom basis, or unit cell dimensions. Some low-enthalpy structures at the end of each GA search were selected for single point DFT calculations, whose energies, force, and stress were used to fit the interatomic potential parameters for the next iteration of GA search. A total of 40 adaptive iterations were performed to explore the complex potential energy surfaces of Ca-O compounds.

C. Enhanced sampling for training set

As an enhanced sampling method, we used Deep Potential GENerator (DP-GEN) [33,34] to collect as many local atomic environments as possible during training. DP-GEN is able to build uniformly accurate DP models on-the-fly while minimizing data generation costs. In DP-GEN, the PES represented by the DP model is denoted by $E_\omega(R)$, where R denotes atomic positions and ω denotes the parameters. A key point of DP-GEN of the potential energy surface (PES) was that an ensemble of models $\{E_{\omega 1}, E_{\omega 2}, \ldots, E_{\omega\alpha}, \ldots\}$, which were trained at the same time, using the same training data but with different parameters for the initialization. In this case, different parameters in the initialization led to different minimizers and thus to different PES models. In certain regions of the PES where there were sufficient training data, different PES models should all be reasonably accurate and therefore produced predictions that were close to each other. That is, if the predictions of different models deviate significantly for a new structure, indicating that the PES here has not been explored comprehensively, the structure would be labeled to be trained in the next iteration. Considering that force is an atomic property that is sensitive to local accuracy, we used the formula to calculate the standard deviation of the predicted maximum force to assess model consistency,

$$\delta_F = \max_i \sqrt{\langle ||F_i - \langle F_i \rangle||^2 \rangle}, \tag{1}$$

where i runs over all atoms and the average of the four models is taken. In this work, those structures generated by DP-based MD (DPMD) with model deviation 0.05–0.20 eV/Å were selected and passed to the training data sets.

The LAMMPS package [35] had been interfaced with DP-GEN to perform DPMD. During the exploration of the PES, the DPMD simulations were applied to generate serials of new snapshots of structures with distortions. We performed the isothermal-isobaric (NPT) simulations [36] under periodic boundary conditions at the pressure of 0.1, 0.5, 1, 5, 10, 30, and 50 GPa, corresponding to the internal pressure of the Earth [1]. In this work, the number of DPMD steps was set to 2000 and the length of the trajectory was set to 3 ps.

D. DP training

For the training process of the DP model, the smooth method of the DEEPMD-Kit software package was used [18,37], which was constructed from all information (both angular and radial) of atomic configurations. As seen in Fig. S1 of the Supplemental Material [31], the local environment, that was the atomic positions of the Ca-O compound within the cutoff radius in relation to other neighboring atoms, would be encoded into the descriptors D_i. These descriptors would then be fed into a high-dimensional neural network to calculate the atomic energy E_i. In this work, the cutoff radius of neighbor atoms was set to 5.0 Å. In the DP model, the relative coordinates of atoms are mapped onto generalized coordinates $\hat{x}_{ji} = \frac{s(r_{ji})x_{ji}}{r_{ji}}$ to preserve the physical symmetry, where $s(r_{ji})$ is a continuous and differentiable scalar weighting function. Here, the inverse distance $1/r$ decayed smoothly from 0.5 to 5.0 Å to remove the discontinuity introduced by the cutoff radius. The filter neural network had three hidden layers which were {25, 50, 100}, and the fitting network was {240, 240, 240}, respectively. The neural network was initialized with random parameters and the total number of training steps was 1 000 000. During the training, the Adam stochastic gradient descent method [38] was applied so that the learning rate decreased exponentially. The decay step and decay rate were set to 5000 and 0.96. The loss function L was defined by [18]

$$L(p_e, p_f, p_\xi) = \frac{p_e}{N}\Delta E^2 + \frac{p_f}{3N}\sum_i |\Delta F_i|^2 + \frac{p_\xi}{9N}\|\Delta \Xi\|^2, \tag{2}$$

where ΔE, ΔF_i, and $\Delta \Xi$ denote root mean square (rms) error in energy, force, and virial, respectively. In this work, the energy prefactor p_e started at 0.02 and ended at 1. At the same

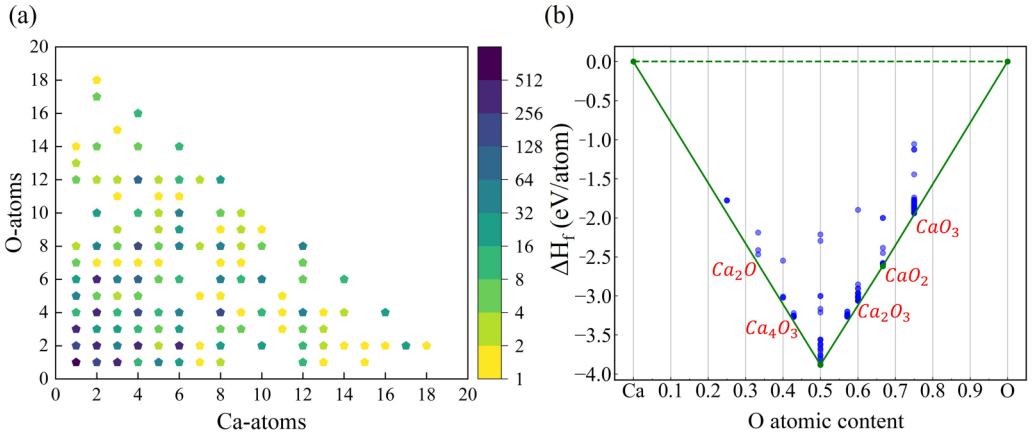

FIG. 1. (a) Binary structures from MP, whose elements replaced by Ca and O atoms. The color bar represents the number of structures. (b) Convex hull diagram of the Ca-O system at 20 GPa by AGA.

time, the force prefactor p_f started at 1000 and went down to 1 in the end. Since there are no virial data included during the training progress, p_ξ is equal to 0.

III. RESULTS AND DISCUSSION

A. Construction of training set

Typically, DP model requires a large number of structures as a training set. However, there are only a few crystalline phases available in the materials projects (MPs) [39] for the Ca_xO_y system (i.e., Ca_2O_3, CaO, CaO_2, CaO_3, and CaO_{10}), which are not sufficient for deep learning. In this work, the initial training set is divided into two parts. The first part is binary compounds from the MPs. Figure 1(a) shows all selected binary compounds from the MP database. Here, structures are selected based on two criteria: (1) the maximum lattice constant is smaller than 10 Å, and (2) there are no more than 20 atoms in the unit cell. The colors represent the number of structures of different stoichiometries. There are a total of 10 980 binary phases from MPs, whose elements are replaced with Ca and O. The horizontal and vertical axes of Fig. 1(a) indicate the number of Ca and O atoms in the unit cell, respectively. It can be found that structures of certain formula units do not exist in the database (i.e., 5 and 7 f.u.).

For the second part, in order to obtain more low-enthalpy structures of Ca-O compounds as training data, different stoichiometries of Ca_xO_y ($x:y$ = 3:1, 2:1, 3:2, 4:3, 1:1, 3:4, 2:3, 1:2, 1:3) with different formula units (i.e., 1, 2, 3, 4, and 6 f.u.) are searched by the AGA method at 20 GPa, which corresponds to pressures between Earth's upper and lower mantle. The relative stability of these predicted Ca-O compounds was investigated at the corresponding pressures based on the calculated enthalpies of formation,

$$\Delta H_f = \frac{[H(Ca_xO_y) - xH(Ca) - yH(O)]}{x+y}, \quad (3)$$

where x, y are the numbers of atoms of Ca and O, and H is the calculated enthalpy of the given structure. Before delving into the stable structures of Ca-O compounds, the crystal structures of elemental Ca and O crystal must be clarified. At 20 GPa, our calculated results suggest that $Im3m$ and epsilon ($C2/m$) phases are the ground state phase for elemental Ca and O, respectively, which agrees well with previous studies [40,41]. Figure 1(b) shows the convex hull of the Ca-O system at 20 GPa. It can be found that CaO with $Fm\bar{3}m$ symmetry and CaO_2 with $P2_1/c$ symmetry lie on the convex hull; these results are also consistent with Refs. [2,3]. Besides the CaO ($Fm\bar{3}m$) and CaO_2 ($P2_1/c$), several low-enthalpy metastable Ca_xO_y stoichiometries, such as Ca_2O ($C2/m$), Ca_4O_3 ($C2/m$), Ca_2O_3 ($C2/m$), and CaO_3 ($Imm2$) are also identified at 20 GPa. The crystal structures of these stable and metastable phases are shown in Fig. S3 and their structural parameters are listed in Table S1 [31].

Based on these stable and metastable Ca-O from the AGA search, structures with their energies and forces away from equilibrium state are also considered: (i) The lattice constant of structures is dilated and compressed uniformly by $a' = a\,(1 \pm 0.02 \times n)$ (n = 0–4) to sample different interatomic distances. (ii) The coordinates of atoms in the cell are shifted by 0.01 Å at random, and a total of 30 perturbations are generated in this step. In these ways, we can generate 270 structures for each predicted Ca-O structure. Moreover, more distorted compounds are also added to the training data sets, which are generated with *ab initio* molecular dynamics (AIMD) simulations from the above predicted structures by AGA. Supercells containing hundreds of atoms are used for these structures. The AIMD simulations are performed with the NVT ensemble at T = 300 K, under periodic boundary conditions with Nosé-Hoover thermostat [42,43], in order to generate serial snapshots of structures with distortions. In summary, for each of the predicted Ca-O structures, we obtained 5400 configurations for the initial data set, which contained the information of atomic positions, energies, and forces. The data that are unconverged should be removed from the data sets.

The Ca_2O ($C2/m$), Ca_4O_3 ($C2/m$), CaO ($Fm\bar{3}m$), Ca_2O_3 ($C2/m$), CaO_2($P2_1/c$), and CaO_3 ($Imm2$) configurations from the previous AGA search are also used as the initial structures for the DP-GEN. Five iterations are performed at 0.1, 0.5, 1, 5, 10, 30, and 50 GPa to enhance the sampling of PES. The DPMD explorations result in the addition of 18 108 new configurations to the data set. Finally, combining all the crystal data sets with different compositions, we have 48 468

TABLE I. The overall information of the training data sets for the Ca-O system. The RMSE of energy predicted by the DP model are the validation RMSE.

Systems		Atoms	Training set (90%)	Validation set (10%)	Energy RMSE (meV/atom)
MP structures			9882	1098	5.6
AGA structures	Ca_2O	108	4500	500	2.0
	Ca_4O_3	112	3240	360	4.8
	CaO	128	3240	360	2.4
	Ca_2O_3	160	3222	358	6.3
	CaO_2	108	3240	360	2.4
	CaO_3	216	3240	360	1.4
DP-GEN structures			16298	1810	7.1

configurations in the training process. A typical split of the data set is 90% for training and 10% for validation. The overall information of the training and validation data sets is summarized in Table I.

B. Reliability and validation of DP

In order to demonstrate the reliability of DP, a comparison of the predicted energies and forces by DFT and DP will be necessary. Taking the above predicted CaO_3 ($Imm2$) system as an example, Fig. 2 shows the predictive accuracy of DP energies and forces per atom compared with that of DFT for the CaO_3 ($Imm2$) system in the validating sets. We can see that the results of DP and DFT are highly correlated. The root-mean-squared errors (RMSEs) are less than 1.5 meV/atom for energies and 100 meV/Å for forces. Also, it can be seen that the RMSEs of the training and validating sets are extremely close, which suggests that the neural network is not overfitted. (See Fig. S2 [31] for details on the training set.) Furthermore, it can be seen in Fig. 2(a) that there is a wide distribution of energy, which may indirectly reflect the complicated configuration spaces in our exploration of PES.

Due to the dependence on the quality of the matrix of force constants force constant matrix (dynamical or Hessian matrix), the phonon density of states (PHDOS) is extremely sensitive to the accuracy of atomic interactions. Therefore, we compared the PHDOS and the phonon dispersion relations for a Ca_2O ($C2/m$) structure at $T = 0$ K and 20 GPa calculated by

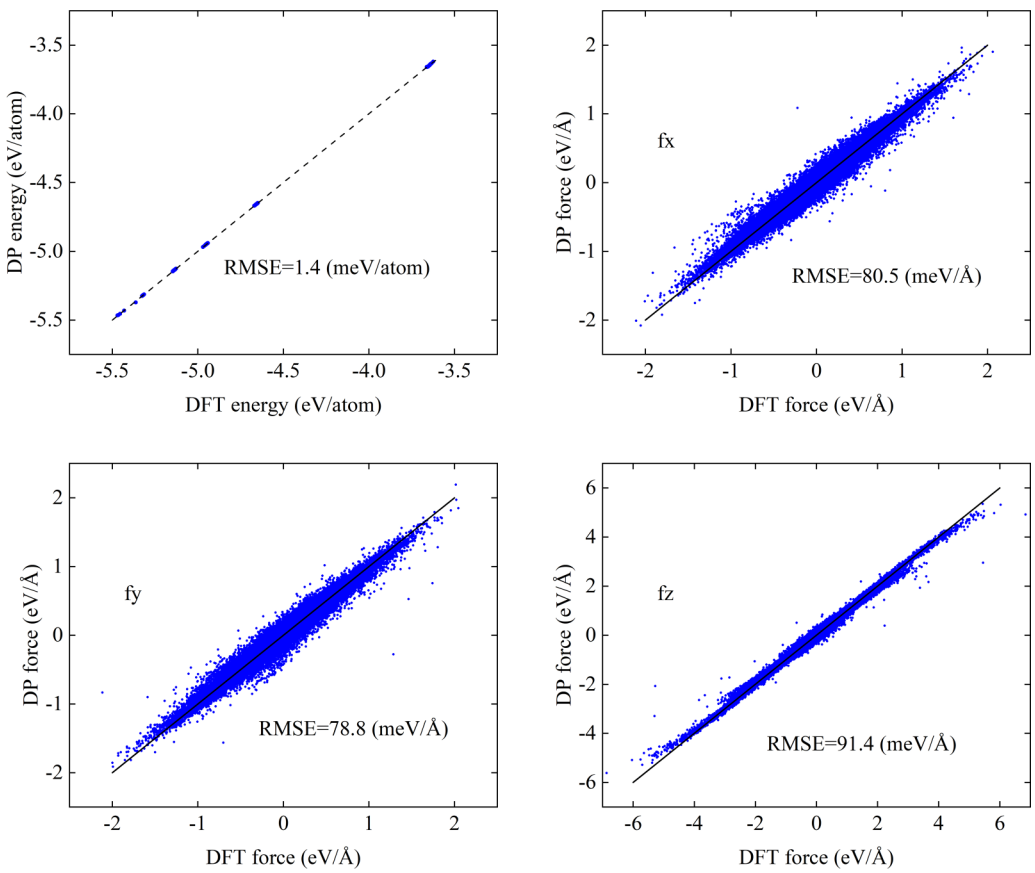

FIG. 2. DP vs DFT energies and forces for CaO_3 compounds in the validation data.

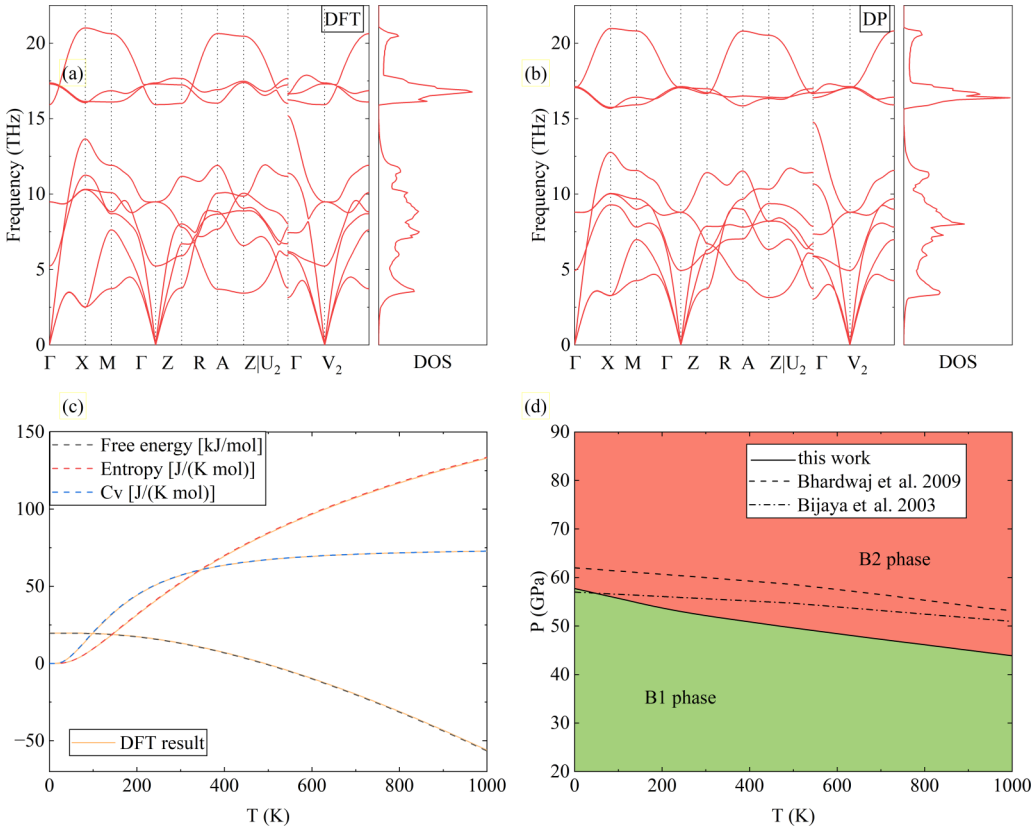

FIG. 3. (a) DFT dispersion relation of the phonons and PHDOS compared to the (b) DP results for the Ca_2O ($C2/m$) structure at $T = 0$ K and 20 GPa. (c) Thermal properties of Ca_2O ($C2/m$): specific heat capacity (C_V), Helmholtz free energy (F) and entropy (S) computed from DFT (solid line) and the DP (color dotted line) as a function of temperature. (d) Pressure-temperature phase diagram of CaO by DP calculations.

DFT and DP. Here, the DP result of PHDOS was calculated by PHONOPY [44] and its LAMMPS interface PhonoLAMMPS [45]. As can be seen in Figs. 3(a) and 3(b), compared to the DFT results, DP is found to accurately predict the location of the PHDOS peaks and reproduce the phonon dispersion relations. Some of the differences in the dispersion of the highest optical phonon modes can be attributed to finite size effects. With the above phonon dispersion relations and the formalism from previous studies [46], it is possible to obtain the thermodynamic properties such as specific heat (C_V), entropy (S), energy (E), and Helmholtz free energy (F), respectively. As shown in Fig. 3(c), the F, C_V, and S are calculated from 0 to 1000 K for the Ca_2O structure. The solid line represents results calculated by DFT and the DP data are given as the dotted line. Our DP results match the DFT calculations to a high degree. Notably, the thermodynamic properties such as C_V are the derivative of energy with respect to temperature, while our training set contains no temperature information. Thus, the accurate prediction of these thermodynamic properties implies that our DP model has a high generalization capability to temperature. Also, the Debye temperature is an important physical quantity and an appropriate parameter to describe the physical phenomena in the solid state related to specific heat, melting point, lattice vibrations, and elastic constants. For the CaO with $Fm\bar{3}m$ symmetry, the DFT calculated Debye temperature was 651.5 K, while the DP calculated value was 670.4 K. The error between the DFT and DP result is about 3%, which implies the accuracy of the DP model. Furthermore, we also studied the pressure-temperature phase diagram by the DP. Usually, the construction of phase diagram by *ab initio* calculation is tough work due to the expensive computational costs [47]. However, with a well-trained DP model, this cost can be reduced from days to minutes to obtain the Gibbs free energy at various pressures and temperatures. To demonstrate its accurately, we calculated the Gibbs free energy (G) using DFT, then compared with our DP results. As shown in Fig. S4 [31], the DP results match the DFT calculations to a high degree. Then we calculated the difference in Gibbs free energy (G) between the two phases of CaO ($B1$ and $B2$) to assess the stability. The difference in Gibbs free energy (G) is calculated by the following formulas:

$$\Delta G(T, p) = G_{B1}(T, p) - G_{B2}(T, p), \tag{4}$$

where $G(T, p)$ is the calculated Gibbs free energy at given temperature T and pressure p for a given structure. A positive value of ΔG indicates that the $B2$ phase is stable relative to $B1$, while a negative value of ΔG indicates the $B1$ phase is more stable. By calculating the ΔG at different temperatures and pressures, the pT diagram of CaO can be obtained, as shown in Fig. 3(d).

The typical CaO structure is the $B1$ (NaCl structure) phase under ambient conditions, which will transform into the $B2$ (CsCl structure) phase under high pressure and high temperature. Our DP calculated results show the transition pressure is

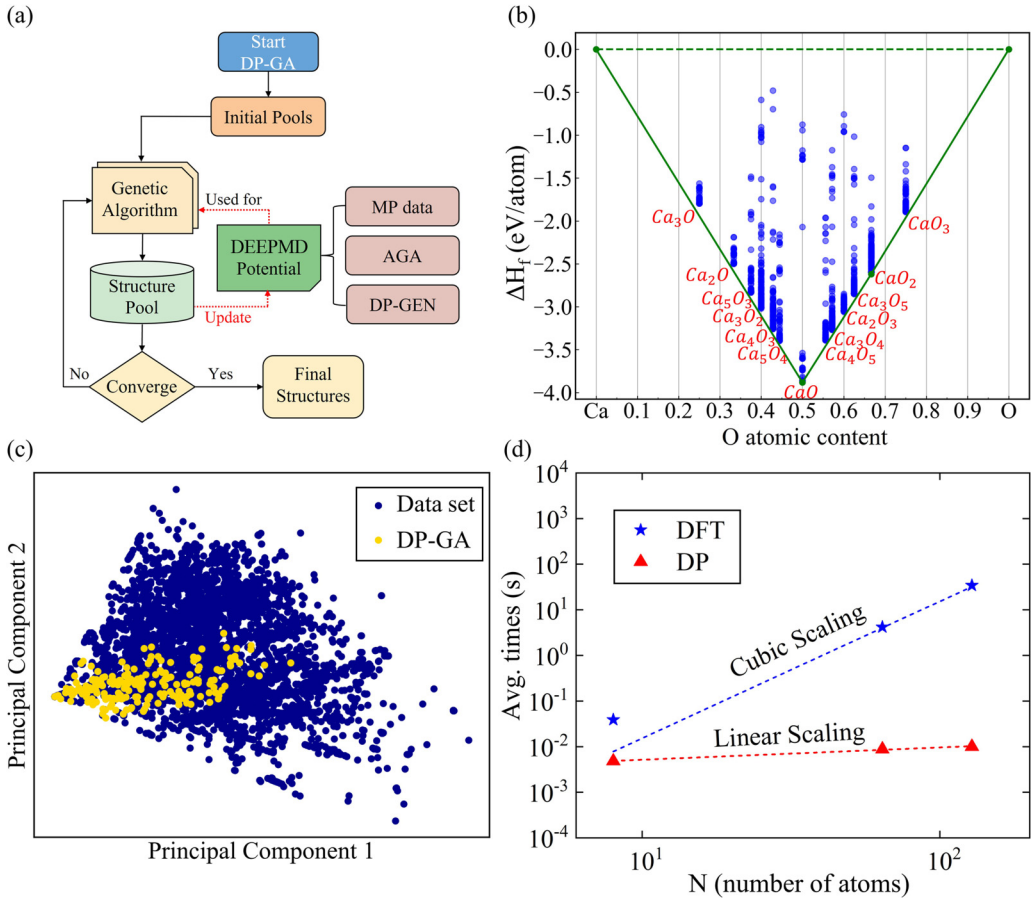

FIG. 4. (a) Flowchart of the DP-GA. (b) Convex hull diagrams of the Ca-O system at 20 GPa by DP-GA. (c) The distribution of G vectors projected onto the first two principal component axes (PC1 and PC2). (d) Computational cost with DFT and DP versus system size.

52 GPa at 300 K, compared to 50–70 GPa from experiments [48–50]. Moreover, the phase transition pressure monotonically decreases when the temperature increases, which is also consistent with DFT results from Bhardwaj *et al.* [51] and Bijaya *et al.* [52,53].

C. Application of DP for crystal-structure search

Local structure optimization is the most expensive part of the crystal structure prediction (CSP) process, as each structure is evaluated using a traditional *ab initio* calculation that demands a significant amount of computational effort [54]. Thus, most searches consider only a limited number of structures and a few stoichiometries. Here, the DP model is suitable for CSP due to its efficiency in calculating free energies and its ability to generalize to large systems. To demonstrate this, we try to perform a more comprehensive search for new phases and compounds in Ca-O systems with a wide range of stoichiometries by interfacing the genetic algorithm with DP (DP-GA). Figure 4(a) depicts the details of the DP-GA scheme. First, the initial structures are randomly generated without assuming the Bravais lattice type, symmetry, atom basis, or unit cell dimensions. During the GA, the evolutionary operators and random generators, such as mutation, heredity, and permutation, are used to produce new structures. Energies of these structures are calculated by the DP model. Here, the DP is trained on the MP, AGA, and DP-GEN data sets (see Sec. III A). It is worth noting that the structures from GA can be used to update the DP model to ensure the accuracy and generalization capability. Finally, convergence is considered when the lowest energy in the structure pool remains unchanged for 300 GA generations.

Using the DP-GA, Ca_xO_y structures with a wide range of atomic ratios ($x:y$ = 1:1, 1:2, 1:3, 2:1, 2:3, 3:1, 3:2, 3:4, 3:5, 4:3, 4:5, 5:3, 5:4) were searched. Different formula units (i.e., 1, 2, 3, 4, and 6 f.u.) were considered. The maximum system size for our crystal structure prediction is 54 atoms. Figure 4(b) presents the convex hull of the Ca-O system at 20 GPa. In this work, in addition to the stable structures CaO ($Fm\bar{3}m$) and CaO_2 ($P2_1/c$), several metastable structures with enthalpy very close to the convex hull (less than 100 meV/atom) were identified, such as Ca_2O ($C2/m$), Ca_5O_3 ($P3m1$), Ca_3O_2 ($C2/m$), Ca_4O_3 ($C2/m$), Ca_5O_4 ($R\bar{3}m$), Ca_4O_5 ($P\bar{1}$), Ca_3O_4 ($C2/m$), Ca_2O_3 ($C2/m$), Ca_3O_5 (Cm), and CaO_3 ($Imm2$). Since the current calculations do not consider the effect of temperature on structural stability, these low enthalpy metastable compounds may become stable at finite temperatures. Here, the structures searched by previous AGA can be reproduced within a few generations. Details of all these structures can be seen in Fig. S3 [31]. For comparison, the corresponding energies calculated by DFT and DP are also investigated. As shown in Table S2 [31], the difference between the DP and DFT

energies of these structures is around 5 meV/atom, indicating that the DP can accurately describe the PES. Moreover, our training set did not contain any Ca_4O_3 structures. The accurate prediction of these unseen compositions implies that our DP model has a high generalization capability, which indirectly reflects our thorough exploration of the local atomic environments.

To further illustrate this point, principal component analysis (PCA) [55] was used to capture characteristics of the local atomic environments. The atom-centered symmetry-functions vector **G** [56] was selected as collective variables instead of real atomic positions in order to maintain the symmetry of structures. As shown in Fig. 4(c), using the PCA, we examined the **G**-vector distributions in the training data set and structures generated by DP-GA for the Ca-O system. It can be seen that most of the **G** points of the structures generated by DP-GA are located within the training set. Thus, we explicitly confirm that our sampling of the local atomic environment is adequate, ensuring the accuracy of the CSP.

We also examined the computational efficiency of the DP method. As computational hardware, we used clusters of AMD® Opteron Processor 6378 2.4 GHz (8 cores per CPU) to build the training set, and a NVIDIA® Tesla® V100 GPU to train the DP model. It took us 3 days to obtain the entire training set, 21 h to train, and just 10 h to complete 1000 DP-GA generations. However, the DFT-based GA can only complete about 500 generations in the same amount of time (4–5 days) and with the same computational resources. Moreover, we compared the average time of DP and DFT to evaluate the energy and force for different system sizes in units of CPU core seconds. As seen in Fig. 4(d), the DFT calculations scales in $O(N^3)$, whereas DP scale almost linearly with $O(N)$, respectively. Thus, computational efficiency of the DP makes it possible to predict new structures of a larger system.

IV. CONCLUSIONS

In summary, we proposed a way to prepare the training set and then developed, validated, and applied a DP for the Ca-O system using the DEEPMD scheme. The DP was developed using a training set including snapshots of the prototype binary structures from MP, as well as Ca-O structures with their molecular dynamic trajectories obtained from the AGA search. In comparison to DFT, this DP model can accurately predict the energies, forces, and vibrational and thermodynamic properties of the Ca-O system at a greatly reduced cost. In addition, due to its ability to fast calculating free energies, our DP model makes it efficient to construct a temperature-pressure phase diagram for CaO at a lower cost. Combining with the genetic algorithm, DP can also be used to investigate stable and metastable Ca-O structures, the accuracy of which is validated by a PCA analysis for the local atomic environments. Our training approach of DP presented in this paper provides an effective strategy to study structural property, which can also be applied to other systems as well.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 11874307). S. Fang and T. Wu from the Information and Network Center of Xiamen University are acknowledged for the help with the graphics processing unit (GPU) computing.

[1] A. E. Doyle, E. D. Young, B. Klein, B. Zuckerman, and H. E. Schlichting, Oxygen fugacities of extrasolar rocks: evidence for an Earth-like geochemistry of exoplanets, Science **366**, 356 (2019).

[2] B. B. Karki and J. Crain, Structure and elasticity of CaO at high pressure, J. Geophys. Res. Solid Earth **103**, 12405 (1998).

[3] J. R. Nelson, R. J. Needs, and C. J. Pickard, Calcium peroxide from ambient to high pressures, Phys. Chem. Chem. Phys. **17**, 6889 (2015).

[4] Y. Wang, M. Xu, L. Yang, B. Yan, Q. Qin, X. Shao, Y. Zhang, D. Huang, X. Lin, J. Lv *et al.*, Pressure-stabilized divalent ozonide CaO_3 and its impact on Earth's oxygen cycles, Nat. Commun. **11**, 4702 (2020).

[5] G. Kresse and J. Hafner, *Ab initio* molecular dynamics for liquid metals, Phys. Rev. B **47**, 558 (1993).

[6] R. Car and M. Parrinello, Unified Approach for Molecular Dynamics and Density-Functional Theory, Phys. Rev. Lett. **55**, 2471 (1985).

[7] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. **285**, 316 (2015).

[8] T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, and R. Ramprasad, A universal strategy for the creation of machine learning-based atomistic force fields, npj Comput. Mater. **3**, 37 (2017).

[9] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[10] J. Behler, Perspective: machine learning potentials for atomistic simulations, J. Chem. Phys. **145**, 170901 (2016).

[11] T. Morawietz, A. Singraber, C. Dellago, and J. Behler, How van der waals interactions determine the unique properties of water, Proc. Natl. Acad. Sci. USA **113**, 8368 (2016).

[12] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, Phys. Rev. Lett. **104**, 136403 (2010).

[13] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, Phys. Rev. Lett. **108**, 058301 (2012).

[14] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks, Nat. Commun. **8**, 13890 (2017).

[15] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields, Sci. Adv. **3**, e1603015 (2017).

[16] J. S. Smith, O. Isayev, and A. E. Roitberg, ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, Chem. Sci. **8**, 3192 (2017).

[17] A. V. Shapeev, Moment tensor potentials: a class of systematically improvable interatomic potentials, Multiscale Model. Simul. **14**, 1153 (2016).

[18] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics, Phys. Rev. Lett. **120**, 143001 (2018).

[19] Y. Wang, J. Wang, A. Hermann, C. Liu, H. Gao, E. Tosatti, H.-T. Wang, D. Xing, and J. Sun, Electronically Driven 1D Cooperative Diffusion in a Simple Cubic Crystal, Phys. Rev. X **11**, 011006 (2021).

[20] L. Zhang, H. Wang, R. Car, and W. E, Phase Diagram of a Deep Potential Water Model, Phys. Rev. Lett. **126**, 236001 (2021).

[21] L. Bonati and M. Parrinello, Silicon Liquid Structure and Crystal Nucleation from *Ab Initio* Deep Metadynamics, Phys. Rev. Lett. **121**, 265701 (2018).

[22] M. F. Calegari Andrade, H.-Y. Ko, L. Zhang, R. Car, and A. Selloni, Free energy of proton transfer at the water–TiO_2 interface from *ab initio* deep potential molecular dynamics, Chem. Sci. **11**, 2335 (2020).

[23] H. Wang, X. Guo, L. Zhang, H. Wang, and J. Xue, Deep learning inter-atomic potential model for accurate irradiation damage simulations, Appl. Phys. Lett. **114**, 244101 (2019).

[24] W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral, and G. Cui, Deep learning for nonadiabatic excited-state dynamics, J. Phys. Chem. Lett. **9**, 6702 (2018).

[25] T. Wen, C. Z. Wang, M. J. Kramer, Y. Sun, B. Ye, H. Wang, X. Liu, C. Zhao, F. Zhang, K. M. Ho, and N. Wang, Development of a deep machine learning interatomic potential for metalloid-containing Pd-Si compounds, Phys. Rev. B **100**, 174101 (2019).

[26] L. Tang, Z. J. Yang, T. Q. Wen, K. M. Ho, M. J. Kramer, and C. Z. Wang, Development of interatomic potential for Al-Tb alloys using a deep neural network learning method, Phys. Chem. Chem. Phys. **22**, 18467 (2020).

[27] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).

[28] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[29] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[30] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. **6**, 15 (1996).

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.6.103802 for structural parameters studied in this work, DP vs DFT energies of metastable structures, ENCUT and K-point convergence test, training process of DP model, validate of DP energies and forces in the training data, details of newly discovered structures, and Gibbs free energy of CaO and Ca_4O_3.

[32] S. Q. Wu, M. Ji, C. Z. Wang, M. C. Nguyen, X. Zhao, K. Umemoto, R. M. Wentzcovitch, and K. M. Ho, An adaptive genetic algorithm for crystal structure prediction, J. Phys.: Condens. Matter. **26**, 035402 (2014).

[33] Y. Z. Zhang, H. D. Wang, W. J. Chen, J. Z. Zeng, L. F. Zhang, H. Wang, and E. Weinan, DP-GEN: a concurrent learning platform for the generation of reliable deep learning based potential energy models, Comput. Phys. Commun. **253**, 107206 (2020).

[34] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, Active learning of uniformly accurate interatomic potentials for materials simulation, Phys. Rev. Mater. **3**, 023804 (2019).

[35] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. **117**, 1 (1995).

[36] G. J. Martyna, M. L. Klein, and M. Tuckerman, Nosé–Hoover chains: the canonical ensemble via continuous dynamics, J. Chem. Phys. **97**, 2635 (1992).

[37] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and W. E, End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems, in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Vol. 31 (Curran Associates, Inc., 2018).

[38] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun (2015).

[39] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The Materials Project: a materials genome approach to accelerating materials innovation, APL Mater. **1**, 011002 (2013).

[40] U. Schwarz, D. Kasinathan, C. Bergner, J. Hunger, K. Meier-Kircher, L. Akselrud, M. Hanfland, M. Mezouar, K. Glazyrin, and G. W. Stinton, Distortions in the cubic primitive high-pressure phases of calcium, J. Phys.: Condens. Matter **31**, 065401 (2018).

[41] S. F. Elatresh and S. A. Bonev, Stability and metallization of solid oxygen at high pressure, Phys. Chem. Chem. Phys. **22**, 12577 (2020).

[42] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Mol. Phys. **52**, 255 (1984).

[43] W. G. Hoover, Canonical dynamics: equilibrium phase-space distributions, Phys. Rev. A **31**, 1695 (1985).

[44] A. Togo and I. Tanaka, First principles phonon calculations in materials science, Scr. Mater. **108**, 1 (2015).

[45] Abel Carreras, phonoLAMMPS: A python interface for LAMMPS phonon calculations using phonopy (0.8.1), Zenodo (2021), https://doi.org/10.5281/zenodo.5668319.

[46] M. T. Dove, *Introduction to Lattice Dynamics* (Cambridge University Press, Cambridge, 1993).

[47] J. Bouchet and F. Bottin, High-temperature and high-pressure phase transitions in uranium, Phys. Rev. B **95**, 054113 (2017).

[48] R. Jeanloz, T. Ahrens, H. Mao, and P. Bell, B1-B2 transition in calcium oxide from shock-wave and diamond-cell experiments, Science **206**, 829 (1979).

[49] J. Mammone, H. Mao, and P. Bell, Equations of state of CaO under static pressure conditions, Geophys. Res. Lett. **8**, 140 (1981).

[50] P. Richet, H.-K. Mao, and P. M. Bell, Static compression and equation of state of CaO to 1.35 mbar, J. Geophys. Res. Solid Earth **93**, 15279 (1988).

[51] P. Bhardwaj and S. Singh, Role of temperature in the numerical analysis of CaO under high pressure, Cent. Eur. J. Chem. **8**, 126 (2010).

[52] B. B. Karki and R. M. Wentzcovitch, Vibrational and quasiharmonic thermal properties of CaO under pressure, Phys. Rev. B **68**, 224304 (2003).

[53] B. B. Karki, R. M. Wentzcovitch, S. de Gironcoli, and S. Baroni, High-pressure lattice dynamics and thermoelasticity of MgO, Phys. Rev. B **61**, 8793 (2000).

[54] C. Hong, J. M. Choi, W. Jeong, S. Kang, S. Ju, K. Lee, J. Jung, Y. Youn, and S. Han, Training machine-learning potentials for crystal structure prediction using disordered structures, Phys. Rev. B **102**, 224104 (2020).

[55] H. Abdi and L. J. Williams, Principal component analysis, Wiley Interdiscip. Rev. Comput. Stat. **2**, 433 (2010).

[56] J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. **134**, 074106 (2011).