WILEY | WIREs
COMPUTATIONAL MOLECULAR SCIENCE

SOFTWARE FOCUS

LASP: Fast global potential energy surface exploration

Si-Da Huang [ID] | Cheng Shang [ID] | Pei-Lin Kang | Xiao-Jie Zhang | Zhi-Pan Liu [ID]

Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai, China

Correspondence
Zhi-Pan Liu, Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China.
Email: zpliu@fudan.edu.cn
Cheng Shang, Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China.
Email: cshang@fudan.edu.cn

Here we introduce the LASP code, which is designed for large-scale atomistic simulation of complex materials with neural network (NN) potential. The software architecture and functionalities of LASP will be overviewed. LASP features with the global neural network (G-NN) potential that is generated by learning the first principles dataset of global PES from stochastic surface walking (SSW) global optimization. The combination of the SSW method with global NN potential facilitates greatly the PES exploration for a wide range of complex materials. Not limited to SSW-NN global optimization, the software implements standard interfaces to dock with other energy/force evaluation packages and can also perform common tasks for computing PES properties, such as single-ended and double-ended transition state search, the molecular dynamics simulation with and without restraints. A few examples are given to illustrate the efficiency and capabilities of LASP code. Our ongoing efforts for code developing and G-NN potential library building are also presented.

This article is categorized under:
    Software > Simulation Methods

KEYWORDS
LASP, neural network, potential energy surface, SSW

1 | INTRODUCTION

Numerous phenomena in physical science involve molecules/materials to make and break chemical bond at the minority/defective sites in nanostructures, such as the crystallization and phase transition of materials (with the change of coordination) and heterogeneous catalytic reactions. The direct atomistic simulation of these processes remains highly challenging even with the modern supercomputing facilities, not least because of the deficiency of theoretical methods and the lack of efficient computational software. For nanomaterial systems with thousands of atoms, the traditional atomistic simulation techniques either suffer from poor computation efficiency or fail in accuracy in describing the potential energy surface (PES). The computational cost for the electronic structure calculations, for example, scales typically over $O(N^2)$ and the simulation is thus routinely limited to hundreds of atoms or less. Empirical force field potentials, on the other hand, has the intrinsic problem in predicting the complex PES owing to the relatively simple function forms that constitute the force field. This software focus introduces our recently-developed LASP (Large-scale Atomistic Simulation with neural network Potential) program, an atomistic simulation package targeted for solving the complex PES problems in large material systems.

LASP project initiates in the early of 2018. The first objective is to merge two major simulation tools developed by Liu group in the past decade into a single package for the better and simpler usage, namely (i) the global PES exploration based
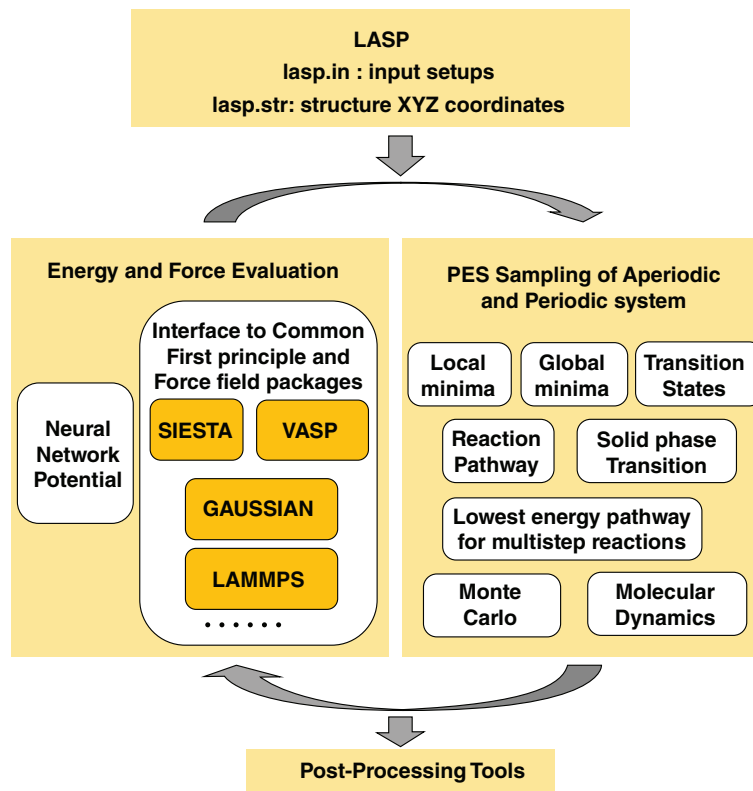
FIGURE 1 Architecture and the modular map of LASP code

on the stochastic surface walking (SSW) method[1,2] and (ii) the fast PES evaluation using global neural network (G-NN) potential.[3,4] The aim is soon expanded to develop a general-purpose atomistic simulation package using G-NN potentials. G-NN potentials utilized in LASP, being different from the NN potentials developed by others, are created by learning the first principles dataset that is obtained from SSW global PES exploration. They are thus more robust and transferable, suitable for investigating global PES problems starting from scratch. The architecture, SSW plus G-NN, has now been utilized to sample unbiasedly and globally the multi-dimensional PES of complex materials,[5,6] leading to solve challenging problems ranging from structure determination to reaction pathway prediction.

While the combination of SSW and G-NN is the key feature to distinguish with other packages, LASP is now shaped more towards a software platform for many purposes, not only for atomistic simulation but also for PES data building and exchange and even for G-NN potential generation. A large set of powerful simulation techniques has been assembled into LASP program (the current version 1.6) to simplify the usage and to enlarge the scope of the current atomistic simulation. This can be seen clearly in Figure 1, which overviews the modules in LASP. Apart from the G-NN computation for energy and force evaluation, LASP implements standard data-exchange interfaces to connect with common PES evaluation packages, which allows for the PES data generation using quantum mechanics calculation. To consider the finite temperature and pressure conditions in PES exploration, LASP also provides the standard and restrained molecular dynamics (MD) functionalities, in addition to the structure local/global geometry optimization and transition state (TS) search. All these functionalities can be switched on and off in LASP with simple keywords, and thus a variety of different tasks can be fulfilled facilely within one platform, such as SSW-NN PES global exploration, PES data generation using quantum mechanics calculations, G-NN potential creation and molecular dynamics based on G-NN potential.

For the PES exploration, LASP has implemented the following techniques:

(i) SSW method for structure search[1,2,7];

(ii) SSW method for reaction pathway sampling (SSW-RS)[8];

(iii) Transition state search based on constrained Broyden dimer method (CBD)[9] and double-ended surface walking method (DESW).[10,11]

(iv) Standard isothermal and/or isobaric molecular dynamics and restrained molecular dynamics for sampling rare events.

To be more specific, all PES exploration methods have two flavors, the fixed cell and the variable cell. The former does not allow the change of simulation box, and thus mainly used for molecular systems, surfaces (slab) calculations. The latter can cope with the variation of lattice (volume), and thus can simulate solid materials with dynamic volume (e.g., in solid phase transition, crystallization and growth).

For the energy and force evaluation, LASP has the access to the following options:

(i) G-NN computation;

(ii) interface to common first principles packages, such as VASP,[12] Gaussian,[13] CP2K[14];

(iii) interface to force field packages, such as LAMMPS[15];

(iv) flexible bias potentials that have the mathematical forms specifiable by users.

In the following sections, we will elaborate main functionalities implemented in LASP. The detail of the related theoretical methods can be found in previous literatures.

2 | GLOBAL OPTIMIZATION AND PATHWAY SAMPLING USING SSW

2.1 | SSW global structure search

The SSW global optimization method was invented in 2013.[1] In the past 5 years, it has been applied to many different PES problems, ranging from clusters to molecules and to solids.[16] The method takes the energy and force (the first derivative) of a structure as input to explore the global PES smoothly and unbiasedly. The SSW method can overcome the high barrier separating minima on PES via an automated and adaptive climbing mechanism, which adds consecutively bias potentials along a random soft mode direction. A pseudocode for performing a single SSW step is show in Figure 2, showing the five sequential steps in SSW PES exploration. In detail, an efficient mode softening technique, namely the constrained Broyden dimer (CBD) with bias potential method, is used to fast optimize the random-generated movement direction. The mode softening is found to be essential for eliminating any unphysical movement towards energetically forbidden area of PES (e.g., the collision of atoms). The Metropolis Monte Carlo method[17,18] is utilized to accept or reject a new minimum obtained from a fully structural optimization. LASP implements two versions of SSW global optimization, the fixed cell and the variable cell, which differs in the treatment of the lattice degrees of freedom.[1,2]

2.1.1 | Fixed cell

The fixed cell SSW only varies the atomic coordinate during the PES exploration.[1,7] Since the lattice is fixed, the method is designed to explore the configuration and reaction space of clusters and molecules in the gas phase or on the surfaces. The PES sampling is controlled mainly by the width and the total number of the Gaussian bias potentials. LASP computes the point symmetry group and the modified Steinhart order parameter[8] for each minimum structure, which can be utilized to distinguish different minima.

2.1.2 | Variable cell

The variable cell SSW is designed to explore the PES of materials in periodic lattice.[2] Both the atom and the lattice degrees of freedom (3 N + 6 in total) are allowed to change during SSW search. The variable cell SSW is often utilized to predict the structure of crystal and to search for defective sites in solids. In the variable cell SSW search, the lattice and atom movement are coupled via a generalized coordinate, six lattice parameters plus the scaled fractional coordinates of atoms in cell. The PES sampling is thus controlled by the width and the total number of the Gaussian bias potentials acting both on atom and on lattice. As the lattice movement is generally more effective in finding stable crystal structures, the lattice-dominated movements (atoms only follow the lattice movement) may be specified in practice. LASP computes the space group and the modified Steinhart order parameter[8] for each minimum structure, which can be utilized to distinguish different minima.

2.1.3 | Rigid body restraint

The rigid body restraint is implemented using axis-angle representation[19,20] in LASP program to accelerate the global structure search. It can be utilized in both the fixed cell and the variable cell tasks. The total number of rigid bodies and the atom indexes in each rigid body need to be specified in simulation. The technique is particularly useful for structures containing rigid molecular groups, such as those in molecular crystals and biomolecules. The rigid components in a structure can be molecule as a whole (e.g., methane, benzene), or a group/fragment (e.g., methyl group). Only the translation and rotation movements are allowed in each rigid body.

2.2 | Transition state location

The transition state (TS) that links initial state and final state is a key measure for reactivity. The location of TS is of significance for reaction prediction. LASP implements two different TS search methods developed by Liu group, the single-ended and the double-ended methods,[9,10] which differ in the input. The single-ended method starts from a guessed structure,

```fortran
call update_energy_force()

select case (SSWstatus)
  case ("NewStart")
    call generate_random_mode()
  case ("CBD")
    call soften_mode()
  case ("moveds")
    call move_ds()
  case ("Climb")
    call climb()
  case ("Allopt")
    call opt_structure()
end select
```

FIGURE 2   Pseudocode for performing a SSW step in LASP code written mainly in Fortran 90. For a SSW simulation, the energy and forces of the current structure (update_energy_force) is first required. Five sequential steps are taken according to the SSW status: ("NewStart") step is to generate a random movement direction; ("CBD") step is to soften the movement direction by using the constrained Broyden dimer with bias potential method; ("moveds") step is to extrapolate atoms/cells coordinates along the direction with the step size of a Gaussian ds; ("Climb") step is to locally relax the structure on the modified PES with added bias potentials; ("Allopt") is to perform the fully geometry relaxation and finally reach a new minimum

presumably close to the TS, and requires a pair of structure to define the reaction coordinate, whereas the double-ended method starts from the initial and the final state pairs.

Since the TS by definition is a saddle point on PES, all TS location methods need to identify the imaginary mode, an eigenvector of Hessian matrix, associated with the desired reaction. Thanks to the finite difference approach utilized to approximate the second derivative, both TS location methods in LASP do not compute Hessian matrix explicitly, and requires only energy and force to identify TS.

2.2.1 | Constrained Broyden dimer method

The single-ended TS search method implemented in LASP is the constrained Broyden dimer method (CBD) developed in 2010,[9] originating from the dimer method[21] and the hybrid eigenvector-following method.[22,23] The CBD method is utilized to identify the reaction mode for a given structure and the Broyden method is then used to translate a structure towards the TS. The reaction mode is initially constructed from a pair of structures from input, where a structure match procedure is utilized to remove the redundant translation and rotation displacements of the pair. As the reaction mode by definition is an imaginary mode along the reaction direction, it is possible to locate the TS by following the mode when the input structure is close enough to the TS. The success of the CBD method to identify a TS depends on how close in geometry the input structure is to the TS. The TS location is also affected by the parameters to control the dimer rotation accuracy to locate the reaction mode, and the stepsize of the translation. Typically, with a good guess of the initial structure, the TS may be identified within 50 energy/force evaluation steps.

2.2.2 | Double-ended surface walking method

The double-ended TS search method implemented in LASP is the double-ended surface waling (DESW) method developed in 2012–2015.[10,24] The double-ended method takes a structure pair as input (no need for the initial guess of TS) and thus can be utilized for automatic pathway sampling once the initial and the final states are provided (see below for SSW pathway sampling). The DESW first builds a pseudopathway connecting the initial and final states and then utilizes the CBD method to identify the TS where the maximum energy image along the pseudopathway is taken as the initial guess structure. In building the pseudopathway, the reaction mode is continually updated from the pair of structures at the two ends of the pathway, where a structure match procedure is utilized to remove the redundant translation and rotation displacements of the pair. The DESW method does not need to predefine the number of images along the pathway and the length of a pseudopathway is self-adaptive to the reaction. For difficult cases where the minimum energy pathway may be required for TS identification, the images along the pseudopathway can be relaxed subject to their force perpendicular to the reaction pathway.

The efficiency of DESW method depends much on the nature of reaction. For elementary reaction with only one TS in between the reaction pairs provided, DESW method can generally find the TS facilely. On the other hand, if the pathway connecting the structure pairs contains more than one TSs, the program will try to locate the TS corresponding to the highest barrier along the pathway. For such situations, the TS search can be more difficult and may even identify the TS that is not directly connected to the structure pairs provided. The extrapolation of TS towards the neighboring minima should be performed to confirm the TS and its associated reaction. To this end, LASP outputs a series of structures (animation snapshots separated by a small distance) along the imaginary eigenvector associated with the identified TS. These structures can be used conveniently to extrapolate the TS structure towards the nearby minima.

2.2.3 | Variable cell TS search

Both TS methods mentioned above have two versions, the fixed cell and the variable cell.[11] The fixed cell TS search is typical for molecular reactions. The variable cell TS location is most useful for finding the homogenous solid phase transition pathway. In the variable cell TS search, a generalized coordinate combining the lattice and atom coordinate is used to couple the lattice and atom degrees of freedom, which can produce the imaginary mode containing both lattice and atom movements. The method has been utilized for identifying a number of solid phase transition reactions, which can yield rich information on the solid phase transition, for example, the coherent interface, if present, between two phases that is observable in experiment.

2.3 | SSW pathway sampling

For its smooth structure perturbation, SSW has the power to identify the geometrically close minimum pairs on global PES. By combining the SSW method with the DESW method, it is now possible to achieve the automatic reaction pathway sampling. SSW pathway sampling can therefore be considered as an extension of the SSW global structure search.[25] They share the same global structure search module. In SSW pathway sampling, a large number of reaction pairs need to be collected first by SSW global search, and each of them then serves as the reaction coordinate for finding the reaction pathways connecting them by DESW. Different from other path sampling approaches, such as the transition path sampling,[26] artificial force induced reaction[27] and discrete path sampling approach,[28,29] the SSW-RS implemented in LASP expands the reaction network by the global PES exploration SSW method, which effectively speeds up the search for a large range of minima on the PES. The DESW method is only utilized for the connection of pathway between interested minima. Same as the SSW structure search, SSW pathway sampling also has two versions, namely the fixed cell and the variable cell.

2.3.1 | Fixed cell pathway sampling

The fixed cell pathway sampling is designed mainly for sampling of the reaction space of molecular reactions. By starting from a given molecule, the SSW pathway sampling is utilized to identify new reaction that can be distinguished by a number of descriptors, including total energy, the bond matrix, the point group, the chirality and the modified Steinhart order parameter. If a molecule differs from the initial molecule, for example, the change of the bond matrix due to the make/breaking of new chemical bonds, the structure pair will be recorded, the so-called reactant/product (R/P) pair. After the SSW pathway sampling, users can use the DESW module in LASP to locate the reaction pathway for all R/P pairs, from which the lowest energy pathway can be identified. For the molecular reactions, the pathway sampling generally needs to collect thousands of R/P pairs in order to identify the lowest energy reaction pathway.

2.3.2 | Variable cell pathway sampling

The variable cell pathway sampling can be utilized to sample the solid–solid phase transition pathways. Different from the fixed cell pathway sampling, the variable cell pathway sampling distinguishes a new minimum according to the total energy, the space group, and the modified Steinhart order parameter. For the subsequent reaction pathway location, the variable cell DESW should be used to find the solid phase transition pathways. To give a rough estimate, the crystal-to-crystal phase transition involving ~24 atoms per cell may require hundreds of R/P pairs from the pathway sampling in order to identify the lowest energy reaction pathway.

3 | MOLECULAR DYNAMICS

MD is a standard method to explore the PES of material at finite temperature and pressure. LASP supports full-fledged MD simulation modules, including the micro-canonical, canonical, isothermal-isobaric ensembles (NVE, NVT, and NPT)[30] and enhanced sampling with specifiable restraints. The Verlet algorithm is utilized to integrate the equation of motion. The thermostats and barostats utilized in NVT and NPT adopt Nose-Hoover[30,31] and Parrinello-Rahman[32] methods, respectively.

To enhance the sampling of rare events, LASP implements a flexible input interface for the addition of bias potentials, which describes the reaction coordinate of desirable reactions. The bias potential can be specified as a combination of simple mathematic operations over geometry parameters, including the atomic Cartesian coordinate, the bond distance, the bond angle, the coordination number, the modified Steinhart order parameters and power type structural descriptors. The mathematic functions implemented can be any forms derived from the sum, exponential, trigonometric functions and so on. Within this framework, one may even construct the complex reaction coordinates, such as the energy gap reaction coordinate, without any coding from users. By using the enhanced sampling with restraints, the free energy profile can be constructed by utilizing umbrella sampling.[33]

4 | COMPUTATION USING G-NN POTENTIAL

The G-NN potential in LASP follows the high-dimensional neural network architecture proposed by Behler and Parrinello in 2007,[34] in which the total energy is decomposed into the sum of individual atom energy. No explicit charge on atoms is involved in our current NN implementation. The long range interaction has been learned effectively by NN based on the local and medium range structural patterns (e.g., within 8 Å). The feedforward NN is utilized to predict the atomic energy, where the input nodes of the NN are a set of coordinate-based structural descriptors that should maximally distinguish structures on PES. The introduction of global PES exploration method (i.e., SSW method) in constructing the global dataset is a major contribution from our group,[3] which provides a practical route for building a general and transferable global NN potential. Our G-NN is established by self-learning the DFT dataset that are generated by SSW-NN simulation. The final size of training set depends very much on the system (e.g., the PES complexity), which typically contain a few ten thousand to a few hundred thousand configurations. During the self-learning procedure, the NN potential is continually tested by SSW simulation: the structural configurations that produce unrealistic energetics, unrealistic forces and second derivatives, and outranged structural descriptors can be identified and added to the dataset for learning. At the moment, the learning cost for a three-element system (e.g., Ti, O, H system) G-NN potential is, typically, 1–2 weeks' computation of ~1,000 CPU cores. Despite the initial painful effort in the potential generation, the NN computation is several orders of magnitude faster than first principles calculations, and thus is paid off for large-scale atomistic simulation. Some examples will be provided in the next section to demonstrate the capability of G-NN for material simulation.

4.1 | Power type structural descriptors

Sensitive structural descriptors are key for constructing accurate and transferable NN potential. To fit the global PES dataset from SSW, LASP utilizes a new set of structural descriptors, namely the power type structural descriptors (PTSDs), which consist of up to four body terms.[4] The PTSDs take the translation- and rotation-invariant geometrical parameters, including the atom pair distance, the angle in three atoms, the torsion angle among four atoms as the variables, and combine them using the power, cosine and spherical functions to construct the n-body (two-, three- and four-body) structure descriptors. The spherical harmonic function is found to increase significantly the quality of structural descriptors to distinguish complex atomic environment. The four-body term with torsion angles is often important for describing molecular configurations, for example, with different molecular packing patterns. The PTSDs have been applied successfully to build the G-NN potentials for different materials (e.g., covalent, metal, oxide solids, and molecules) with typical root mean square errors of energy ~10 meV/atom and force ~0.2 eV/A over the global training dataset.

4.2 | Parallel implementation

The G-NN computation has been optimized in LASP towards massively parallelization.[35] From the NN architecture, the G-NN calculation should scale linearly with the number of atoms. Our recent efforts have been devoted to the G-NN computation for situations when CPU/core numbers exceed significantly the number of atoms. This is done via a hierarchy parallelization scheme based on Message Passing Interface (MPI). A three-level strategy is designed to distribute the computational task into processors as even as possible, namely firstly over atom, then over PTSDs and finally over the n-body functions (group unit) in PTSDs (Figure 3). The optimal parallelization scheme can be invoked automatically by LASP according to the number of atoms in system, the total number and types of PTSDs and the available processors.

4.3 | Element-based G-NN potential library

It is our goal to build a large library of G-NN potentials in the coming years, which should cover common materials and catalytic reactions on them. Our tests for a large number of different systems have shown that the current NN architecture plus PTSD is able to achieve a good accuracy for the global PES of a wide range of materials, ranging from covalent, to metallic, to ionic and to molecules. The most time-consuming part to date is the generation of the global PES dataset from accurate quantum mechanics calculations.

The current G-NN potentials in LASP are trained based on SSW global optimization data that are computed from first principles DFT (GGA-PBE) calculations using VASP program with projector augmented wave pseudopotential (PAW).[36] By default, all datasets are computed using the same kinetic cutoff (450 eV) with automatic K-point generation scheme at 25 Å^{-1} lengths. This facilitates greatly the dataset reuse considering that the complex multi-element materials share the same element or compositions with single- or dual-element materials. For example, one can generate the global dataset for water (H_2O) and the dataset can be directly utilized in generating the G-NN potential for TiO_2 material in water solution.
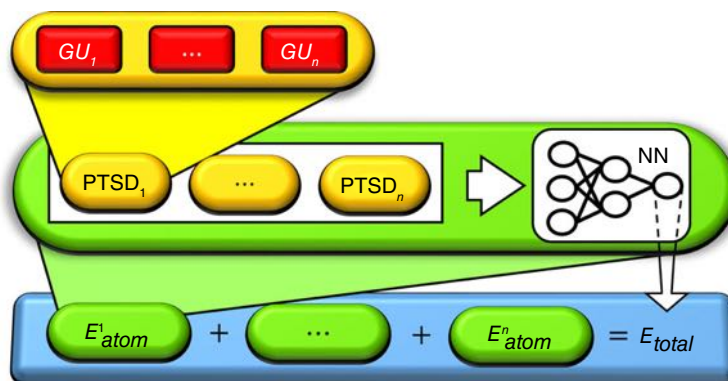
FIGURE 3 Scheme of the parallel computation hierarchy in the PTSD based HDNN. The total energy of a system (blue) is a sum over all atomic energy (green). The atomic energy involves descriptor computation (orange) and an NN evaluation (white). The descriptor computation includes contributions from each n-body function (group unit) computation (red). (Copyright: 2018, John Wiley and Sons)

To date, the G-NN potentials have been generated for more than 50 different systems, ranging from single element (C, Si etc.) to metals/oxides (TiO, PtO, NiO, etc.) and to molecules in reactions up to four elements (CHON, CuCHO, and PtCHO). Each G-NN potential should in principle work for any combination of the elements in the potential. For example, an TiO potential can describe the PES for Ti metal, TiO_x oxide (Ti_2O_3 and TiO_2) and the mixture of metal and oxide. The chemistry for different oxidation state can be learned by the NN training provided with the suitable global dataset.

5 | EXAMPLES

5.1 | Efficiency of G-NN computation

The typical G-NN computation efficiency is illustrated in Figure 4. Taking the G-NN of boron crystal as example, which involves intensive NN computation (a large set of PTSDs and a large network), we find that the G-NN computation can be at least two orders of magnitude faster than the DFT calculations based on the plane wave functional even for small systems (28 atoms per cell). Our testing results further show that for system more than 1,000 atoms, the NN calculations can be at least three orders of magnitude faster than DFT.[35]
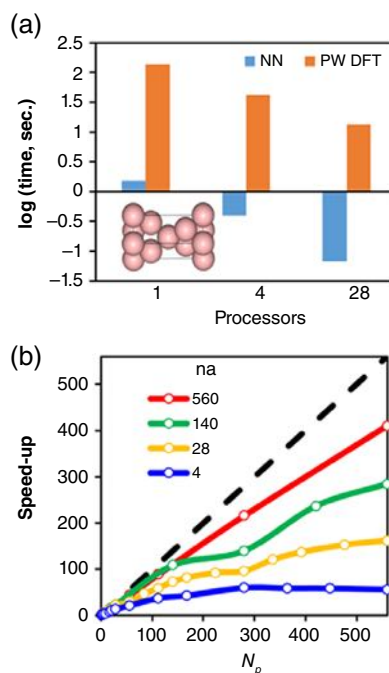

FIGURE 4 (a) Comparison of the computation time for a 28-atom boron crystal using NN and DFT with plane wave basis set (PW DFT). The x axis is the logarithm of time in the unit of seconds. The inset shows the structure of the boron crystal. (b) Speed-up of G-NN computation of boron crystals. The x axis (N_p) is the number of processors and the y axis is the speed-up. (Copyright: 2018, John Wiley and Sons)
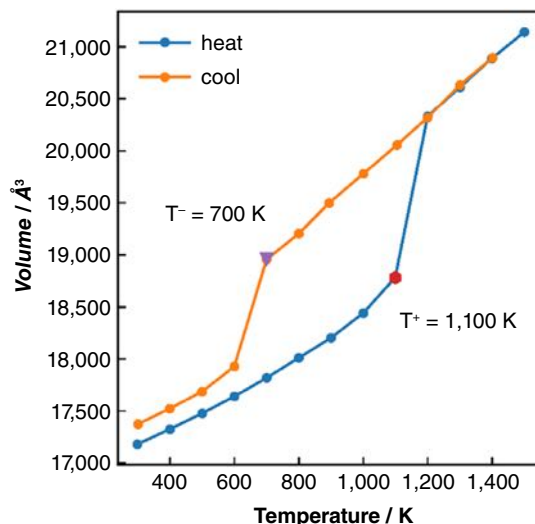
FIGURE 5 Volume versus temperature plot from molecular dynamics simulation of the melting (blue) and refreezing (orange) of Al metal. A complete hysteresis of volume forms during the continuous heating-cooling process. The transition temperatures T^+ is 1,100 K and T^- is 700 K

In addition, the G-NN calculation can be significantly speeded up if many processors are available. For a very small periodic system, that is, 4 boron atoms per cell, the maximum speed-up is 60 with 280 processors, in which the computation for each atom is distributed over 70 processors. For a large system, that is, 560 atoms in the unit cell, the speed up on 560 processors is 411, equivalent to a parallel efficiency of 73%, suggesting a good linear scaling ability for the atom-based parallelization.

To demonstrate the numerical accuracy of NN parallel computation, we have performed a micro-canonical (NVE) MD simulation using LASP for a 1,120-atom boron crystal for 0.1 ns (1 fs per timestep) on 560 cores (referred to Intel(R) Xeon(R) CPU E5-2695 v3 @ 2.30GHz CPU hereafter). The task takes 6 hr and 17 min in total and a good numerical stability is achieved.

5.2 | Melting point of Al

The emergence of G-NN computation allows for simulating complex materials with much longer time and much larger size. Here we demonstrate that G-NN can be used to directly simulate the melting/crystallization of material with the straightforward heating–cooling procedure.

Taking Al metal as the example, we have computed its melting point using the isothermal-isobaric ensemble (NPT) MD simulation in LASP. We start from the faced-centered cubic Al, most stable form of Al at room temperature, using a 1,024-atoms orthorhombic supercell (lattice parameters a = b = 16.1165 Å, c = 64.6962 Å). The heating process sweeps from 300 to 1,500 K, and that for cooling process is from 1,400 to 300 K, with the step temperature of 100 K. The external pressure is set to zero, the MD time step is 1.0 fs, and for each temperature the simulation runs 40 ps. The final equilibrium volume versus temperature curves is shown in Figure 5, where the transition temperature for the first-order phase transition occurs at 700 K in the cooling curve (T^-) and 1,100 K in the heating curve (T^+). Using the homogeneous nucleation melting theory, the melting point T_m can be estimated as 922.5 K ($= T^+ + T^- - \sqrt{T^+ T^-}$).[37,38] The result is only slightly lower than that from experiment (933.47 K), and outperforms the widely tested Mei-Davenport potential (800 K).[39] The whole calculation takes only 6.5 hr using 364 CPU cores, involving 10^6 energy/force evaluations.

5.3 | SSW global optimization for β-B

The exploration for the global PES of material has been a formidable task due to the huge computational cost. For example, because of the self-doping of B atoms at the interstitial lattice sites, the possible configuration of β-B crystal, the most stable form of boron, reaches to an astronomical number and thus the exact atomic structure of β-B crystal was uncertain for decades. With the advent of G-NN computation, it is now possible to resolve the thermodynamically stable configurations of β-B crystal. We have utilized SSW global optimization based on G-NN to establish the global PES of β-B crystal. In detail, massive SSW runs (more than 20) were carried out independently, with 10,000 minima to visit in each run. The initial structures are chosen from likely β-B configurations in rhombohedral cell containing either 105, 106 or 107 atoms per cell. By removing the duplicated minima, a set of 15, 293 distinct β-B minima is obtained to build the global PES. Figure 6 shows the PES contour
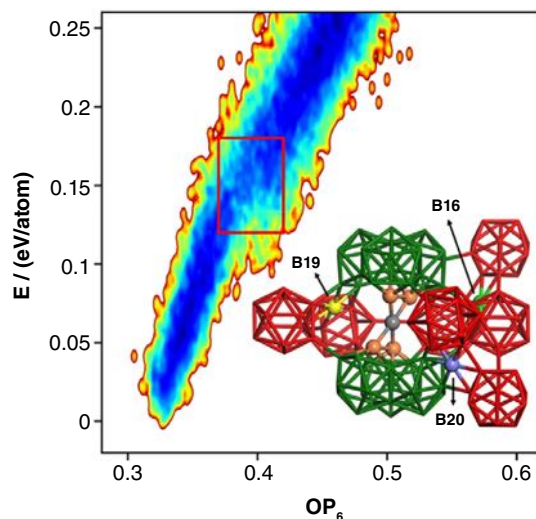
FIGURE 6 The energy-OP$_6$ contour map for the β-B minima from the SSW global optimization using the NN PES. OP$_6$ is the distance-weighted Steinhardt order parameter with degree set to 6, and the density of states is indicated by color. The energy of β-B global minimum is set as zero. The red box indicates the boundary above which the B$_{28}$ cages start to melt. The inset shows a typical structural fragment of β-B, including B$_{12}$ (red) and B$_{28}$ (green) cages, and important partial occupied sites, that is, B16 (green ball), B19 (yellow ball), B20 (purple ball)

plot using the structural order parameter (OP$_6$) against energy for all these distinct minima, and a representative atomic structure of β-B is also shown in the inset.

As shown, the β-B global PES is separated to two high density of states regions (deep blue zones) by an energy gap at ~0.15 eV/atom. It was found that all structures below the energy gap contains intact B$_{12}$ (red) and B$_{28}$ (green) cages as the skeleton, and the partially occupied sites (POS) are spread around the tetrahedral void surrounded by every four neighboring B$_{12}$/B$_{28}$ cages. The dominant POSs in β-B are labeled out in Figure 6, including B16 (green ball), B19 (yellow ball) and B20 (purple ball). The further DFT refinement of these low energy structures from global NN PES results in an accurate energy spectrum for β-B configurations, from which the general rules for stable permutation isomers were revealed and occupancy rates for the POSs at finite temperatures were finally obtained.[4]

It should be mentioned that ~4×10^7 NN energy/force evaluations are required for building the β-B global PES (200 energy evaluation steps per SSW step) and the overall PES scanning takes 2 days in total of 560 CPU cores.

6 | CONCLUSION

LASP code represents the first atomistic simulation software that combines the state-of-the-art global optimization techniques with global neural network potential. The software aims to become a material simulation platform for multiple purposes, including quantum mechanics PES data generation, NN potential generation and large-scale material simulation from global optimization to long-time MD. While the current version (LASP 1.6) has equipped with a large family of functionalities as introduced above, our group is continuing to make efforts to fulfill the software. Some short-term objectives are listed as follows.

1. To build a large G-NN potential library for a wide range of materials and for catalytic reactions on surface/interfaces.
2. To implement GPU-based G-NN computation.
3. To implement the on-the-fly rigid-body restraints in the global PES search.
4. To implement a standard move class for materials, which can be utilized for SSW and MC modules.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

RELATED WIREs ARTICLE

[GENESIS: A hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecularand cellular simulations](#)

ORCID

Si-Da Huang https://orcid.org/0000-0002-0055-1510
Cheng Shang https://orcid.org/0000-0001-7486-1514
Zhi-Pan Liu https://orcid.org/0000-0002-2906-5217

FURTHER READING

Further information can be found online at http://www.lasphub.com/

REFERENCES

1. Shang C, Liu Z-P. Stochastic surface walking method for structure prediction and pathway searching. *J Chem Theory Comput.* 2013;9(3):1838–1845.
2. Shang C, Zhang X-J, Liu Z-P. Stochastic surface walking method for crystal structure and phase transition pathway prediction. *Phys Chem Chem Phys.* 2014;16 (33):17845–17856.
3. Huang S-D, Shang C, Zhang X-J, Liu Z-P. Material discovery by combining stochastic surface walking global optimization with a neural network. *Chem Sci.* 2017; 8(9):6327–6337.
4. Huang S-D, Shang C, Kang P-L, Liu Z-P. Atomic structure of boron resolved using machine learning and global sampling. *Chem Sci.* 2018;9(46):8644–8655.
5. Guan S-h, Shang C, Huang S-D, Liu Z-P. Two-stage solid-phase transition of cubic ice to hexagonal ice: Structural origin and kinetics. *J Phys Chem C.* 2018;122 (50):29009–29016.
6. Ma S, Huang S-D, Fang Y-H, Liu Z-P. TiH hydride formed on amorphous black Titania: Unprecedented active species for Photocatalytic hydrogen evolution. *ACS Catal.* 2018;8(10):9711–9721.
7. Zhang X-J, Shang C, Liu Z-P. From atoms to fullerene: Stochastic surface walking solution for automated structure prediction of complex material. *J Chem Theory Comput.* 2013;9(7):3252–3260.
8. Zhang X-J, Shang C, Liu Z-P. Pressure-induced silica quartz amorphization studied by iterative stochastic surface walking reaction sampling. *Phys Chem Chem Phys.* 2017;19(6):4725–4733.
9. Shang C, Liu Z-P. Constrained Broyden minimization combined with the dimer method for locating transition state of complex reactions. *J Chem Theory Comput.* 2010;6(4):1136–1144.
10. Zhang X-J, Shang C, Liu Z-P. Double-ended surface walking method for pathway building and transition state location of complex reactions. *J Chem Theory Comput.* 2013;9(12):5745–5753.
11. Zhang X-J, Liu Z-P. Variable-cell double-ended surface walking method for fast transition state location of solid phase transitions. *J Chem Theory Comput.* 2015; 11(10):4885–4894.
12. Kresse G, Furthmüller J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comp Mat Sci.* 1996;6(1): 15–50.
13. Frisch MJ, GWT HBS, Scuseria GE, et al. *Gaussian 09, revision A02.* Wallingford, CT: Gaussian, Inc, 2016.
14. Hutter J, Iannuzzi M, Schiffmann F, VandeVondele J. cp2k: Atomistic simulations of condensed matter systems. *WIREs Comput Mol Sci.* 2014;4(1):15–25.
15. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys.* 1995;117(1):1–19.
16. Shang C, Zhang X-J, Liu Z-P. Crystal phase transition of urea: What governs the reaction kinetics in molecular crystal phase transitions. *Phys Chem Chem Phys.* 2017;19(47):32125–32131.
17. Li Z, Scheraga HA. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc Natl Acad Sci USA.* 1987;84(19):6611–6615.
18. Wales DJ, Doye JPK. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J Phys Chem A.* 1997;101(28):5111–5116.
19. Wales DJ. The energy landscape as a unifying theme in molecular science. *Philos Trans Royal Soc A.* 2005;363(1827):357–377.
20. Rühle V, Kusumaatmaja H, Chakrabarti D, Wales DJ. Exploring energy landscapes: Metrics, pathways, and Normal-mode analysis for rigid-body molecules. *J Chem Theory Comput.* 2013;9(9):4026–4034.
21. Henkelman G, Jonsson H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J Chem Phys.* 1999;111 (15):7010–7022.
22. Munro LJ, Wales DJ. Defect migration in crystalline silicon. *Phys Rev B: Condens Matter.* 1999;59(6):3969–3980.
23. Zeng Y, Xiao P, Henkelman G. Unification of algorithms for minimum mode optimization. *J Chem Phys.* 2014;140(4):044115.
24. Shang C, Liu Z-P. Constrained Broyden dimer method with bias potential for exploring potential energy surface of multistep reaction process. *J Chem Theory Comput.* 2012;8(7):2215–2222.
25. Zhang X-J, Liu Z-P. Reaction sampling and reactivity prediction using the stochastic surface walking method. *Phys Chem Chem Phys.* 2015;17(4):2757–2769.
26. Dellago C, Bolhuis PG, Csajka FS, Chandler D. Transition path sampling and the calculation of rate constants. *J Chem Phys.* 1998;108(5):1964–1977.
27. Maeda S, Morokuma K. Communications: A systematic method for locating transition structures of a + B → X type reactions. *J Chem Phys.* 2010;132(24): 241102.
28. Wales DJ. Discrete path sampling. *Mol Phys.* 2002;100(20):3285–3305.
29. Röder K, Wales DJ. Energy landscapes for the aggregation of Aβ17–42. *J Am Chem Soc.* 2018;140(11):4018–4027.
30. Frenkel D, Smit B. Chapter 6 - molecular dynamics in various ensembles. In: Frenkel D, Smit B, editors. *Understanding molecular simulation.* 2nd ed. San Diego, CA: Academic Press, 2002; p. 139–163.

31. Shuichi N. Constant temperature molecular dynamics methods. *Prog Theor Phys Suppl.* 1991;*103*:1–46.

32. Parrinello M, Rahman A. Crystal structure and pair potentials: A molecular-dynamics study. *Phys Rev Lett.* 1980;*45*(14):1196–1199.

33. Kästner J. Umbrella sampling. *WIREs Comput Mol Sci.* 2011;*1*(6):932–942.

34. Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys Rev Lett.* 2007;*98*(14):146401.

35. Shang C, Huang S-D, Liu Z-P. Massively parallelization strategy for material simulation using high-dimensional neural network potential. *J Comput Chem.* 2019; *40*(10):1091–1096.

36. Kresse G, Joubert D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B.* 1999;*59*(3):1758–1775.

37. Luo S-N, Ahrens TJ, Çağın T, Strachan A, Goddard WA, Swift DC. Maximum superheating and undercooling: Systematics, molecular dynamics simulations, and dynamic experiments. *Phys Rev B.* 2003;*68*(13):134206.

38. Luo S-N, Strachan A, Swift DC. Nonequilibrium melting and crystallization of a model Lennard-Jones system. *J Chem Phys.* 2004;*120*(24):11640–11649.

39. Mei J, Davenport JW. Free-energy calculations and the melting point of Al. *Phys Rev B.* 1992;*46*(1):21–25.