structure. As a direct tool to correlate atomic structure with energy, theoretical simulations have been widely practiced in modelling catalyst structure and even predicting activity, especially with the advent of density functional theory (DFT) calculations.[5] However, DFT simulation is generally limited to hundreds of atoms, and thus fails to explore exhaustively the phase space of complex catalysts, such as amorphous structures, or multicomponent oxide/alloys.[6–8]

To identify the active site of a catalyst from theory, it is essential to explore the complex PES of exposed surfaces. This requires not only a fast and reliable approach to evaluate the energetics of structures but also an efficient method to explore the structural phase space. As for PES calculation methods, most are based on quantum mechanics as represented by DFT calculations, while the empirical force field calculations, and more recently artificial neural network (NN)[9] potential method (see Figure 19.1), despite their limitations in transferability, are also often utilized in material applications.

As catalytic conversion involves chemical bond making and breaking, DFT calculations have been the most popular way to provide an accurate
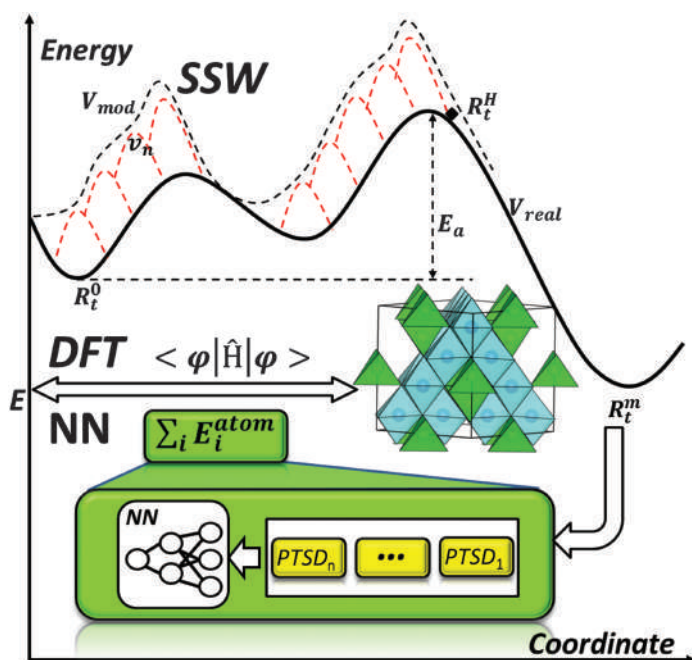

Figure 19.1 Combination of NN potential with SSW global optimization. The total energy, which is the sum of all atomic energies, can be obtained from NN potential. The atomic energy involves power-type structural descriptor (PTSD) computation and an NN evaluation. The SSW global optimization is utilized to explore the PES and produce new structures. Reproduced from ref. 27, https://doi.org/10.1063/1.5113673, under the terms of the CC BY 4.0 license, http://creativecommons.org/licenses/by/4.0/.

description of reactions with a reasonable computational cost. The computation speed of DFT is sensitive to the complexity of the employed density-functional (*e.g.* PBE,[10] HSE06[11]) and has a poor scaling (at least *O(NlnN)*). It has great difficulty in exploring the reaction network in large catalytic systems (*e.g.* 4 100 atoms). As a promising alternative, the NN potential method developed in the past decade demonstrates its power in treating complex PES problems, from gas-phase reactions to material dynamics.[12–15] Recently, it was also utilized to solve the structures of heterogeneous catalysts, for example, Pd(O),[16] Pt(H),[17] CuAu,[18,19] CuCeO,[20] and CuZnO.[21] Unlike traditional force field potentials, the NN potential is capable of describing chemical reactions with high accuracy as long as the training dataset contains reactive data, such as the transition state (TS).

Paralleling progress in PES evaluation methods, many PES exploration methods have been developed in past decades. As catalytic reactions and structural reconstruction generally occur above ambient temperatures involving high barrier processes, traditional molecular dynamic (MD) simulation is often not appropriate for active site identification. Instead, global optimization methods, which can overcome the high barrier on PES, are desirable. Common global optimization methods include simulated annealing,[22] genetic algorithm,[23] basin hopping[24] and stochastic surface walking (SSW)[25,26] *etc.*

This chapter outlines recent developments in catalyst PES exploration and active site identification. We will show that the combination of a global NN potential with the global optimization SSW method provides a powerful platform to resolve the catalyst structure, which can finally lead to the prediction of catalyst activity from the first principles dataset.[27]

19.2   Methods

Unlike classical force field potentials that utilize physical models to describe interatomic interactions, the NN potential is a numerical function fitted by a large number of parameters. Consequently, the NN potential has a limited predictability beyond the training dataset. The key to improving the quality of the NN potential, therefore, relies heavily on the PES representability of the dataset. To overcome this limitation of the NN potential, we proposed in 2017 a global-to-global scheme to generate a reliable and robust global-NN (G-NN) potential for material simulation.[28] This scheme combines the SSW global sampling method,[29] the high-dimensional NN (HDNN) for PES description, and a self-learning procedure to expand the dataset and upgrade the G-NN potential. In the following, we elaborate each part of the scheme.

19.2.1   High-dimensional NN Architecture

The artificial NN method was firstly developed to understand signal processing in the brain.[30] In the following decades, NN evolved into a class of powerful algorithms applied to a variety of fields from numerical prediction

and pattern recognition to data classification. It is most renowned for the powerful ability to establish the functional relationship between independent variables and target (dependent) values *via* non-linear "black box" data processing.

As an important tool for function fitting, the standard feed-forward NN (FFNN) has been utilized naturally for constructing the PES in small molecular systems, *e.g.* for gas-phase dynamics.[12–15] A major drawback of the traditional FFNN potential is that it is difficult to extend to systems of different size (variable number of atoms). A breakthrough was achieved in 2007 by Behler and Parrinello who implemented an HDNN architecture for material simulation. The method is applicable to high-dimensional systems containing thousands of atoms.[31–33] In the approach, the total energy E_{tot} of the system is written as the summation over all atoms (see eqn (19.1)), see Figure 19.2.

Each atom is represented by an atomic FFNN, where the input layer is a series of structural descriptors to represent the atomic bonding environment and the output layer yields the atomic energy E_i (i indexes atoms). This HDNN scheme is applicable to systems with any number of atoms as the total energy is a linear summation of all atomic energy.

$$E_{tot} = \sum_i E_i(\mathbf{R}) \tag{19.1}$$

The input nodes that are used to distinguish the atomic chemical environment are critical for the performance of the NN potential. They utilize the radial and angular information from neighbouring atoms to represent the multi-dimensional interaction between atoms. A good structural
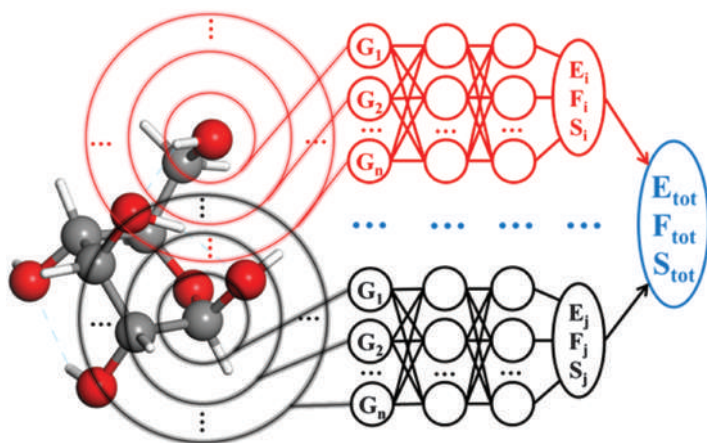

Figure 19.2 The atom-centred structure description scheme in the high-dimensional NN architecture. Each element has a different NN, and each atom is distinguished by its element and structure descriptors centred on it.

descriptor set must be a single-valued function so that different atomic environments yield different values, and they must be continuous and differentiable to enable the calculation of analytic derivatives of energy.[34–36] More details of structure descriptors are discussed in the next subsection.

For one set of structural descriptors, the atomic force and static stress tensor matrix element s_{ab} can be analytically derived according to eqn (19.2), where the force component $F_{k,a}$, $a \frac{1}{4} x$, y or z, acting on the atom k is the derivative of the total energy with respect to its coordinate $R_{k,a}$. By combining with eqn (19.2), the force component can be further related to the derivatives of the atomic energy with respect to jth structural descriptors of atom i, $G_{j,i}$.

$$F_{k;a} \frac{1}{4} \quad \frac{@E^{tot}}{@R_{k;a}} \frac{1}{4} \quad \overset{X}{\underset{i;j}{}} \quad \frac{@E_i}{@G_{j;i}} \frac{@G_{j;i}}{@R_{k;a}} ; \qquad (19.2)$$

Similarly, the static stress tensor matrix element s_{ab} can be analytically derived using the pairwise distance according to eqn (19.3).

$$s_{ab} \frac{1}{4} \quad \frac{1}{V} \overset{X}{\underset{i;j;d}{}} \quad \frac{\check{o}r_d \flat \check{o}r_d \flat}{r_d} \frac{@E_i}{@G_{j;i}} \frac{@G_{j;i}}{@r_d} ; \qquad (19.3)$$

where r_d and r_d are the Cartesian distance vector and its module between atom i and j, respectively, and V is the volume of the structure.

19.2.2 Structural Descriptors

As NN potential correlates the geometry of structure with its energetics, one would wonder what kinds of structural descriptors can achieve the best performance to describe the multi-dimensional PES. Behler and Parrinello suggested a series of rotation-invariant symmetry functions as the structural descriptors, which are the mathematic functions of intrinsic coordinates (pair distance, angles) multiplying the cutoff radial function.[31] Following the same terminology of an empirical forcefield, these functions can be distinguished as two-body terms, three-body terms, and so on. As the radial parts of these functions are Gaussian functions, we label them as Gaussian-type structure descriptors (GTSD). The mostly used two-body G^2 and three-body G^4 functions are described in eqn (19.4)–(19.6).

$$f_c\ R_{ij}\ \frac{1}{4} \quad \left(\begin{matrix} 0{:}5 & \tan h^3\ 1 & \frac{r_{ij}}{r_c} \\ 0\ \text{for}\ r_{ij}\ 4\ r_c \end{matrix}\right); \quad \text{for}\ r_{ij} \quad r_c \qquad (19.4)$$

$$G_i^2\ \frac{1}{4} \quad \overset{X}{\underset{j a\ i}{}} \quad e^{\ Z\check{o}\ r_s\ \flat}\ f_c\ r_{ij} ; \qquad (19.5)$$

$$G_i^4\ \frac{1}{4}\ 2^{1\ z} \quad \overset{Xll}{\underset{j;ka\ i}{}} \quad 1\ \flat\ l\ cosy_{ijk} \quad {}^z\ e^{\ Z\ r_{ij}^2\ \flat\ r_{ik}^2\ \flat\ r_{jk}^2}\ f_c\ r_{ij}\ f_c\check{o}r_{ik}\ \flat\ f_c\ r_{jk} ; \qquad (19.6)$$

where r_{ij} is the inter-nuclear distance between atom i and j, y_{ijk} is the angle centred at i atom with j, k being neighbours (i, j, k are atom indices). The key ingredients in GTSD are the cutoff function f_c that decays to zero beyond the r_c (see eqn (19.4)), the Gaussian-type radial function, and the trigonometric angular functions. By changing five parameters, r_c, r_s, ζ, z, and l, a set of G^2 and G^4 functions can then be generated; these serve to distinguish the atomic environment of the centre atom i.

To improve the sensitivity of structural descriptors in learning global PES, we have designed a series of new structural descriptors, power-type structural descriptors (PTSD), see eqn (19.7)–(19.12).

$$S_i^1 = \sum_{j \neq i} R^n(r_{ij}) ; \tag{19.7}$$

$$S_i^2 = \left[\sum_{m=-L}^{L} \left(\sum_{j \neq i} R^n(r_{ij}) Y_{Lm}(r_{ij}) \right)^2 \right]^{\frac{1}{2}} \tag{19.8}$$

$$S_i^3 = 2^{1-\zeta} \sum_{j,k \neq i} (1 + \lambda \cos y_{ijk})^\zeta \, R^n(r_{ij}) \left[R^m(r_{ik}) + R^p(r_{jk}) \right] ; \tag{19.9}$$

$$S_i^4 = 2^{1-\zeta} \sum_{j,k \neq i} (1 + \lambda \cos y_{ijk})^\zeta \, R^n(r_{ij}) R^m(r_{ik}) ; \tag{19.10}$$

$$S_i^5 = \left[\sum_{m=-L}^{L} \left(\sum_{j,k \neq i} \left[R^n(r_{ij}) R^m(r_{ik}) + R^p(r_{jk}) \right] \left[Y_{Lm}(r_{ij}) + Y_{Lm}(r_{ik}) \right] \right)^2 \right]^{\frac{1}{2}} ; \tag{19.11}$$

$$S_i^6 = 2^{1-\zeta} \sum_{j,k,l \neq i} (1 + \lambda \cos d_{ijkl})^\zeta \, R^n(r_{ij}) R^m(r_{ik}) \left[R^p(r_{il}) \right] ; \tag{19.12}$$

In PTSD, S^1 and S^2 are two-body functions, S^3, S^4, and S^5 are three-body functions and S^6 is a four-body function. The replacement of the Gaussian function in GTSD by the power function in PTSD has several advantages: (i) the computational cost in numerical calculations is reduced; (ii) the adjustable parameters are reduced from two (r_s, ζ) to one (n), which simplifies the search for optimal parameters for the two-body function; (iii) the power function when combining with the decaying cutoff function can create radial distributions with flexible peak and shape, which fulfils a similar purpose to the Gaussian function; (iv) the introduction of different powers (n, m, p) in the three-body function can couple conveniently different radial distributions.

19.2.3 Neural Network Training

Once the network architecture is settled, the next step is to determine the weights and biases (NN parameters) in each NN subnet, a process known as the training of the NN. The number of weights and biases is typically

10^4B 10^6 for the standard FFNN with two hidden layers and 40 nodes in each layers. To train such a large number of NN parameters, a performance function J_{tot} is defined (see eqn (19.13)),[28,37] which measures the deviation of the NN output with respect to the training set properties. The training process will minimize J_{tot} until the accuracy of the NN predicted properties reaches the preset criteria.[36,38,39] The training set is a large structure dataset on PES with accurate energetics and forces, most often computed from first principles calculations.

$$J_{tot} = J_E + r J_F + t J_s = \frac{1}{2N} \left(E^{NN} - E^{real} \right)^2 + \frac{r}{6N} \left(F_{k;a}^{NN} - F_{k;a}^{real} \right)^2 + \frac{t}{18} \left(s_{ab}^{NN} - s_{ab}^{real} \right)^2$$

(19.13)

where $r = 1$B 100 and $t = 0.1$B 1. This J_{tot} allows the fitting of all three properties simultaneously. For the global optimization of solids, both forces and stresses must be accurate; the most convenient way to do this is to allow the NN training to fit all three terms, either simultaneously or independently (depending on the adjustable parameters, r and t, in eqn (19.13)).

Many gradient-based optimization algorithms have been used to optimize the network weights and biases, such as stochastic gradient descent, conjugate gradient, Levenberg–Marquardt (LM) *etc.* It is generally accepted that quasi-Newton second-order methods, such as Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) and LM, converge more rapidly to the true minimum for large data sets. It should be mentioned that, compared to the dominant computational effort required to generate the data set using quantum mechanics calculations, the training of the NN is, in fact, not the rate-determining step in the whole NN PES construction procedure.

19.2.4   Data Set Generation and SSW Global Optimization

Undoubtedly, the data set used for training the NN largely determines the quality of the NN PES. Previous work utilizes either local PES data, *e.g.* the MD trajectories of a single reaction, or the combination of local PES data from different structural sources, *e.g.* solids, slabs, and clusters. To increase the transferability of the NN potential, as well as to simplify data set generation, the structures produced from global optimization trajectories appear to be the better and more natural choice, which should represent as differently as possible the chemical environment of atoms.

There are currently a variety of global optimization methods for structure prediction, including, for example, simulated annealing,[22] genetic algorithm,[23] basin hopping,[24] and SSW[25,26] *etc.*, which should be able to search a wide area on the PES, and identify unbiasedly the global minimum, even starting from random structures. Among them, basin-hopping and genetic algorithms transform the PES by overlooking the transition region between minima and thus they are likely to miss important structural patterns at the transition region. On the other hand, simulated annealing and minimum-

hopping are based on extensive MD calculations at elevated temperatures, and both methods are rarely utilized for global search in combination with quantum mechanics (even DFT) calculations. Furthermore, the structure patterns from the MD trajectories being closely related could be overwhelmingly redundant for NN training and an iterative structure selection scheme may be required to produce a compact data set. In our approach, we use the SSW global optimization method for data set generation. In the following, we introduce the method in detail.

The central idea of the SSW method combines bias-potential-driven dynamics[40] and Metropolis Monte Carlo (MC).[41] It manipulates smoothly the structural configuration from one minimum to another on PES and relies on Metropolis MC at a given temperature to decide acceptance of the move. A series of consecutive minimum structures, both local and global, is generated in an SSW simulation, forming a continuous trajectory which can provide key information on the reaction pathways between minima. The method was initially developed for aperiodic systems,[26] such as molecules and clusters, and has been extended to periodic crystals.[29]

Each step in SSW, also termed an MC step, comprises three independent parts: (i) climbing, (ii) relaxation, and (iii) Metropolis MC. As schematically illustrated in a one-dimensional PES in Figure 19.1, such an MC step utilizes the climbing module to move uphill and the relaxation module to locate the minimum. Once a minimum is reached, the Metropolis MC is used to judge whether the structure will be accepted or refused. The climbing procedure lies at the heart of the SSW method and is elaborated below in detail.

The climbing module of the SSW involves repeated bias-potential-driven structure extrapolation and local geometry optimization, which gradually drags $\mathbf{R}_t^0$ to a high energy configuration $\mathbf{R}_t^H$, where "t" is the index of the current MC step (see Figure 19.1). Starting from the current minimum $\mathbf{R}_t^0$, SSW first generates a random direction $\mathbf{N}_t^0$, a normalized vector pointing to a direction to change the current geometry. To enable an unbiased exploration of PES, the initial direction $\mathbf{N}_t^0$ by design combines two randomly generated vectors, the so-called global random mode $\mathbf{N}_t^g$ and the local random mode $\mathbf{N}_t^l$, as eqn (19.14).

$$\mathbf{N}_t^0 = \frac{\mathbf{N}_t^g + k\mathbf{N}_t^l}{\mathbf{N}_t^g + k\mathbf{N}_t^l} \qquad (19.14)$$

where the mixing parameter k controls the relative proportion of these two displacement directions. Specifically, $\mathbf{N}_t^g$ in our implementation is set as a randomly generated normalized vector that obeys the Maxwell–Boltzmann velocity distribution at 300 K, as utilized in standard molecular dynamics, to generate the initial random velocity. As $\mathbf{N}_t^g$ distributes over a group of atoms, it represents a gentle, global displacement of the atoms. By contrast, $\mathbf{N}_t^l$ describes a stiff, local atomic move, and in our implementation, it is set as a collision movement between two distant atoms. For example, the $\mathbf{N}_t^l$ associated with an atom A (*e.g.*, the first atom in a system) and an atom B

(the second atom in the system) can be derived as eqn (19.15), using their coordinates q_A and q_B. The atom pair A and B can be either chosen randomly or learned from previous trajectories, as long as the two atoms are not in close contact (*i.e.*, their distance $4\ 3$ Å).

$$\mathbf{N}_t^I = \begin{pmatrix} 0 \\ q_B \\ q_A \\ 0 \\ \vdots \end{pmatrix} \begin{pmatrix} 0 \\ q_A \\ q_B \\ 0 \\ \vdots \end{pmatrix} \tag{19.15}$$

Because reactions with low barriers generally involve soft normal mode directions, it would be ideal to soften the randomly generated $\mathbf{N}_t^0$ towards one eigenvector of the Hessian matrix with small eigenvalues (not necessarily the lowest one). However, the direct computation of the Hessian is expensive and not affordable in a PES search. To solve this, we have developed a numerical approach, the bias-potential constrained Broyden dimer (BP-CBD) method,[42] to soften $\mathbf{N}_t^0$ using eqn (19.16)–(19.19).

$$R_1 = R_0 + N_m \cdot \Delta R \tag{19.16}$$

$$C = \frac{(F_1 \quad F_0) \cdot N_m}{\Delta R} \tag{19.17}$$

$$V_{R_1} = V_{\text{real}} + V_N \tag{19.18}$$

$$V_N = \frac{a}{2} (\Delta R \cdot N_m \quad N_t^0)^2 \tag{19.19}$$

Following the unbiased dimer rotation method,[43] we have defined two images separated by a fixed distance of R on PES, namely, $\mathbf{R_0}$ and $\mathbf{R_1}$ (see eqn (19.16)). The rotation of the dimer by their force ($\mathbf{F_0}$ and $\mathbf{F_1}$) perpendicular to the dimer vector $\mathbf{N_m}$ using constrained Broyden optimization (CBD)[44] will converge to the softest eigenvector of Hessian, and the local curvature (C in eqn (19.17)) can be computed according to the finite difference equation. The softest eigenvector is, however, often not the desired one for reaction as it corresponds to the translational and rotational modes ($C = 0$) in the basin region. Therefore, a biased rotation scheme as implemented in BP-CBD is developed, in which the potential of $\mathbf{R_1}$ is modified as eqn (19.18), where V_N is the bias potential added to the real PES V_{real} of $\mathbf{R_1}$ that is a quadratic function of coordinates $\mathbf{R_1}$ along $\mathbf{N}_t^0$ (see eqn (19.19)). As long as the parameter a (see eqn (19.19)) is large enough, the biased rotation can guarantee the rotation of dimer will not deviate far from $\mathbf{N}_t^0$. The force due to the bias potential in constraining the dimer rotation can be evaluated straightforwardly.

$$V_{\text{mod}} = V_{\text{real}} + \sum_{n=1}^{G} v_n = V_{\text{real}} + \sum_{n=1}^{G} w_n \exp\left[\frac{(R_t \quad R_t^n \cdot N_t^n)^2}{2 \quad ds^2} \right] \tag{19.20}$$

$$F_{\text{mod}} = F_{\text{real}} + \sum_{n=1}^{N_G} w_n \exp\left[-\frac{\left(\frac{R_t - R_t^n \cdot N_t^n}{2 \, ds^2}\right)^2}{}\right]^{\#} \left(\frac{R_t - R_t^n \cdot N_t^n}{2 \, ds^2}\right) N_t^n$$

(19.21)

On moving from R_t^n to a high energy configuration R_t^H, a modified PES V_{mod} is utilized, as shown in eqn (19.20), in which a series of bias Gaussian potentials v_n (n is the index of the bias potential, $n = 1,2\ldots H$) is added one by one consecutively along the direction N_t^n and thus creates a series of local minima R_t^n along the moving trajectory on the modified PES. Similar techniques have been used in Metadynamics.[40] The local minimum R_t^n is identified using local geometry optimization, where the force can be evaluated according to eqn (19.21). The w and ds in eqn (19.20) control the height and the width of the Gaussian function v_n, respectively. While w can be computed on-the-fly to guarantee the success of the uphill move, ds is left as an adjustable parameter in the SSW simulation. It should be mentioned that N_t^n will always be updated from the initial random direction N_t^0 at each R_t^n and is then refined using BP-CBD rotation.

To sum up, the uphill movement from R_t^n to R_t^H is a repeated procedure containing (i) updating the direction N_t^n at R_t^n; (ii) adding a new Gaussian function v_n and displacing R_t^n along the direction N_t^n by a magnitude of ds; and (iii) relaxing to R_t^{n+1} on the modified PES.

The overall efficiency of the SSW method depends on the choice of the step length of surface walking ds and the maximum number of Gaussian potentials H. A typical value of ds ranges from 0.2 to 0.6 Å, being 10% to 40% of a typical bond length. The maximum number of Gaussian potentials (H) is also system-dependent and in general a number in the range of 6–15 is a sensible choice. With large ds and large H, a large scope of PES can be explored rapidly, but at the expense of resolution on the reaction pathway between minima. Typically, for complex Lennard-Jones clusters (*e.g.*, LJ$_{75}$ with multiple funnels and high barrier), the SSW simulation requires 300–400 energy/force evaluation per MC step with $ds = 0.6$ and $H = 14$. It takes typically 70–80% computational efforts (energy and gradient evaluation) on the climbing and the remaining 20–30% on relaxation to a minimum.

19.2.5 Self-learning Procedure

The self-learning procedure is described briefly as follows, also see Figure 19.3. In the first stage, an initial dataset is constructed by performing in parallel short-time SSW sampling based on first-principle DFT calculations. These DFT calculation are often restricted to small systems (typically below 20 atoms) and with a low accuracy calculation setup to speed up the global PES sampling. After the PES data is obtained from SSW, a small dataset is randomly selected and computed using DFT with a high accuracy calculation setup. This stage produces a dataset with the most common atomic environment for the target PES.
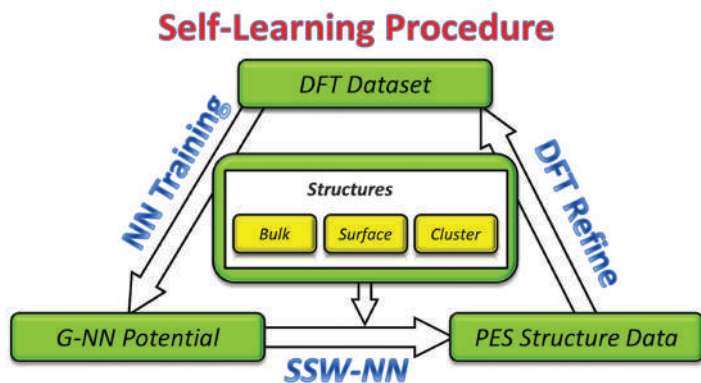
Figure 19.3   Self-learning procedure for the generation of global NN potential
(G-NN).
Reproduced from ref. 27, https://doi.org/10.1063/1.5113673, under the terms
of the CC BY 4.0 license, http://creativecommons.org/licenses/by/4.0/.

Table 19.1   A prototype of hyperparameters used in our previous works.

Hyperparameters	Values
Optimization method	L-BFGS
Weight ratio $1:r:t$	$1:5:0.05$
Number of hidden layers	2B 3
Neurons number in input layer	100B 400
Neurons number in each hidden layer	50B 120
Weight initialization method	Xavier initialization
Type of neuron activation function	tan h^3
Size of training data	20 000B 100 000
Training epochs	B 20 000
Number of weights and bias	10^4B 10^6

The second stage generates an NN potential using the current first
principles dataset. The involved hyper-parameters utilized in our NN
training can be found in Table 19.1. Our G-NN potential typically has two or
three hidden layers and 50B 120 neurons for each layer, which leads to the
number of network parameters (weights and biases in NN) in the range of
10^4B 10^6. The optimization of such a large parameter space is achieved *via*
the quasi-Newton L-BFGS method by simultaneously fitting the total en-
ergy, force, and stress from first principles that constitutes the cost func-
tion J_{tot} in eqn (19.13). The training procedure is controlled by a number of
hyper-parameters, such as the relative weight among energy, force, and
stress ($1:r:t$ in eqn (19.13), network weight initialization method, type of
neuron activation function, training epochs *etc.*[45,46] Due to the structural
variety of the global PES, the typical accuracy for G-NN is 5B 10 meV per
atom for root mean square errors (RMSE) of energy and 0.1B 0.2 eV Å^{-1} for
RMSE of force.

The third stage expands the dataset by carrying out long-time SSW global PES sampling using the current NN potential, the so-called SSW–NN method. These SSW–NN simulations start from a variety of initial structures with different morphology, including bulk, surface and clusters, different chemical compositions, and different numbers of atoms per unit cell. A small additional dataset is thus obtained from the SSW sampling trajectories, containing the structures on PES either randomly selected or exhibiting a new atomic environment (*e.g.* out-of-bounds in structural descriptor, unrealistic energy/force/curvature). After calculating these additional data by DFT, they are added into the training dataset and the whole self-learning procedure returns back to the previous stage.

Typically, after B 100 iterations, a robust and accurate NN potential can be obtained with a compact training set that contains the most representative structures. We emphasize that it is better to adopt consistent and high accuracy calculation setups in constructing the training dataset using first principles calculations, which can benefit greatly data transferability and compatibility between systems; this also helps to reduce the NN fitting error.

19.3 Applications

19.3.1 PES Exploration

The thermodynamics and kinetics of a material are determined by the underlying PES. This can be conveniently and efficiently established using an SSW–NN simulation. In recent years, we have mapped out the PESs for a number of systems, *e.g.* single element crystal (boron),[28] molecular crystal (ice),[47] metal oxide (TiO_2, $ZnCr_2O_4$).[48,49] The PESs of TiO_2 and $ZnCr_2O_4$ are projected onto a two-dimensional contour map, as shown in Figure 19.4, where the structural order parameter (**OP**) and the relative energy are the **x**- and **y**-axis. The density of the map represented by different colours indicates the density of states on PES, showing the energy degeneracy of structural configurations at the same structural **OP**.

The **OP** is defined by eqn (19.22).[50]

$$OP_l = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^{l} \left| \frac{1}{N_{bonds}} \sum_{i \ne j} e^{-\frac{r_{ij}-r_c}{2r_c}} Y_{lm}(\mathbf{n}) \right|^2 \right]^{1/2} \tag{19.22}$$

where Y_{lm} is the spherical harmonic function of degree l and order m; $\mathbf{n}$ is the normalized direction between all bonded atoms; i and j are atoms in the lattice, r_{ij} is the distance between atom i and j, and r_c is set at 60% of the typical single bond length for i and j atoms. N_{bonds} is the number of bonds. By choosing a suitable degree, the order parameter can measure the short- and medium-range ordering of atoms in the lattice and thus distinguish important crystal structures and amorphous structures. For example, it is often straightforward to tell the coordination number from the OP value,
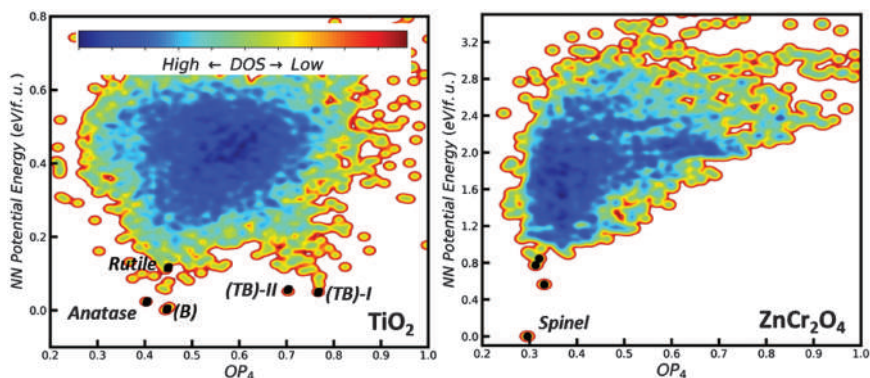
Figure 19.4   Global PESs of TiO_2 and $ZnCr_2O_4$ as determined from SSW–NN global
search. The x axis is the common structure fingerprint for the crystal,
namely, the distance-weighted Steinhardt-type order parameter (**OP**)
with different angular moment $L¼2$, 4, 6, which can distinguish
different minima, and the y axis is the relative energy (DE) of minima.
Reproduced from ref. 48 with permission from American Chemical
Society, Copyright 2018, and from ref. 79 with permission from
Springer Nature, Copyright 2019.

e.g. six-coordinated Ti atom with $OP_2 ¼ 0.3$ꓹ 0.5 and five-coordinated Ti
atom with $OP_2 ¼ 0.6$ꓹ 0.8 in TiO_2 PES.

As shown in Figure 19.4, the PESs of the TiO_2 and $ZnCr_2O_4$ have rather
different shapes: for TiO_2 the PES is shaped somewhat like a butterfly, while
for $ZnCr_2O_4$ the PES is funnel-shaped. At the bottom of the PES, where the
low energy structures lie, the TiO_2 PES demonstrates that there are many
energetically similar TiO_2 crystal phases but with different structural pat-
terns, having multiple funnels in the butterfly shape. By contrast, the
$ZnCr_2O_4$ PES has only a well-defined global minimum, a spinel type struc-
ture; all other structures are much higher in energy. It is not surprising that
more than 10 different TiO_2 crystal structures have been synthesized to date
(*e.g.* anatase, rutile, (R) and (H) *etc.*), while only a spinel phase is known for
$ZnCr_2O_4$.

In addition to the global minimum, the global PES can also reveal un-
known metastable structures, which may act as important phase-transition
intermediate states or exhibit attractive physicochemical properties. From
the global PES of ice, we have identified the low energy ice phase with va-
cancies, which are the key intermediate in the lowest energy pathways for the
cubic-to-hexagonal solid phase transition, a reaction central to ice nucle-
ation.[47] We found that the solid transition is facilitated by water vacancy at
the ice basal plane interface. The results have profound significance for
understanding some everyday phenomena that we observe, ranging from
humidity in clouds to the unusual slipperiness of ice. For the TiO_2, it is
known in general chemistry that the $[TiO_6]$ octahedron is the common
building block for TiO_2 crystals. However, a new class of unprecedented

microporous TiO_2 crystalline phases, named TiO_2(TB) in Figure 19.4, have been identified from the global PES, which features a special [TiO_5] trigonal bipyramid building block, a large pore size (5Ð 7 Å), and high thermal stability.[48] These microporous materials are predicted to be candidate anode materials for Li-ion and Na-ion batteries.[51]

19.3.2 Ultrasmall Au Supported on CeO_2

The atomic structure of supported catalysts must rank top in PES exploration problems. The dual-complexity of nanoparticles/support and the large size of the interfacial system are two key challenges to global structure search. With the advent of SSW–NN, it becomes feasible to answer some long-standing puzzles about supported catalysts, for example, the nature of cationic Au on ceria support. The past two decades have witnessed tremendous research interest in Au-based catalysts that may exhibit outstanding low-temperature catalytic ability for a wide range of reactions.[52,53] Among the catalysts, ceria supported gold (Au/CeO_2) can catalyze the water-gas-shift (WGS) reaction starting from Ð 350 K to produce high-grade hydrogen; it has a high stability.[54,55] Owing to the good oxygen storage ability,[56] it was long assumed that the oxygen vacancies on the CeO_2 support play key roles in the Au/CeO_2 system to stabilize the Au nanoparticles.[57] However, seminal work by Fu *et al.* suggested that oxidized gold (*e.g.* Au^1 or Au^{31}) is the active site for the WGS reaction.[54] The origin of the high catalytic activity of Au/CeO_2 has been a hot subject of debate ever since.

To explore the configurations of ultrasmall Au clusters on CeO_2 and determine their thermodynamics and kinetic stability, we have performed the SSW–NN and enhanced molecular dynamics (MD)[58,59] simulation to obtain the AuCeO global dataset and to build a robust and accurate AuCeO G-NN potential. Our simulations contain structures from 14 to 444 atoms per cell and cover different Au : Ce : O ratios with different morphology forms *e.g.* bulks, layers, and clusters. The final AuCeO global dataset contains 33 654 structures. The AuCeO G-NN potential contains 186 PTSDs for each element, and compatibly, the network involves three-hidden layers (186-60-50-50-1 net), equivalent to Ð 51 000 network parameters in total. The final RMSE of energy and force are 6.115 meV per atom and 0.152 eV Å^{-1} respectively.

By establishing the ternary Au-Ce-O G-NN potential based on a first principles global dataset, we have searched for the global minima for a series of Au/CeO_2 systems, the Au cluster size from four to twelve atoms and the CeO_2 support including pristine, the surface with O vacancy and the surface with structural defects (missing O-Ce-O, CeO_2-SD), see Figure 19.5a–i. The energetics of Au clusters on ceria can be evaluated by relating the supported system to the unsupported surface, Au bulk metal and the O_2 molecule (if necessary). As shown, the formation energy (E_{form}^{Au}) of Au clusters on pristine CeO_2 drops gradually from þ 0.78 eV to þ 0.69 and to þ 0.64 eV (eV/Au atom). They are generally poorer compared to them on the other two CeO_2 surfaces, suggesting the ultrasmall clusters do not prefer to anchor on a pristine

surface, and the larger cluster is thermodynamically more favoured. The presence of the O vacancy, interestingly, does not much increase the $E_{\text{form}}^{\text{Au}}$ for the Au clusters. The similar energetics for Au on pristine CeO_2 and CeO_2 with O vacancy indicates that the O vacancy does not induce much additional charge transfer, *e.g.* from CeO_2 support to Au in forming the negatively charged Au cluster. In fact, the additional electrons due to the missing of a surface O are stored in nearby surface Ce atoms, forming two Ce^{31}. This explanation is supported by DFT calculations that the net magnetic spin for Au_x/CeO_2-O_v systems is generally larger than 2, suggesting that the Au adsorption does not quench the spin on Ce atoms. On the other hand, the oxidized Au clusters on CeO_2-SD are much stabilized as reflected by the lower $E_{\text{form}}^{\text{Au}}$. In particular, Au_4O_2/CeO_2-SD has a negative formation energy,

0.085 eV/Au atom, which suggests that the CeO_2-SD can act as the anchoring site for ultrasmall oxidized Au clusters below 4 Au atoms. With the increase of the cluster size, the energetic preference at the CeO_2-SD sites disappears gradually.

Our results show that the ultrasmall cationic Au clusters attaching to surface structural defects are the only stable structural pattern, *e.g.* Au_4O_2/CeO_2-SD, where Au clusters can donate electrons to the underlying Ce mediated by the Au–O bonding (see Figure 19.5j). The Au_4O_2/CeO_2-SD is believed to be the potential stable active structure for an ultrasmall Au cluster on the CeO_2 surface. The other clusters on different CeO_2 surfaces have a strong energy preference to grow into bulky Au metal. We demonstrate that the global PES exploration is critical for understanding the morphology of metal clusters on oxide support, which now can be realized *via* the NN method.

19.3.3 Hydrogen Evolution Reaction on Amorphous TiO_xH_y

Black titania (TiO_2), which absorbs strongly across the entire visible-light spectrum, has been found to exhibit hydrogen evolution reaction (HER) activity several orders higher than convention TiO_2 material.[60–65] Synthesized by hydrogenating pristine TiO_2, the as-synthesized black TiO_2 commonly exhibits a core–shell structure with the amorphous shell, a few nanometers thick, coated on the anatase crystals.[61,66–68] The amorphous shell is believed to provide the catalytic active site and to be responsible for the enhanced HER activity. However, structure determination for the amorphous shell is challenging both experimentally and theoretically, without even mentioning the need to explain the high HER activity.

To resolve the HER active sites on the amorphous TiO_xH_y shell, we have performed SSW–NN simulations to obtain the TiO_xH_y global dataset with 143 786 structures and build a robust and accurate TiO_xH_y G-NN potential. The structures in the database cover different Ti : O : H ratios, mainly Ti_4O_7, Ti_4O_8, $Ti_4O_8H_x$ x ¼ 1B 4, $Ti_8O_{16}H_x$, x ¼ 1B 4 with bulk, layer, and cluster forms, and also contain large surface systems such as $Ti_{56}O_{112}H_x$ and molecular systems (H_2, H_2O). The TiO_xH_y G-NN potential utilizes 201 PTSDs for each element, *i.e.* 77 two-body PTSDs, 108 three-body PTSDs and
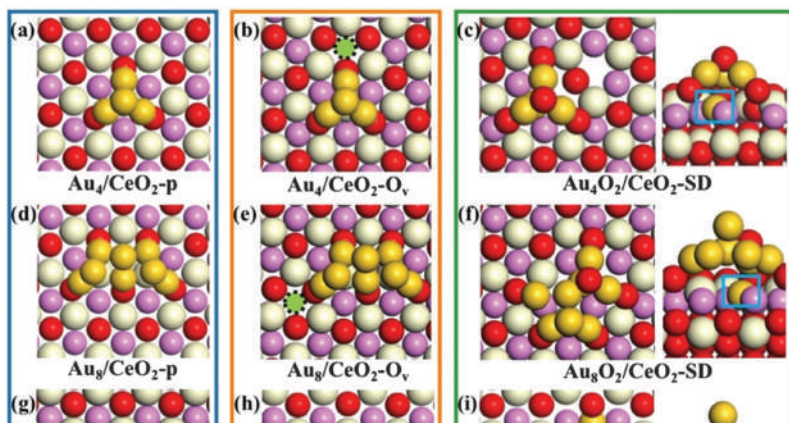
Figure 19.5 Global minimum structures for Au clusters on CeO_2 determined by SSW–NN. (**a–i**) Structure snapshots of Au clusters on three different CeO_2 surfaces. For Au_x/CeO_2-SD, both the top and the side views are shown for clarity. O atoms are in red/pink, and Ce atoms are in white. The surface layer O and the subsurface O are in red (darker) and pink (lighter), respectively. Au atoms are in yellow (for example, the central cluster in (g)). The oxygen vacancy is indicated by dashed circles in Au_x/CeO_2-O_v. The embedded Au atom is highlighted by a blue rectangle. (**j**) The formation energy of Au cluster relative to the CeO_2 surface and Au bulk.

CHAPTER 19

Machine Learning for Heterogeneous Catalysis: Global Neural Network Potential from Construction to Applications

SICONG MA(0000-0001-5894-5910), PEI-LIN KANG(0000-0003-2147-2472), CHENG SHANG(0000-0001-7486-1514) AND ZHI-PAN LIU(0000-0002-2906-5217)*

Collaborative Innovation Centre 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

19.1   Introduction

Heterogeneous catalysts are notable for their great complexity of material composition and surface structures. There is a long history of both experimental and theortical struggling to resolve the active site of catalysts that are responsible for activity and selectivity.[1–4] Recent experimental progress is represented by high spatial resolution techniques, *e.g.* the spherical aberration-corrected high-resolution transmission electron microscope, and synchrotron-based measurement, *e.g.* extended X-ray absorption fine