

# UCLA

## UCLA Previously Published Works

### Title

Ensemble-Average Representation of Pt Clusters in Conditions of Catalysis Accessed through GPU Accelerated Deep Neural Network Fitting Global Optimization

### Permalink

<https://escholarship.org/uc/item/4dt970ww>

### Journal

Journal of Chemical Theory and Computation, 12(12)

### ISSN

1549-9618

### Authors

Zhai, Huanchen  
Alexandrova, Anastassia N

### Publication Date

2016-12-13

### DOI

10.1021/acs.jctc.6b00994

Peer reviewed

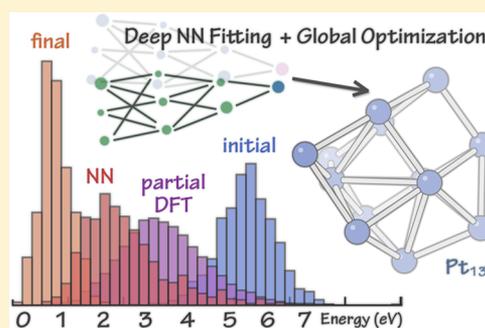
# Ensemble-Average Representation of Pt Clusters in Conditions of Catalysis Accessed through GPU Accelerated Deep Neural Network Fitting Global Optimization

Huanchen Zhai<sup>†</sup> and Anastassia N. Alexandrova<sup>\*,†,‡</sup><sup>†</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States<sup>‡</sup>California NanoSystems Institute, Los Angeles, California 90095, United States

## Supporting Information

**ABSTRACT:** We first report a global optimization approach based on GPU accelerated Deep Neural Network (DNN) fitting, for modeling metal clusters at realistic temperatures. The seven-layer multidimensional and locally connected DNN is combined with limited-step Density Functional Theory (DFT) geometry optimization to reduce the time cost of full DFT local optimization, which is considered to be the most time-consuming step in global optimization. An algorithm based on bond length distribution analysis is used to efficiently sample the configuration space and generate random initial structures. A structure similarity measurement method based on depth-first search is used to identify duplicates. The performance of the new approach is examined by the application to the global minimum searching for Pt<sub>9</sub> and Pt<sub>13</sub>.

The ensemble-average representations of the two clusters are constructed based on all geometrically different isomers, on which the structure transition is predicted at low and high temperatures, for Pt<sub>9</sub> and Pt<sub>13</sub> clusters, respectively. Finally, the ensemble-averaged vertical ionization potential changes when temperature increases, and the performance in conditions of catalysis can be different from that evaluated at the global minimum structure.



## 1. INTRODUCTION

The catalytic activity of platinum clusters has been intensely investigated experimentally and computationally for a long time.<sup>1–3</sup> Since structure and property are related, finding the correct geometry of the global minimum and low energy local minima of these clusters becomes an essential part of the catalytic property studies. However, this requires a significant amount of computational efforts. Over decades many efficient global optimization techniques have been developed, including Genetic Algorithm (GA) based methods,<sup>4–8</sup> Particle Swarm Optimization (PSO),<sup>9,10</sup> Simulated Annealing (SA),<sup>11</sup> and Basin Hopping (BH).<sup>12</sup> Many of these methods incorporate the Density Functional Theory (DFT) level local geometry optimization procedure as an internal step of global optimization.<sup>13</sup> This important technique is a great compromise between quality of the solution and a reasonable computational expense and accelerates the global minimum searching. However, for medium- to large-sized transition metal clusters, the DFT local optimization step itself is very time-consuming, especially when hybrid functionals are used. In addition, most of the aforementioned global optimization methods, although spectacularly successful in many cases, are less favorable in a highly paralleled computational environment, because of their evolutionary feature.

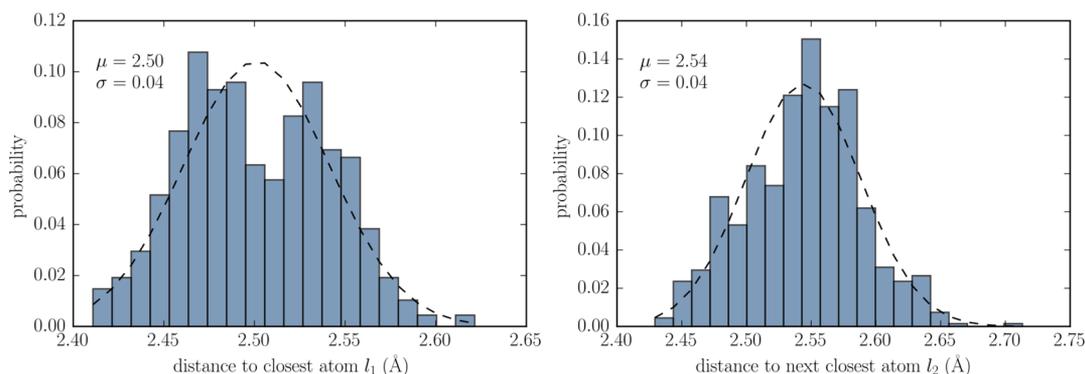
Force Field (FF) or Neural Network (NN) fitting, on the other hand, can generate an approximate PES so that the efforts on performing DFT level local optimization would be

alleviated. It is shown in our previous work<sup>14</sup> that the FF fitting can be used to assist global searching. However, the restricted form of the FF function may prevent its extension to larger systems. NN has been widely applied to the fitting of molecular PES for many years.<sup>15</sup> Nevertheless, most studies of this fitting approach focus on small-sized clusters and molecules, such as Si<sub>5</sub>,<sup>16</sup> BeH<sub>3</sub>,<sup>17</sup> and FH<sub>2</sub>O.<sup>18</sup> Recently, a high-dimensional NN fitting method (atomistic NN) has been proposed, which is based on expressing the total energy as the sum of atomic energies.<sup>19,20</sup> The new atomistic NN approach has been successfully applied to the PES fitting of a variety of systems, including Zn<sub>N</sub>O<sub>N</sub> (N = 1–40).<sup>21</sup> Very recently, it has been shown that a combination of BH and atomistic NN approaches (NN-BH) can be used for global optimization for large-sized metal clusters, such as Au<sub>58</sub><sup>22</sup> and Na<sub>20–40</sub>.<sup>23</sup> Despite the success of the atomistic NN method, the many-body expansion method has also been combined with NN for PES fitting, and some applications to small molecules, such as H<sub>2</sub>O<sub>2</sub><sup>24</sup> and C<sub>2</sub>H<sub>3</sub>Br,<sup>25</sup> are reported.

To the best of our knowledge, all aforementioned NN PES fitting approaches are based on the traditional NN with fully connected and shallow structure (with typically three to four layers). However, it is shown<sup>26</sup> that NN with insufficient depth may require a large number of training samples to tune the

Received: October 12, 2016

Published: November 7, 2016



**Figure 1.**  $l_1$  and  $l_2$  distribution of 113  $\text{Pt}_6$  low-energy local minima (with energy lower than 1.0 eV with respect to global minimum energy). The initial structures are generated using CK, and the DFT optimization is done with Turbomole 6.6, the PBE0 hybrid functional, and def2-TZVP basis. The dashed lines are corresponding normal distribution curves fitted based on the mean and standard deviation of  $l_1$  and  $l_2$  data.

parameters. On the other hand, large scale fully connected NN is difficult to train and can potentially cause overfitting issues. During recent years, several novel approaches have been proposed to solve the Deep Neural Network (DNN) training problem. For example, the large Convolutional Neural Network (CNN), as an example of locally connected DNN, is shown to have an impressive performance on image classification.<sup>27</sup> Alternatively, the Deep Belief Network (DBN)<sup>28</sup> is an example of DNN that can be pretrained using unsupervised learning techniques. In addition, benefiting from recent progress on GPU computing,<sup>29</sup> the training of DNN can be largely accelerated.

In this work, we discuss a general approach to find global and low energy local minima of metal clusters at the DFT level, using locally connected DNN fitting combined with a new random structure generation algorithm based on bond length distribution analysis. Unlike some other NN based global optimization methods, the training data are not obtained by performing full local optimization. Instead, the limited-step local optimization is performed so that the PES can be moderately explored, and meanwhile the time cost of local optimization with DFT can be greatly reduced. The NN is constructed in a truncated many-body expansion way, but several mixing layers are added so that the NN potential can represent a more general function than the original truncated many-body one. The fitted NN is then used to optimize some newly generated random structures. As a final step, automatically selected NN-relaxed structures are fully optimized using DFT to yield the final results. The performance of this new approach is illustrated by the application to  $\text{Pt}_9$  and  $\text{Pt}_{13}$  gas phase clusters. The properties of the clusters at catalysis conditions are then explored by constructing an ensemble-average representation of local minima.

## 2. METHOD

### 2.1. Bond Length Distribution Algorithm: Motivation.

One of the major differences between the global optimization in a computational chemistry context and general global optimization is that for chemical problems some restrictions can be applied to generate start points that are very close to minimum, rather than totally random. A common practice of choosing such kind of restriction, namely, the Bound Checking (BC) approach, is setting the lower and upper bounds of the distance between each atom and its nearest neighbor (denoted as  $l_1$ ).<sup>23</sup> The resulting initial structures are less probably to suffer from a Self-Consistent Field (SCF) convergence

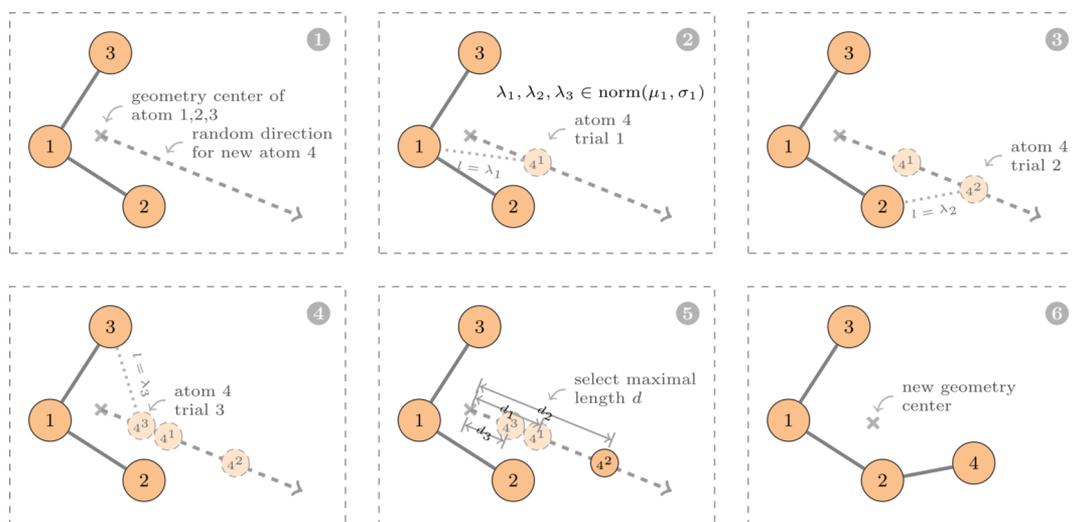
problem. In addition, the time cost for searching in a chemically unfavorable configuration space can be greatly reduced.

Despite the fact that the BC approach for initial structure generation has been successfully applied in many evolutionary global optimization methods, we note that this may not be the most natural or efficient way to sample the chemically relevant configuration space. In contrast, the Coalescence Kick (CK) method<sup>30</sup> generates initial structures in a dynamical way: the randomly positioned atoms are pushed toward the center of mass step by step until all connected. The CK method is equivalent to setting an upper bound of  $l_1$  and obtaining a more natural distribution of  $l_1$  at the shorter bond length side. We found that CK produces structures that are generally better than those generated by BC. Nevertheless, the  $l_1$  distribution of CK structures is still unnatural to some extent.

We may improve this by requiring that the initial structures have a similar distribution to that of the expected local minima. Figure 1 shows the  $l_1$  and  $l_2$  distribution of low energy local minima of  $\text{Pt}_6$  clusters found using the CK method, where  $l_2$  denotes the distance between each atom and its next nearest neighbor. It is obvious that for  $\text{Pt}_6$  clusters both distances can be fitted to normal distribution. In addition, the mean values of  $l_1$  and  $l_2$  (2.50 and 2.54 Å) are very close to the sum of covalent radii of Pt atoms ( $1.28 \times 2 = 2.56$  Å). In light of this observation, we may require that the  $l_1$  and/or  $l_2$  parameters of initial structures be selected from normal distribution. We call the generation algorithm based on this statistical restriction the Bond Length Distribution Algorithm (BLDA).

When only  $l_1$  is restricted, the algorithm is referred to as first-order BLDA (F-BLDA). When both  $l_1$  and  $l_2$  are restricted, the algorithm is referred to as second-order BLDA (S-BLDA). It is clear that when  $l_2$  is restricted to obey a normal distribution, each atom in the generated cluster will have a coordination number of at least two. While this is true for most metal clusters, for other clusters (like Boron clusters<sup>14</sup>) this assertion is very likely to fail. Therefore, we note that F-BLDA can be used for most gas phase atomic cluster systems, and S-BLDA is more focused on metal clusters, which tend to have compact geometries in which their delocalized bonding is optimized. In this work, S-BLDA is used as the initial structure generation method for  $\text{Pt}_9$  and  $\text{Pt}_{13}$ .

**2.2. Bond Length Distribution Algorithm: Implementation.** In F-BLDA, the cluster is created by each time adding one atom to the current structure. The position of that atom is determined by first selecting a random direction, to which the new atom will be placed. After that, we only need to figure out



**Figure 2.** One intermediate step of first order Bond Length Distribution Algorithm (F-BLDA): adding an atom to the existing structure. Before placing atom 4, a random direction is first selected. For each old atom  $i$ , a distance  $\lambda_i$  is picked from the required distribution  $\text{norm}(\mu_i, \sigma_i)$ . Then a trial atom 4 is placed along the selected direction with a distance  $\lambda_i$  from atom  $i$ . Finally, among all trial atoms, the one with maximal distance from the geometry center is accepted.

the distance between geometry center and the new atom. With the distance and direction, we can thus construct a vector, along which the new atom can be placed. The length of the vector is the minimal length such that all of the distances between this new atom and any other atom already present in the cluster are greater than the lengths randomly picked from the required normal distribution  $l_i$ . The core step of adding one atom to the existing structure is illustrated in Figure 2.

It should be noted that since a lot of value comparison and selection are involved in this generation procedure, strictly speaking the bond lengths resulting from this procedure will not be exactly normal distributed. Nevertheless, in practice, we have found that it gives a very good approximation to the desired normal distribution.

Comparing to F-BLDA, the implementation of S-BLDA is more complicated. Nevertheless, it is still based on the procedure of adding one atom at a time. When placing the first two atoms, we follow the same procedure as that in F-BLDA. Starting from the third atom, the step of adding one new atom is divided into four substeps:

1. **Selection.** Select any two atoms  $A, B$  from existing atoms. At the same time, pick two lengths,  $\lambda_1$  and  $\lambda_2$  from  $\text{norm}(\mu_1, \sigma_1)$  and  $\text{norm}(\mu_2, \sigma_2)$ , respectively. These two normal distributions are the required  $l_1$  and  $l_2$  distributions, respectively.

2. **Circling.** The new atom is required to have a distance of the length  $\lambda_1$  to atom  $A$  and a distance of the length  $\lambda_2$  to atom  $B$ . This restricts the new atom to pick a position on a circle in three-dimensional space. The normal vector and radius of the circle plane can then be determined.

3. **Projection.** Select a random direction vector in three-dimensional space. Then project this random vector to the circle plane so that we can determine the exact position on the circle for the new atom to pick.

4. **Checking.** For each of the old atoms other than  $A$  and  $B$ , denoted as  $C_i$ , pick a length  $\lambda_i$  from  $\text{norm}(\mu_i, \sigma_i)$ . The distance between the new atom and atom  $C_i$  is required to be greater than  $\lambda_i$ . If for any atom  $C_i$  this is not satisfied, go back to substep (3) and select another random direction. If it still fails, try at most 10 times from substep (3) to (4). If it fails more

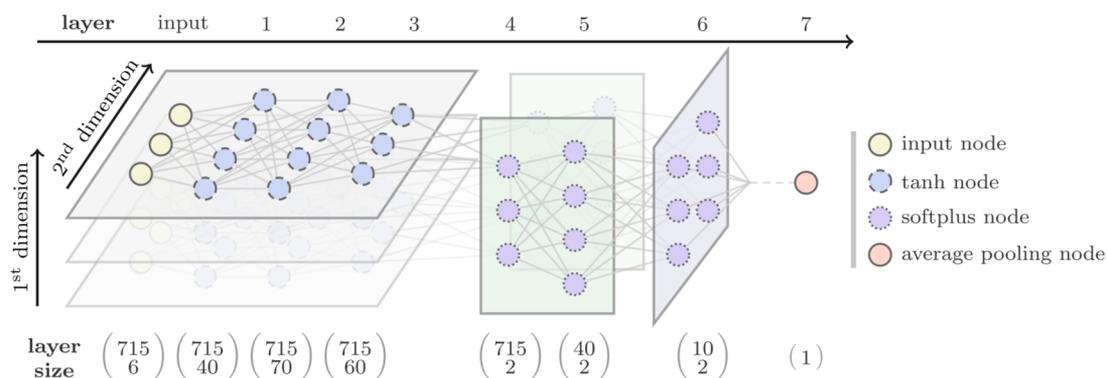
than 10 times, go back to substep (1) and try another two atoms  $A$  and  $B$ .

**2.3. Deep Neural Network: Approximation.** The NN architecture used in this work is designed based on the truncated many-body expansion approach,<sup>31</sup> with only the  $k$ -body terms considered. In this work,  $k = 4$  is used. Based on our experience, NN constructed using higher order terms can make the training very time-consuming and may cause overfitting issues. On the other hand, if only lower order terms are included, poor fitting accuracy may be expected.

For simplicity, consider a homogeneous atomic cluster, say,  $\text{Pt}_N$  ( $N \geq k$ ). The Cartesian coordinates of the atoms are denoted as  $r_i$  ( $i = 1, 2, \dots, N$ ). Given any set  $X$ ,  $\mathcal{S}_k(X)$  will be the set of all subsets of  $X$  that are composed of  $k$  elements. Obviously, if the number of elements in  $X$ , denoted as  $|X|$ , is equal to  $N$ , then  $|\mathcal{S}_k(X)| = C(N, k)$ , which is the number of combinations of  $k$  elements from  $X$ .

The original PES can be expressed as a function of  $\{r_i\}$ , namely,  $F(r_1, r_2, \dots, r_N)$ . Using the many-body expansion, if only the  $k$ -body terms are retained, then the original PES is approximated by  $F'(\{r_i\}) = \text{sum}(F_k(\{r_j\}))$ , where  $k$  is a fixed parameter,  $\{r_j\} \in \mathcal{S}_k(\{r_i\})$ , and the sum is over all  $\{r_j\}$  in  $\mathcal{S}_k(\{r_i\})$ . Now the undetermined function  $F_k$  can be fitted using the traditional fully connected NN.

In practice, when  $N$  is large and  $k$  is kept to be a moderate number (as in our case, four), this procedure may cause a large fitting error since many high order terms are truncated. To alleviate this problem, we replace the summation by an undetermined function  $G$ . This will introduce some nonlinear mixing among  $k$ -body terms, by which the fitting accuracy can be improved. The original PES is then more accurately approximated by  $F''(\{r_i\}) = G(\{F_k(\{r_j\})\})$ , where  $G$  is a function of  $C(N, k)$  variables. Since  $F_k$  is merely an intermediate variable, it no longer has to be a scalar function. Therefore, in our approach it is generalized to be a vector function  $F_k$ , and this will introduce more flexibility in the constructed neural network. Therefore, our final neural network framework to approximate the real PES is  $F'''(\{r_i\}) = G(\{F_k(\{r_j\})\})$ , where  $F_k$  and  $G$  can be fitted simultaneously, by locally connected NN.



**Figure 3.** Architecture of the deep neural network used in this work. The actual numbers of NN nodes (layer sizes) are larger than or equal to the number of nodes shown in the figure. The layer sizes for the  $\text{Pt}_{13}$  case ( $N = 13$ ,  $k = 4$ ) are indicated in the bottom. There is no node connection across the stacked planes.

We call this approach the Many-Body Expansion Neural Network plus Mixing (MBE-NN+M) approach.

It is known that Cartesian coordinates are not suitable to be directly used as input of NN.<sup>19</sup> Therefore, two transformations are performed on the input sample data in Cartesian coordinates  $\{r_j\}$ . The original coordinates  $\{r_j\}$  (after  $k$ -body selection procedure) are first transformed to interatomic distances  $\{a_m\}$ , where  $m = 1, 2, \dots, C(k, 2)$  are the indices of all interatomic distances in the  $k$ -body fragment. After that, an exponential function is applied to help better describe the PES:<sup>32</sup>  $b_m = \exp(-a_m/L)$ , where  $\{b_m\}$  is the input of NN and  $L$  is a fixed parameter. In this work, we choose  $L = 4.0 \text{ \AA}$ . The output of NN is a single value representing the energy. A linear transformation is used so that the energies are mapped to interval  $[0, 1]$  for NN training, and the corresponding inverse transformation is performed when interpreting NN predictions.

**2.4. Deep Neural Network: Architecture.** The DNN architecture used in this work consists of three parts: MBE part, mixing part, and pooling part. Local connectivity, parameter sharing, and average-pooling<sup>33</sup> are three main features used in this NN, which make it different from the traditional fully connected NN.

The MBE part of NN is used to fit  $F_k$ . Based on the MBE-NN+M approach, the input of NN is a second-order tensor (or two-dimensional array) of the size  $(C(N, k), C(k, 2))$ , where the second dimension represents  $C(k, 2)$  variables of the function  $F_k$ , and the first dimension indicates that there are  $C(N, k)$   $k$ -body terms, each denoted as  $F_k(\{r_j\})$ . To fit  $F_k$ , the NN connection is localized in the second dimension, and the layer size of the first dimension is kept fixed. The weights and biases of NN connection are shared among different indices of the first dimension, so that the fitted function form of  $F_k$  is kept consistent among different  $k$ -body terms. Since  $F_k$  is a vector function, the size of the output value (as a vector) of each  $F_k$  term, denoted as  $f$ , must be determined when constructing NN. In this work, we choose  $f = 2$ . The MBE part is composed of four layers with the following sizes:  $(C(N, k), C(k, 2)) - (C(N, k), 40) - (C(N, k), 70) - (C(N, k), 60) - (C(N, k), 2)$ .

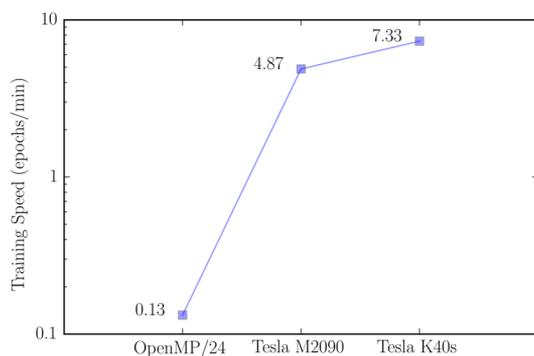
The mixing part is used to fit  $G$ . Within this part the NN connection is localized in the first dimension, and the size of the second dimension is kept fixed. The parameters of NN connection in this part are shared among different indices of the second dimension. In this work, the mixing part is composed of two layers with the following sizes:  $(C(N, k), 2) - (40, 2) - (10, 2)$ .

The last part is used to transform the output of mixing part to a single value, representing the energy. The average-pooling is used, which means that we take the average value of all elements in the matrix of the previous layer as the final output. In this work, the pooling part is composed of one layer of the size:  $(10, 2) - (1)$ .

To describe the nonlinearity of the unknown function, suitable activation functions must be used in NN architecture. In previous NN PES fitting work by other authors, common choices are the hyperbolic tangent and the sigmoid function.<sup>20</sup> However, it has been found that these traditional activation functions are less efficient in DNN.<sup>34</sup> It has also been observed that in some DNNs, the earlier layers tend to learn less efficiently compared with later layers, when trained by backpropagation learning algorithm.<sup>35</sup> Therefore, in this work, the first three hidden layers are activated by the hyperbolic tangent, while the remaining layers except for the output layer are activated by the softplus function. We found that this hybrid approach is able to overcome the intrinsic disadvantage of both activation functions and increase the training efficiency. The overall architecture of NN used in this work is shown in Figure 3.

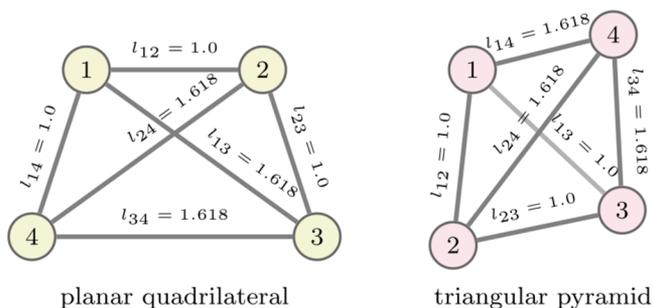
**2.5. Deep Neural Network: Implementation.** The NN used in this work is constructed using *theano* framework,<sup>36</sup> which is a *python* library that allows for the acceleration of the evaluation and differentiation of mathematical expressions both numerically and symbolically. Based on this framework, efficient C/OpenMP codes and/or CUDA codes for GPU can be automatically generated at runtime. We note that this feature is significant for saving time during training our large-scale deep NN. It is illustrated in Figure 4 that for a typical DNN PES fitting task in this work, acceleration of computation with GPU can be more than 50 times compared to that done with only CPU. Based on this observation, in this work, the GPU Tesla K40s is used for all NN training tasks.

The Mini-Batch Stochastic Gradient Descent with Momentum (MB-SGD-M) method<sup>37</sup> is used for NN training. This is a widely tested training method for large-scale NNs. The step decay approach is used to improve convergence. The step length for parameter updating at epoch  $i$  is determined by  $s_0 r / (r + i)$ , where  $s_0$  is the initial step length and  $r$  is the step decay factor. In our work, we choose  $s_0 = 0.1$  and  $r = 60$ . When NN training finished, the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method<sup>38</sup> is used for NN based geometry optimization.



**Figure 4.** Training speed measured in number of epochs per minute under different machine architectures. The test case is a 7-layer NN PES fitting for  $\text{Pt}_{13}$  energy data. During each epoch, 50 000/5 000 random structures and their corresponding energies are used for training/testing, respectively. For the first architecture (OpenMP/24), the CPU model is Intel Xeon E5-2697v2 (2.7 GHz), and the 24 cores OpenMP parallelism scheme is used. For the other two architectures, a single CPU core and one indicated GPU accelerator are used.

**2.6. Structure Similarity Measurement.** In order to reduce the computational time for DFT-level optimization, an automatic and reliable scheme for structure similarity measurement is required. The simplest way is comparing two structures based on their DFT energies.<sup>6</sup> Since energy is not available for initial structures, and for some large-sized clusters different structures may have very close energies, this approach is not always applicable. Another approach is based on the difference of sorted atom–atom adjacency matrices<sup>39</sup> or, similarly, integrals over the distribution of atomic distances and angles.<sup>40</sup> However, we note that two different structures can have exactly the same atomic distance distribution or, equivalently, the same sorted adjacency matrices, as illustrated in Figure 5, and this is also true for the angle distribution, especially when many distances or angles in the cluster are identical. Other approaches use some fingerprint functions or descriptors, such as spherical harmonics,<sup>41</sup> spherically averaged scattering intensity,<sup>42</sup> and experimental elemental or molecular properties. The effectiveness of fingerprint approaches will largely depend



**Figure 5.** An example of two different structures with the same interatomic distance distribution. Note that when the distances are matched, the corresponding atoms cannot be matched at the same time. In this particular case, for planar quadrilateral the three “1.0” distances connect atoms 1, 2, 3, and 4, but for a triangular pyramid the three “1.0” distances connect atoms 1, 2, and 3. Therefore, the requirement of a one-to-one atomic matching can be used to distinguish the two structures. The AM-DFS method will first establish such an atomic matching, then based on which differences of distances are measured. This ensures an accurate measurement of structure similarity.

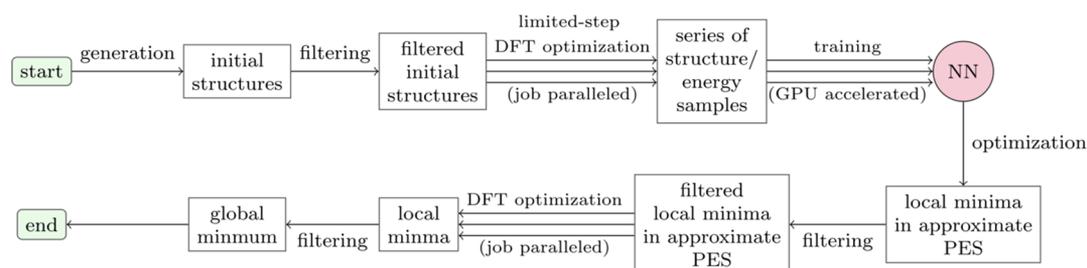
on the selected functions and may not be equally applicable to all cases. In particular, elemental and molecular properties are not generally applicable or available for metal clusters. Therefore, there is a need to develop a simple, reliable, and nonenergy-dependent structure similarity measurement method.

The example shown in Figure 5 indicates that, in order to distinguish two isomers with exactly the same interatomic distance distributions, we need to also find a one-to-one atomic matching. Based on a best atomic matching, the corresponding interatomic distances can be compared between two structures. If any of the matched pairs of two interatomic distances have a difference larger than a threshold value, then we can assert that the two structures are different. We note that the interatomic distances based method cannot distinguish structures that are mirror images of each other. However, the mirror images also have the same energy at the DFT level, thus are unnecessary to be separated.

Now the problem is how to find the best atomic matching between two structures. A direct way is to enumerate all possible sequences of atom indices for one structure and match them to the sequence of the other. Since there are numerous permutations of atom indices for clusters composed of ten or more atoms, this is computationally too expensive. However, given that the threshold value  $d$  for confirming similarity is small, the comparison can be performed when an incomplete atomic matching is available. If the maximum difference of matched interatomic distances based on the given atomic matching fragment is larger than the threshold, then all atomic matching with that fragment will definitely fail. Using this rule, only a small portion of permutation space is needed to be checked. In practice, we use the Depth-First Search (DFS)<sup>43</sup> algorithm to enumerate the sequences of atoms for one structure and match each sequence to the ordered atomic sequence of the other structure. Only atoms of the same element will be matched. Whenever a new testing atomic matching fragment is formed, the corresponding difference of matched interatomic distances is checked against the threshold value. If the difference is larger than the threshold value, all sequences involving the current fragment will be excluded from the searching space. Otherwise, we may find one atomic matching, based on which all differences of matched interatomic distances are smaller than the threshold value. In this case, we assert that the two structures are similar. If no such kind of atomic matching can be found, we claim that the current structure is unique. Based on our testing, using this approach on a single CPU core, thousands of structure comparisons for  $\text{Pt}_{13}$  can be accomplished within seconds. We denote this new, fast, and deterministic structure comparison method as the Atomic Matching Depth-First Search (AM-DFS) method.

**2.7. Global Optimization.** Combining the aforementioned generating, comparing, and fitting methods, we are able to propose a new NN fitting-based and highly paralleled global optimization scheme shown in Figure 6. The scheme (NN-PGOPT) consists of three major steps:

- 1) **Initial structure preparation.**
  - a. Random structure generation (S-BLDA)
  - b. Filtering out duplicates (AM-DFS)
- 2) **Local optimization.**
  - a. Limited-step geometry optimization (DFT)
  - b. PES fitting (MBE-NN+M and MB-SGD-M)
  - c. PES optimization (L-BFGS)



**Figure 6.** Overall flow chart of the NN-PGOPT scheme. The paralleled algorithms are indicated by multiple arrows, and the parallelism models are shown in parentheses. The DFT optimization is paralleled in two levels. Different atomic configurations are independently paralleled at job level. For every atomic configuration, the OpenMP or MPI parallelism is utilized as implemented in the DFT package.

**Table 1.** Number of Epochs, Average Fitting Errors Per Structure and Total Time Used in Training NN for Pt<sub>9</sub> PES<sup>44</sup>

NN	parameter guess	epochs	training error (meV)	testing error (meV)	validation error (meV)	total time (hour:min)
reference	random	1400	236	262	252	3:57
singlet	nonet	500	230	243	246	1:25
triplet	reference	500	151	158	153	1:25
quintet	triplet	500	137	147	146	1:25
septet	quintet	500	136	146	142	1:25
nonet	septet	500	129	131	132	1:25

<sup>44</sup>The parameter guess column indicates that the final parameters of which NN are used as the initial guess of parameters. The errors are measured in Root Mean Square Error (RMSE).

- d. Filtering out duplicates (AM-DFS)
  - e. Full-step geometry optimization (DFT)
- 3) **Global minimum identification.**
- a. Filtering out duplicates (AM-DFS)

### 3. APPLICATIONS

**3.1. Gas Phase Pt<sub>9</sub> Global Optimization.** *3.1.1. DFT Parameters and Fitting Errors.* The global optimization is performed for Pt<sub>9</sub> as an example of an application of the NN-PGOPT scheme. A separate direct searching is also performed, where the steps 2a–2d mentioned in section 2.7 are skipped, in order to test the performance without NN fitting.

The DFT optimization is done using Turbomole 6.6<sup>44,45</sup> with the Tao, Perdew, Staroverov, and Scuseria hybrid (TPSSH)<sup>46</sup> functional, since the nonempirical TPSSH hybrid functional has been found to have a good performance for late transition metals clusters.<sup>47</sup> To save DFT optimization time, the randomly generated structures are first optimized with def2-SV(P) and then with the def2-TZVP basis set,<sup>48</sup> in the direct searching case. For the NN-PGOPT scheme, the NN is fitted for def2-SV(P) energies, and then the full-step DFT optimization is performed with def2-TZVP.

500 unique initial structures are generated using the S-BLDA approach with parameters  $\mu_1 = 2.50 \text{ \AA}$ ,  $\mu_2 = 2.55 \text{ \AA}$ , and  $\sigma_1 = \sigma_2 = 0.04 \text{ \AA}$  for singlet, triplet, quintet, septet, and nonet multiplicities, respectively. The threshold value  $d = 0.25 \text{ \AA}$  is used to filter the S-BLDA generated structures. Ten DFT optimization steps are performed on these initial structures to yield 4240, 4770, 4670, 4210, and 4080 structure/energy samples for each multiplicity, respectively. For each multiplicity, the samples span an energy range of around 5.0 eV. Five independent NNs are fitted for different multiplicities, and an additional reference NN is first fitted for singlet to generate an initial guess of NN parameters. 82%, 9%, and 9% of samples are used for training, testing, and validation, respectively. Since the cluster structure will be the same when the coordinates of any two atoms of the same element are swapped, these atoms can

be randomly shuffled to generate more samples, so that the NN can learn the permutation invariance of atoms of the same element. In this way, 200 000, 20 000, and 20 000 samples are generated for training, testing, and validation, respectively. The minibatch size and the momentum are set to be 50 and 0.7, respectively, as the parameter of the MB-SGD-M method.

Since  $N = 9$  and  $k = 4$ , the layer sizes of the NN used for Pt<sub>9</sub> PES fitting are (126, 6) - (126, 40) - (126, 70) - (126, 60) - (126, 2) - (40, 2) - (10, 2) - (1). Therefore, the total number of parameters (including weights and biases) used in this NN is  $40(6 + 1) + 70(40 + 1) + 60(70 + 1) + 2(60 + 1) + 40(126 + 1) + 10(40 + 1) = 13022$ . The number of epochs, fitting errors, and time are summarized in Table 1. We note that within each epoch, all training samples are used once for training the NN with random order, and then all testing samples are used once to give the testing error. After all epochs, the validation samples are used once to give the validation error. The parameters with the lowest testing error are selected as the final parameters.

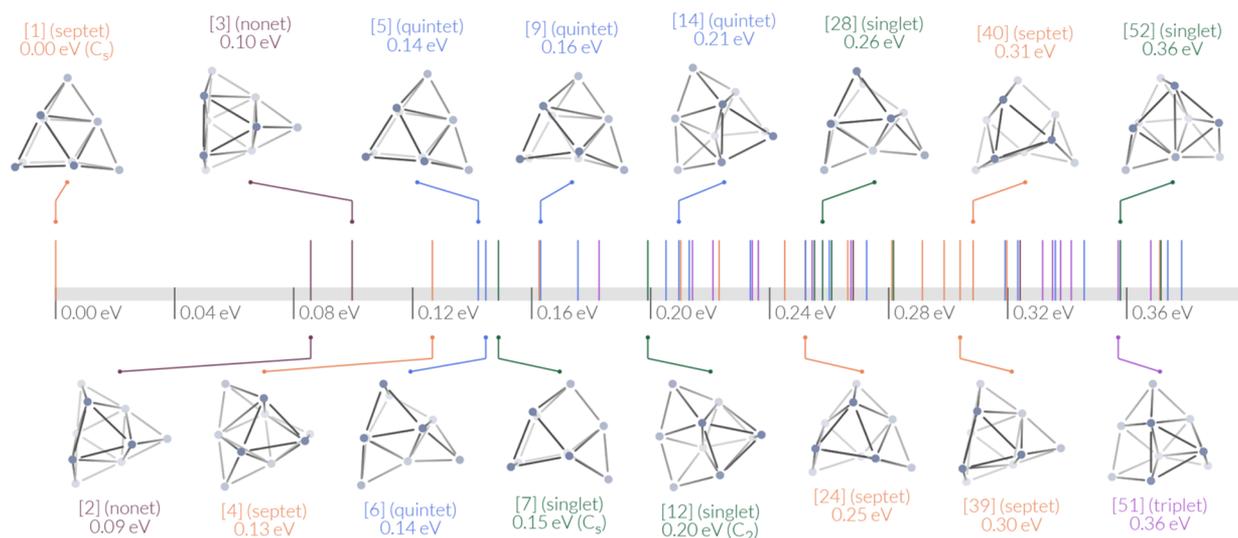
After the PES fitting, 5000 unique initial structures are generated with the same parameters as those used in the first step, for each multiplicity, and then optimized to their local minima using the corresponding NN. Some NN-relaxed local minima with interatomic distances that are too long or too short are considered as extrapolated structures and are excluded. Then the remaining NN-relaxed local minima are filtered with threshold value  $d = 0.25 \text{ \AA}$ . Finally, 300 filtered NN-relaxed local minima with relatively lower energies are selected to perform the full-step DFT optimization with the def2-TZVP basis set.

*3.1.2. Time Efficiency.* A separate direct global optimization without NN fitting is also performed for Pt<sub>9</sub> for comparison. During the direct approach, 500 unique initial structures are generated for each multiplicity using the same method and parameters as those in the NN-PGOPT scheme. Thirty additional converged local minima from singlet, triplet, quintet, and septet are added to triplet, quintet, septet, and nonet initial structures, respectively, since the converged minima of one multiplicity can be a good initial guess to be relaxed for a

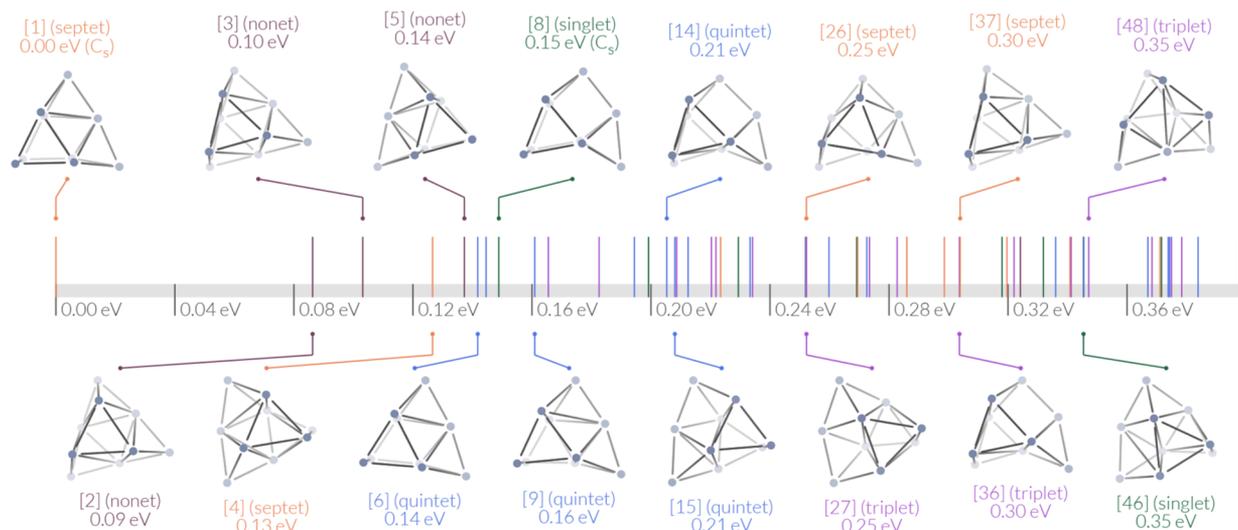
Table 2. CPU Core Hours Used during the DFT Optimization for the Direct Scheme and the NN-Assisted Scheme<sup>a</sup>

approach	basis set	singlet	triplet	quintet	septet	nonet	total
direct	def2-SV(P)	11096	7794	8769	8750	13939	50348
	def2-TZVP	6066	4129	4039	4339	2819	21392
	<b>total</b>						<b>71740</b>
NN-PGOPT	def2-SV(P)	4344	2895	3271	3493	5182	19185
	def2-TZVP	9515	6787	6324	6735	6369	35730
	<b>total</b>						<b>54915</b>

<sup>a</sup>The NN training and optimization time is about 10 CPU hours (with GPU acceleration) in total for all structures, which is order of magnitudes smaller than DFT hours, and not included in this table.



**Figure 7.** Energy and some of the structures of 58 low energy optima of Pt<sub>3</sub> found at the TPSSh/def2-TZVP level of theory using a direct approach. Only structures with energy less than 0.40 eV relative to the putative global minimum (the leftmost structure) are listed. Green, purple, blue, orange, and brown lines represent the energies of singlet, triplet, quintet, septet, and nonet multiplicities, respectively. The point group symmetry (if not C<sub>1</sub>) is indicated for each structure.

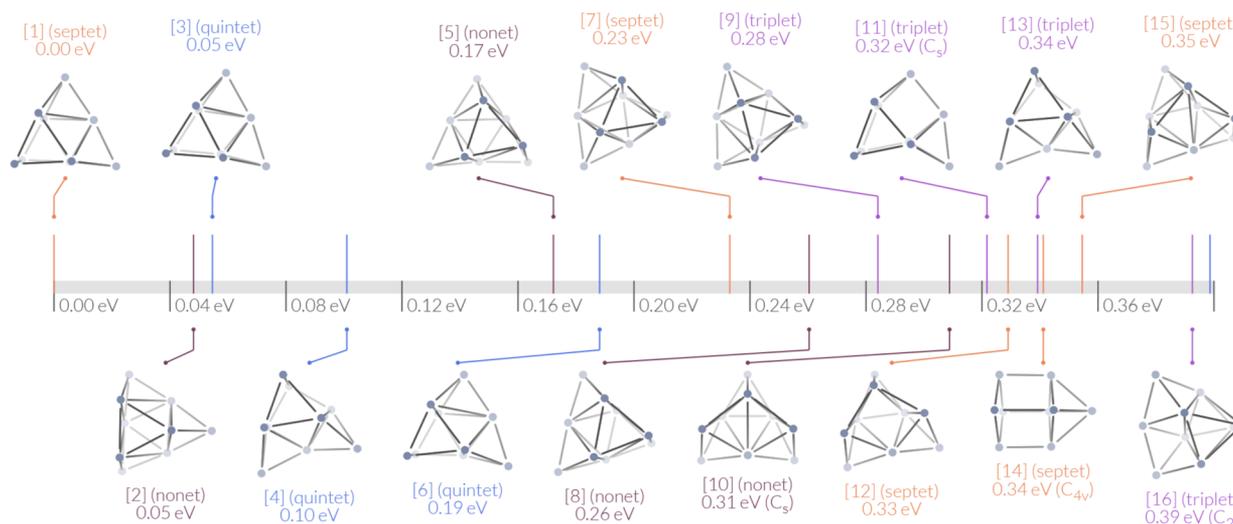


**Figure 8.** Energy and some of the structures of 58 low energy optima of Pt<sub>3</sub> found at the TPSSh/def2-TZVP level of theory using the NN-PGOPT approach (after final full-step DFT optimization). Notations are the same as that of the previous figure.

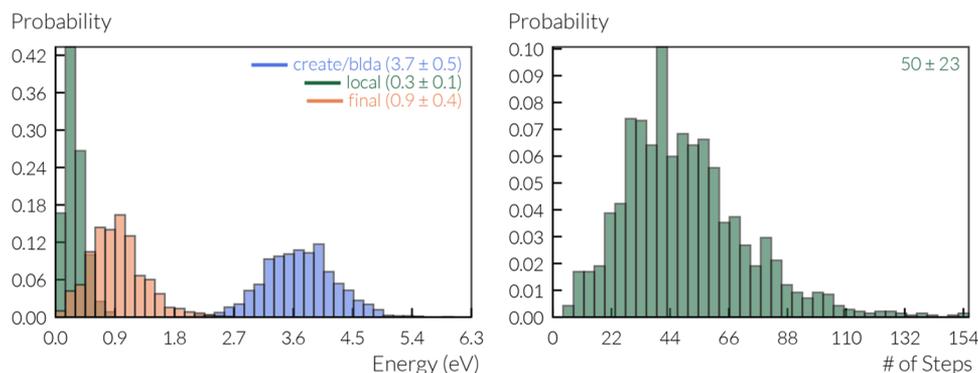
different multiplicity. Full-step DFT optimization with the def2-SV(P) basis set is then performed on these initial structures. The threshold value  $d = 0.10$  Å is used to filter the found local minima for each multiplicity individually. After the filtering, 264, 300, 300, 294, and 212 unique lower energy local minima are selected for DFT optimization with the def2-TZVP basis

set, respectively. The final results are filtered with the threshold value  $d = 0.25$  Å.

The CPU hours used during the DFT optimization for the direct scheme and the NN-assisted scheme are summarized in Table 2. We note that the total time for NN optimization and filtering for all structures and multiplicities is about 10 h using a



**Figure 9.** NN-relaxed low energy structures of Pt, using the NN-PGOPT approach (before final full-step DFT optimization). The single point energies (relative to the leftmost structure, which is 0.38 eV higher in energy than the putative global minimum in Figure 8) are re-evaluated at the TPSSh/def2-TZVP level of theory. Notations are the same as that of the previous figure.



**Figure 10.** The energy and step distribution during the first stage of the direct approach for Pt, global optimization. Left: the energy distribution of S-BLDA generated initial structures (blue), additional initial structures borrowed from local minima of another multiplicity (green), and relaxed local minima at the TPSSh/def2-SV(P) level of theory (orange). Right: the TPSSh/def2-SV(P) full-step local optimization step distribution. The mean and standard deviation of the distribution are also indicated in the figure.

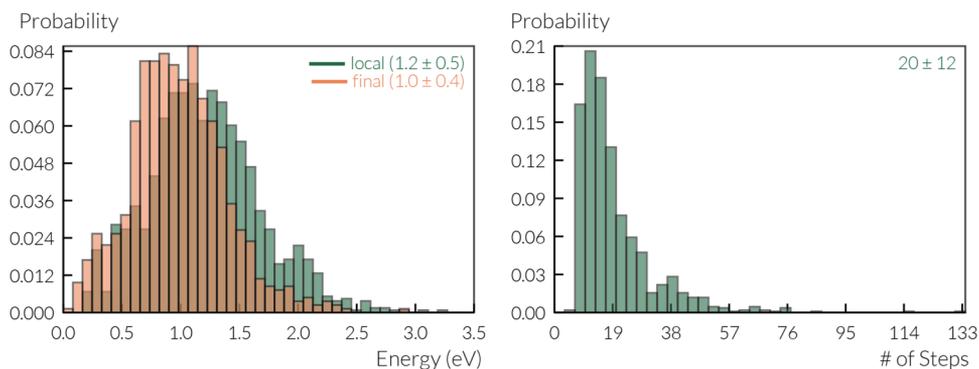
single CPU core. The total time for GPU accelerated NN fitting is shown in Table 1. Therefore, the time cost for NN fitting, optimization, and filtering is negligible compared to the DFT optimization time. In total, 23.5% CPU hours can be saved using the NN-PGOPT scheme for Pt, global optimization, compared to the direct approach. We note that at the first stage with the def2-SV(P) basis set, since only the first 10 steps of DFT optimization are performed using the NN-PGOPT scheme, 62% CPU hours are saved. However, because the quality of NN-relaxed structures is worse than that of the fully DFT-relaxed structures, at the second stage with the def2-TZVP basis set, 67% more CPU hours are required for the NN-PGOPT scheme. Nevertheless, in terms of the total time, NN-PGOPT is faster, given that the same number of initial structures is used.

**3.1.3. Performance of the Local Optima Search.** The final low energy isomers found using the direct approach and NN-PGOPT are listed in Figure 7 and Figure 8, respectively. We note that the number of isomers in the selected energy range (0.0–0.4 eV), which corresponds to a catalyst condition related region, is the same for the two approaches, and the first four isomers with lower energies are also consistent. The inconsistency in the higher energy region is due to the

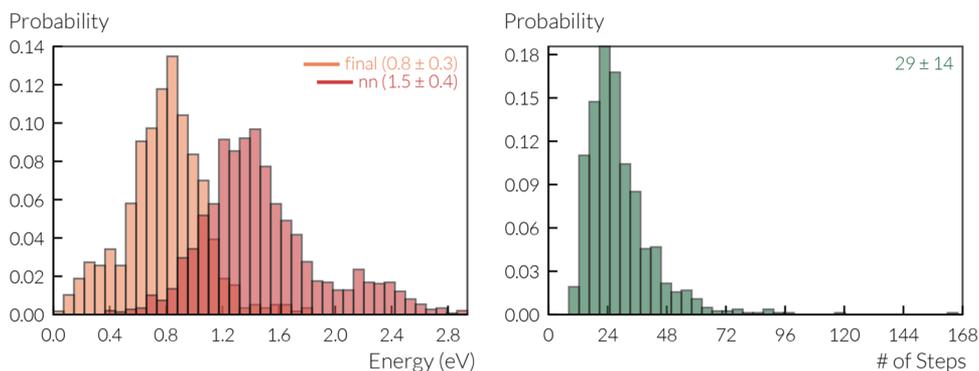
complexity of PES and that 500 initial structures are not enough to generate a complete list of local minima. However, the requirement of completeness is both unnecessary and computationally too expensive.

Figure 9 shows some NN-relaxed structures with relatively low energy re-evaluated at the DFT level. Since these structures are not real local optima at the DFT level, their energies are generally higher than the DFT-relaxed optima shown in Figures 7 and 8. However, the geometry of these NN-relaxed structures is already very close to their DFT-relaxed counterparts. For example, the NN-relaxed structures 1, 5, 2, 7, 8, and 3 in Figure 9 are very similar to the local optima 1, 2, 3, 4, 5, and 6 in Figure 8. We note that since the distances between these NN-relaxed structures and their corresponding DFT optima vary, the energy order is not expected to be consistent.

Figure 10 shows the energy and step distribution during the first stage (with the def2-SV(P) basis set) of the direct approach. The average energy of randomly created initial structures is 3.7 eV higher than the energy of the global minimum, while the average energy of the additional initial structures borrowed from local minima of other multiplicities is only 0.3 eV higher than the energy of the global minimum. This indicates that the effect of multiplicity change on energy is



**Figure 11.** The energy and step distribution during the second stage of the direct approach for  $Pt_9$  global optimization. Left: the energy distribution (at the def2-TZVP level) of def2-SV(P)-relaxed local minima structures (green) and rereaxed local minima at the TPSSh/def2-TZVP level of theory (orange). Right: the TPSSh/def2-TZVP local optimization step distribution.

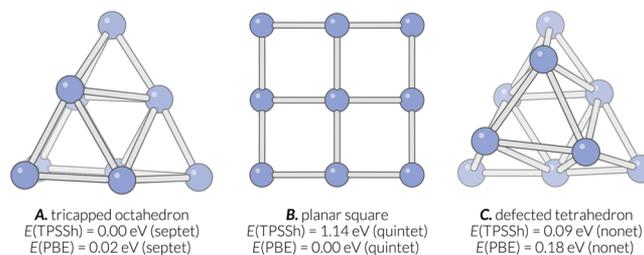


**Figure 12.** The energy and step distribution during the second stage of the NN approach for  $Pt_9$  global optimization. Left: the energy distribution (at the def2-TZVP level) of NN-relaxed local minima structures (red) and rereaxed local minima at the TPSSh/def2-TZVP level of theory (orange). Right: the TPSSh/def2-TZVP local optimization step distribution.

much smaller than that of geometry change. The average number of steps is 50 for the full optimization with the def2-SV(P) basis set. Note that in the NN-PGOPT approach only the first 10 of these steps are performed, but we can only save 62% of the time. This is because the beginning steps generally take longer time.

Figures 11 and 12 show the energy and step distribution during the second stage (with the def2-TZVP basis set) of the direct and NN approaches, respectively. The average relative energy of NN-relaxed and def2-SV(P) fully relaxed structures are 1.5 and 1.2 eV, respectively, which means that the quality of NN-relaxed structures is slightly worse. However, the average relative energy of final rereaxed structures for the NN and direct approaches are 0.8 and 1.0 eV, respectively. This indicates that in general the NN assisted approach is able to generate final structures that are more concentrated in low energy region.

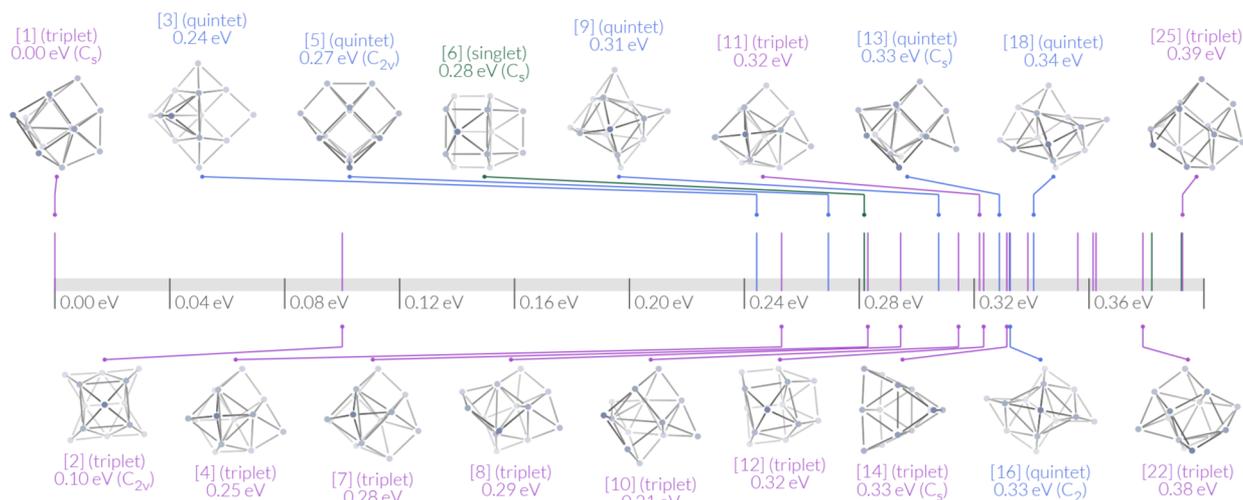
**3.1.4. Putative Global Minimum.** In this work the septet tricapped octahedron (A), shown in Figure 13, is found to be the putative global minimum for  $Pt_9$ . Singh and Sarkar<sup>49</sup> have also reported this structure (with nonet multiplicity) as the lowest energy isomer. However, Kumar and Kawazoe<sup>50</sup> and Chaves et al.<sup>51</sup> have reported the planar structure with four squares (B) as the lowest. Winczewski et al.<sup>52</sup> have reported the defected tetrahedron (C). We reoptimized the three structures using both TPSSh and PBE functionals. The structures and their DFT energies with these two functionals are shown in Figure 13. Based on the calculated energies, the difference in



**Figure 13.** Putative global minima for gas phase  $Pt_9$  cluster proposed in this work (A) and other literature (A, B, C). C is also found in this work. The TPSSh and PBE energies relative to the energies of respective lowest energy structure are also listed. The TPSSh energy is evaluated using Turbomole 6.6 with the def2-TZVP basis set. The PBE energy is evaluated using VASP 5.4.1 with 20 Å cubic cell and energy cutoff 500 eV.

global minimum structures is due to the different choice of the DFT functional.

**3.2. Gas Phase  $Pt_{13}$  Global Optimization.** **3.2.1. DFT Parameters and Fitting Errors.** The  $Pt_{13}$  global optimization is performed as another example of the application of the NN-PGOPT scheme. The DFT optimization is done using the spin-polarized DFT program implemented in Vienna Ab-initio Simulation Package (VASP)<sup>53</sup> 5.4.1 with the Projector Augmented-Wave (PAW) method<sup>54</sup> and the Perdew–Burke–Ernzerhof (PBE)<sup>55</sup> functional. The cubic cell of size  $15 \times 15 \times 15 \text{ \AA}^3$  is used. The multiplicity is automatically relaxed during the geometry optimization. Only the gamma point is sampled.



**Figure 14.** Energy and some of the structures of 25 low energy optima of  $\text{Pt}_{13}$  found at the DFT/PBE level of theory using the NN-PGOPT approach. Notations are the same as that of Figure 7.

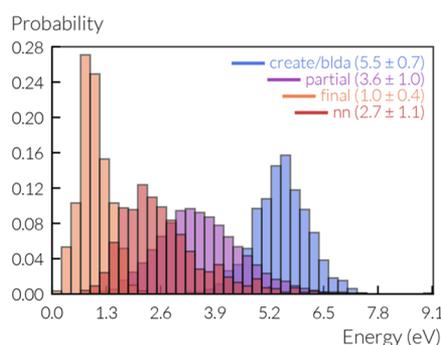
To save time, the DFT optimization at both the first and second stage of the NN-PGOPT scheme is performed with energy cutoff 280 eV. The final local optima are reoptimized with energy cutoff 400 eV, but no significant geometry or energy change is observed.

1000 unique initial structures are generated using the S-BLDA approach with the same parameters as the  $\text{Pt}_9$  case. Twenty DFT optimization steps are performed to yield 19 860 structure/energy samples. After discarding some high energy structures, 19 741 samples spanning an energy range of 5.44 eV are used. 82%, 9%, and 9% of samples are atomically shuffled to generate 200 000, 15 000, and 15 000 samples for training, testing, and validation, respectively. The minibatch size and the momentum are set to be 50 and 0.7, respectively, as the parameter of the MB-SGD-M method.

Since  $N = 13$  and  $k = 4$ , the layer sizes of the NN used for the  $\text{Pt}_{13}$  PES fitting are (715, 6) - (715, 40) - (715, 70) - (715, 60) - (715, 2) - (40, 2) - (10, 2) - (1). Therefore, the total number of parameters (including weights and biases) used in this NN is  $40(6 + 1) + 70(40 + 1) + 60(70 + 1) + 2(60 + 1) + 40(715 + 1) + 10(40 + 1) = 36582$ . The NN is trained for 700 epochs, but no error improvement has been observed since the 410th epoch. The final training, testing, and validation errors are 268, 298, and 303 meV, respectively.

After the PES fitting, 5000 unique initial structures are generated and then optimized to their local minima using the corresponding NN. Extrapolated structures and duplicates are excluded. Finally, 500 NN-relaxed local minima with relatively lower energies are selected to perform the full-step DFT optimization.

**3.2.2. Local Optima Searching Performance.** Some of the final low energy isomers found using the NN-PGOPT approach are listed in Figure 14. Although the PES of  $\text{Pt}_{13}$  is expected to be much more complicated than that of  $\text{Pt}_9$ , fewer low energy isomers are found within the 0.4 eV region. This indicates that the energy of the global minimum is significantly lower than most of other local isomers. This is different from the situation of  $\text{Pt}_9$  and may have important practical consequences, which is investigated in detail in the next subsection. Figure 15 shows the energy distribution of initial, partially DFT-relaxed, NN-relaxed, and fully DFT-relaxed structures. From the average values shown in the figure, we

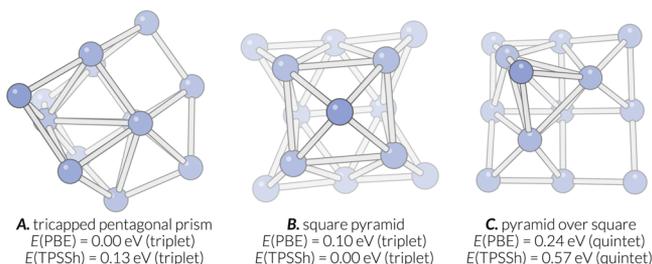


**Figure 15.** Energy distribution of S-BLDA generated initial structures (blue), all intermediate structures during the 20-step DFT optimization (purple), NN-relaxed local minima structures (red), and DFT rereaxed local minima (orange) during  $\text{Pt}_{13}$  global optimization. All energies are evaluated at the DFT/PBE level of theory.

can infer that the constructed NN has the ability to predict new lower energy structures than its input, because the average energy of NN-relaxed structures is 0.9 eV lower than that of the partially relaxed. Therefore, DFT local optimization time is saved by the combination of partial DFT optimization and NN fitting.

**3.2.3. Putative Global Minimum.** In this work the tricapped pentagonal prism (A) is found to be the putative global minimum for  $\text{Pt}_{13}$ , shown in Figure 16. Many other publications have also reported this structure as the putative global minimum.<sup>51,56,57</sup> However, Sun et al. and Zhang et al. have reported the square pyramid (B) as the global minimum.<sup>40,58</sup> As shown in Figures 14 and 16, the square pyramid structure is 0.10 eV higher in energy (with PBE functional) than the lowest one. Since the same DFT parameters are used, it may suggest that our search is more complete.

**3.3. Ensemble-Average Representation of Pt Clusters.** Both Figures 13 and 16 show that different DFT functionals could result in a different relative order of isomer energies and thus change the geometry of putative global minima. This may be partially due to the complexity of PES of both  $\text{Pt}_9$  and  $\text{Pt}_{13}$ . Nevertheless, at a realistic temperature, it is expected that the



**Figure 16.** Putative global minima for gas phase  $\text{Pt}_{13}$  cluster proposed in this work (A) and other literature (A, B) and a third-lowest isomer found in this work (C). The shown PBE relative energies are evaluated using VASP 5.4.1 with 15 Å cubic cell and energy cutoff 400 eV. Single point energies at the TPSSh/aug-cc-PVTZ level of theory are calculated for each of these structures using MOLPRO 2015.1.<sup>59</sup>

global minimum will become less important; the energy separations between local minima, however, will have a significant effect on properties. The energy distribution of the two clusters investigated in this work indeed show different patterns, especially if we only focus on PBE energies. Specifically, there are many more  $\text{Pt}_9$  isomers distributed in the low energy range (0 to 0.25 eV). In view of this, a quantitative analysis of the contributions of the isomers and the changes of properties in realistic temperature is necessary. This can be achieved by an ensemble-average representation of isomers. In what follows, we assume that the interconversion of isomers is not strongly kinetically hindered and that at high temperatures, eventually, the thermodynamic equilibrium would be reached, with every isomer being populated according to its free energy. Kinetic trapping will be the subject of future studies that will allow for the expansion of our model.

At finite temperature, the ensemble average of any property  $A$  can be approximately computed by taking the weighted average of that property of each geometrically unique isomer  $A_k$

$$\langle A \rangle \approx \sum_k P_k A_k$$

where  $P_k$  are the occurrence probabilities of isomer  $k$ . The probabilities  $P_k$  are obtained from the superposition approximation,<sup>60</sup> where the partition function of the system is expressed as the sum over individual partition function of isomers

$$Z = \sum_k Z_k, \quad P_k = \frac{Z_k}{Z}$$

In order to compute  $Z_k$ , the electronic, vibrational, and rotational degrees of freedom of the system will be considered.<sup>61</sup> The translational contribution is almost the same for each isomer, which is ignored here. Therefore, the partition function of isomer  $k$  is written as

$$Z_k = Z_{\text{elec},k} Z_{\text{vib},k} Z_{\text{rot},k}$$

where the electronic partition function is

$$Z_{\text{elec},k} = g_{\text{spin},k} e^{-\beta E_k}$$

where  $g_{\text{spin},k}$  is the spin degeneracy, which is equal to the multiplicity of the isomer.  $E_k$  is the ground state electronic energy of the isomer. It is assumed that for small enough clusters the quantum confinement effect is large enough to make the ground and excited electronic states well-separated in energy and contributions to the electronic entropy due to the

population of excited states being insignificant. This fact is generally supported by the lack of multireference character of the cluster wave functions in many selected Pt clusters that we tried (CASSCF test).

The vibrational partition function is (using quantum harmonic oscillator approximation)

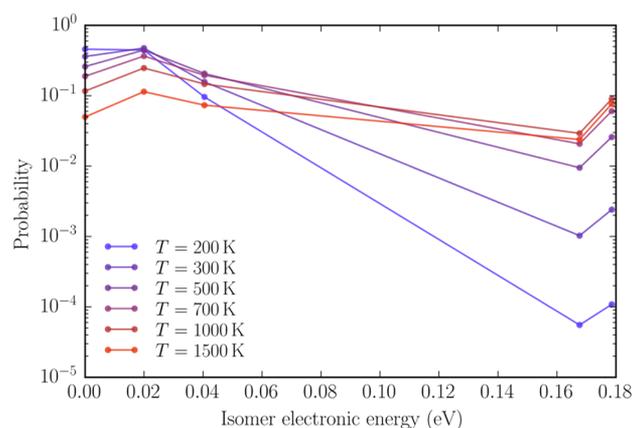
$$Z_{\text{vib},k} = \prod_i \frac{e^{-\beta \hbar \omega_{ki}/2}}{1 - e^{-\beta \hbar \omega_{ki}}}$$

where  $\omega_{ki}$  are  $3N - 6$  vibrational frequencies of the isomer  $k$ . Note that the Zero-Point Energy (ZPE) and vibrational entropy effect are included in this term. The rotational partition function is (in high-temperature approximate form)

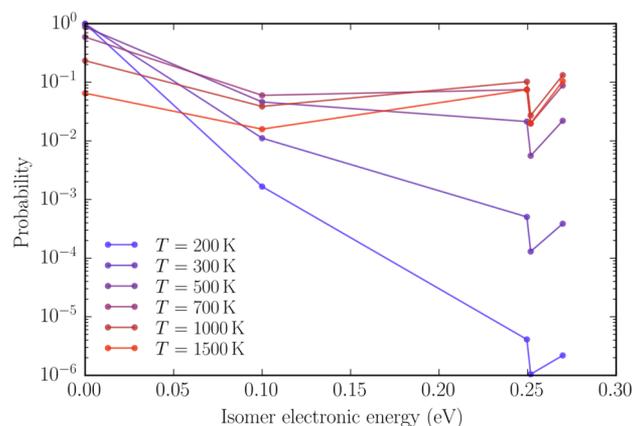
$$Z_{\text{rot},k} = \frac{\sqrt{\pi}}{\sigma_k} \left( \frac{2}{\beta \hbar^2} \right)^{3/2} \sqrt{I_{k1} I_{k2} I_{k3}}$$

where  $\sigma_k$  is the order of the rotational subgroup of the point group symmetry of isomer  $k$ , and  $I_{k1}$ ,  $I_{k2}$ , and  $I_{k3}$  are the three eigenvalues of the inertia tensor.

Figures 17 and 18 show the calculated probabilities of first five low energy isomers of  $\text{Pt}_9$  and  $\text{Pt}_{13}$  clusters at finite



**Figure 17.** Occurrence probabilities of the first 5 low energy isomers of  $\text{Pt}_9$  clusters at different temperatures, with energy and frequencies evaluated using the PBE functional.



**Figure 18.** Occurrence probabilities of the first 5 low energy isomers of  $\text{Pt}_{13}$  clusters at different temperatures, with energy and frequencies evaluated using the PBE functional.

temperatures, respectively. For Pt<sub>9</sub>, the PBE global minimum at zero temperature (planar square, quintet) is no longer the most populated one at  $T > 200$  K, mainly due to the fact that the second lowest isomer (tricapped octahedron, septet) has a higher multiplicity and lower symmetry, which are favored at high temperature. For Pt<sub>13</sub>, the structure favors its global minimal shape (tricapped pentagonal prism, triplet) until  $T = 1000$  K. After that, the third lowest isomer (pyramid over square, quintet) becomes more populated. We note that the second lowest isomer is not favored because of its slightly higher symmetry.

The probability change between different temperature indicates that there can be a structure transition for Pt<sub>9</sub> (with the PBE functional) at low temperature ( $T < 200$  K) and for Pt<sub>13</sub> at relatively high temperature ( $T \approx 1000$  K). The structural stability over a wide temperature range, predicted here for Pt<sub>13</sub>, may play an important role in the explanation of catalyst selectivity.

To confirm the existence of structure transition at a certain temperature of the two structures, the heat capacity can be computed as the following

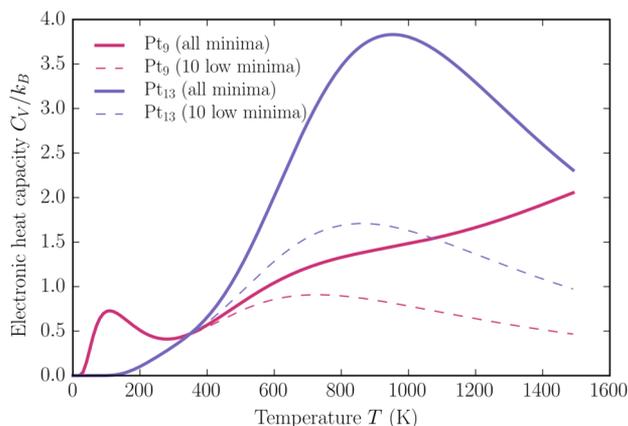
$$C_V = k_B \beta^2 \left[ -\frac{1}{Z^2} \left( \sum_k W_k Z_k \right)^2 + \frac{1}{Z} \sum_k (W_k^2 Z_k + V_k Z_k) \right]$$

where

$$W_k = \frac{1}{Z_k} \frac{\partial Z_k}{\partial \beta} = -\frac{3}{2\beta} - \sum_i \frac{\hbar \omega_{ki}}{e^{\beta \hbar \omega_{ki}} - 1} - \frac{1}{2} \sum_i \hbar \omega_{ki} - E_k$$

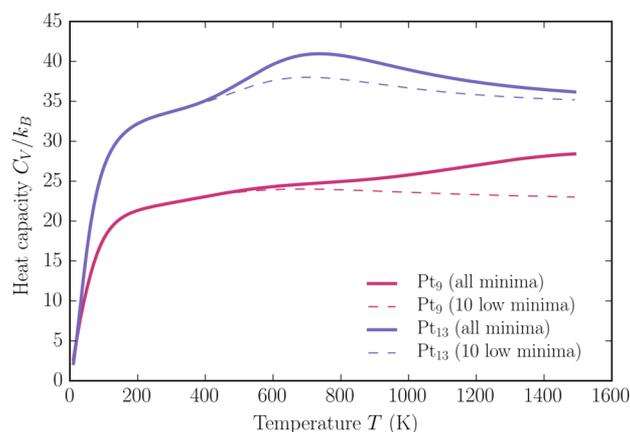
$$V_k = \frac{\partial W_k}{\partial \beta} = \frac{3}{2\beta^2} + \sum_i \frac{(\hbar \omega_{ki})^2 e^{\beta \hbar \omega_{ki}}}{(e^{\beta \hbar \omega_{ki}} - 1)^2}$$

Figures 19 and 20 show the heat capacity of the two cluster isomer ensembles as a function of temperature, for electronic



**Figure 19.** Heat capacity contributed from the electronic degree of freedom for isomer ensemble of Pt<sub>9</sub> (red) and Pt<sub>13</sub> (blue), when all found local minima (solid line) and only the first ten low energy isomers (dashed line) are considered.

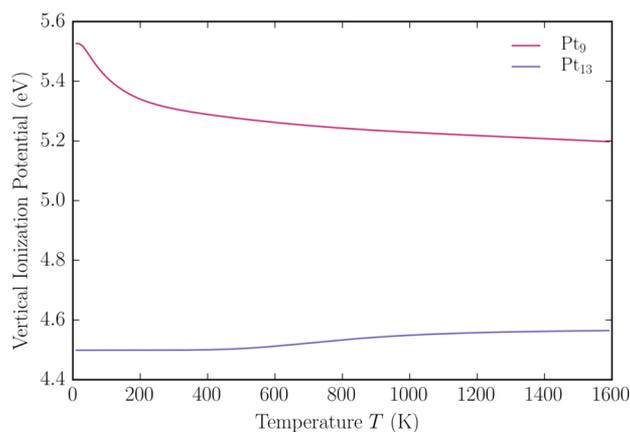
and all degrees of freedom, respectively. From the heat capacity contributed from the electronic degree of freedom, we can clearly identify a peak at  $T = 100$  K for Pt<sub>9</sub> and a peak at  $T = 950$  K for Pt<sub>13</sub>. This agrees with our previous observation on probabilities. We note that the peak for Pt<sub>9</sub> is too small to be observed in the total heat capacity, and the peak for Pt<sub>13</sub> in total heat capacity is shifted to  $T = 720$  K. This can be ascribed to



**Figure 20.** Heat capacity contributed from the electronic, vibrational, and rotational degree of freedom for isomer ensemble of Pt<sub>9</sub> (red) and Pt<sub>13</sub> (blue), when all found local minima (solid line) and only the first ten low energy isomers (dashed line) are considered.

the vibrational and rotational entropy effects. The dashed lines in Figures 19 and 20 show that a significant underestimation of heat capacity (and also some other properties) would be expected if we only consider several low energy isomers. Nevertheless, the first ten low energy isomers will be a good approximation if one focuses on phenomena at room temperature, for these systems. Figure 20 shows that the high temperature limits of total heat capacity for Pt<sub>9</sub> and Pt<sub>13</sub> are close to 22.5 and 34.5, respectively. This is related to the total number of degrees of freedom that we have considered, namely,  $3N - 6$  from the vibrational and 1.5 from the rotational ones, where  $N$  is the total number of atoms. Electronic degrees of freedom, however, do not contribute at high  $T$ .

Finally, in Figure 21 we note that the electronic property of the clusters at finite temperatures can be different from the



**Figure 21.** Ensemble-averaged vertical ionization potential of Pt<sub>9</sub> (red) and Pt<sub>13</sub> (blue), evaluated at different temperatures.

ones evaluated on just the global minimum. For example, The Vertical Ionization Potentials (VIPs,  $E(\text{Pt}_n^+) - E(\text{Pt}_n)$ ) of Pt<sub>9</sub> and Pt<sub>13</sub> show different trends when temperature increases. The VIP of Pt<sub>13</sub> changes relatively slowly, due to the fact that the energies of Pt<sub>13</sub> isomers are sparsely distributed.

#### 4. CONCLUSIONS

We proposed a DNN fitting based parallel global optimization scheme NN-PGOPT and demonstrated that this new scheme is

able to successfully find the global minima of Pt<sub>9</sub> and Pt<sub>13</sub>, which are in agreement with some other literature, as well as low energy optima, which are important at realistic temperature. Using the S-BLDA structure generation method, the configuration space can be randomly sampled in an efficient and natural way to generate initial structures. With the help of NN based local optimization combined with limited-step DFT optimization, the global optimization is proved to be faster than the traditional full-step DFT optimization embedded way. In addition, the training of DNN can be greatly accelerated by modern GPU accelerators. An efficient DFS based structure similarity measurement algorithm has also been proposed, and duplicates can be excluded at different stages of the global search. The tricapped octahedron structure is found to be the putative global minimum of Pt<sub>9</sub> using the TPSSh functional, but a planar structure has even lower energy when the PBE functional is used. The tricapped pentagonal prism structure is found to be the putative global minimum of Pt<sub>13</sub> using the PBE functional, while the TPSSh functional favors a square pyramid structure. However, at catalysis relevant temperature, the low symmetry and high multiplicity structures are predicted to be more populated, which is found to be a functional independent fact, for the system investigated in this work. Particularly, the structure transitions for Pt<sub>9</sub> and Pt<sub>13</sub> clusters can be identified, at relatively low and high temperatures, respectively, based on the ensemble-average representation of local minima. The energy separation of isomers can also have a significant influence on the properties observed at finite temperature.

Finally, we note that our DNN approach can also be combined with evolutionary algorithms, such as GA, when necessary. For larger systems, the time saving for DFT optimization may become even more important, where the global optimization can greatly benefit from the limited-step DFT optimization and the NN fitting approach. However, in this case the architecture of DNN should be carefully designed so that the number of samples needed to train the NN can be minimized. For practical applications, when the anharmonicity of PES is important, the ensemble average can be more accurately constructed using NN PES, in lieu of the harmonic approximation.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00994.

XYZ coordinates of Pt<sub>9</sub> and Pt<sub>13</sub> low energy isomers (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: ana@chem.ucla.edu.

### ORCID

Anastassia N. Alexandrova: 0000-0002-3003-1911

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Air Force Office of Scientific Research under a Basic Research Initiative grant (AFOSR FA9550-16-1-0141) and NSF Career Award (No. CHE1351968). CPU and GPU resources at the DoD High

Performance Computing Modernization Program (the US Air Force Research Laboratory DoD Supercomputing Resource Center (AFRL DSRC), the US Army Engineer Research and Development Center (ERDC), and the Navy DoD Supercomputing Resource Center (Navy DSRC)), supported by the Department of Defense, and the UCLA-IDRE cluster were used to conduct this work.

## ■ REFERENCES

- (1) Alonso-Vante, N. Platinum and non-platinum nanomaterials for the molecular oxygen reduction reaction. *ChemPhysChem* **2010**, *11*, 2732–2744.
- (2) Dadras, J.; Jimenez-Izal, E.; Alexandrova, A. N. Alloying Pt Sub-nano-clusters with Boron: Sintering Preventative and Coke Antagonist? *ACS Catal.* **2015**, *5*, 5719–5727.
- (3) Kou, Z.; Cheng, K.; Wu, H.; Sun, R.; Guo, B.; Mu, S. Observable Electrochemical Oxidation of Carbon Promoted by Platinum Nanoparticles. *ACS Appl. Mater. Interfaces* **2016**, *8*, 3940–3947.
- (4) Alexandrova, A. N.; Boldyrev, A. I. Search of the Li<sub>n</sub><sup>0/+1/-1</sup> ( $n = 5-7$ ) Lowest-Energy Structures Using the ab-initio Gradient Embedded Genetic Algorithm (GEGA). Elucidation of the Chemical Bonding in the Lithium Clusters. *J. Chem. Theory Comput.* **2005**, *1*, 566–580.
- (5) Alexandrova, A. N. H·(H<sub>2</sub>O)<sub>n</sub> clusters: Microsolvation of the hydrogen atom via molecular ab initio gradient embedded genetic algorithm (GEGA). *J. Phys. Chem. A* **2010**, *114*, 12591–12599.
- (6) Kanters, R. P. F.; Donald, K. J. CLUSTER: Searching for Unique Low Energy Minima of Structures Using a Novel Implementation of a Genetic Algorithm. *J. Chem. Theory Comput.* **2014**, *10*, 5729–5737.
- (7) Davis, J.; Shayeghi, A.; Horswell, S. L.; Johnston, R. L. The Birmingham Parallel Genetic Algorithm and its application to the direct DFT global optimization of Ir<sub>N</sub> ( $N = 10 - 20$ ) clusters. *Nanoscale* **2015**, *7*, 14032–14038.
- (8) Bandow, B.; Hartke, B. Larger water clusters with edges and corners on their way to ice: Structural trends elucidated with an improved parallel evolutionary algorithm. *J. Phys. Chem. A* **2006**, *110*, 5809–5822.
- (9) Call, S. T.; Zubarev, D. Y.; Boldyrev, A. I. Global minimum structure searches via particle swarm optimization. *J. Comput. Chem.* **2007**, *28*, 1177–1186.
- (10) Avendaño-Franco, G.; Romero, A. H. Firefly Algorithm for Structural Search. *J. Chem. Theory Comput.* **2016**, *12*, 3416–3428.
- (11) Wang, J.; Ma, L.; Zhao, J.; Jackson, K. A. Structural growth behavior and polarizability of Cd<sub>n</sub>Te<sub>n</sub> ( $n = 1-14$ ) clusters. *J. Chem. Phys.* **2009**, *130*, 214307.
- (12) Zhai, H.-J.; Zhao, Y.-F.; Li, W.-L.; Chen, Q.; Bai, H.; Hu, H.-S.; Piazza, Z. a; Tian, W.-J.; Lu, H.-G.; Wu, Y.-B.; Mu, Y.-W.; Wei, G.-F.; Liu, Z.-P.; Li, J.; Li, S.-D.; Wang, L.-S. Observation of an all-boron fullerene. *Nat. Chem.* **2014**, *6*, 727–731.
- (13) Oganov, A. R. *Modern Methods of Crystal Structure Prediction*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010; pp 133–134.
- (14) Zhai, H.; Ha, H.-A.; Alexandrova, A. N. AFFCK: Adaptive Force-Field-Assisted ab initio Coalescence Kick method for global minimum search. *J. Chem. Theory Comput.* **2015**, *11*, 2385–2393.
- (15) Sumpter, B. G.; Noid, D. W. Potential energy surfaces for macromolecules. A neural network technique. *Chem. Phys. Lett.* **1992**, *192*, 455–462.
- (16) Malshe, M.; Narulkar, R.; Raff, L. M.; Hagan, M.; Bukkapatnam, S.; Komanduri, R. Parameterization of analytic interatomic potential functions using neural networks. *J. Chem. Phys.* **2008**, *129*, 044111.
- (17) Le, H. M.; Raff, L. M. Molecular Dynamics Investigation of the Bimolecular Reaction BeH + H<sub>2</sub> → BeH<sub>2</sub> + H on an ab Initio Potential-Energy Surface Obtained Using Neural Network Methods with Both Potential and Gradient Accuracy Determination. *J. Phys. Chem. A* **2010**, *114*, 45–53.
- (18) Li, J.; Jiang, B.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems. *J. Chem. Phys.* **2013**, *139*, 204103.

- (19) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.
- (20) Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.
- (21) Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *83*, 153101.
- (22) Ouyang, R.; Xie, Y.; Jiang, D. Global minimization of gold clusters by combining neural network potentials and the basin-hopping method. *Nanoscale* **2015**, *7*, 14817–14821.
- (23) Chiriki, S.; Bulusu, S. S. Modeling of DFT quality neural network potential for sodium clusters: Application to melting of sodium clusters (Na<sub>20</sub> to Na<sub>40</sub>). *Chem. Phys. Lett.* **2016**, *652*, 130–135.
- (24) Manzhos, S.; Carrington, T. A random-sampling high dimensional model representation neural network for building potential energy surfaces. *J. Chem. Phys.* **2006**, *125*, 084109.
- (25) Manzhos, S.; Carrington, T. Using neural networks, optimized coordinates, and high-dimensional model representations to obtain a vinyl bromide potential surface. *J. Chem. Phys.* **2008**, *129*, 224104.
- (26) Bengio, Y. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127.
- (27) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012: *Neural Information Processing Systems*; **2012**, Lake Tahoe, Nevada.
- (28) Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy Layer-Wise Training of Deep Networks Yoshua. *Adv. Neural Inf. Process. Syst.* **2007**, *19*, 153.
- (29) Chetlur, S.; Woolley, C.; Vandermersch, P.; Cohen, J.; Tran, J.; Catanzaro, B.; Shelhamer, E. *cuDNN: Efficient Primitives for Deep Learning*. 2014; arXiv:1410.0759v3. arXiv.org ePrint archive. <https://arxiv.org/abs/1410.0759> (accessed Aug 6, 2016).
- (30) Sergeeva, A. P.; Averkiev, B. B.; Zhai, H.-J.; Boldyrev, A. I.; Wang, L.-S. All-boron analogues of aromatic hydrocarbons: B<sub>17</sub><sup>-</sup> and B<sub>18</sub><sup>-</sup>. *J. Chem. Phys.* **2011**, *134*, 224304.
- (31) Murrell, J. N.; Carter, S. Approximate single-valued representations of multivalued potential energy surfaces. *J. Phys. Chem.* **1984**, *88*, 4887–4891.
- (32) Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1989**, *39*, 5566–5568.
- (33) Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. What is the best multi-stage architecture for object recognition? In *12th Int. Conf. Comput. Vision*; IEEE: 2009; pp 2146–2153.
- (34) Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In *Int. Conf. Artif. Intell. Stat.* April, **2011**; *15*, 315–323.
- (35) Hochreiter, S.; Bengio, Y.; Frasconi, P.; Schmidhuber, J. Gradient Flow in Recurrent Nets: The Difficulty of Learning Long Term Dependencies. In *A Field Guide to Dynamical Recurrent Networks*; IEEE: 2009; 237–243.
- (36) Al-Rfou, R.; Alain, G.; Almahairi, A.; Angermueller, C.; Bahdanau, D.; Ballas, N.; Bastien, F.; Bayer, J.; Belikov, A.; et al. *Theano: A Python framework for fast computation of mathematical expressions*. 2016; arXiv:1605.02688v1. arXiv.org ePrint archive. <https://arxiv.org/abs/1605.02688> (accessed Aug 11, 2016).
- (37) Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, **2013**; Vol. 3, 1139–1147.
- (38) Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* **1997**, *23*, 550–560.
- (39) Ponce, Y. M. Total and Local Quadratic Indices of the Molecular Pseudograph's Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds. *Molecules* **2003**, *8*, 687–726.
- (40) Sun, Y.; Zhang, M.; Fournier, R. Periodic trends in the geometric structures of 13-atom metal clusters. *Phys. Rev. B* **2008**, *77*, 075435.
- (41) Lv, J.; Wang, Y.; Zhu, L.; Ma, Y. Particle-swarm structure prediction on clusters. *J. Chem. Phys.* **2012**, *137*, 084104.
- (42) Abraham, N. L.; Probert, M. I. J. Improved real-space genetic algorithm for crystal structure and polymorph prediction. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 134117.
- (43) Even, S. *Graph Algorithms*; Even, G., Ed.; Cambridge University Press: Cambridge, 2011; pp 46–48.
- (44) TURBOMOLE V6.6 2014, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007. Available from <http://www.turbomole.com> (accessed Nov 14, 2016).
- (45) Treutler, O.; Ahlrichs, R. Efficient molecular numerical integration schemes. *J. Chem. Phys.* **1995**, *102*, 346.
- (46) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129–12137.
- (47) Soini, T. M.; Genest, A.; Nikodem, A.; Rosch, N. Hybrid Density Functionals for Clusters of Late Transition Metals: Assessing Energetic and Structural Properties. *J. Chem. Theory Comput.* **2014**, *10*, 4408–4416.
- (48) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (49) Singh, N. B.; Sarkar, U. Structure, vibrational, and optical properties of platinum cluster: a density functional theory approach. *J. Mol. Model.* **2014**, *20*, 2537.
- (50) Kumar, V.; Kawazoe, Y. Evolution of atomic and electronic structure of Pt clusters: Planar, layered, pyramidal, cage, cubic, and octahedral growth. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 205418.
- (51) Chaves, A. S.; Rondina, G. G.; Piotrowski, M. J.; Tereshchuk, P.; Da Silva, J. L. F. The Role of Charge States in the Atomic Structure of Cu<sub>n</sub> and Pt<sub>n</sub> (n = 2–14 atoms) Clusters: A DFT Investigation. *J. Phys. Chem. A* **2014**, *118*, 10813–10821.
- (52) Winczewski, S.; Rybicki, J. Structure of Small Platinum Clusters Revised. *Comput. Methods Sci. Technol.* **2011**, *17*, 75–85.
- (53) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169–11186.
- (54) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953–17979.
- (55) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (56) Chou, J. P.; Hsing, C. R.; Wei, C. M.; Cheng, C.; Chang, C. M. Ab initio random structure search for 13-atom clusters of fcc elements. *J. Phys.: Condens. Matter* **2013**, *25*, 125305.
- (57) Bunău, O.; Bartolomé, J.; Bartolomé, F.; Garcia, L.-M. Large orbital magnetic moment in Pt<sub>13</sub> clusters. *J. Phys.: Condens. Matter* **2014**, *26*, 196006.
- (58) Zhang, M.; Fournier, R. Density-functional-theory study of 13-atom metal clusters M<sub>13</sub>, M = Ta–Pt. *Phys. Rev. A: At., Mol., Opt. Phys.* **2009**, *79*, 043203.
- (59) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. et al. *MOLPRO, version 2015.1, a package of ab initio programs*. See <http://www.molpro.net> (accessed Nov 14, 2016).
- (60) Ballard, A. J.; Stevenson, J. D.; Das, R.; Wales, D. J. Energy landscapes for a machine learning application to series data. *J. Chem. Phys.* **2016**, *144*, 124119.
- (61) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; 2007; pp 429–433.