IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Advanced Machine Learning Based Global Optimizations

Dr. N. Prasad Babu,

Principal, CSTS Government Kalasala,

Jangareddigudem, Andhra Pradesh, 534447, India

ABSTRACT

Pt-nanoclusters have attracted attention due to their extensive use as catalysts in various sectors and their catalytic capabilities, instigating a theoretical investigation to correlate structure and property. On the other hand, it is challenging to find stable and reliable structures to support experimental results at the nanoscale due to their fluxional nature at ambient temperature. The major objective of this work is to test the capability of stable and reliable structure findings at the nanoclaster region by Gaussian Process Regression (GPR) model potentials on the-fly within the evolutionary framework using the Bayesian optimization approach. The entire algorithm is called Global Optimizations by GPR (GO-GPR) learning. In this regard, the GO-GPR algorithm examined the potential energy surfaces of bare P t a^ * nanoclusters of sizes (n = 3 - 6, 7, 8, 10, 13) GO-GPR identified new low- lying isomers and global minimum structures are in correlation . In the case of P*t_{13} and Pts nano-clusters, the global minimum structure is close to the second lowest energy structure, implying these clusters can have fluxional nature. In fact, a few experimental studies have shown that P*t_{B} and P*t_{13} are effective in catalyzing reactions.

Keywords: Global optimizations (GO) Bayesian optimizations approach Gaumian regression proces (GPR) potentials Pr-nanoclusters.

1. Introduction

An Atomic level understanding of a wide range of disciplines, for example nano-catalysis, and molecular biology. It requires atomic structure search methods like global optimizations approaches for searching minimum energy structure in a multidimensional energy landscape. Random search [1], basin and minima hopping (BH) [2,3], and evolutionary algorithms [4-8] are examples of such approaches. These atomic structure search methods demand large-scale computa- tions due to the supplement of energy and forces depending on the method used for calculation. The accuracy of results mainly depends on the accuracy of energy and forces supplied and also the effectiveness of the algorithm being used. First principal methods and density functional theory (DFT) calculations are used for reproducing experimentally [9] relevant structure but they are computationally demanding.

Machine Learning (ML) algorithms have been successfully used since 2007 to mimic expensive DFT energies and forces by fitting networks such as artificial neural networks (ANN) [10-15], and kernel-based regression [16-19]. These ML-based algorithms offer huge speedups in predicting DFT energy and forces with just minor accuracy loss. Global optimizations are ramped up utilising ML. models rather than expensive

DFT approaches, allowing for search at a fraction of the initial cost. ML. models are learned on-the-fly and adjust its accuracy by correcting the network weights [20,21]. Inexpensive ML. models allow the scientific community to explore potential energy surfaces at experimentally relevant sizes. Hence, integrating ML. algorithms with established global optimization algorithms speeds up the search for global atomistic structures.

Metal nanoclusters received a lot of attention because of their unique physicochemical, optoelectronic, and magnetic properties. This is due to the fact that nanoclusters have a higher surface to volume ratio and a finite energy level difference than their bulk counterparts.

Especially, small Pt-nanoclusters have attracted much attention due to their catalytic properties [22,23]. For example, supported Pt-nanoclusters were used as catalyst for water gas shift reaction. (WGSR) [24]. In the studies based on the fuel cell, the oxygen reduction reaction is accelerated by small Pt-nanoclusters [25]. Pty and a few other clusters had a significant impact on oxygen reduction in acid conditions when the electrodes were investigated for electrolytic oxygen reduction and ethanol oxidation processes [25]. P*t_{R} nanoclusters are capable of catalyzing the reduction of NO due to its low symmetry structure.

Recent theoretical investigations on nanoclusters for studying adsorption strength of CO, NO, and OH molecules showed that the Pt nanocluster has high adsorption tendency to the adsorbents when compared to the pure Cu13 cluster and these appear to be equally competitive in their bimetallic form when doped with suitable compo- sition of Cu metal [27]. As a result, understanding the structure-property relationship in these nanoclusters is of critical relevance.

Several global optimization studies have been carried out to inves tigate the structure of Pt nanoclusters at various levels of theory [28-32]. The methods are either expensive in terms of computational time or less accurate to locate relevant low-lying structures. Therefore, we have implemented the Bayesian optimization algorithm within the evolutionary frame work according to the Bisbo et al., [33,34]. The evolutionary algorithm generates new genes from population and Bayesian optimization provides relevant data for active learning process. Algorithm utilizes GPR potentials to understand Pt-nanoclusters at molecular level.

The current paper is organized in implementing Bayesian optimiza- tion technique with GPR potentials in the methodology section followed by demonstrating the low-lying isomers of the Pt nanoclusters of 3-13 atoms in gas phase in the results section.

2. Methodology

2.1. Computational methods

Global optimizations are carried out employing GO-GPR ML algo- rithm [33]. DFT computations are performed in the Atomic Simulation Environment (ASE) package [35] using Grid-Based Projector Augmented Wave (GPAW) method [36,37]. For global optimizations, the linear combination of atomic orbitals (LCAO) approach was utilised, while for local relaxations, the plane wave (PW) method with spin-polarized (spinpol True) calculations was utilised. During DFT computations, 10 A distance was maintained in the x, y, and z di- mensions to avoid super cell interactions. The Brillouin zone sampling was done using k (1, 1, 1) points during global optimization and k (2, 2, 2) points were used while local relaxations. The Perdew-Burke-Ernzerhof (PBE) [38] functional was utilised in the case of the PW method.

2.2. Gaussian process regression (GPR)

Regression is the process of determining the best fitness function to represent a group of points. The Gaussian process is a probabilistic approach for locating an unknown point in the data by using prior knowledge about the function. GPR is a fitting of these data points to a normal distribution. In essence, each data point has a normal distribu- tion, and the aggregate distribution of all data points becomes Gaussian. As a result, the mean (4) and covariance matrix (2) can be used to describe this multivariate Gaussian distribution. The mean vector a in- dicates the expected value of each data point. The covariance matrix (2) represents the variance of each data point as well as the correlation with the others. They matrix is a symmetric, positive semi-definite matrix. The diagonal values represent the variance (6) of the ith data point, whereas the off-diagonal components represent the correlation of the ith data point with the jth data point. As stated in equation (1), the covariance matrix is defined in terms of expectation value (E). Where X and X, are data points.

2=Coυ (X. X.) = E(X-6) (X)-0)

Eq(1)

GPR is a supervised machine learning algorithm that takes minimal data points to learn and often requires less data points than the ANN algorithm. GPR is appropriate when data generation is expensive. Along with predictions, it also provides the uncertainty associated with each prediction. They have been applied successfully to regression and clas sification problems in several disciplines [39]. A set of hyper parameters, such as the autocorrelation coefficient, overall scale of the function, and gaussian noise level, are tuned to adjust the system-dependent properties.

2.3. Bayesian global optimization with GPR

Bayesian optimization is a strategy for solving problems of unknown functional form by applying sequence of strategies with evaluation of function using black box. This particular form of optimization is employed for either maximisation or minimization of an unknown function. Also, this is best suitable, if an unknown functional form is highly expensive to evaluate in terms of both function value and de rivatives. The following important steps are necessary for constructing the Bayesian optimization.

1) If the function is unknown, it should be treated as a random function, and prior distribution is used when collecting data.

2) Predict the posterior function assessment using surrogate model training, such as GPR, utilising the previous data.

3) Placing the acquisition function will allow to combine unknown. function data with existing data.

4) The acquisition function always allows the black box to investigate the desired data with predictable behaviour.

The aim of GPs is to estimate energy E (Xnew) and variance (r) of a new structure as shown in Equations (2) and (3) using a Gaussian normal distribution over the data P (Eg X, E) of training set. Training set consisting of structures X = (x1, x2,...,x) and corresponding en- ergies E (e, eд ex). To feed geometrical data to the GPR, where X (1 to N) is used as a feature vector rather than cartesian coordinates.

E(X)=(k+1) (E-(X)) + (X) Eq(2)

(X)=(XX)-(+) Eq(3)

Where Kk(X.X) and konk(X.X) and the k and are covari ance functions defined between data and new data point.

The GPR is updated on-the-fly as shown in Fig. 1. By controlling the fitness function, the expected uncertainty will help us in exploring critical surfaces. The model starts with few data points and iteratively data points are increased based on the acquisition function defined in Equation (4). The data is then collected using the acquisition function (g(X)), which takes into account both expected energy and data uncer- tainty. The degree of exploration in the search is determined by this tuneable parameter kappa (x)..

(X)E(X)-(X)

Fig. 2, shows a schematic diagram for GO-GPR method as discussed above. The approach begins with 10 randomly generated structures that are relaxed at the DFT level of theory. These randomly generated pop- ulation as shown in Fig. 3 is then utilised to fit GPR model. Following that, new child structures are created with mutations incorporated while the evolutionary process is operating. The GPR potential that was fitted in the earlier steps is then used to relax these mutated structures. The relaxed structures are then utilised to find the best fit structure by uti- lising the data base's energy and uncertainty as shown by Equation (4),

The best suited structure is subsequently processed for single point

Eq(4)

DFT calculations before being added to the population and data base. In subsequent cycles, GPR model is fitted with the new data set and the accuracy is improved as the number of steps increases. The low-lying energy structures are then updated to population as shown in Fig. 3. The global minimum structure is discovered after a few iterations. To confirm the global minimum structure, 10 parallel runs are submitted simultaneously.

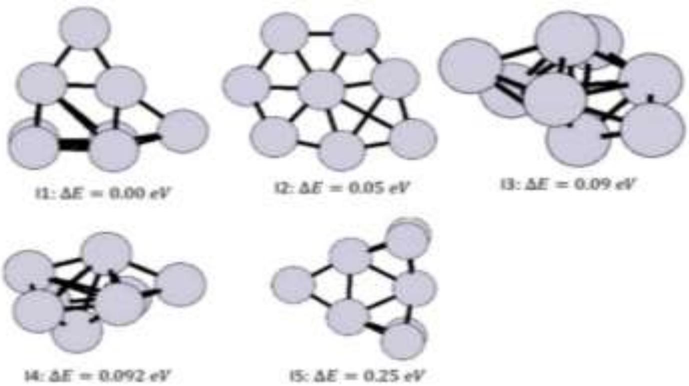

Fig. 5. Putative low-lying isomers for Pt_8 nanoclusters.

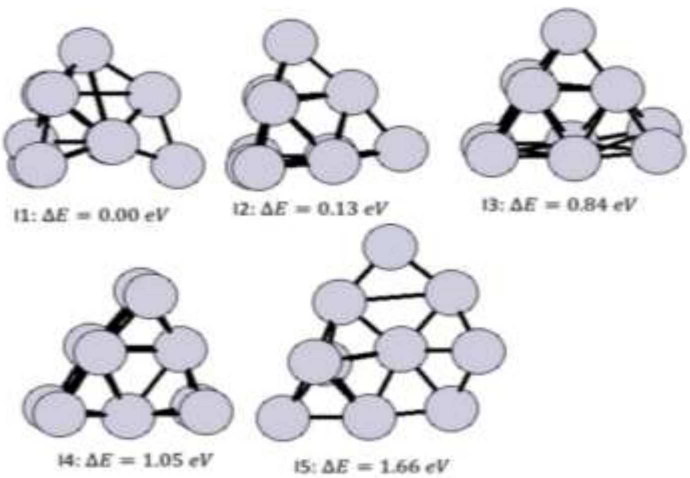

Fig. 6. Putative low-lying isomers for Pt_{10} nanoclusters.

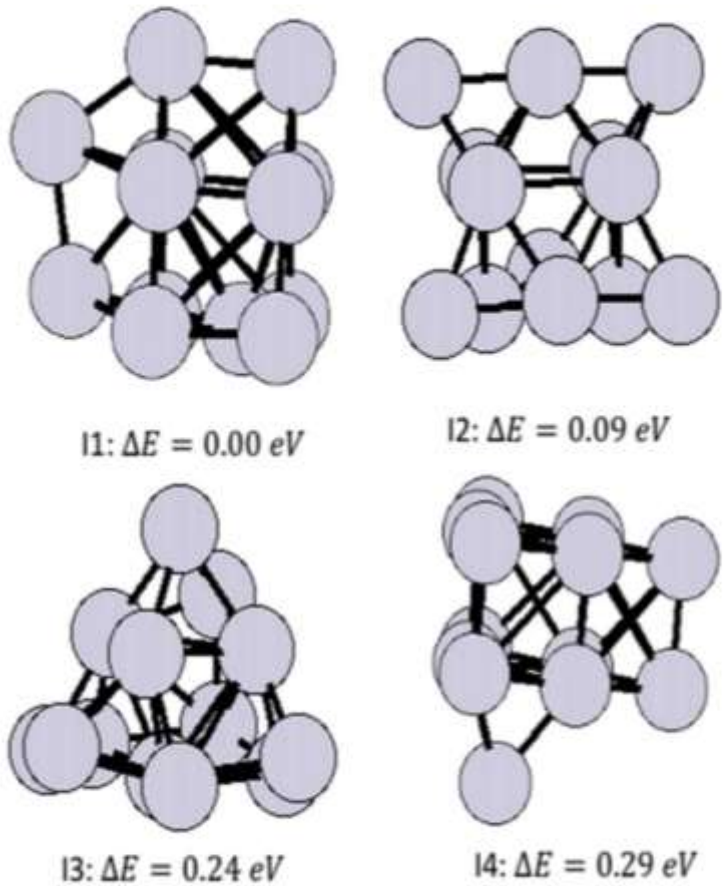

Fig. 7. Putative low-lying isomers for Pt_{13} nanoclusters.

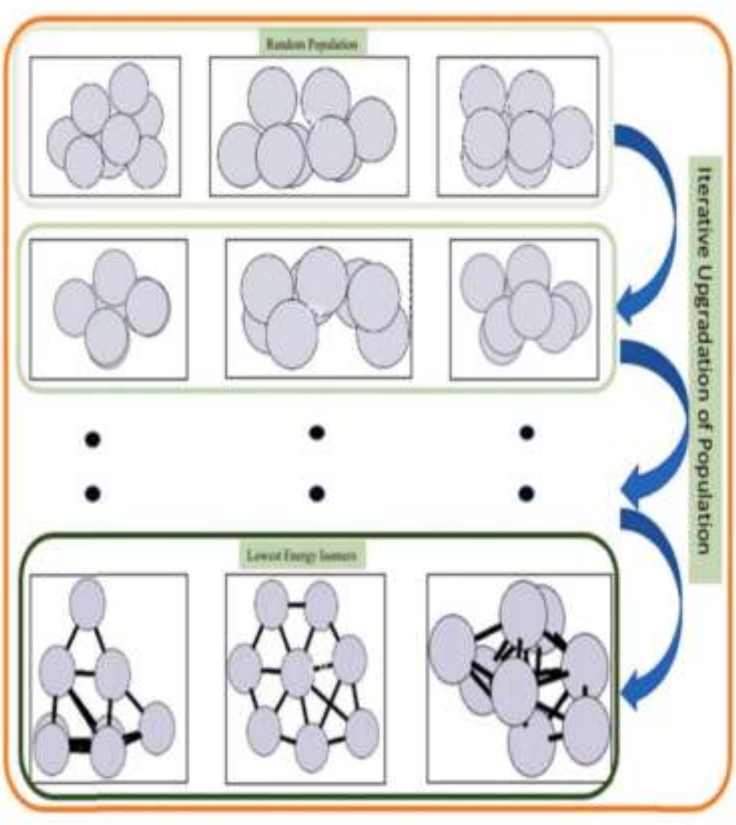

Fig. 3. Iterative Population Updating on-the-fly.

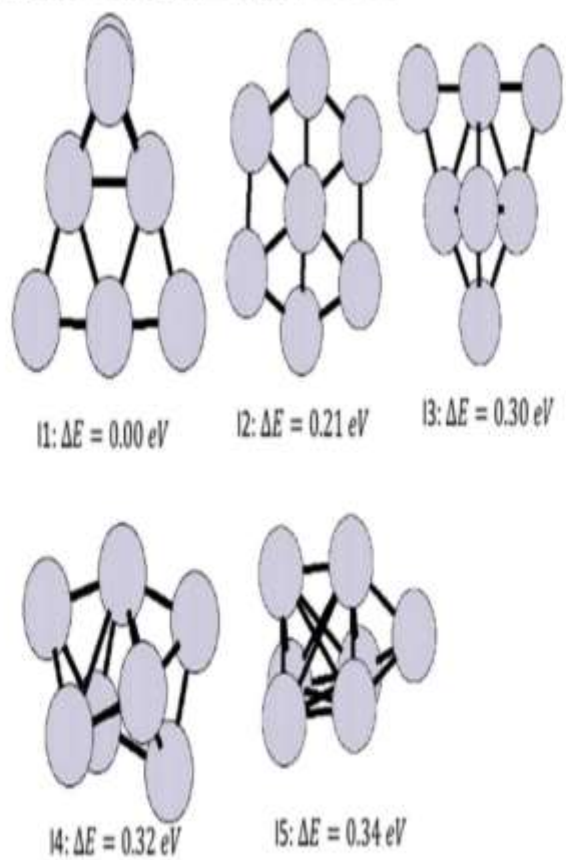

Fig. 4. Putative low-lying isomers for bare Pt_7 nanoclusters.

3. Results and discussion

3.1. Very small P t pi^ - nanoclusters (n = 3 - 6) In one of the study, Jennings and co-workers used genetic algorithm (GA) coupled with DFT to ascertain the global minimum (GM) structures of small platinum clusters (Pt fp ,n=3-6) . The results show that Pt clusters have non-singlet ground states and that spin multiplicity can change the geometry of GMs [40]. A global minimum search was, therefore, conducted on P*t_{n} with n = 3 - 6 The global minimum for P*t_{3} to P*t_{6} were consistent with the GA-DFT study.

3.2. Global optimization of P*t_{tau} nanoclusters

Fig. 4 depicts the lowest energy configurations of P*l_{7} nanoclusters. The GM structure is a quasi-2D triangular planar shape with an addi- tional add atom in the corner. According to Fig. 4, 3D structures (13, 14, 15) are the least stable, whereas quasi-2D and planar topologies (11, 12) are the most stable for P*t_{7} nanoclusters. Interestingly, in one of the study, P*L_{7} nanocluster on A*l_{2}*O_{3} support, quasi-2D geometries play an important role in the activation of alkene while dehydrogenation [41].

3.3. Global optimization of P*t_{6} nanoclusters

The low-lying P*t_{s} isomers are shown in Fig. 5. Fig. 5 shows that although 3D, 2D, and quasi-2D geometries are all feasible, 3D geometry is the global minimum. Only a 50 meV energy difference separates the global minimum 3D structure (11) from the planar hexagon structure (12), making both planar and 3D type arrangements are competitive. It indicates that a structural transition from 2D to 3D occurs at Pt Bn but there is no straightforward method to predict experimentally, the cata lytic properties of this shift. Scanning tunnelling microscopy research revealed that 2D to 3D transition occurs at n = 8 and that the activation energy of the CO oxidation reaction decreases [42]. Another DFT study examined the 2D and 3D geometries of a P*t_{6} cluster on an anatase T * 1O_{2}(101) surface [43,44]. In the C*O_{2} photo reduction reaction, 3D geometries are critical for activating C*O_{2} through the formation of bent C*O_{2} This is because C*O_{2} adsorbs more strongly at interfaces and at the edges of 3D geometries. 3D geometries are more fluxional and undergo geometrical changes due to C*O_{2} adsorption, whereas 2D geometries are strongly bound on the surfaceFig. 5 supports the fluxional nature of P*t_{8} nanocluster because it can exist in three different forms such as 11, 12 and 13 as shown in Fig. 5 at ambient temperatures due to less energy difference (90meV).

3.4. Global optimization of P*t_{10} nanoclusters

Fig. 6 depicts the low-lying isomers of the P*t_{10} nanoclusters. The majority of the P*t_{10} low-lying isomers were made of a triangle with six atoms on the plane and four more atoms on the top. As all of the structures are three-dimensional, a flat configuration is less desirable. In terms of energy, the global minimum (11) separates the next lowest energy structure (12) by 0.13 eV and no other isomer found till 0.84 eV energy difference, implying that the P*t_{10} nanocluster is a stable cluster with just one or two isomeric forms ambient temperature.

3.5. Global optimization of nanoclusters

Fig. 7 depicts four probable low-lying isomers for the P*t_{13} nano- cluster. All of the low-lying isomers have 3D structures. The most common basic structure in all formations is a square planar shape, on top of which the remaining atoms are organized. Structures with high- symmetry (11 and 12) are preferable, while structures with low symmetry (13 and 14) are less so. Deformed structures (13 and 14), on the other hand, formed at 0.24 eV P*t_{13} nanocluster can interconvert from 11 to 12 at ambient temperatures to less energy difference (90 meV). This interconversion of nanocluster structures can be good for selective catalysis of important organic reactions.

Conclusion:

By adapting the advanced machine learning algorithm "GO-GPR", low-lying isomers are sampled for bare Pt-nanoclusters with sizes (n = 3-6, 7, 8, 10, 13) and putative global minimums are found. Planar structures are seen as the global minimum in P*t_{7} nanoclusters, whereas 3D structures are seen as the global minimum in higher nanoclusters. In case of Pts, both 2D and 3D structures are competitive. The triangle arrangement is most typical in all P*t_{a} cluster formations, and on top of that, the extra add-on atoms were positioned to make as many bonds feasible. In case of P*t_{13} nanocluster two distinct structures with a 90 mev energy difference from global minimum were found. In the instance of Pta nanoclusters, the energy difference between two low-lying iso- mers is 50 meV. As a result, P*t_{8} and P*t_{13} nanoclusters may be best suitable for catalysis as in accordance with experimental studies. In the future, it is planned to extend on this work by performing critical re-actions like WGSR and CO-oxidation on supported Pt nanocluster cata- lysts with reaction path generators. And it is intended to combine this Go-GPR with the reaction path generating algorithms on which we are currently working.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

(1) CJ. Pickant, .. Needs, Ab bitin Random Structure Seasching 2011.
(2) DJ Wales, LPK. Dove, Global Optimization by Basis Hopping and the Lewest Energy Structures of Lennard-Jones Centers Containing np ο το 110 Αταση, 1996
(3) 1. Zhang J. Han, I Wang [a scalable model with the accuracy of quantum mechanics. Phys. Bes Lett. 120 (14) (2018)
(4) M. Ganegger, L. Schmiedrak, M. Bittermann, F. Berzsenyi, P. Marquetand, WACSF weighted storm contered symmetry functions as descriptors in machine lesening potensials, J. Chem. Phys. 148 (24) (2018)
(5) KT. Schän. LE Saucoda, P.1. Kindermans A. Thatchenko, K.R. Miller, Schiet- [deep leaming architecture for mulecules and materials. J. Chem. Phys. 148 (24) (2018).
(6) . Chirk, Bulu Modeling of DFT quality arunal network potential fur sodium clusters application to melting of sodium cheaters (Na20 m Nat), Chem. Phys Lett 652 (2016) 130-135.