Received: 14 November 2024 | Revised: 22 December 2024 | Accepted: 03 January 2025

https://doi.org/10.56946/jce.v4i1.589

Journal of Chemistry and Environment

Research Article

Optimization of Silver Nanocluster Geometries: A Deep Reinforcement Learning Approach to Identifying the Most Stable Configurations in Ag_{15} Cluster

Malik Ahmed Mubeen[1,2*], Fuyi Chen[1,2], Khalid Mehmood Ur Rehman[3]

[1]State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, Xi'an 710072, China.
[2]School of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China.
[3]Department of Physics, Ghazi University, DG Khan, Pakistan.
*Correspondence E-mail: malikahmedmubeen@mail.nwpu.edu.cn

Abstract

Silver nanoclusters (Ag_{15}) are of huge interest due to their unique electronic, optical, and catalytic properties, which all strongly depend on their atomic geometries. Finding their most stable configurations is an important step toward understanding and exploiting such properties. Traditional optimization methods, including genetic algorithms and basin hopping, usually perform poorly for systems with complex potential energy surfaces. This paper reports on a Deep Reinforcement Learning based strategy to explore and optimize Ag_{15} clusters. We show that the navigation of the potential energy surface of such silver nanoclusters through deep reinforcement learning allows us to identify the most stable configurations more efficiently than through conventional methods. This method has been demonstrated to perform much better than previous approaches regarding computational efficiency and quality of the identified configurations, hence providing new insights into nanocluster stability.

Keywords: Nanoclusters, global minimum, deep reinforcement learning, effective medium theory, actor critic.

1. Introduction

The unique properties of silver nanoclusters, radically different from bulk silver, are determined by quantum effects [1-6]. Indeed, most of its properties, such as catalytic activity [7], optical response [8], and electronic structure [9], strongly depend on atomic configuration. With this aim, to take full advantage of such properties, it becomes mandatory to determine the most stable atomic geometries corresponding to the global minimum GM configuration of the Potential Energy Surface-PES. Finding the GM configuration is a challenging task since, for such clusters, PES is typically rugged containing many local minima. GAs and BH are two conventional methods that have been traditionally employed for structure optimization of nanoclusters. However, these techniques often suffer from slow execution and are often trapped by local minima [10-14]. Also, these techniques are rather computationally expensive for larger clusters. Normally, the geometrical optimization of a nanocluster intends the attainment of the total-energy minimum of the system in question, which usually appears to be a function of all the atomic positions.

Some usual procedures, such as GA methods, rely on simulating evolution processes-mutation and cross-choice with selection-to probe PES. Basin-hopping has also been performed-complementing random steps by the application of local optimizations-with problems in the scale for potential energy surfaces. Both of these approaches are very sensitive to the initial configuration and easily get stuck into a local minimum, particularly for larger clusters. As far as using these techniques goes, most challenges present themselves; because of this, alternative recent approaches include the application of machine learning methods (ML) and

reinforcement learning methods (RL). Reinforcement Learning (RL), specifically Deep Reinforcement Learning (DRL), has emerged as an effective approach to addressing optimization problems in high-dimensional domains. DRL use neural networks for approximation of policies that integrate states (configurations) into actions (atomic shifts). The model learns through interaction with the environment, obtaining feedback (rewards) that corresponds to the energy of the configuration, and systematically optimizing its strategy over time. Recent studies have shown that DRL is capable of effectively exploring complex PES for small molecules and nanoclusters. This method addresses particular challenges of conventional techniques by acquiring optimal action strategies via trial and error, thereby facilitating enhanced efficiency and precision in optimization. Silver nanoclusters, exemplified by Ag_{15}, are significant in a range of applications, encompassing catalysis, sensing, and optical devices.

The properties of these clusters depend on their size, shape, and atomic arrangement, making the determination of their stable configurations critical. Recent advancements in computational chemistry and machine learning have enabled more accurate predictions of nanocluster geometries, but there is still a need for faster and more efficient methods for larger systems. For metal clusters, Pt_9 and Pt_{13} clusters, Zhai and Alexandrova previously presented a GPU-accelerated global optimization method that integrates deep neural network (DNN) fitting with limited-step density functional theory (DFT) optimization, drastically cutting down on computational cost compared to full DFT local optimization [15]. To speed up the process of finding the global minimum configurations for nanoclusters that are either pure or alloyed, Raju et al. designed a framework for active learning genetic algorithms [16]. Hansen and associates used a symmetry constrained GA with a neural network potential to estimate the energies of Pt−Ni nanoalloys [18], while Wang et al. [17] used on-the-fly machine learning to speed up the genetic algorithm (GA) search for aluminum nanoclusters. Results from investigations of carbon clusters in gas-phase and supported environments show that the GOFEE algorithm, created by Bisbo and Hammer [19], an evolutionary algorithm improved with a machine-learned surrogate model and Bayesian statistics, effectively finds low-energy structures in complicated energy landscapes defined by first-principles methods. In response to these difficulties, this research presents a new method for detecting GM configurations in nanoclusters by PES scanning that makes use of deep reinforcement learning (DRL). To enable the model to produce a wide variety of geometries, this method starts with randomly configured nodes. For efficient PES scanning and precise GM configuration localization, such diversity is critical. An important step forward, our DRL-based method provides a more effective and efficient way to investigate the intricate terrain of nanocluster topologies and may be able to circumvent the shortcomings of conventional evolutionary algorithms. The framework is sufficiently versatile to be used with nanoclusters made of a single metal element as well as those made of an alloy.

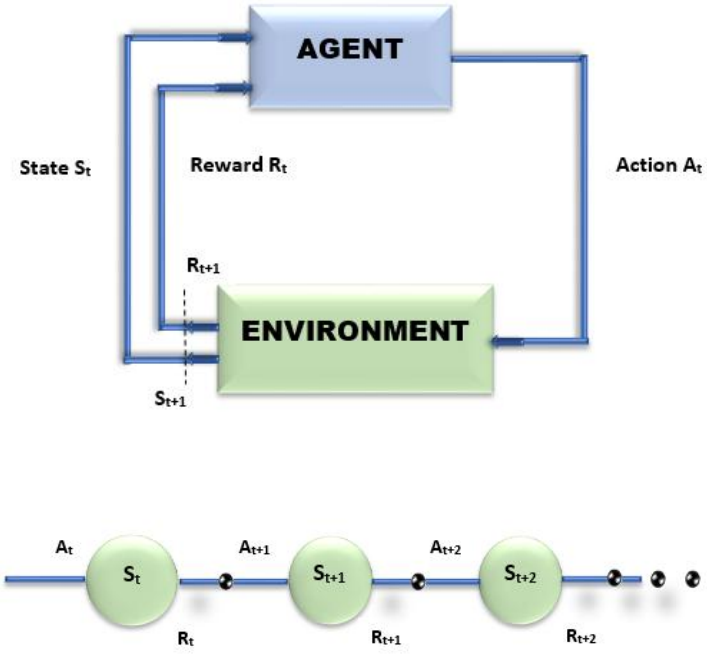

Figure 1. Reinforcement learning mechanism

The problem in this research addresses the optimization of the geometries of silver nanocluster (Ag_{15}) to identify their most stable configurations efficiently and accurately. The limitations of traditional optimization methods call for the application of more advanced techniques that can handle the complex PES of nanoclusters and swiftly find the lowest energy configurations and the global minimum. This study aims to apply Deep Reinforcement Learning (DRL) to the optimization of Ag_{15} nanocluster. The primary objective is to Employing Deep Reinforcement Learning (DRL) to efficiently explore the PES of Ag_{15} nanocluster, identify their GM configuration and the most stable configurations (low-energy states).

2. Methodology

2.1 Representation of Ag15 nanoclusters

Ag_{15} nanoclusters are described by atomic coordinates in Cartesian space, where each of the 15 atoms is represented by its position in 3D space. This configuration provides a complete structural representation of the nanocluster, which will serve as the basis for energy calculations and further optimizations. Atom-Centered Symmetry Functions (ACSFs) will be used to encode the structural features of Ag_{15}. ACSFs monitor pairwise distances and angular connections among atoms, serving as a reliable descriptor for machine learning models [33]. This, in turn, will also introduce the descriptors that will have properties enabling the projection of the atomic arrangement of a cluster to a high-dimensional feature space, allowing a reinforcement learning agent to represent an understanding of a cluster architecture. Finally, the nanocluster energies for Ag_{15}-first are calculated using one of the efficient computational ways-known for metallic systems-EAM calculations. The method which is applied here is one called the Embedded Atom Method, computing the preliminary approximation of energy for the whole nanocluster. This would involve the use of DRL to explore the Ag_{15} nanocluster for the PES. A generally assigned agent, with randomly sampled or changed initial configurations, takes an action in interaction with the environment to reduce further energy levels. An acting agent, depending on a cluster state, executes one operation, such as translations and rotations of atoms to identify the global minimum or other low-energy configurations. The agent balances exploration and exploitation in such a way that it uncovers the most optimal structure of the nanocluster. Regarding increasing diversity in initial configurations of the Ag_{15} cluster, "move," "rotate," and "rattle" are some mutation operations that can be made. Those will result in changing the structure of the cluster into diverse starting states an agent could learn from. The objective is to avert premature convergence of the agent to local minima and to assure comprehensive exploration of the potential energy surface.

2.2 Deep reinforcement learning framework

DRL is an efficient way to explore high-dimensional configuration spaces, which naturally emerge due to the huge number of possible atomic arrangements. This is especially true for nanoclusters such as Ag_{15}. Indeed, DRL is much more appropriate for dealing with these complex spaces since it will efficiently navigate through intricate energy landscapes and optimize molecular structures. Most of the traditional methods get stuck with huge numbers of possible atomic configurations that DRL can overcome, given its ability to refine their strategies by interacting with the system. That makes up one alternative in effectively finding GM. Applications from DRL in molecular design have already proven their capability by efficiently optimizing chemical reaction paths and molecular structures while outperforming many previous conventional methods. DRL is a subcategory of Artificial Intelligence that incorporates both Reinforcement Learning and Deep Learning [20]. An RL system learns through interactions with its environment by the actions executed to attain rewards or incur penalties. Such learning thus enables an agent to achieve an objective, which is to maximize cumulative rewards over time. DRL effectively merges RL with deep neural networks such that the agent can manage high-dimensional state spaces and complicated environments. By using neural networks to represent decision-making policies, DRL can learn and discover innovative strategies that are not predefined.

In the case of Ag_{15} nanoclusters, the formulation of the DRL frame is such that it optimizes the possibility of PES and explores the GM configuration. The balanced exploration-exploitation ability of DRL is important in researching new configurations and refining the known ones that are effective to give the best atomic arrangement in the cluster [21]. This immediately offers a more significant concept than that of traditional variational methods when the exploration of a nanocluster configuration would be required. The RL problem is couched, in this instance, in the exploration of PES for Ag_{15} clusters, seeking the most stable configuration, or GM. That framework follows the MDP model wherein, at every step in time, the agent chooses an action given its current state and receives a reward, changing its state. The agent learns an optimal policy to maximize expected cumulative rewards.

To mitigate the challenges posed by high-dimensional state spaces, DRL uses deep learning techniques to represent policies, enabling the agent to process complex environments effectively. We applied the Trust Region Policy Optimization (TRPO) algorithm to stabilize policy learning. TRPO combines both value-based and policy-based methods, using an actor-critic network where the actor determines actions and the critic evaluates them. This approach reduces the variance of policy gradients and enhances training stability. In our study, we have illustrated how to formulate an RL problem for exploring the PES of nanoclusters, with the goal of detecting GM structures. Reinforcement learning mechanism is presented in figure 1. This formulation encompasses the state (S_t), action (A_t), reward function (R_t), Transition Function $T(S_{t+1}/S_t, A_t)$ and Policy (π_θ), each tailored to the specific challenges of nanocluster PES analysis [22-26].

(a) State Space S_t

The state space S_t represents all possible configurations of the Ag_{15} cluster. Each state corresponds to a specific atomic arrangement, and the configuration is encoded using Atom-Centered Symmetric Functions (ACSFs). ACSFs serve as descriptors of the local atomic environment and capture the essential structural features needed for DRL to evaluate the configuration's energy. Binary vectors are also employed to record certain events, such as whether the agent encounters overlapping atoms, dissociation, or revisits a previously identified minimum. These encoded states are processed by multilayer perceptrons and used as input for the critic network.

(b) Action space A_t

The action space A_t consists of two types of actions that the agent can perform at each step. In action 1 the agent randomly selects an atom from the Ag_{15} cluster. In the action 2 selected atom is then moved by a distance of either +2.0 Å or −2.0 Å, with the direction of movement determined by a vector from the atom's initial position to the center of mass of the cluster. This action space allows the agent to explore different atomic arrangements by selecting atoms and repositioning them, effectively searching for local minima and eventually the global minimum.

(c) Reward function R_t

The reward function R_t is designed to encourage the agent to find low-energy configurations. Positive rewards are given for configurations that have lower energy than the previous configuration. Negative rewards penalize configurations that cause atom overlap or cluster dissociation. The goal of the agent is to maximize cumulative rewards by progressively discovering configurations with lower energy. The reward structure for the Ag_{15} nanocluster is as follows. If the agent's action causes an atom to dissociate from the cluster, a significant penalty of −10 is imposed. If the agent's actions result in overlapping atoms, another penalty of −10 is applied. If the agent revisits a previously identified local minimum in the same episode, it receives a penalty of −10. When the agent finds a new local minimum with lower energy than the initial configuration, it receives a reward proportional to the energy difference. The reward is calculated as following equation 1.

$$reward = 1000 \times \Delta E \qquad \text{Eq (1)}$$

where ΔE is the energy difference in eV between the new

minimum and the initial configuration.

If the agent finds a configuration with higher energy than the initial setup, no reward is given. These rewards encourage the agent to find stable, low-energy configurations while avoiding configurations that result in atom overlap or dissociation. The termination criteria for each episode are as follows. The episode ends when the agent identifies five new lower-energy minima. Alternatively, the episode also concludes when the maximum number of steps is reached.

(d) Transition function $T(S_{t+1}/S_t, A_t)$

The transition function determines the new state S_{t+1} after taking action A_t in state S_t, representing a new atomic configuration.

(e) Policy (π_θ)

Policy π_θ is a strategy or mapping from states to actions. The policy determines the agent's behavior at any time. It can be deterministic (one action for each state) or stochastic (a distribution over actions for each state).

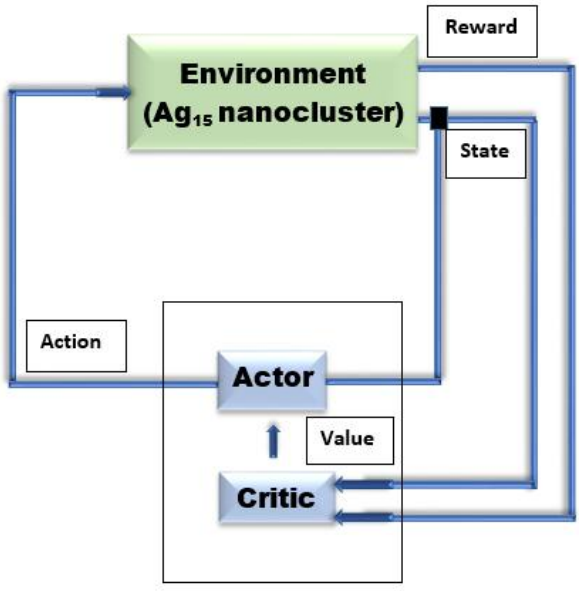

Figure 2. Actor-Critic RL architecture.

2.3 Actor critic architecture

The DRL model uses an actor-critic architecture [30-32], where the actor selects actions and the critic evaluates the quality of those actions based on the rewards received. The Trust Region Policy Optimization (TRPO) [29] algorithm is used to optimize the agent's policy during training, balancing exploration (trying new configurations) and exploitation (refining known good configurations). This actor-critic architecture process is shown in figure 2. The actor decides on an action based on current policy and environment provides feedback in the form of reward and new state. The critic evaluates the selected action's value and computes a learning signal. The actor updates the policy based on the feedback from the critic. This process repeats iteratively until the policy converges to an optimal solution.

2.4 Computational methods

In the geometry optimization and energy-force calculations of the nanoclusters, we have used the effective medium theory potentials. EMT is a semiempirical interatomic potential widely used in materials science, especially for the study of metallic systems and alloys, because of its computational efficiency. EMT represents the potential energy of an atom within a material using an "effective medium" that reflects the influence of surrounding atoms. The total energy is a combination of pairwise interactions and contributions from the electron density of neighboring atoms, capturing key aspects of metallic bonding. The EMT has found wide applications in the investigations of structural properties, formation energies, surface energies, and defect structures of metals, alloys, and nanoclusters, serving in this respect as an effective tool for large-scale simulations where quantum mechanical methods cannot be applied.

2.5 Computational tools

This research use the computational tools specified in the technique provided:

2.5.1 OpenAI Gym

OpenAI Gym is one of the common toolsets for the creation and evaluation of reinforcement learning algorithms. In this work, OpenAI Gym is utilized as a basis for the implementation and testing of a DRL model applied to the optimization of nanocluster structures. It provides the possibility to work in a unified way with different environments in a very convenient manner while designing experiments and comparing algorithms. The Gym environment of the nanocluster's potential energy surface, coupled with the DRL,

allows the agent to better explore and optimize the atomic topologies [36].

2.5.2 Tensorforce

Tensorforce is an extended version of TensorFlow that allows creating and elaborating on state-of-the-art deep reinforcement learning to model RL. As part of the model, in the present paper, it is adopted for training the DRL agent with its exploration of PES regarding nanoclusters. By dealing with basic reinforcement learning complications, it involves the learning based on neural networks and policy optimization. Therefore, Tensorforce improves nanocluster topologies according to energy evaluations on the model and refines the process of decisionmaking of an agent [37].

2.5.3 Embedded atom method (EAM):

The work applied the Embedded Atom Method to obtain the basic energy calculation for the nanoclusters. The EAM model is a semi-empirical approach that has been widely used to model atomic interactions in metallic systems within material science and technology. The EAM expresses the total energy of a system in terms of a contribution from each atom, taking into consideration the local atomic environment surrounding every atom. EAM represents a computationally fast way of estimating the interatomic forces and energy that are necessary in large-scale simulations, such as the optimization of nanocluster architectures [27].

2.6 Deep reinforcement learning experiments

In this respect, we experimentally demonstrate the proposed DRL framework on Ag_{15} nanoclusters by using OpenAI Gym and Tensorforce-known for their robustness and flexibility to solve complex reinforcement learning problems [28]. Each experimental episode was designed in such a way that there should be a maximum execution of 200 action steps in order to comprehensively explore the action space without wasting computing resources. Other important parts of our research involved the energy and force calculation for nanoclusters using the effective medium theory potentials. EMT can be regarded as a very reliable and computationally efficient estimation of interatomic forces and energy in metallic systems [38-40]. By embedding EMT into our DRL framework, we guaranteed that the model's inferences were based on physically correct data, hence enhancing the reliability of the optimization process. In the quest for preferred nanocluster configurations, the reinforcement learning agent was tasked with the objective of finding five different lower-energy minima in each episode. The training methodology started with randomly generated nanocluster structures that were relaxed to local minima. These relaxed setups therefore formed the basis for a DRL framework that enabled further training.

2.7 Generation of initial random configurations

In this study related to nanocluster geometries using the DRL framework, there is a requirement to generate several initial states for each training-again and again-either by randomly generating a single geometry or generating a couple of random geometries with some stochastic process. A newly obtained hybrid configuration, akin to "offspring," is generated in this process by the use of crossover operations, as commonly called "mating operations.". The methodologies used in the crossover tactics are taken from the BPGA [34-35]. In order to increase the diversity of these initial states and include a wider range of possible configurations, we have implemented various mutation processes from the BPGA in the creation of the initial random geometries. The idea here is that the methodology would involve a random creation of configurations in the first phase, followed by a mutation operation in the second, so as to ensure that training starts with truly diverse initial configurations.

2.8 Identification of global minimum and low-energy stabled configurations

In each episode, the model observed all low-energy configurations and separately archived the lowest energy configurations. A collection of the lowest energy configurations is generated, usually comprising a certain quantity (10 configurations). As new minima are identified during the training process, each new minimum supplants the highest energy configuration in the pool. This iterative procedure persisted until the training wrapped up. Upon completion of the training, the model discerned the global minimum (GM) and other low-energy configurations from the archived minimum

configurations throughout the episodes. Alternatively, upon completion of the training phases, a specialized analysis is performed to ascertain the ground state and the lowest energy configurations. This research entails scrutinizing the retained minimal configurations gathered throughout the training episodes.

3. Results and discussion

In this study we have elucidated the detailed analysis of our DRL experiment on Ag_{15} cluster. The DRL model successfully identified the global minimum and the most stable configurations for Ag_{15}, significantly reducing the time to convergence compared to traditional methods. The agent was able to find low-energy configurations with high accuracy, often outperforming GAs and BH in terms of the number of steps required to reach the GM. Figure 3 represents the progression of episodic rewards during the training of a Deep Reinforcement Learning (DRL) model applied to the optimization of Ag_{15} nanocluster.

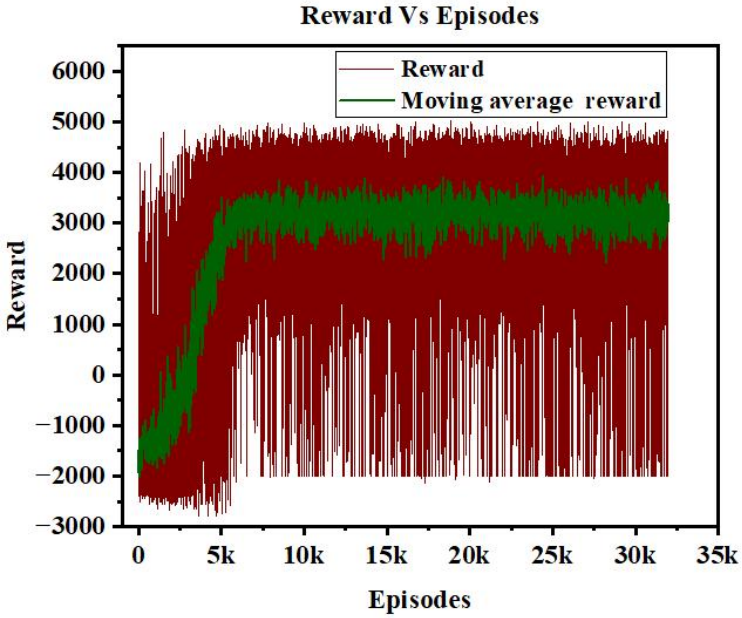

Figure 3. Rewards throughout the training session.

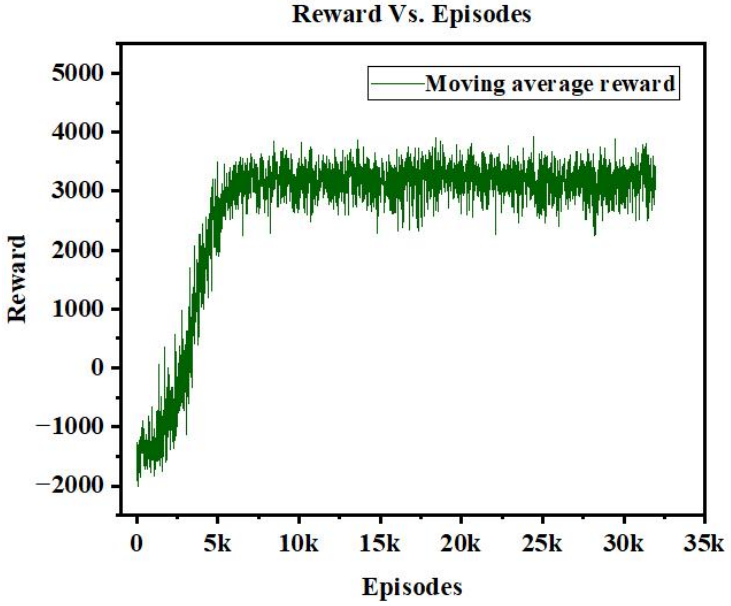

Figure 4. Moving average reward.

The graph contains two key elements: Brown Lines – These represent the episodic rewards the agent received after each action or step during the training episodes. In reinforcement learning, the agent learns to make decisions based on feedback, which is given as a reward for its actions in a given state.

Green Curve – This represents the moving average of the episodic rewards, which is calculated over a specified window of episodes to smooth out the fluctuations in the rewards and show the general trend of the agent's learning progress. The moving average at each episode t, denoted as MA_t, is calculated as the average of the last N episodic rewards (equation 2).

$$MA_t = \frac{1}{N}\sum_{i=t-N+1}^{t} R_i \qquad \text{Eq (2)}$$

This formula means that the moving average at time t is the average of the rewards from episodes $t - N + 1$ to t. MA_t is the moving average of the rewards at time step t, R_i is the episodic reward at step i, N is the window size, which is the number of episodes over which the average is calculated. The brown lines represent the episodic rewards at each episode, which are given to the agent based on its actions during that episode. In optimizing the Ag_{15} nanocluster, the agent's goal is to explore various configurations and find the most stable (lowest-energy) structure. Moving average reward is represented in figure 4.

3.1.1 Exploration vs. Exploitation:

In early training, the agent is more likely to explore the Potential Energy Surface (PES) of the Ag_{15} nanocluster, which results in higher fluctuations in the episodic rewards. This stage is essential for discovering different configurations. As the agent learns which configurations are more stable, it starts to exploit the better solutions it has discovered. This reduces fluctuations and increases the episodic rewards, resulting in a more consistent upward trend. The green curve (moving average) highlights the balance between exploration and exploitation. The agent initially explores new configurations, but over time, it focuses on refining its knowledge to exploit already discovered low-energy configurations.

3.1.2 Convergence to global minimum:

As the green curve starts to plateau, this likely indicates that the agent has identified a set of low-energy configurations and is no longer making significant improvements. The curve is flattening at a 6500 episodic rewards value, it suggests that the agent has successfully optimized the Ag_{15} nanocluster, possibly converging to its global minimum state near-optimal configuration.

3.1.3 Learning efficiency:

The rate at which the green curve rises gives insight into how quickly the agent is learning. A sharp increase is indicating the rapid learning and improvement in identifying stable configurations.

3.1.4 Training Stability:

A flattening or plateau of the green curve at 6500 episodic rewards is an indication of stability in the learning process, where the agent has likely achieved convergence. In optimization of Ag_{15}, this indicates that the optimization process is successful in identifying stable, low-energy atomic configurations.

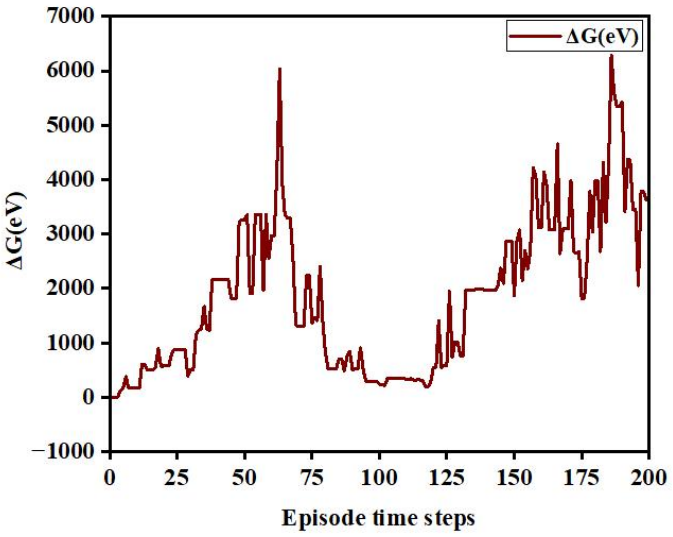

Figure 5. Energy profile before stable policy.

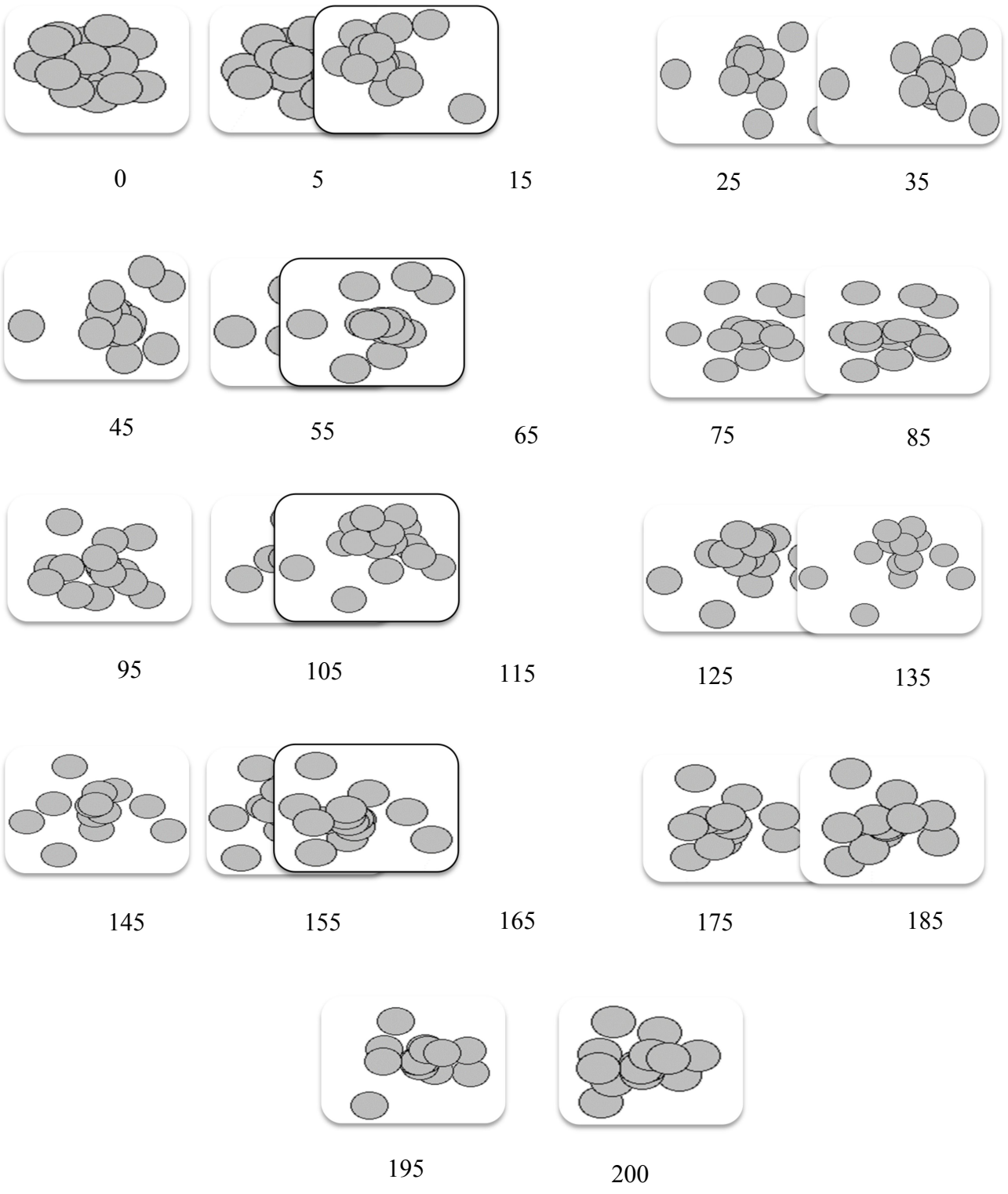

Figure 6. Configurations at different time steps before stable policy.

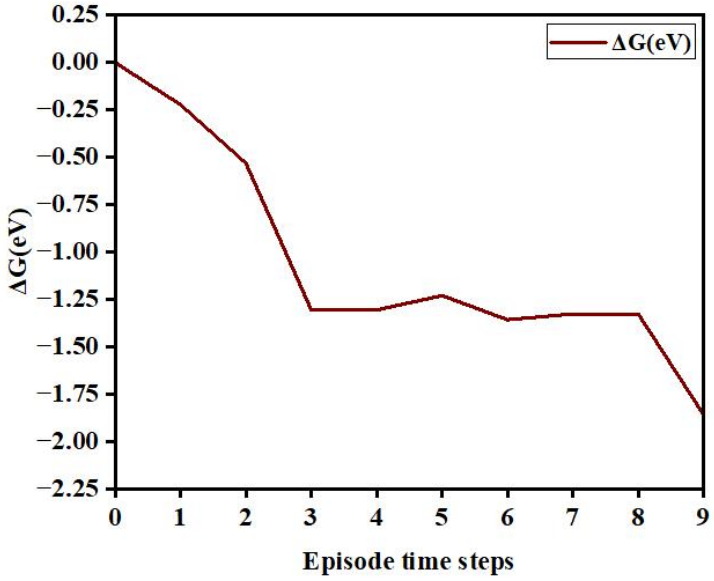

Figure 7. Energy profile after stable policy.

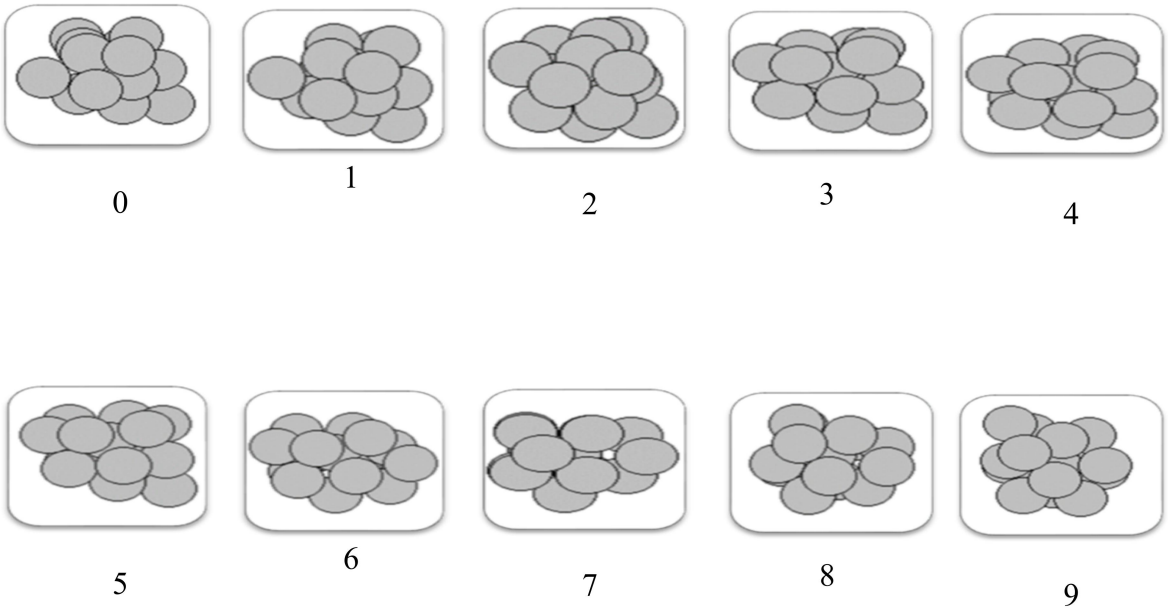

Figure 8. Configurations at different time steps after stable policy.

3.2 Energy profile analysis before development of stable policy

Figure 5 represents the energy profile obtained from a representative episode in the initial phases of training for Ag_{15} nanoclusters, prior to the agent achieving a stable policy and these configurations are represented in figure 6. During the initial training phase, the configurations are primarily disassociated or overlapped. Consequently, the agent does not execute relaxation on these configurations, leading to immediate negative feedback. Furthermore, the displacement of atoms from

these elevated energy states is resulting in increasingly unfavorable configurations, thereby exacerbating the negative rewards. The energy profile depicted in figure 5 indicates that the episode concludes after 200 time steps, with the agent failing to identify five distinct lower-energy minima. However, after approximately 6500 episodes, the agent began to learn an effective policy, and the rewards eventually stabilized.

3.3 Energy profile analysis after development of stable policy

Figure 7 represents the configurations and energies of nanoclusters subsequent to the agent acquiring a stable policy. In this representative episode, the agent performed the task in 10 time steps, successfully identified five different minimum configurations and four similar minimum configurations. These configurations are represented in figure 8. It reflects that this agent implemented an effective policy without generating overlapped or dissociated configurations. Although the goal of each episode is to find five different local minima, in the process, the agent also explores other higher-energy configurations that allow it to cover more space by moving atoms from those high-energy states. After training, we collected all the minimal configurations found and performed a deep analysis to identify the distinct configurations.

3.4 Global minimum and lowest energy configuration analysis

In start, policy is not stabled and agent is unable in exploring minimum configurations even after competing 200 steps per episode and agent encountered the negative reward as it produced overlapped cluster configurations. This period is of exploration and agent is in learning policy phase, after approximately 6500 episodes, policy is stabilized and agent started to explore minimum energy configurations successfully even after 5 to 6 steps per episode. After the training DRL successfully identified the global minimum and three other lowest energy configurations. Fig. 7 is representing the global minimum and three lowest energy configurations of Ag_{15} nanocluster, identified by the DRL.

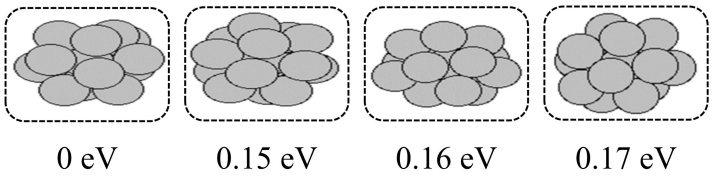

Figure 9. Global minimum and three lowest energy configurations identified by DR.

In start, policy is not stabled and agent is unable in exploring minimum configurations even after competing 200 steps per episode and agent encountered the negative reward as it produced overlapped cluster configurations. This period is of exploration and agent is in learning policy phase, after approximately 6500 episodes, policy is stabilized and agent started to explore minimum energy configurations successfully even after 5 to 6 steps per episode. After the training DRL successfully identified the global minimum and three other lowest energy configurations. Fig. 7 is representing the global minimum and three lowest energy configurations of Ag_{15} nanocluster, identified by the DRL.

3.5 Descriptive statistics from DRL experiments

Metrics of DRL experiments for 32000 episodes are represented in figures 10 (a)-(c). There are maximum steps of 200 for each episode to explore the minimum energy configurations. Total number of overlapped configurations in whole experiment is illustrated in Figure 10(a), in starting when policy is in learning phase and trying to explore the minimum energy configurations, there are more number of overlapped configurations. After approximately 6500 episodes, agent has started to minimize the overlapped configurations and agent has successfully started to explore the minimum energy configurations in perspective episodes. Figure 10 (b) is describing the total number of nonbonded configurations. In early phase, when policy is learning there are more number of nonbonded configurations. When policy is stabilized, number of nonbonded configurations is minimized. Figure 10(c) is illustrating the total number similar minimum configurations. In early episodes, less number of similar minimum energy configurations are explored by the agent. After 6500 episodes, when policy is stabilizing, agent has started

in exploring the more minimum energy configurations in perspective episodes. A high count of overlapped configurations in figure 11 indicates initial difficulties in avoiding geometrically improbable configurations. A substantial reduction in overlapped configurations shows that the model's policy learned to avoid these configurations, further optimizing the search for lower energy states. A high count of overlapped configurations in figure 11 indicates initial difficulties in avoiding geometrically improbable configurations. A substantial reduction in overlapped configurations shows that the model's policy learned to avoid these configurations, further optimizing the search for lower energy states.

In pre policy learning phase of figure 13, fewer relaxations to local minima led to limited encounters with similar configurations, suggesting an initial lack of diversity in the configurations reached.

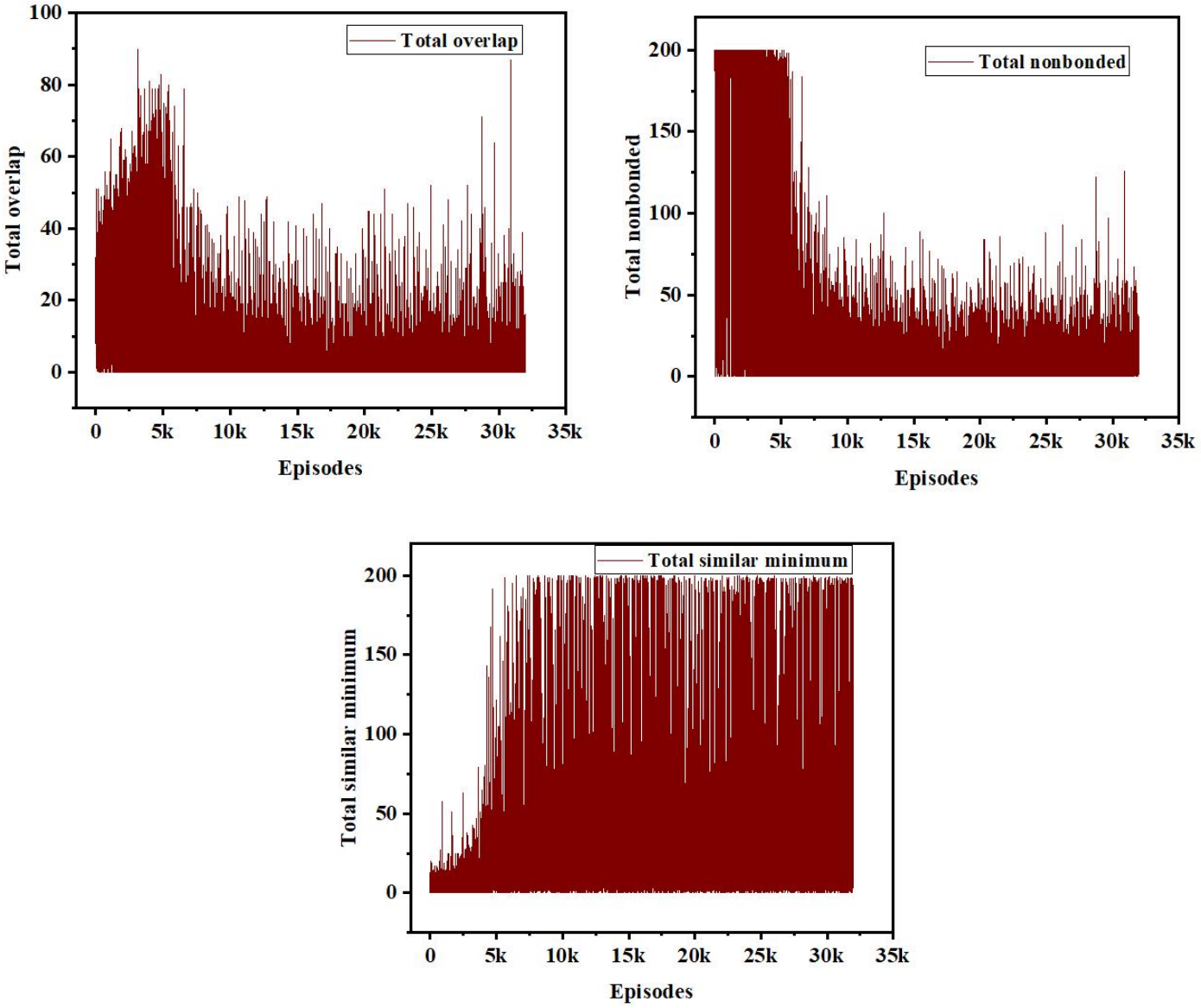

Figure 10. Total number of (a) overlapped, (b) nonbonded and (c) similar minimum configurations in DRL experiments

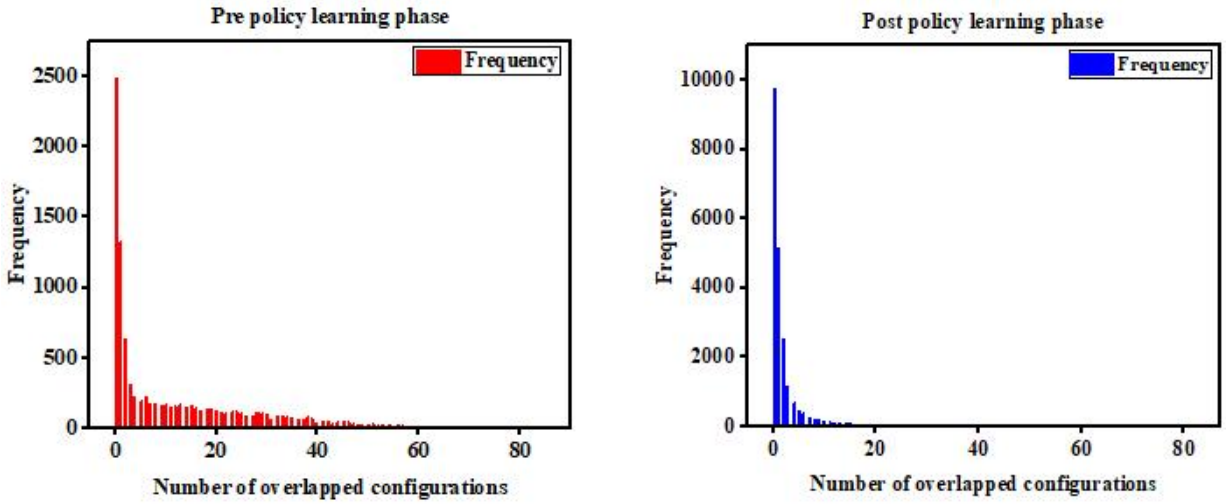

Figure 11. Overlapped configurations in pre and post policy learning phase.

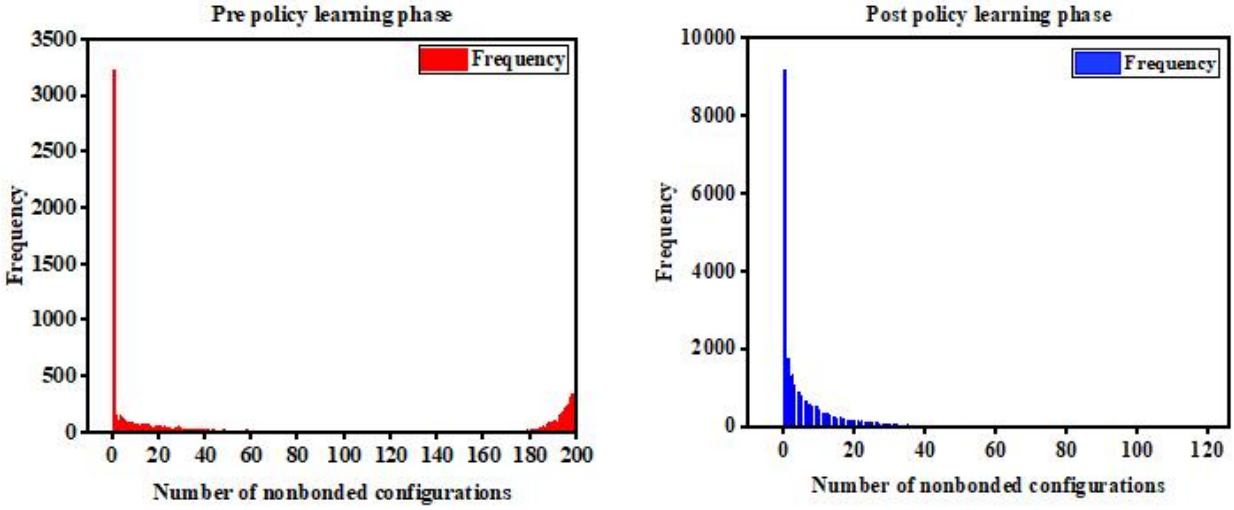

Figure 12. Number of nonbonded configurations in pre and post policy learning phase.

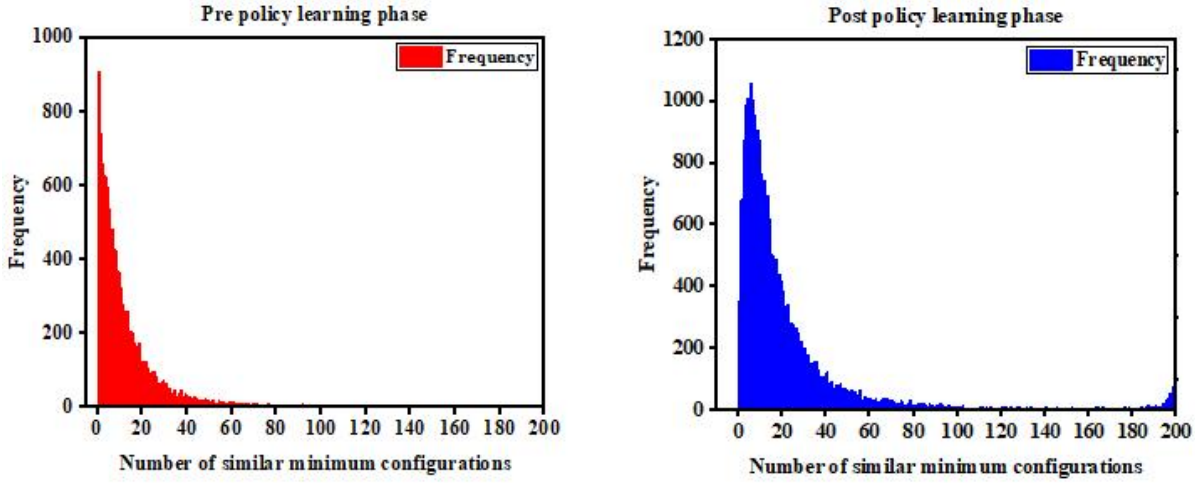

Figure 13. Similar minimum configurations in pre and post policy learning phase.

In the post learning phase of figure 13, with the model concluding episodes more quickly and finding lower energy configurations more efficiently, the instances of encountering similar configurations diminished, indicating a more focused and effective search strategy.

The Deep Reinforcement Learning (DRL) approach has demonstrated significant efficacy in optimizing the search method for determining the most stable configurations of nanoclusters. The model initially operated through a process of trial and error, which proved to be inefficient. Subsequently, upon acquiring a solid policy, there was a notable shift towards a more informed and strategic exploratory methodology. This transition represents a significant enhancement of the model's ability to navigate the complex energy landscapes of nanoclusters. The framework exhibited a great enhancement in avoiding structures that are less stable, nonbonded, dissociated, or overlapping and energetically unfavorable. The ability to avoid such configurations is important for the effective search of global minimum energy states. Additionally, subsequent to learning the policy, the DRL framework indeed guides the agent to reach the lower energy states rapidly. The efficiency shows that the model really learned its policy with success in arriving at the lower energy configuration quickly. These pieces of data all point and affirm together that the proposed DRL framework is indeed effective in improving the search approach toward the attainment of the optimum configurations of nanoclusters, succeeding in traversing complex potential energy landscapes in view of realizing the most stable arrangement of nanoclusters.

5. Conclusion

The work illustrates that DRL is a very effective and efficient optimization technique for the geometries of Ag_{15} silver nanoclusters. We have identified the stable configurations with much speed and preciseness by applying Deep Reinforcement Learning compared to conventional optimization techniques such as genetic algorithms and basin hopping, which are suffering frequently from ineffective convergence rates and a potential for prematurely converging on suboptimal solutions. Because DRL can change its approach dynamically based on feedback from the environment, it naturally performs much better in the investigation of complex PES for nanoclusters. Due to the nature of this method, it manages to balance exploration and exploitation without getting stuck into a local minimum; thus, it accelerates the process toward the GM. This implies that the new technique, DRL, will provide more rapidity and accuracy in determining stable atomic configurations compared to the conventional methods applied in the case of silver nanoclusters.

This innovative technique develops not only the knowledge of nanocluster stability, bringing new visions of atomic configuration and energy landscapes, but also provides a prospect for future applications in the design of nanomaterials and in materials science. The application of DRL to Ag_{15} clusters will better optimize their structural stability, something highly relevant for new nanomaterials developments with specified properties. Furthermore, the flexibility in the DRL method also allowed the attainment of new advanced materials in such cases as catalysis, nanomedicine, and other electronic devices. Thus, the obtained effectiveness of DRL on optimum nanocluster structures demonstrates that this is another impressive step forward in applying machine learning to materials exploration and design.

Authors contribution

Malik Ahmed Mubeen: investigation, visualization, writing, validation, formal analysis and methodology. Fuyi Chen: conceptualization, writing, review and supervision. Khalid Mehmood Ur Rehman: formal analysis and investigation.

Conflicts of Interest

There are no conflicts of interest reported by the writers.

Acknowledgment

I, Malik Ahmed Mubeen, the first author of this manuscript, would like to express my sincere gratitude to all those who contributed to the successful completion of this research. Special thanks to my doctoral supervisor, Professor Fuyi Chen,

for their valuable guidance and support throughout the study. I also appreciate the assistance provided by Dr. Khalid Mehmood Ur Rehman, for their expertise. Lastly, I am grateful to my family and friends for their unwavering encouragement.

Data Availability statement

The data presented in this study are available on request from the corresponding author.

Funding: Not applicable(N/A).

REFERENCES

1. Olveira, S.; Forster, S. P.; Seeger, S. Nanocatalysis: Academic discipline and industrial realities. J. Nanotechnol. 2014, 2014, 324089.
2. Somwanshi, S. B.; Somvanshi, S. B.; Kharat, P. B. Nanocatalyst: A Brief Review on Synthesis to Applications. J. Phys.: Conf. Ser. 2020, 1644, 012046.
3. Dai, Y.; Wang, Y.; Liu, B.; Yang, Y. Metallic nanocatalysis: An accelerating seamless integration with nanotechnology. Small 2015, 11, 268−289.
4. Piccolo, L. Restructuring effects of the chemical environment in metal nanocatalysis and single-atom catalysis. Catal. Today 2021, 373, 80−97.
5. Zhai, H.; Alexandrova, A. N. Fluxionality of Catalytic Clusters: When It Matters and How to Address It. ACS Catal. 2017, 7, 1905−1911.
6. Astruc, D. Introduction: Nanoparticles in Catalysis. Chem. Rev. 2020, 120, 461−463.
7. Ferrando, R.; Jellinek, J.; Johnston, R. L. Nanoalloys: From theory to applications of alloy clusters and nanoparticles. Chem. Rev. 2008, 108, 845−910.
8. Shan, S.; Luo, J.; Wu, J.; Kang, N.; Zhao, W.; Cronk, H.; Zhao, Y.; Joseph, P.; Petkov, V.; Zhong, C.-J. Nanoalloy catalysts for electrochemical energy conversion and storage reactions. RSC Adv. 2014, 4, 42654−42669.
9. Jellinek, J. Nanoalloys: Tuning properties and characteristics through size and composition. Faraday Discuss 2008, 138, 11−35.
10. Johnston, R. L. Evolving better nanoparticles: Genetic algorithms for optimizing cluster geometries. J. Chem. Soc., Dalton Trans. 2003, 3, 4193−4207.
11. Wales, D. J.; Doye, J. P. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem. A 1997, 101, 5111−5116.
12. Hohl, D.; Jones, R. O.; Car, R.; Parrinello, M. Structure of sulfur clusters using simulated annealing: S2 to S13. J. Chem. Phys. 1988, 89, 6823−6835.
13. Zhang, J.; Dolg, M. ABCluster: The artificial bee colony algorithm for cluster global optimization. Phys. Chem. Chem. Phys. 2015, 17, 24173−24181.
14. Lv, J.; Wang, Y.; Zhu, L.; Ma, Y. Particle-swarm structure prediction on clusters. J. Chem. Phys. 2012, 137 (8), 084104.
15. Zhai, H.; Alexandrova, A. N. Ensemble-Average Representation of Pt Clusters in Conditions of Catalysis Accessed through GPU Accelerated Deep Neural Network Fitting Global Optimization. J. Chem. Theory Comput. 2016, 12, 6213−6226.
16. Raju, R. K.; Sivakumar, S.; Wang, X.; Ulissi, Z. W. Cluster-MLP: An Active Learning Genetic Algorithm Framework for Accelerated Discovery of Global Minimum Configurations of Pure and Alloyed Nanoclusters. J. Chem. Inf. Model. 2023, 63, 6192−6197.
17. Wang, Y.; Liu, S.; Lile, P.; Norwood, S.; Hernandez, A.; Manna, S.; Mueller, T. Accelerated prediction of atomically precise cluster structures using on-the-fly machine learning. npj Comput. Mater. 2022, 8 (1), 64−66.
18. Han, S.; Barcaro, G.; Fortunelli, A. Unfolding the structural stability of nanoalloys via symmetry-constrained genetic algorithm and neural network potential. npj Comput. Mater 2022, 8, 121.
19. Bisbo, M. K.; Hammer, B. Global optimization of atomic structure enhanced by machine learning. Phys. Rev. B 2022, 105, 245404.

20. Zeng, Fanyu & Wang, Chen & Ge, Shuzhi. (2020). A Survey on Visual Navigation for Artificial Agents With Deep Reinforcement Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3011438.
21. Würz, Valentin & Weißenfels, Christian. (2024). Inverse Material Design using Deep Reinforcement Learning and Homogenization. 10.1016/j.cma.2024.117617.
22. Modee, Rohit & Verma, Ashwini & Joshi, Kavita & Priyakumar, U. (2023). MeGen - Generation of gallium metal clusters using Reinforcement Learning. Machine Learning: Science and Technology. 4. 10.1088/2632-2153/acdc03.
23. Elsborg, Jonas & Bhowmik, Arghya. (2023). Equivariant Graph-Representation-Based Actor-Critic Reinforcement Learning for Nanoparticle Design. Journal of chemical information and modeling. 63. 10.1021/acs.jcim.3c00394.
24. Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. Adv. Mater. 2019, 31 (46), 1902765.
25. Schu¨ tt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Mu¨ ller, K.-R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. J. Chem. Theory Comput. 2019, 15, 448−455.
26. Zhou, Z.; Li, X.; Zare, R. N. Optimizing Chemical Reactions with Deep Reinforcement Learning. ACS Cent. Sci. 2017, 3, 1337−1344.
27. Gupta, Ambesh & Dahale, Chinmay & Maiti, Soumyadipta & Goverapet Srinivasan, Sriram & Rai, Beena. (2024). Development of an embedded-atom method potential of Ni-Mo alloys for electrocatalysis / surface compositional studies.
28. Binti Agos Jawaddi, Siti Nuraishah & Ismail, Azlan. (2024). Integrating OpenAI Gym and CloudSim Plus: A simulation environment for DRL Agent training in energy-driven cloud scaling. Simulation Modelling Practice and Theory. 130. 102858. 10.1016/j.simpat.2023.102858.
29. Schulman, J.; Levine, S.; Abbeel, P.; Jordan, M.; Moritz, P. Trust Region Policy Optimization. In Proceedings of the 32nd International Conference on Machine Learning Lille, France; JMLR, 2015; pp.1889-1897.
30. Sutton, R. S.; McAllester, D.; Singh, S.; Mansour, Y. Policy Gradient Methods for Reinforcement Learning with Function Approximation. Advances in Neural Information Processing Systems; MIT Press, 1999, 1057−1063.
31. Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; Kavukcuoglu, K. Asynchronous Methods for Deep Reinforcement Learning. In Proceedings of The 33rd International Conference on Machine Learning New York, New York, USA; PMLR, 2016; pp. 19281937.
32. Haarnoja, T.; Zhou, A.; Hartikainen, K.; Tucker, G.; Ha, S.; Tan, J.; Kumar, V.; Zhu, H.; Gupta, A.; Abbeel, P., et al. Soft Actor-Critic Algorithms and Applications; arXiv, 2018. preprint arXiv:1812.05905.
33. Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. Phys. Rev. Lett. 2007, 98, 146401.
34. Davis, J. B.; Shayeghi, A.; Horswell, S. L.; Johnston, R. L. The Birmingham parallel genetic algorithm and its application to the direct DFT global optimization of IrN (N = 10−20) clusters. Nanoscale 2015, 7, 14032−14038.
35. Jäger, M.; Schäfer, R.; Johnston, R. L. GIGA: A versatile genetic algorithm for free and supported clusters and nanoparticles in the presence of ligands. Nanoscale 2019, 11, 9042−9052.
36. Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; Zaremba, W. Openai gym; arXiv, 2016, 1−4

37. Kuhnle, A.; Schaarschmidt, M.; Fricke, K. Tensorforce: a TensorFlow Library For Applied Reinforcement Learning, 2017. https://github.com/tensorforce/tensorforce (accessed January 01, 2023).
38. Tadmor, E. B.; Elliott, R. S.; Sethna, J. P.; Miller, R. E.; Becker, C. A. The potential of atomistic simulations and the knowledgebase of interatomic models. JOM 2011, 63, 17−17.
39. Jacobsen, K. W.; Norskov, J. K.; Puska, M. J. Interatomic interactions in the effective-medium theory. Phys. Rev. B 1987, 35, 7423−7442.
40. Jacobsen, K. W.; Stoltze, P.; Nørskov, J. K. A semi-empirical effective medium theory for metals and alloys. Surf. Sci. 1996, 366 (2), 394−402.

How to cite this article:
Mubeen M.A., Chen F., Rehman K.M. (2024). Optimization of Silver Nanocluster Geometries: A Deep Reinforcement Learning Approach to Identifying the Most Stable Configurations in Ag_{15} Cluster. *Journal of Chemistry and Environment*. 4(1). p. 1-17.