Toward full *ab initio* modeling of soot formation in a nanoreactor

Qingzhao Chu[a,b], Chenguang Wang[c,d,e], Dongping Chen[a,b,*]

[a] State Key Laboratory of Explosion Science and Technology, Beijing 100081, China
[b] Explosion Protection and Emergency Disposal Technology Engineering Research Center of the Ministry of Education, Beijing, 100081, China
[c] Guangzhou Institute of Energy Conversion, Chinese Academy of Sciences, Guangzhou 510640, China
[d] CAS, Key Laboratory of Renewable Energy, Guangzhou 510640, China
[e] Guangdong Provincial Key Laboratory of New and Renewable Energy Research and Development, Guangzhou 510640, China

---

Abstract

A neural network (NN)-based model is proposed to construct the potential energy surface of soot formation. Our NN-based model is proven to possess good scalability of O(N) and retain the *ab initio* accuracy, which allows the investigation of the entire evolution of soot particles with tens of nm from an atomic perspective. A series of NN-based molecular dynamics (NNMD) simulations are performed using a nanoreactor scheme to investigate critical processes in soot formation, acetylene polymerization, and inception of PAH radicals. This shows that NNMD can capture the dynamic process of acetylene polymerization into PAH precursors. The simulation of PAH radicals reveals that physical interaction enhances chemical nucleation, and such enhancement is observed for clusters of π- and σ-radicals, which is distinct from the dimer. We also observed that PAH radicals of ~ 400 Da can produce core-shell soot particles at a flame temperature, with a disordered core and outer shell of stacked PAHs, suggesting a potential physically stabilized soot inception mechanism.

Keywords: Soot cluster; PAH radicals; Neural network; Molecular dynamics

---

*Corresponding author.

1. Introduction

Soot emitted from human activities leads to respiratory diseases and global warming [1]. A significant number of studies are ongoing to understand the mechanisms responsible for soot formation, which involves precursor chemistry, particle nucleation, and mass/size growth [2,3]. The critical stage of soot formation is the inception process, in which gas-phase aromatics form condensed clusters, resulting in carbonaceous nanoparticles.

Polycyclic aromatic hydrocarbons (PAHs) are widely accepted as precursors for soot. Two major pathways of soot nucleation initiated from PAHs [2,4] are known, namely, physical nucleation of PAHs into stacked clusters [5,6] and chemical nucleation of PAHs into an aromatic network via cross-linking reactions [7–9]. Physical nucleation of PAHs is only possible for large PAHs (>667 Da), but these species fail to be observed in experimental studies due to their low concentrations [10]. Chemical nucleation is a process where reactive sites on PAHs are chemically combined to form a stable structure. These reactive sites can be connected by acetylene, which refers to the well-known hydrogen abstraction acetylene addition (HACA) growth mechanism [9]. Another chemical pathway is the formation of a π-bonding network where π-radicals directly react with PAH σ-radicals without the requirement of hydrogen abstraction [11]. These π-radicals can be stabilized due to delocalization, also known as long-lived resonantly stabilized radicals (RSRs). The existence of PAH π-radicals in sooting flames has been reported by high-resolution atomic force microscopy (HR-AFM) and photoionization mass spectrometry experiments [11,12]. The reactivity of different sites on a range of PAH radicals was examined by Martin et al. [13]. They found that σ-radicals are the most reactive, forming bonds with other radicals. Partially saturated rim-based pentagonal rings were found to form localized π-radicals with high reactivities. The forward rate constants of these radicals are calculated by Menon et al. [14]. The maximum rate was 10^{12} $cm^3mol^{-1}s^{-1}$ for aryl-radical recombination, which is still insufficient to trigger the nucleation flux for nanoparticle formation [15].

Physical nucleation is rapid but too weak to be stable at flame temperatures, while most chemical nucleation rates are slow. Frenklach and Mebel [15] recently suggested an enhancement of polyaromatic polymerization induced by internal rotors. These rotors were argued to provide multiple opportunities for an aryl radical to attach to the reactive edge of a PAH, such as a rim-based pentagonal ring. A dimerization reaction of π-diradicals was reported to confirm that the physical interaction between PAHs could accelerate chemical nucleation [16]. Such a mechanism, known as physically stabilized soot inception, shows a feasible pathway to explain the fast growth of soot. However, most previous work focuses on the dimerization process. The dynamic evolution between multiple reactive species is still not well understood.

The complexity of the chemical and physical reactions during soot nucleation makes it both numerically and experimentally challenging. *Ab initio* molecular dynamics (AIMD) simulations have been widely applied to reveal the atomic insights of complex reactive systems based on fundamental equations of quantum and classical mechanics. Although AIMD has been successfully used for polyyne inception [17] and PAH reactive dimerization [16], AIMD calculations of large PAH molecules/radicals are still challenging due to the high computational cost despite impressive progress in computing hardware and software in recent decades. Over the past decade, many empirical potentials (or force fields) have been developed to mimic electronic structure calculations' potential energy surface (PES). Such empirical potentials, including ReaxFF [18,19] and REBO [20], trade accuracy for a lower computational expense, making it possible to extend simulation scales to orders of magnitude beyond AIMD methods. ReaxFF is a bond order-based force field that describes reactive systems without a priori knowledge of the predefined reactive sites. It has been a powerful tool to study kinetic mechanisms for systems of large molecules and complex reactions. Recently, machine learning-based tools, especially neural networks (NNs), have been applied to construct PES models in an entirely data-driven manner, where the PES is abstracted from a well-selected training dataset using suitable functional expressions automatically [21]. NN models constitute a very flexible class of mathematical functions, which enable the development of PES models with the efficiency of the empirical potentials and the accuracy of the DFT method. Different NN-based PES models have been proposed for materials and biomolecules [22,23]. Zeng et al. [24] implemented an NN-based model to reveal the mechanisms of methane combustion.

In addition to developing PES models, many other methods have been developed to accelerate AIMD calculations. The "nanoreactor" is a recently introduced AIMD-accelerating simulation method [25]. The rate of reactions can be highly accelerated with a virtual piston periodically pushing molecules toward the center of the simulation. Several new pathways for glycine synthesis from primitive compounds proposed to exist on early Earth have been successfully revealed [26]. Nanoreactors would provide new insights into complicated chemical processes and discover elementary steps suitable for complex systems such as soot.

Here, we develop an NN-based model to explore the soot formation mechanism with *ab initio* accuracy. The NN-based model is first trained and validated against the DFT database with wide-ranging precursors involved in the soot formation process. Molecular dynamics simulations are performed to reveal the soot inception mechanism of PAH radicals using a nanoreactor scheme. The products are analyzed to understand the effects of different radicals on soot structures.

2. Computational methods

2.1 Development of the neural network potential

The NN-based potential was constructed following the Deep Potential (DP) scheme recently developed by Zhang et al. [22], which can accurately reproduce the interatomic forces and energies predicted by *ab initio* calculations in condensed matter systems. Figure 1 shows the illustration of DP scheme. The PES is represented by a deep neural network model that interprets the atomic coordination (R) into interatomic forces (F) and energies (E). The deep neural network contains a filter (embedding) network with three layers (25, 50, 100 nodes/layer) and a fitting net with three layers (240 nodes/layer). The loss function (L) is defined as,

$$L = \frac{p_e}{N}\Delta E^2 + \frac{p_f}{3N}\sum_i |\Delta F_i|^2 \quad (1)$$

where p_e and p_f are the weight for the energy and force terms, respectively. N represents the number of atoms in the structure. Similar to a classical neural network, the DP scheme trains the model by computing the gradient of the loss function using the back-propagation algorithm [27]. The NN model is trained for 1.0×10^6 iterations with an exponentially decaying learning rate from 1.0×10^{-3} to 5.0×10^{-8}.

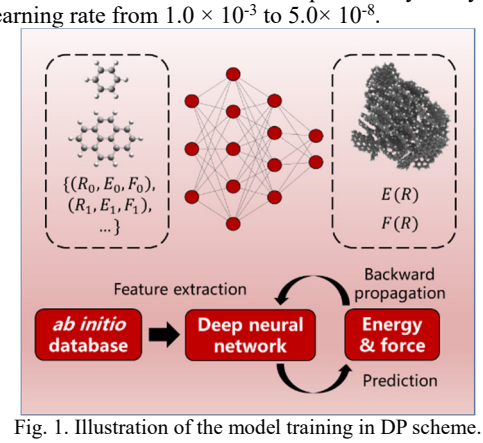

Fig. 1. Illustration of the model training in DP scheme.

The performance of NN potential highly depends on the quality of the training dataset [28,29]. Although NNs are good at interpolating between training points, they cannot predict the energies and forces of configurations distant from those of the training set. Therefore, it is critical to ensure that the dataset covers the PES of interest. The formation of molecular soot precursors starts from the fuel pyrolysis process, where acetylene and propargyl are the key intermediates. These intermediates would further react to form the first aromatic ring – benzene. Subsequent reactions lead to the formation of larger PAHs via the HACA mechanism. This paper constructs an *ab initio* database to cover the entire evolution from acetylene to large PAH molecules. In Fig. 2, there are 47 subsystems in the database, including alkanes, alkenes, alkynes, gaseous radicals, PAH molecules, PAH radicals, and carbon materials. Each subsystem consists of MD trajectories of a specific molecule. Details about the molecular configuration of each system are listed in Table S1. The trajectories are generated from ReaxFF MD simulation under an NVT ensemble at temperatures of 300, 1000, 2000, 3000, 4000, 5000, and 6000 K.

To obtain accurate energies and forces, 1000 configurations are randomly selected from a 40-ps ReaxFF MD simulation, and high-level DFT calculations are further conducted. DFT calculations were performed using the CP2K package [30]. Core electrons are treated using Goedecker–Teter–Hutter (GTH) pseudopotentials and the Perdew Burke Ernzerhof generalized gradient approximation method [31,32]. The Grimme DFT-D3 method [33] is used to account for dispersion interactions. A double-zeta Gaussian basis set plus polarization (DZVP-MOLOPT) [34] is considered. In addition to MD simulations, configurations are also obtained using active learning sampling as implemented in the DP-GEN package [35] with a temperature range from 300 to 4000 K. With the above method, an ab initio database with energy and force information is constructed to train the NN potential.

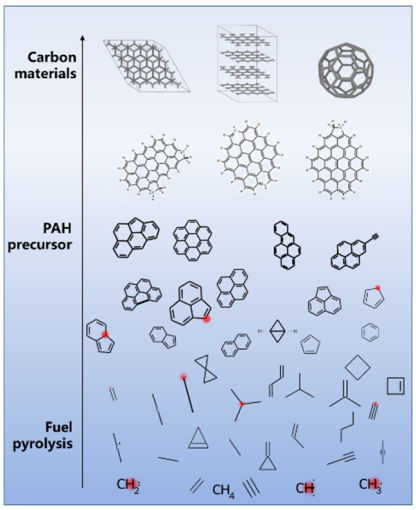

Fig. 2. Illustration of the 47 species included in the *ab initio* database. The red shadow represents the position of reactive sites in radicals.

2.2 Nanoreactor based MD simulations

The soot formation process is investigated with our NN potential. Two kinds of systems are studied: acetylene for their polymerization and the formation of aromatic rings and PAH π-radicals and σ-radicals for physically stabilized soot inception. Figure 3 shows the structure of the PAH radicals studied in this work. PAH-a, -c, and -e are σ-radicals involved in soot inception, while PAH-b, -d, -f, and -g are π-radicals reported in simulations and experiments [12,16]. Before MD simulations, monomers of acetylene and PAH radicals were optimized using the DFT method with the B3LYP/6-311G(d,p) level of theory. The optimized monomers were duplicated 1000 times for acetylene and 50 for PAH radicals and then randomly distributed in a cubic box. The equations of motion are integrated by the velocity Verlet method using periodic boundary conditions. A Nose–Hoover thermostat is applied with an equilibrium temperature of 1500 K and a dump parameter of 20 fs. A one ns MD simulation with a time step of 0.1 fs is performed for both acetylene and PAH radical systems.

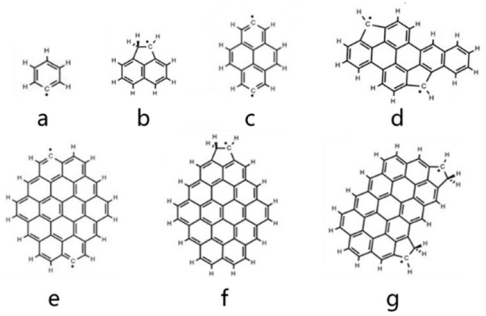

Fig. 3. PAH radicals studied in this work.

The nanoreactor method is used to accelerate the simulation, where a virtual piston in a sphere is imposed on the system during MD simulations. The virtual piston is defined by a time-dependent boundary potential [25]:

$$V(r,t) = f(t)U(r,r_1,k) + \big(1 - f(t)\big)U(r,r_2,k) \quad (2)$$

$$U(r,r_0,k) = \frac{mk}{2}(r - r_0)^2\theta(r - r_0) \quad (3)$$

$$f(t) = \theta(\left\lfloor \frac{t}{T} \right\rfloor - \frac{t}{T} + \frac{\tau}{T}) \quad (4)$$

where k=0.01 kcal mol^{-1} Å, τ = 1.0 ps, T = 2.0 ps. r_0 is the sphere radius, m is the atomic mass, $\lfloor\ \rfloor$ is the floor function and θ is the Heaviside step function. To minimize the temperature fluctuations in nanoreactors, the compression process is divided into three stages. The boundary potential oscillates between a large value of r_1, a median value of r_2, and a small value of r_3. The r values are determined by the system density. We use three values of 0.1, 0.4 and 0.7 g/cm^3 to calculate the corresponding r_1, r_2, and r_3 values. Detailed model settings are listed in Table S2.

3. Results and discussion

3.1 Validation of NN-model

The performance of the NN potential is tested against the *ab initio* database. Figure 4 shows the prediction of DFT energies and forces using the NN potential trained by the DP scheme. The results predicted from the ReaxFF forcefield [36] are also listed for comparison. Detailed mean absolute errors (MAEs) are listed in Table S2. We found that the MAE values increase with the system temperature in MD sampling. The DP model shows a lower error than the ReaxFF model (0.67 *v.s.* 1.18 kcal/mol/atom), suggesting the good fitting ability of the DP scheme. Surprisingly, we find that the DP model shows excellent performance for force prediction, while the ReaxFF model seriously underestimates the interatomic forces (38.62 kcal/mol/Å). This can be explained by the parameterization of ReaxFF, as only energy information is taken as the training targets. By considering the forces in loss function (Eq. 1), the DP model significantly improves the model prediction and, therefore, accurately captures the dynamic evolution of the soot formation process.

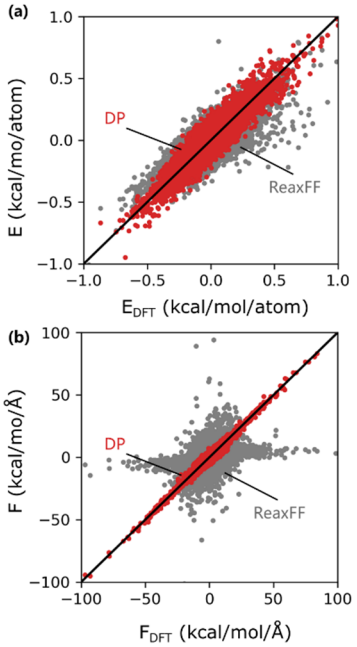

Fig. 4. The mean absolute errors of (a) energy and (b) force for DP and ReaxFF models on the *ab initio* database.

We also test the computational cost of DP and ReaxFF potentials on C_2H_2 systems with 20 to 160,000 atoms. The NNMD simulations of the DP model are performed with one NVIDIA V100 GPU, and the ReaxFF MD simulations are on a 64-processor server with two AMD EPYC 7452 CPUs. Figure 5 shows that our DP model follows an O(N) scaling rule, while ReaxFF exhibits an O(NlogN) scaling rule. The better scaling performance of the DP model enables the exploration of a system with tens of thousands or even millions of atoms at *ab initio* accuracy, providing a possibility to investigate the entire evolution of soot particles from an atomic perspective.

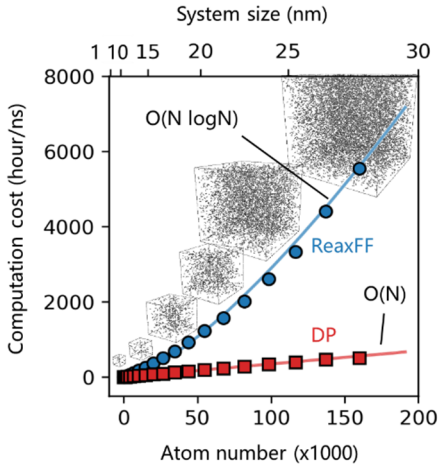

Fig. 5. The computational cost of DP (red square) and ReaxFF (blue circle) models on C_2H_2 systems with 20 to 160,000 atoms. The system lengths are from 1 to 28 nm. The insert shows snapshots of the corresponding systems.

3.2 Acetylene polymerization into PAH

First, we analyze the polymerization of acetylene, which corresponds to the formation of soot precursor – PAH molecules. The reaction pathway from acetylene to benzene is shown in Fig. 6a. The polymerization starts with the collision and addition reaction between two acetylene molecules at 1.5 ps to generate an HC≡CH-CH≡CH molecule, which further grows into a long chain molecule (e.g., HC≡CH-CH≡CH-CH≡CH) with the addition of another acetylene molecule at 3.6 ps. This species further forms a benzene via an isomerization reaction at 6.4 ps, and this pathway is consistent with a previous work [37]. Figure 6b shows the pathway from benzene to the 2nd aromatic ring. The hydrogen atom in benzene is abstracted by other radicals to generate a reactive site. Then, a CCH• radical attacks the reactive site at 11.1 ps to form a benzene with an acetylene chain, and this structure is expected to be stable at flame temperatures [38]. At 12.4 ps, the second addition reaction occurs near the first acetylene chain and forms the 2nd aromatic ring after long isomerization (~5.1 ps). In Fig. 6c, we also observe the formation of a large PAH with four aromatic rings at 101.7 ps, which follows similar acetylene addition and isomerization pathways.

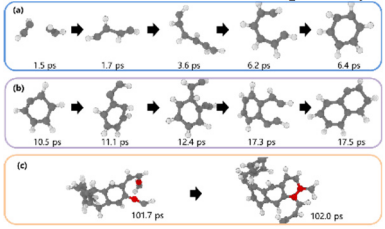

Fig. 6. Snapshots for the reaction pathways to produce the (a) 1st, (b) 2nd, and (c) large aromatic rings. Red atoms in (c) represent the reactive sites.

The dynamic evolution of the main species during the polymerization process is shown in Fig. 7. Since there are a large number of isomers for C_xH_y species, the products are grouped by their carbon atom numbers. It shows that the C_2 group is rapidly consumed within the first 10 ps. The insert in Fig. 7a illustrates the main species of each group, which is extracted from MD trajectories using the ReacNetGenerator package [39]. The main composition includes acetylene, but there are also hydrogen transfer and elimination reactions to produce ·CCH_2 and ·CCH radicals. C_4 and C_6 groups are the main products of C_2 polymerization in the first 20 ps. In Fig. 7b, the $C_{7\text{-}10}$ group represents the growth of the 2nd ring, which reaches a maximum at ~ 30 ps and starts to decrease. Groups with larger carbon atom numbers follow the same trends but at a slow rate. For example, the time needed to reach the maximum number of molecules was ~40 ps for the $C_{11\text{-}15}$ group and ~200 ps for the $C_{15\text{-}20}$ group. In the above simulation, the growth of large PAHs from small acetylene is observed, proving the feasibility of studying soot formation using our NN model.

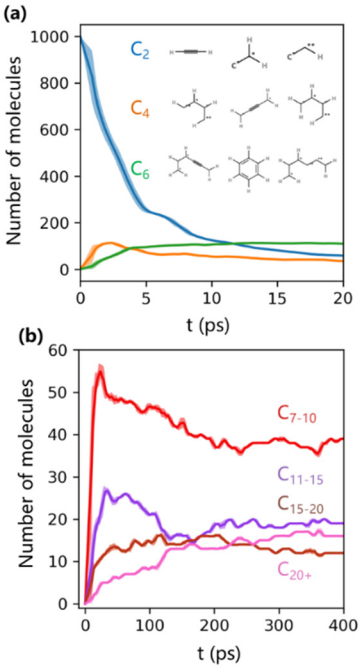

Fig. 7. Evolution of species during the simulation for (a) C_2, C_4, C_6, (b) $C_{7\text{-}10}$, $C_{11\text{-}15}$, $C_{15\text{-}20}$, and C_{20+} groups. The inserts in (a) show the main species for the C_2, C_4, and C_6 groups.

3.2 Physically stabilized soot inception of PAH radicals

The inception process of PAH radicals in Fig. 5 is investigated using a nanoreactor scheme, where a sphere boundary potential with periodic changes in radius is imposed on the PAH radicals to accelerate the collision and reaction process. Seven PAH species are considered (Fig. 3). For each type of PAH, the atom number in the maximum cluster is calculated and normalized by PAH size for comparison. Figure 8a shows the normalized max cluster size of PAH-b and -f in the first 40 ps. In Fig. 8b, the reactor radius (r) shows a variation with the time-dependent rectangular waveform. The period of each cycle is 4 ps. It starts with a relaxing stage of r=24.8 Å within the first 2 ps, followed by two compression stages with r=15.6 Å and r=13.0 Å for 1 ps each. Later, a new relaxing and compression cycle continues until simulations end. A reversible inception process is observed in the nanoreactor. During the compression stage, the reactor is compressed into a small volume (1/7 of the volume in the relaxing stage). PAH radicals collide with each other to form PAH clusters via physical interactions or chemical bonds. The reactor expands out with cluster decomposition in the next relaxation stage. The evolutions of cluster size are different for PAH-b and PAH-f. Cluster sizes in PAH-b fluctuate with the variation of reactor radius and remain a small value after 10 cycles (*e.g.*, 40 ps). In contrast, the cluster for PAH-f gradually grows to a size of 44, although a decomposition stage is also observed in each relaxing stage. The clustering behaviors of PAH-b are somehow unexpected, as PAH-b is reported [13] to form a cross-linked dimer. However, within the simulation time in the nanoreactor, it cannot further grow into large clusters.

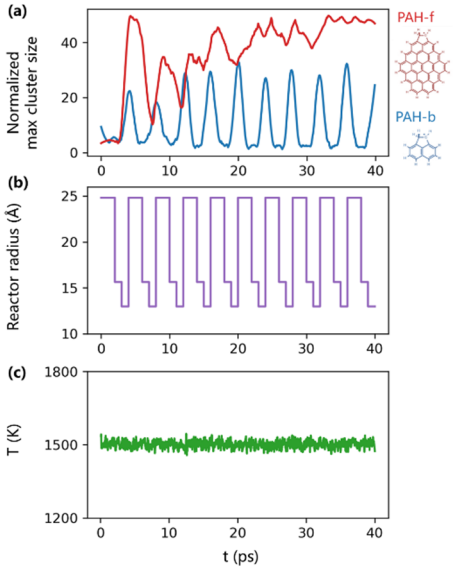

Fig. 8. (a) Evolution of normalized maximum cluster size for PAH-b and -f and in the first 40 ps. The evolution of the (b) radius of the spherical boundary and (c) temperature.

The evolution of cluster numbers (N_{cluster}) and normalized maximum cluster size (S_{cluster}) is listed in Fig. 9. To inhibit fluctuations induced by the compression and relaxing process, only the maximum of N_{cluster} and the minimum of S_{cluster} in each relaxing stage are plotted, which represent the actual clustering state. For small radicals (<200 Da), such as PAH-a and PAH-b, the number of clusters decreases to ~40 at 100 ps, while the maximum cluster size remains lower than 5. As the physical interaction is proportional to molecule size, it is expected that the inception of these radicals is only induced by a chemical bond. The fraction of reactive collisions (F_{reac}) for these PAH radicals has been systematically examined by Martin et al. [16], which indicates the possibility of forming covalent bonds between two PAH radical monomers in a collision. PAH-a has a high fraction of reactive collisions (*e.g.*, 0.023) among these PAH radicals. However, chemical bonding does not generate a large number of clusters within the simulation time. For medium-sized radicals (200-400 Da), such as PAH-c and -d, cluster sizes continue to grow at a flame temperature (1500 K). These radicals are much smaller than the minimum requirement for physical inception (*e.g.*, 667 Da [10]), indicating the feasibility of physically stabilized soot inception. For large radicals

(>400 Da), such as PAH-e, -f, and -g, the growth rates of cluster size are faster than those for medium-sized PAH radicals due to the strong physical interaction. The fraction of reactive collisions has a relation of PAH-g (F_{reac}=0.024) > PAH-f (F_{reac}=0.013) > PAH-e (F_{reac}=0.008) but shows the same clustering trends (see Fig. 9b), suggesting that physical interactions play a more important role than chemical bonds for large radicals. Chemical nucleation is also observed in these radicals, but their rates are much lower than those of physical nucleation, which agrees with the calculations by Menon et al. [14].

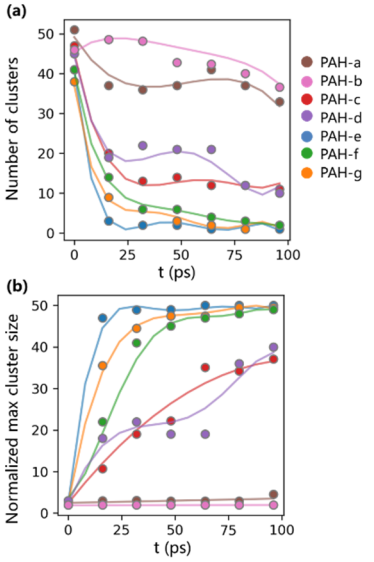

Fig. 9. Evolution of (a) cluster numbers and (b) normalized maximum cluster size for all PAH radicals.

In addition, we also note that the chemical nucleation in the large cluster is slightly different from that in the dimer. Since the radicals in the dimer are always in a parallel stacked structure, Martin et al [16]. reported that σ-radical dimers have weaker physical enhancement than π-radicals because the C-C bond prefers to be collinear with the aromatic planes. However, there are multiple stacks in a large cluster, which enable chemical nucleation between σ-radicals in the same plane. For example, chemical cross-links between different stacks of σ-radicals are observed in PAH-e clusters. In Fig. 10a, we summarize the physical enhancement of the nucleation mechanism of dimers and large clusters. In a dimer, PAH radicals are stacked in parallel due to physical interactions, which can accelerate the cross-linking between π-radicals but does not affect σ-radicals. In a large cluster, there are multiple stacks, and σ-radicals can form cross-links with adjacent stacks.

Figure 11 includes the final products collected from different precursors in the nanoreactors, which are mapped according to the precursor properties. These products are postprocessed using the TEMSIM package [40] to obtain simulated TEM images, which can be directly compared to the structures seen in HRTEM experiments [41,42]. For PAH-a and PAH-b, the PAHs are arranged in a gaseous form. Most radicals form cross-linked dimers or trimmers, and no large cluster is observed. For PAH-c and PAH-d, a prominent cluster is produced in the box center. The structure of PAH-c is disordered and loosely packed, similar to the covalently bound incipient particle predicted by Johansson et al. [11]. For PAH-d, a core-shell-like structure is observed, where the core region shows disordered and cross-linked structures, and the shell region (upper right part) is ordered and close-packed stacks. For large radicals, such as PAH-e, -f, and -g, the final products are spherical stacked structures with cross-links between different layers.

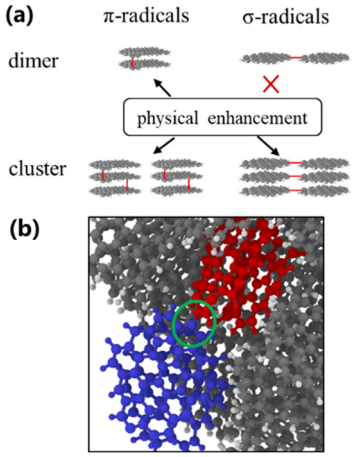

Fig. 10. (a) Illustration of the physical enhancement for dimers and clusters of σ- and π-radicals. (b) A snapshot of the chemical nucleation of PAH-e (σ-radical) between different stacks. The green circle represents the location of a new formed cross-link bond.

These results suggest that the morphology of the final products is highly related to the size of the precursors. At flame temperatures, no PAH cluster can form when the molecular mass of precursors is lower than 200 Da. This critical mass is lower than the value of 666 Da for physical nucleation of PAH molecules simulated by Mao et al. [36]. This can be attributed to the higher reactivity of PAH radicals, which can form cross-links and lead to chemical nucleation. For PAH molecules without reactive sites, a higher temperature (2500 K) is required for chemical nucleation, which is too high for typical sooting flames. As the size of the precursors increases, the product particle changes from a disordered cluster into an ordered and close-packed structure. It is interesting to note that the product of PAH-d has a core-shell structure, which includes a disordered core and outer shell of stacked PAHs [41,42]. Such structures are also reported in HRTEM experiments [41,42]. Another piece of evidence is from a recent mass spectrum experiment [43], which shows that PAHs of 239-838 Da are the main component of soot particles, suggesting that physically stabilized soot inception of PAH radicals larger than 200 Da is a potential inception mechanism.

In summary, the inception mechanism of PAH radicals can be divided into three groups: chemical, “physical+chemical”, and physical, as shown in Fig. 10. When the radical molecular mass is lower than 200 Da, the physical interaction is too weak at flame temperature, and chemical nucleation is the dominant mechanism. With increasing radical mass, the physical interaction becomes stronger, and a stack cluster structure can be formed with the combined effect of physical and chemical nucleation. For large radicals (>400), the physical interaction becomes the dominant nucleation mechanism.

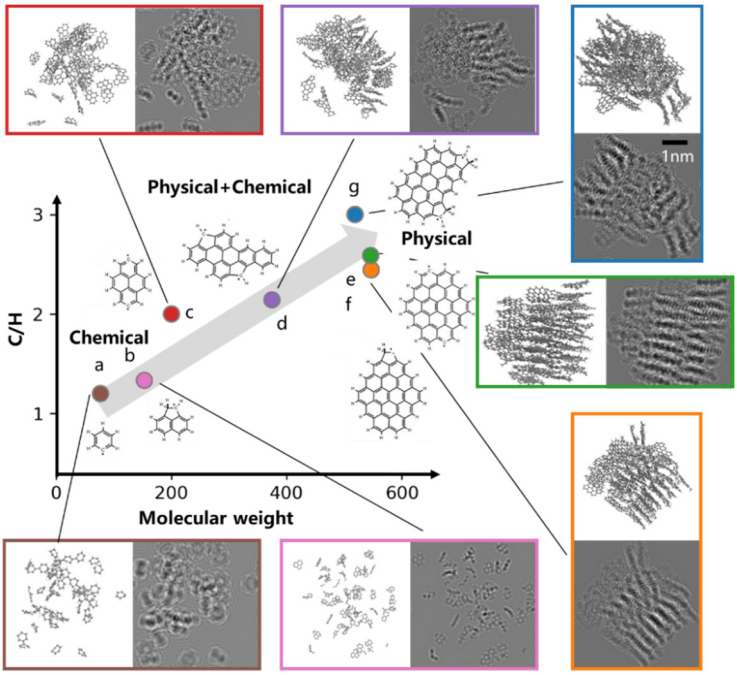

Fig. 11. The final products taken from a nanoreactor at 1 ns with different PAH precursors. The inserts show snapshots and simulated TEM images.

4. Conclusions

In this work, we develop a neural network (NN)-based model to explore the soot formation mechanism with *ab initio* accuracy. NN-based molecular dynamics simulations (NNMDs) of acetylene and PAH radicals of different sizes are performed using a nanoreactor scheme to investigate the mechanism of soot inception. The product structures of PAH radicals are analyzed to interpret the potential nucleation mechanism for different sized radicals.

Our NN model considers both energy and force information from high-level DFT calculations, which has been proven to have significantly higher accuracy than the ReaxFF model. In particular, the ReaxFF model is shown to underestimate the atomic force of DFT calculations, which affects the model accuracy when dealing with the reaction dynamics process. Our NN model also indicates that good scaling performance follows O(N), providing the possibility to investigate the entire evolution of soot particles from an atomic perspective.

The inception mechanism of PAH radicals is investigated with our NN model. We find that the physical interaction enhances chemical nucleation, and such enhancement is observed for clusters of π- and σ-radicals, which is distinct from the dimer. It is also found that PAH radicals can produce soot particles at flame temperatures (1500 K), and the particle morphology is strongly affected by PAH size. For PAH radicals of 374 Da, a core-shell soot particle with a disordered core and outer shell of stacked PAHs is observed, and similar structures are also observed in HRTEM experiments. These results suggest that physically stabilized soot inception of PAH radicals larger than 200 Da is a potential nucleation mechanism.

Acknowledgments

This paper is supported by the project of State Key Laboratory of Explosion Science and Technology (Beijing Institute of Technology). The project number is ZDKT21-01. DC also acknowledges support from the National Natural Science Foundation of China (No. 21961122007 and 51806016). QC acknowledges support from the National Natural Science Foundation of China (Grant No. 52106130).

Supplementary material

Table S1. The ab initio database configurations; Table S2. Nanoreactor configurations for PAH radicals; Table S3. Model prediction for the DP model and the ReaxFF model against DFT calculations.

References

[1] J.R. McConnell, R. Edwards, G.L. Kok, M.G. Flanner, C.S. Zender, E.S. Saltzman, J.R. Banta, D.R. Pasteris, M.M. Carter, J.D. Kahl, 20th-century industrial black carbon emissions altered arctic climate forcing, Science. 317 (2007) 1381–1384.

[2] J.W. Martin, M. Salamanca, M. Kraft, Soot inception: Carbonaceous nanoparticle formation in flames, Prog. Energy Combust. Sci. 88 (2022) 100956.

[3] Y. Wang, S.H. Chung, Soot formation in laminar counterflow flames, Prog. Energy Combust. Sci. 74 (2019) 152–238.

[4] H. Wang, Formation of nascent soot and other condensed-phase materials in flames, Proc. Combust. Inst. 33 (2011) 41–67.

[5] T.S. Totton, A.J. Misquitta, M. Kraft, A quantitative study of the clustering of polycyclic aromatic hydrocarbons at high temperatures, Phys. Chem. Chem. Phys. 14 (2012) 4081.

[6] J.W. Martin, K. Bowal, A. Menon, R.I. Slavchov, J. Akroyd, S. Mosbach, M. Kraft, Polar curved polycyclic aromatic hydrocarbons in soot formation, Proc. Combust. Inst. 37 (2019) 1117–1123.

[7] J.Y. Lai, P. Elvati, A. Violi, Stochastic atomistic simulation of polycyclic aromatic hydrocarbon growth in combustion, Phys. Chem. Chem. Phys. 16 (2014) 7969–7979.

[8] Q. Wang, J.C. Saldinger, P. Elvati, A. Violi, Molecular structures in flames: A comparison between SNapS2 and recent AFM results, Proc. Combust. Inst. 38 (2021) 1133–1141.
[9] M. Frenklach, Reaction mechanism of soot formation in flames, Phys. Chem. Chem. Phys. 4 (2002) 2028–2037.
[10] F. Carbone, M.R. Canagaratna, A.T. Lambe, J.T. Jayne, D.R. Worsnop, A. Gomez, Exploratory analysis of a sooting premixed flame via on-line high resolution (APi–TOF) mass spectrometry, Proc. Combust. Inst. 37 (2019) 919–926.
[11] K.O. Johansson, M.P. Head-Gordon, P.E. Schrader, K.R. Wilson, H.A. Michelsen, Resonance-stabilized hydrocarbon-radical chain reactions may explain soot inception and growth, Science. 361 (2018) 997–1000.
[12] M. Commodo, K. Kaiser, G. De Falco, P. Minutolo, F. Schulz, A. D'Anna, L. Gross, On the early stages of soot formation: Molecular structure elucidation by high-resolution atomic force microscopy, Combust. Flame. 205 (2019) 154–164.
[13] J.W. Martin, D. Hou, A. Menon, L. Pascazio, J. Akroyd, X. You, M. Kraft, Reactivity of Polycyclic Aromatic Hydrocarbon Soot Precursors: Implications of Localized π-Radicals on Rim-Based Pentagonal Rings, J. Phys. Chem. C. 123 (2019) 26673–26682.
[14] A. Menon, J.W. Martin, J. Akroyd, M. Kraft, Reactivity of Polycyclic Aromatic Hydrocarbon Soot Precursors: Kinetics and Equilibria, J. Phys. Chem. A. 124 (2020) 10040–10052.
[15] M. Frenklach, A.M. Mebel, On the mechanism of soot nucleation, Phys. Chem. Chem. Phys. 22 (2020) 5314–5331.
[16] J.W. Martin, L. Pascazio, A. Menon, J. Akroyd, K. Kaiser, F. Schulz, M. Commodo, A. D'Anna, L. Gross, M. Kraft, π-Diradical Aromatic Soot Precursors in Flames, J. Am. Chem. Soc. 143 (2021) 12212–12219.
[17] J. Bai, X. Liu, T. Lei, B. Teng, X. Wen, A combined DFTB nanoreactor and reaction network generator approach for the mechanism of hydrocarbon combustion, Chem. Commun. 57 (2021) 11633–11636.
[18] K. Chenoweth, A.C. Van Duin, W.A. Goddard, ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation, J. Phys. Chem. A. 112 (2008) 1040–1053.
[19] T.P. Senftle, S. Hong, M.M. Islam, S.B. Kylasa, Y. Zheng, Y.K. Shin, C. Junkermeier, R. Engel-Herbert, M.J. Janik, H.M. Aktulga, others, The ReaxFF reactive force-field: development, applications and future directions, Npj Comput. Mater. 2 (2016) 1–14.
[20] Y. Xiao, W. Dong, H. Busnengo, Reactive force fields for surface chemical reactions: A case study with hydrogen dissociation on Pd surfaces, J. Chem. Phys. 132 (2010) 014704.
[21] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A.V. Shapeev, A.P. Thompson, M.A. Wood, S.P. Ong, Performance and Cost Assessment of Machine Learning Interatomic Potentials, J. Phys. Chem. A. 124 (2020) 731–745.
[22] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, W. E, End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems, in: Adv. Neural Inf. Process. Syst., Curran Associates, Inc., 2018.
[23] S. Chmiela, A. Tkatchenko, H.E. Sauceda, I. Poltavsky, K.T. Schütt, K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields, Sci. Adv. 3 (2017) e1603015.
[24] J. Zeng, L. Cao, M. Xu, T. Zhu, J.Z.H. Zhang, Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation, Nat. Commun. 11 (2020) 5713.
[25] L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V.S. Pande, T.J. Martínez, Discovering chemistry with an ab initio nanoreactor, Nat. Chem. 6 (2014) 6.
[26] T.J. Martínez, Ab initio reactive computer aided molecular design, Acc. Chem. Res. 50 (2017) 652–656.
[27] S. Harmeling, A. Ziehe, M. Kawanabe, K.-R. Müller, Kernel-based nonlinear blind source separation, Neural Comput. 15 (2003) 1089–1124.
[28] J.A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, A. Tkatchenko, Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems, Chem. Rev. (2021) acs.chemrev.1c00107.
[29] F. Musil, A. Grisafi, A.P. Bartók, C. Ortner, G. Csányi, M. Ceriotti, Physics-Inspired Structural Representations for Molecules and Materials, Chem. Rev. 121 (2021) 9759–9815.
[30] G. Lippert, J. Hutter, M. Parrinello, The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations, Theor. Chem. Acc. 103 (1999) 124–140.
[31] S. Goedecker, M. Teter, J. Hutter, Separable dual-space Gaussian pseudopotentials, Phys. Rev. B. 54 (1996) 1703.
[32] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865.
[33] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. 132 (2010) 154104.
[34] J. VandeVondele, J. Hutter, Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases, J. Chem. Phys. 127 (2007) 114105.
[35] Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, W. E, DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models, Comput. Phys. Commun. (2020) 107206.
[36] Q. Mao, A.C.T. van Duin, K.H. Luo, Formation of incipient soot particles from polycyclic aromatic hydrocarbons: A ReaxFF molecular dynamics study, Carbon. 121 (2017) 380–388.
[37] C.J. Pope, J.A. Miller, Exploring old and new benzene formation pathways in low-pressure premixed flames of aliphatic fuels, Proc. Combust. Inst. 28 (2000) 1519–1527.
[38] P. Liu, Z. Li, A. Bennett, H. Lin, S.M. Sarathy, W.L. Roberts, The site effect on PAHs formation in HACA-based mass growth process, Combust. Flame. 199 (2019) 54–68.
[39] J. Zeng, L. Cao, C.-H. Chin, H. Ren, J.Z.H. Zhang, T. Zhu, ReacNetGenerator: an automatic reaction network generator for reactive molecular dynamics simulations, Phys. Chem. Chem. Phys. 22 (2020) 683–691.
[40] E. Kirkland, Advanced Computing in Electron Microscopy Plenum Press, N. Y. (1998).
[41] B. Apicella, P. Pré, M. Alfè, A. Ciajolo, V. Gargiulo, C. Russo, A. Tregrossi, D. Deldique, J.N. Rouzaud, Soot nanostructure evolution in premixed flames by High Resolution Electron Transmission Microscopy (HRTEM), Proc. Combust. Inst. 35 (2015) 1895–1902.
[42] M.L. Botero, D. Chen, S. González-Calera, D. Jefferson, M. Kraft, HRTEM evaluation of soot particles produced by the non-premixed combustion of liquid fuels, Carbon. 96 (2016) 459–473.
[43] R.S. Jacobson, A.R. Korte, A. Vertes, J.H. Miller, The Molecular Composition of Soot, Angew. Chem. Int. Ed. 59 (2020) 4484–4490.