Internal Structure of Incipient Soot from Acetylene Pyrolysis obtained via Molecular Dynamics Simulations

Khaled Mosharraf Mukut,[†] Anindya Ganguly,[‡] Eirini Goudeli,[‡] Georgios A. Kelesidis,[¶,§] and Somesh P. Roy[*,†]

†*Department of Mechanical Engineering, Marquette University, Milwaukee, WI 53233, U.S.A.*

‡*Department of Chemical Engineering, University of Melbourne, Parkville, VIC 3052, Australia*

¶*Nanoscience and Advanced Materials Center (NAMC), Environmental and Occupational Health Science Institute, School of Public Health, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, U.S.A.*

§*Particle Technology Laboratory, Institute of Process Engineering, Department of Mechanical and Process Engineering, ETH Zürich, 8092 Zürich, Switzerland*

E-mail: somesh.roy@marquette.edu

Abstract

A series of reactive molecular dynamics simulations is used to study the internal structure of incipient soot particles obtained from acetylene pyrolysis. The simulations were performed using ReaxFF potential at four different temperatures. The resulting soot particles are cataloged and analyzed to obtain statistics of their mass, volume, density, C/H ratio, number of cyclic structures, and other features. A total of 3324 incipient soot particles were analyzed in this study. Based on their structural characteristics, the incipient soot particles are classified into two classes, referred to as type 1 and type 2 incipient soot particles in this work. The radial distribution of density, cyclic (5-, 6-, or 7-member rings) structures, and C/H ratio inside the particles revealed a clear difference in the internal structure between type 1 and type 2 particles. These classes were further found to be well represented by the size of the particles, with smaller particles in type 1 and larger particles in type 2. The radial distributions of ring structures, density, and C/H ratio indicated the presence of a dense core region in type 2 particles. In contrast, no clear evidence of the presence of a core was found in type 1 particles. In type 2 incipient soot particles, the boundary between the core and shell was found to be around 50%–60% of the particle radius of gyration.

1 Introduction

Soot is a harmful carbonaceous nanoparticle generated during the combustion of hydrocarbon fuels. Soot, also known as black carbon, can cause serious health issues[1,2] and acts as a major forcing factor in climate change.[3,4] The exact mechanism of the formation of soot particulates from gaseous precursors is still unknown due to the complex chemical nature of the hydrocarbon reaction network and time- and length-scale of the soot formation processes. According to the present understanding, soot formation occurs by a series of complex physicochemical events such as the formation of gas-phase soot precursors (including, but not limited to, polycyclic aromatic hydrocarbons or PAHs), nucleation of incipient

soot particles, growth, maturation of incipient soot particles due to surface reactions, and aggregation by coagulation or coalescence, and decay of the particles by fragmentation and oxidation.[5–9] The inception of soot particles is arguably the least understood phenomenon among these processes and the exact chemical reaction pathways of soot inception are not completely known yet. Researchers agree that soot formation starts with production of small gas-phase precursor molecules such as acetylene, which leads to PAHs like benzene, pyrene, and coronene.[10–12] These freshly formed PAHs then combine to form the solid or liquid-like incipient soot particles.[13–15] These particles then start to grow by surface reactions and coalescence to form larger soot particles.[13,16–19]

Due to the complexity and scale of incipient soot particles, their exact internal structures are not very well characterized yet. Recent studies have shown young soot particles tend to have a condensed core of ring-like large molecules while surrounded by a shell of less stacked smaller molecules.[20] As these incipient particles mature, their internal structures evolve, affecting their physical and chemical properties.

There have been some recent breakthroughs in the experimental exploration of the internal structure of incipient soot. For example, Chang et al.[21] employed high-resolution transmission electron microscopy (HRTEM) and scanning electron microscopy (SEM) to investigate the structural evolution and fragmentation of coal-derived soot and carbon black particles under high-temperature air oxidation conditions. They also explored the onset of micropores and the internal graphitic microcrystals using X-ray diffraction (XRD) and Raman spectra. Morajkar et al.[22] utilized HRTEM, XRD, Raman spectroscopy, and inductively coupled plasma mass spectrometry (ICPMS) to examine the transmission of trace metals from biodiesel fuels to soot particles and the nanostructural irregularities of the soot. In their study, Gleason et al.[23] indicated that the formation of soot nuclei in an ethylene/nitrogen flame can be attributed exclusively to aromatic compounds comprising one or two rings. Carbone et al.[24] conducted a comprehensive investigation of soot inception in a laminar premixed ethylene flame and found that soot particles undergo an aging transformation – from

being nearly transparent in the visible spectrum to a more graphitic-like composition. Using low-fluence laser desorption ionization (LDI) in conjunction with HRTEM, Jacobson et al.,[25] investigated the molecular composition of soot particles to determine the PAH concentration in them. The capabilities of atomic force microscopy (AFM) were exploited by Barone et al.[26] to calculate particle size distribution functions under different sampling conditions. Schulz et al.[27] investigated the initial phases of soot formation using AFM and observed the presence of multiple aromatic compounds, some of which displayed noticeable aliphatic side chains. Commodo et al.[28] investigated the initial phases of soot formation using X-ray, ultraviolet photoemission spectroscopy (UPS), UV-visible, and Raman spectroscopy to show the coexistence of sp^3 carbon and a more advanced graphitic structure with a slightly larger aromatic island, a reduced band gap, and an increased density of states. In another study, Commodo et al.[29] identified a noteworthy occurrence of aliphatic pentagonal rings in the early stages of soot formation, particularly near the outer region of aromatic soot molecules, and it has been suggested that the elimination of hydrogen from these molecules can result in the creation of resonantly stabilized π-radicals.[30] This phenomenon has also been theorized by Johansson et al.,[31] Gentile et al.,[32] and Rundel et al.[33]

Even with such recent advancements in experimental findings, there is still a lot of unknowns about the internal structure of soot. The limitations of experimental methods can be compensated and complemented by first-principle modeling such as molecular dynamics. With the development of high-performance computational resources, reactive molecular dynamics (RMD) simulation has become more affordable for studying complex reactive networks. For soot-relevant RMD studies, the reactive force field (ReaxFF) potential developed by van Duin et al.[34] for carbon, hydrogen, and oxygen chemistry (CHO-parameters[35,36]) is a popular choice. The ReaxFF potential can capture the physicochemical evolution of hydrocarbon systems in an extensive range of temperatures and pressures. It is based on the bond order between different atoms, which carry information related to bond breakage and formation. In recent years, RMD simulations have been used to investigate soot nucleation

by pyrene dimerization,[37] to shed light on the nucleation and growth of incipient soot from PAHs, such as naphthalene, pyrene, coronene, ovalene and circumcoronene,[16] to explore the initial mechanism of soot nanoparticle formation[38] and to examine the effect of oxygenated additives on the reduction of diesel soot emissions.[39]

Since an RMD simulation provides detailed structural information at the atomic scale, it can be an excellent tool for analyzing the internal structure of incipient soot particles. For example, Pascazio et al.[40] recently looked into the internal structure and the mechanical properties of incipient soot particles using RMD simulation and quantified the amount of cross-linking in the core and shell region of developing and mature soot particles. Mature soot primary particles exhibit a distinct core-shell structure with a disorderly condensed core of ring-like structures surrounded by a shell of chain-like structures.[20,41]

Process temperature plays an important role in the development and aging of soot particles. For example, in a recent study, Pathak et. al.[42] studied graphitization induced structural transformation of candle soot at different temperatures and found that increasing the temperature increases the rate of graphitization that leads to more spherical and mature soot and weakening of the correlation between graphitic nanostructure and surface functional groups (SFGs). SFGs have been found to be connected to the characteristics of soot aggregates, including the fractal dimension.[43] Since soot morphology, maturity, and reactivity are expected to be influenced by temperature during acetylene pyrolysis,[44] it is important to study the internal structure of incipient soot particles at different temperatures.

In this work, we used acetylene pyrolysis as our target configuration. Acetylene (C_2H_2) is often used as a model fuel in studies of soot formation due to its simplicity and well-documented pyrolysis pathways, which make it an ideal candidate for detailed atomistic simulations. While acetylene pyrolysis does not encompass all the complexities of typical hydrocarbon combustion in flames, it provides a controlled environment that captures key aspects of the chemical mechanisms involved in soot nucleation and growth. Acetylene pyrolysis has also been used extensively to investigate soot formation experimentally (e.g. in

shock tubes[45]) and develop kinetic models for soot inception.[46–48] These studies have shown that soot formation during acetylene pyrolysis is driven by hydrogen abstraction and carbon addition, similar to soot formation in fuel rich flame conditions.[46] Based on the modeling of soot inception in both aromatic and aliphatic premixed flames, it has also been suggested[49] that the pathways leading to soot are similar across different aliphatic hydrocarbon types. This supports the qualitative validity of using acetylene as a model fuel in our current work.

In the present study, a series of isothermal RMD simulations using the ReaxFF potential is conducted by mimicking acetylene pyrolysis at different temperatures (1350, 1500, 1650, and 1800K). A variety of physicochemical features of these RMD-generated soot particles are then analyzed to shed light on different types of incipient soot particles and to characterize the internal structure of these particles obtained from RMD simulations.

2 Numerical Methodology

2.1 Simulation configurations

Following the methodology described by Sharma et al.,[50] 1000 acetylene molecules are randomly placed in a cubic domain (75Å × 75Å × 75Å) at four different temperatures, i.e., 1350, 1500, 1650, and 1800 K. The density of the overall system is 0.01 g/cm^3. The temperatures are chosen to capture soot particles from various thermally activated systems. For statistical significance, at each temperature, simulations are performed for at least five times with different initial configurations. In total, 24 RMD simulations were performed for four different temperatures. The RMD simulations are performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[51] software. ReaxFF potential for hydrocarbons[34,52] is used to capture the chemical changes (bond breakage and formation) due to molecular collisions during acetylene pyrolysis. The bond length between individual atoms is calculated at each timestep (0.25 fs) based on the changes in the chemical environment to describe bond cleavage and formation accurately.[36] This helps the model capture

the chemical reactions leading to the radical formation during soot nucleation. Periodic boundary conditions are assumed in all three dimensions. The coordinates of each atom are calculated and updated using the velocity-Verlet algorithm[53] in conjunction with the Nose-Hoover thermostat.[54] A constant number, volume, and temperature (NVT) ensemble strategy is used to run each simulation up to 10 ns. The simulation results are probed every 0.05 ns, and the clusters of hydrocarbons that resemble primary soot particles are isolated, tabulated, and analyzed. Each of these extracted clusters has at least 20 carbon atoms and at least one 5-, 6-, or 7-membered ring structure following an earlier study by Mukut et al.[55] Features such as surface area and volume of primary particles are calculated using MSMS software developed by Sanner,[56] and other physicochemical characteristics are analyzed mostly using MAFIA-MD.[57] The open visualization tool (OVITO)[58] is used for visualization of the molecular clusters.

2.2 Workflow

The workflow adopted in this study can be broken down in four steps. First, a series of RMD simulations are carried out at different temperatures with different initial configurations. The incipient soot particles generated in these simualtions are then extracted for analysis in the next step. The simulation configuration and identification of incipient soot particles are discussed in Sec. 2.1 and 2.3. Second, each of the extracted incipient soot particles are analyzed to extract their chemical and morphological features such as atom counts, carbon-to-hydrogen (C/H) ratio, radius of gyration, atomic fractal dimension, density, surface area, and volume as discussed in Sec. 2.3. In the third step, an unsupervised machine learning approach is used to properly classify these incipient soot particles based on the features extracted in the second step. This is discussed in detail in Sec. 3.2. Finally, the radial distributions of important features of each class of incipient soot particles are analyzed to understand the internal structure of the particles. This analysis is presented in Sec. 3.4.

2.3 Extraction of physicochemical properties

From the RMD simulations, we extract the coordinates of each atom present in the simulation box at regular time intervals via the trajectory file. Each timestep is investigated separately by analyzing the atom coordinates within the entire simulation domain, which contains large molecular clusters and small molecules. The large soot-like molecular clusters are identified as the ones that have more than 20 carbon atoms and have at least one 5-, 6-, or 7-member ring.[55] In our case, the smallest such cluster was found to have 65 carbon atoms. These clusters are isolated using the cluster analysis tool from the OVITO Python module[58] implemented in a unified Python script developed in-house. Then, the isolated clusters are analyzed individually to calculate their physical, morphological, and chemical attributes. Some attributes are obtained trivially from the trajectory files, e.g., number of atoms (N), carbon to hydrogen ratio ($\Theta_{C/H}$), mass (M_p), and molar mass (M). Some other attributes like the radius of gyration (R_g), atomic fractal dimension (D_f), and density (ρ) are extracted by simple algebraic and geometric analysis or by using empirical correlations proposed in the literature. These equations and/or correlations are listed in Appendix B. The volume and the surface area of incipient particles are calculated using MSMS software[56] using a probe radius of 1.5 Å.

The identification and analysis of 5- /6- /7-member ring structures are done using MAFIA-MD.[57] MAFIA-MD can analyze RMD trajectory files to identify cyclic/ring structures in an atomic cluster. Not all cyclic structures identified are necessarily aromatic. As discussed in,[57] it is difficult to exactly confirm which cyclic structures are aromatic as the information about aromaticity requires some approximations regarding the bond order of aromatic bonds and establishment of planarity. To remove any confusion, therefore, we used the terms "ring" or "cyclic" in this work instead of aromatic when discussing these internal structures in the soot clusters. The numbers of 5-, 6-, and 7-member rings are denoted as N_5, N_6, N_7, respectively, and the total number of rings is denoted as $N_{\bigcirc}$. Similarly, the

number of carbons in rings is denoted as $N_{©}$, and the number of non-cyclic carbons in a particle is denoted as $N_{¢}$.

A sample of two particles and a list of their properties are provided in Appendix C. This entire set of features is used in the classification of particles, as discussed in Sec. 3.2. It must be noted here that while all the properties mentioned above were evaluated for each particle, this article only focuses on the internal structure of the particles, which are characterized as discussed in Sec. 2.4. Therefore, beyond their use in classifying particles, the detailed analysis of physical and morphological features such as volume, surface area, radius of gyration, and atomic fractal dimension are not the focus of this work.

2.4 Characterization of internal structure

We analyzed the internal structure of soot particles via the radial distribution of carbon atoms, C/H ratio, and density inside the particle. To compare different-sized particles on the same scale, we first normalized the radius of particles by scaling each particle by its radius of gyration (R_g). Then, each particle is divided into an equal number of radial bins. Each radial bin creates a spherical shell or strip, as shown by the shaded yellow region of interest in Fig. 1. We calculate various internal features in these spherical strips and present them as a function of the normalized radial distance from the center of mass of each spherical strip ($r^* = r/R_g$).

For example, let's consider a strip of width Δr, whose midplane is distance r away from the center of mass. The number of cyclic and non-cyclic carbon atoms in this strip (i.e., within a radial distance of $r \pm \Delta r/2$) are counted and represented as a function of the normalized radius of the midplane ($r^* = r/R_g$), as $n_{©}(r^*)$ and $n_{¢}(r^*)$, respectively. The radial distribution of cyclic and non-cyclic carbon per unit area (indicated by $''$) at a normalized distance r^* are then, respectively:

$$N''_{©}(r^*) = \frac{n_{©}(r^*)}{4\pi(r^* \times R_g)^2}; \qquad N''_{¢}(r^*) = \frac{n_{¢}(r^*)}{4\pi(r^* \times R_g)^2} \tag{1}$$

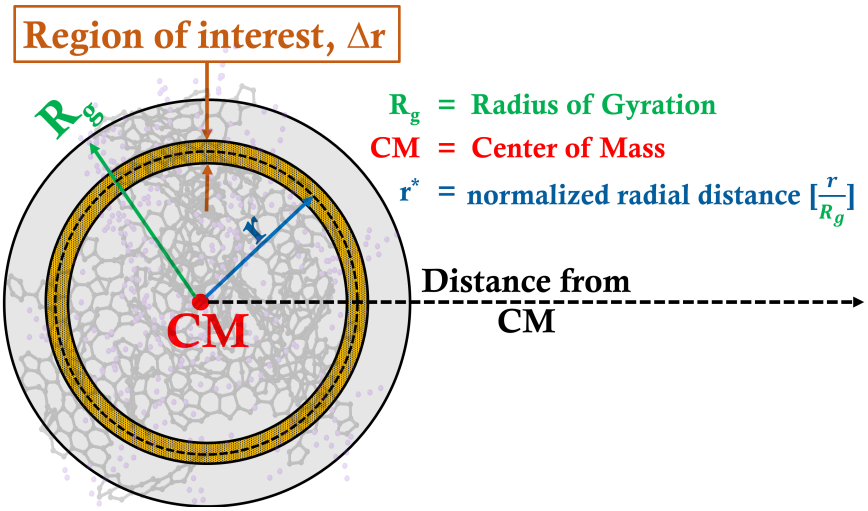


Figure 1: Schematic representation of the calculation of the radial distribution of internal features in the soot particle.

Similarly, the radial distribution of C/H ratio is also analyzed for each particle. The C/H ratio of the entire particle is calculated as $\Theta_{C/H} = {}^{N_C}/_{N_H}$, where N_C and N_H are number of carbon and hydrogen atoms in the entire particle, respectively. This C/H ratio is termed as the *particle* C/H ratio ($\Theta_{C/H}$) to differentiate from the *local* C/H ratio ($\theta_{C/H}$), which is calculated using the number of carbon and hydrogen atoms in the spherical strips, as shown in Fig. 1. The local C/H ratio ($\theta_{C/H}$) is determined by calculating the number of carbon ($n_C(r^*)$) and hydrogen ($n_H(r^*)$) atoms in a spherical strip with the midplane at a normalized distance r^* from the center of mass ($\theta_{C/H}(r^*) = {}^{n_C(r^*)}/_{n_H(r^*)}$). Finally, the local C/H ratio is normalized by the corresponding particle C/H ratio

$$\theta^*_{C/H}(r^*) = \frac{\theta_{C/H}(r^*)}{\Theta_{C/H}} \tag{2}$$

Similarly, the radial distribution of local density is also evaluated by dividing the simulated density of the thin spherical strip using Eqn. B.3 for the strip (referred as *local* density, $\varrho(r^*)$) by the simulated density of the particle (ρ_s) as

$$\rho^*_s(r^*) = \frac{\varrho(r^*)}{\rho_s} \tag{3}$$

Here the local density, $\varrho(r^*)$, is determined by dividing the mass of particles within a spherical strip of radius (r^*) and thickness (Δr) by the volume of that shell. Mathematically, this is expressed as shown in Eqn. 4.

$$\varrho(r^*) = \frac{\sum_{i \in strip} m_i}{\frac{4}{3}\pi((r^* R_g + \frac{\Delta r}{2})^3 - (r^* R_g - \frac{\Delta r}{2})^3)}, \tag{4}$$

where m_i represents the mass of each atom within the spherical strip.

3 Results and Discussion

3.1 Formation of incipient soot particles in RMD

During the RMD simulations, the system of atoms goes through different chemical and physical interactions, resulting in the formation of larger atomic clusters due to the pyrolysis of acetylene. The evolution of one of these atomic clusters is depicted in Fig. 2. Carbon and hydrogen atoms are represented using black and red dots, respectively. First, the acetylene molecules combine to form small linear chains (Fig. 2B: linearization) and then transform into cyclic structures (Fig. 2C: cyclization). After cyclization, the small clusters start growing due to both bond formation at the surface and internal reorganization. These larger atomic clusters resemble incipient soot particles (Fig. 2D–F). It is important to note that the collisions are stochastic in nature and the time required for an event, i.e. linearization, cyclization, surface growth. etc., varies based on the initial configurations, and therefore are omitted from the figure for generality. A similar formation mechanism is also reported by Zhang et al.[59] for carbon-black simulations and Sharma et al[50] for acetylene pyrolysis simulations.

The incipient soot clusters are extracted from the RMD simulations at different timesteps to capture the growth. Each simulation is run at least five times with a velocity field initialized randomly in each case to generate soot particles with different evolutionary histories.

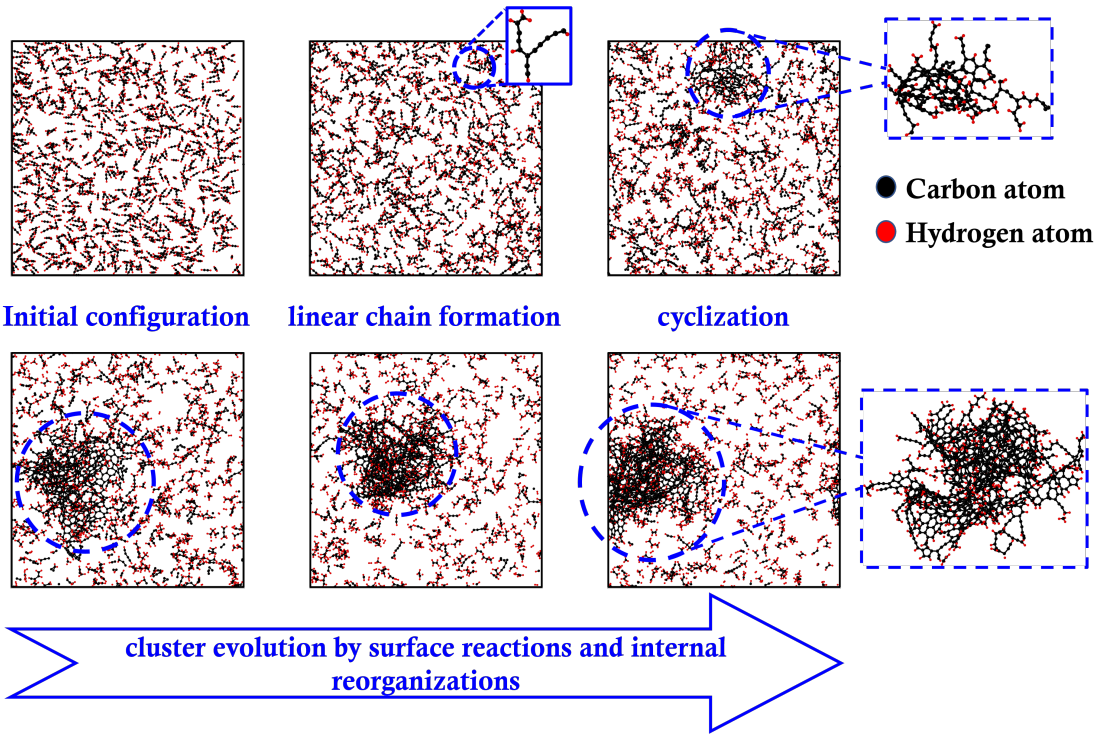


Figure 2: A general representation of steps during the formation and evolution of incipient soot cluster during acetylene pyrolysis (from a simulation performed at 1650 K).

In total, 3324 individual soot clusters are isolated from the RMD simulations (number of carbon atoms ranging from 65 to 1503). The C/H ratios ($\Theta_{C/H}$) of these particles are compared to the theoretical limits for PAHs in Fig. 3. Based on the compactness, the PAHs can be classified into two categories: (a) peri-condensed PAHs, where more than two aromatic rings can share the carbon atoms in the aromatic structures and (b) cata-condensed PAHs, where the carbon atoms in the aromatic structure can be shared by at most two aromatic rings. Siegmann and Sattler[60] proposed a relationship between the number of carbon and hydrogen atoms for both peri-condensed and cata-condensed PAHs. Fig. 3 presents the C/H ratio and molar mass of the soot clusters from different temperatures and compares it with the peri-condensed and cata-condensed PAH zones derived from.[60] As observed from Fig. 3, the soot clusters fall between the peri-condensed and cata-condensed boundaries, indicating an intricate network of different types of aromatic and aliphatic structures in incipient particles. Using atomic force microscopy Commodo et. al.[29] showed that smaller aromatic

clusters (number of carbon atoms ranging from 6 to 55, lower than what studied in this work) in the early stage of soot formation in a slightly sooting premixed ethylene flame tend to be close to the peri-condensed line. However, in this work, we find that large clusters lie closer to cata-condensed limit than peri-condensed limit, potentially indicating significant presence of non-aromatic (i.e., aliphatic and alicyclic) structures.

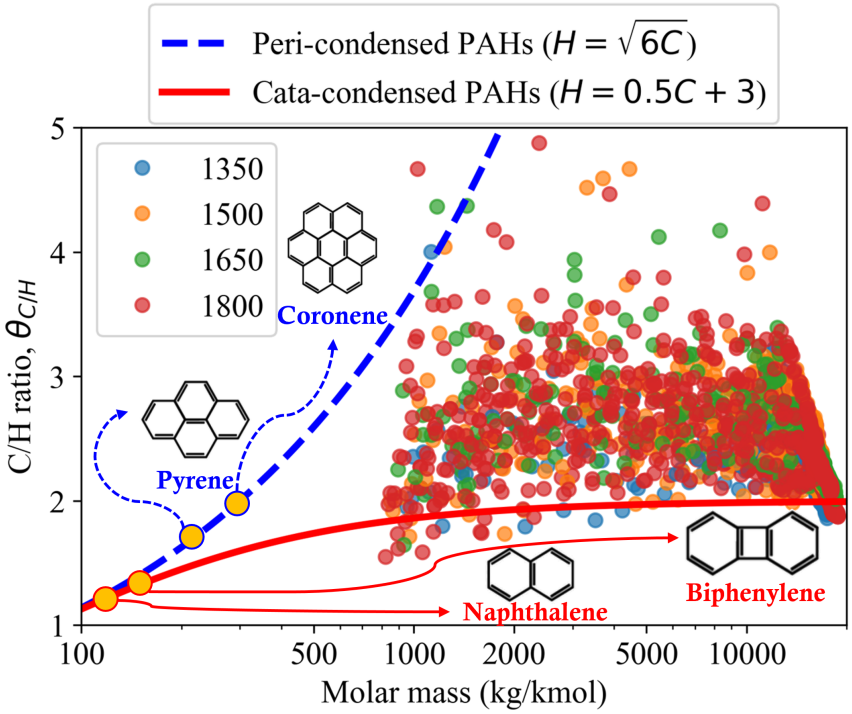
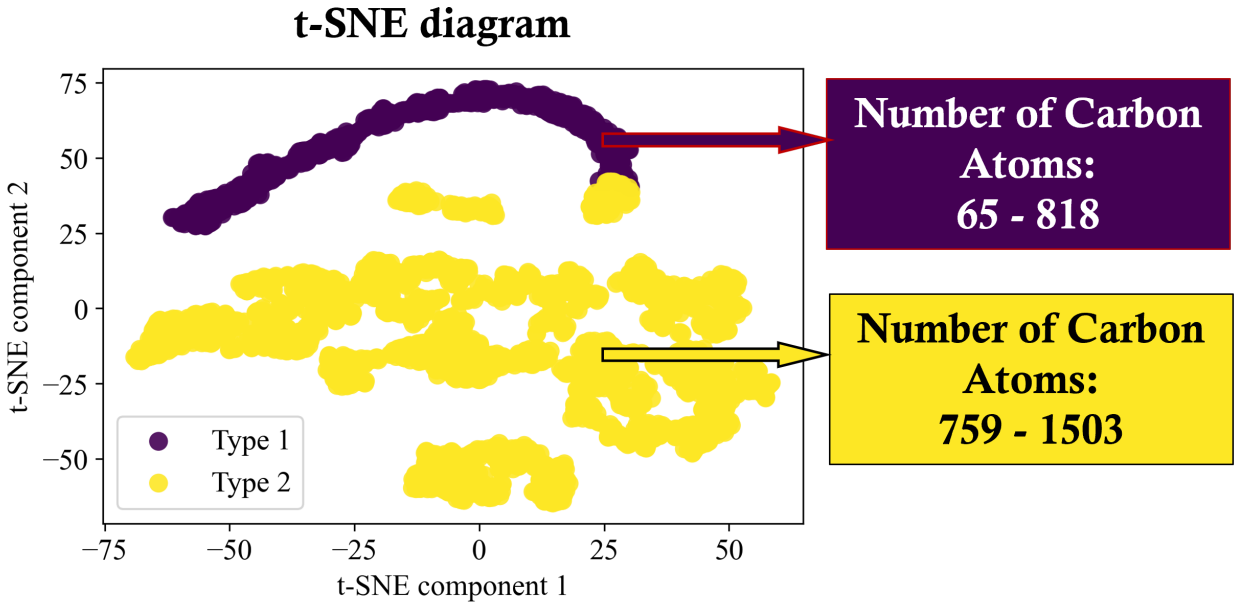


Figure 3: C/H ratio vs. molar mass of soot clusters at different temperatures.

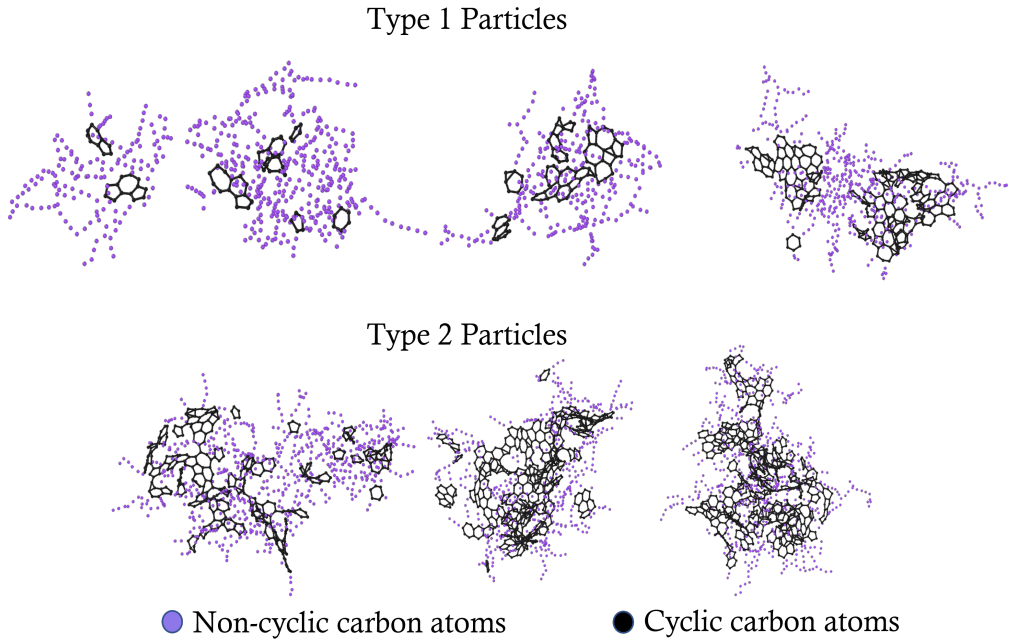
3.2 Classification of incipient soot particles

We tagged, extracted, or calculated the physicochemical features (such as Temperature, number of carbon atoms, number of hydrogen atoms, and molar mass) for each particle. The complete set of features used in this study is listed in C.1, and two sample particles with the entire feature set are shown in C.2.

Our initial observation of trends of various internal and physicochemical features revealed a wide variation, indicating that these particles can be classified into multiple groups based



(a) A t-SNE diagram generated using all the 3324 incipient soot clusters with two different k-means clusters.



(b) Some example particles from each class obtained from RMD simulations. The non-cyclic carbon atom structures are shown in purple dots, and the cyclic structures are shown in black. Hydrogen atoms are omitted from the visualization for clarity.

Figure 4: Classification of incipient soot particles

on their features. The classification of the incipient soot particles is depicted in Fig. 4. We attempted two unsupervised machine-learning techniques to extract unique classifications that may exist in the incipient particle sample space. The first method is the k-means clustering algorithm,[61] which is used to label particles of different classes based on all extracted features of the particles. Then, the t-dispersion stochastic neighbor embedding (t-SNE)[62] plot is used to display the particle properties on a 2D map. In the t-SNE diagram, similar clusters (i.e., potentially belonging to the same class) are expected to be close to each other. Although the number of classes was not known a priori, trial and error with k-means clustering revealed good results with two classes. For identification purposes, these two classes are referred to as "`type 1`" and "`type 2`" particles, respectively. The resulting t-SNE diagram is shown in Figure 4(a).

Looking closely at the two classes, we see that the particles exhibiting similarity fall into a nearly continuous size range. For example, in the first class, (`type 1`), the incipient particles have a lower total number of carbon atoms ($65 - 818$), whereas in the second class (`type 2`), the particles have a higher number of total atoms ($759 - 1503$). This essentially points to the fact that the characteristics of the incipient particles change after a certain level of growth: smaller particles (type 1) show different features and trends than larger particles (type 2). It should be noted here that the number of total carbon atoms is not a unique or absolute marker of the threshold between type 1 and type 2—as indicated by a slight overlap in the number of carbon atoms range between the types. The number of carbon atoms acts as a good approxiamte surrogate for threshold identifier. In total, we have obtained 670 type 1 and 2654 type 2 incipient particles from a total of 3324 particles. Fig. 4(b) depicts some example particles from the analyzed sample space. Here, the non-cyclic carbon atom structures are shown in purple dots and the cyclic structures are shown in black, while hydrogen atoms are omitted from the visualization for clarity.

3.3 Comparison with experimental data

The lack of detailed experimental data on the internal structure of incipient soot particles makes direct comparisons with the simulated soot particles obtained in this study challenging. Nonetheless, averaged physical properties such as density and C/H ratios, which have been more thoroughly investigated in experiments, can be used to validate the simulated soot particles. A detailed validation of these simulations based on average particle properties is presented in Mukut et al,[63] which explored the averaged particle properties for the same ensemble of particles investigated in this work. Since discussion of average particle property is beyond the scope of this work, which only focuses on the internal structure, only a brief summary of validation is reported here. The calculated density of our particles is 1.53 ± 0.08 g/cm^3 which closely matches with the empirical density of 1.51 g/cm^3 reported by Johansson et al.[64] Additionally, the mean C/H ratio of the particles obtained in our simulation is 2.36, which falls within the range of 2.33 ± 0.16 as measured by Schulz et al.[27] for soot particles from a premixed flame.

3.4 Internal structure of incipient soot particles

The radial distributions of various quantities relevant to the analysis of the internal structure of the incipient particles are presented as box and whisker plots in subsequent figures. In this visualization, every box includes the data points within the second and third quartiles, the horizontal line inside the box indicates the median value, and whiskers represent the range of the data. The statistics were found to be insensitive to the process temperature. Hence, aggregate data for all temperatures is presented here.

The radial distribution of cyclic and non-cyclic carbon atoms per unit area for type 1 (Figs. 5(a)-5(b)) and type 2 (Figs. 5(c)-5(d)) incipient particles as a function of the normalized radial distance from the center of mass is shown in Fig. 5. The blue vertical line depicts the location where the radial distance becomes equal to the radius of gyration (R_g) of individual particles. For type 1 particles, an abundance of non-cyclic carbon atoms is observed in the central (i.e., less than 50% of R_g) region. The number of carbon atoms

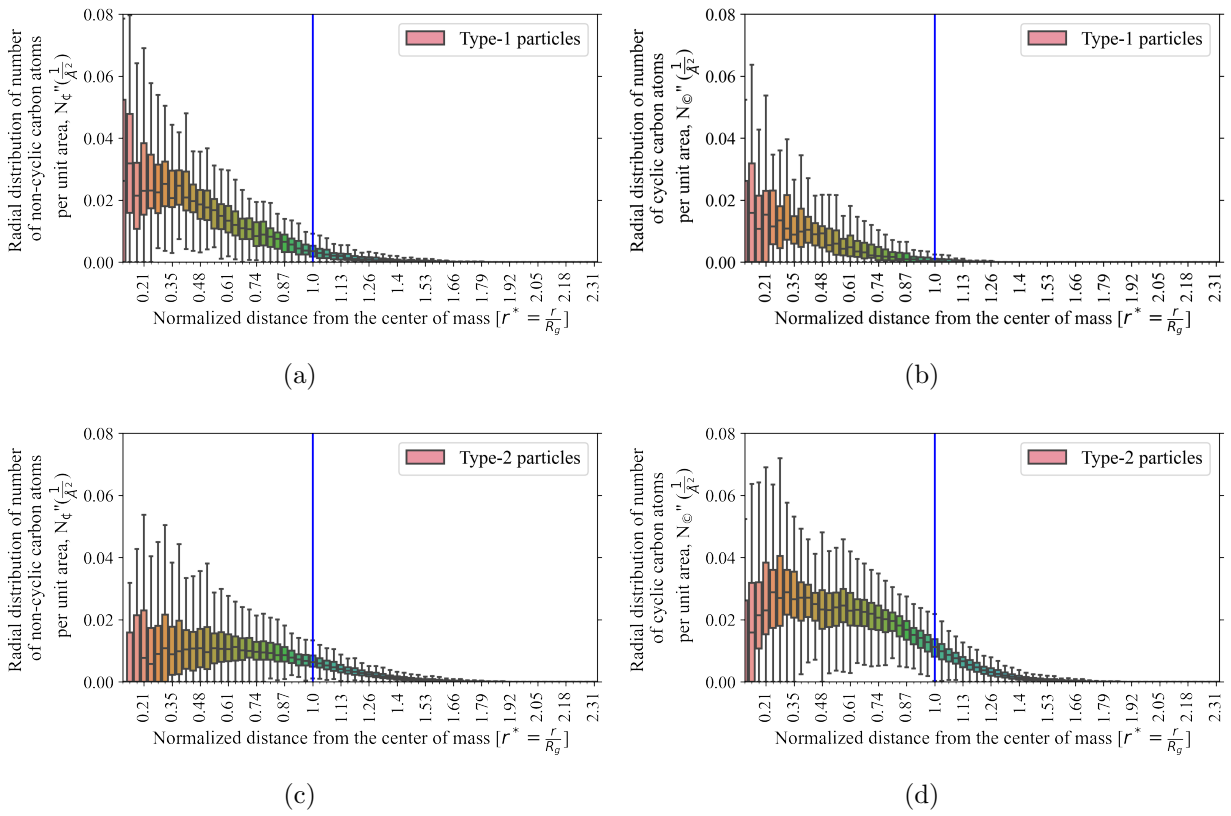


Figure 5: Radial distribution of (a,c) non-cyclic and (b,d) cyclic carbon atoms ($N''_{¢}(r^*)$ and $N''_{©}(r^*)$) in (a,b) type 1 and (c,d) type 2 soot particles as a function of the normalized radial distance ($r^* = \frac{r}{R_g}$) from the center of mass. The blue vertical line is at a radial distance equivalent to R_g.

in non-cyclic structures is almost two times more than that in cyclic structures in type 1 particles. Almost all the carbon atoms reside near the central region of type 1 particles, and the number of carbon atoms quickly drops to zero as we move away from the center of mass. This indicates that in type 1 particles, i.e., at the very early stages of soot formation, the number of non-cyclic structures is significantly higher than that of cyclic structures.

The type 2 particles, on the other hand, show a different trend where a very concentrated region of cyclic carbon atoms is observed in the central region of the particles. The concentration of cyclic carbon atoms slowly decreases as the distance from the center of mass increases. The number of non-cyclic carbon atoms increases from a very low value in the central region, then reaches a steady value near the radius of gyration and then gradually

drops to zero as the distance increases beyond the radius of gyration. This indicates that the non-cyclic carbon atoms are more likely to be present in the outer region (what can be presumably considered near the particle surface) of the type 2 incipient particles, while the central region is dominated by the cyclic carbon atoms.

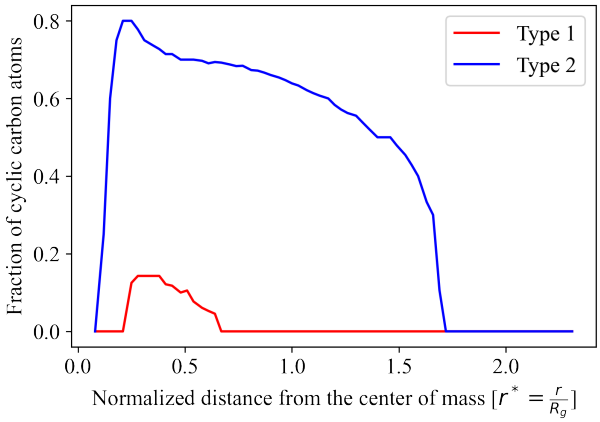
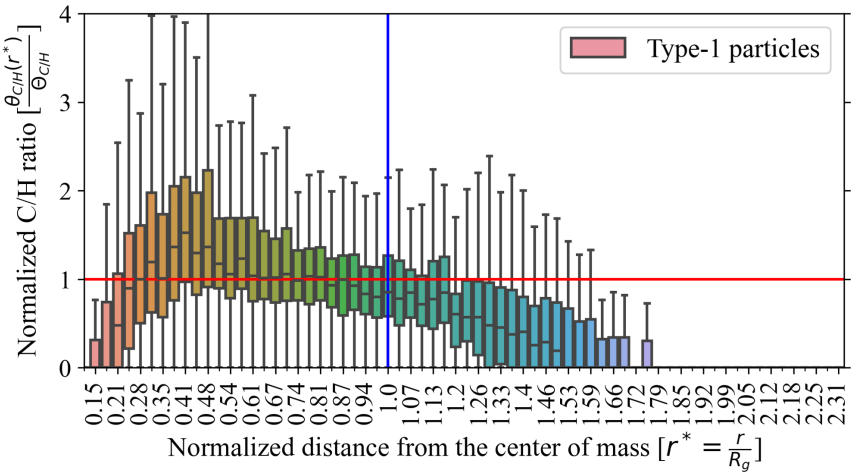


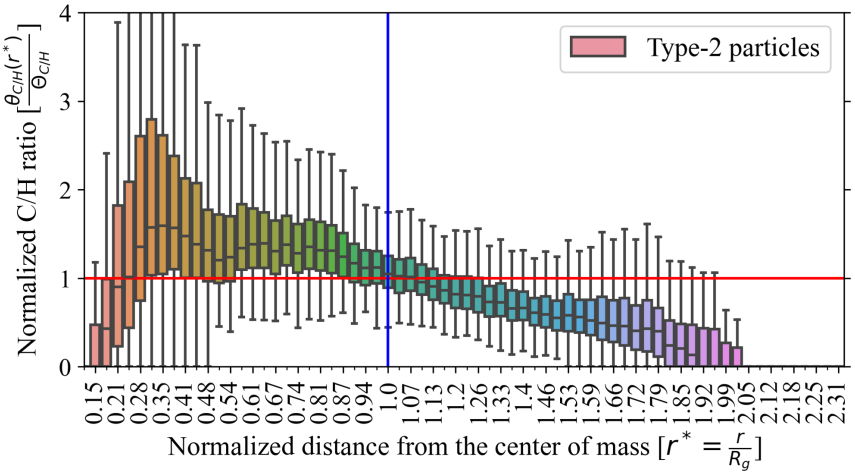
Figure 6: Radial distribution of the median fraction of cyclic carbon.

If the radial distribution of the median value of the fraction of cyclic carbon atoms to the total number of carbon atoms is analyzed (shown in Fig. 6), the dominance of cyclic carbon atoms in type 2 particles becomes very clear. Figure 6 shows a prominent central region where more than half of the carbon atoms belong to a ring structure. These evidences suggest that the internal chemical structure of incipient soot particles changes as the particles transition from type 1 to type 2. This indicates the development of core-shell structures as the incipient soot grows and matures. The presence of such a core-shell structure has been theorized in the literature. For example, Michelsen et al.[65] used a fractal core-shell model to explain the changes in the structure of soot aggregates and primary particles at different heights of a laminar co-flow ethylene-air flame. Kholghy et al.[66] proposed a surface shell formation model to predict the maturity of soot primary particles. More directly, recently, Botero et al.[20] studied the internal structure of soot particles using high-resolution transmission electron microscopy (HRTEM) to identify the PAH structures in the core and shell regions and suggested the presence of a stabilized core region, indicating nano-structural mobility.

Pascazio et al.[40] utilized RMD simulation and identified different cross-linking levels in the core and shell in hypothetical soot particles. Kelesidis et al.[67] also investigated oxidation dynamics of carbonaceous nanoparticles having various core-shell structures using lattice Monte Carlo simulations. The apparent difference in the radial distribution of cyclic and non-cyclic carbon atoms between type 1 and type 2 particles supports these findings.



(a)



(b)

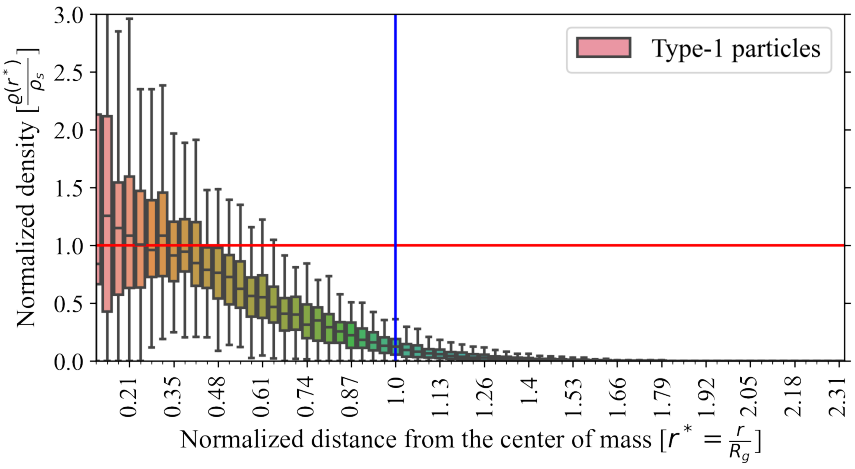
Figure 7: Radial distribution of normalized C/H ratio ($\theta^*_{C/H}(r^*) = \frac{\theta_{C/H}(r^*)}{\Theta_{C/H}}$) in incipient particles as a function of normalized radial distance ($r^* = \frac{r}{R_g}$) from the center of mass. The blue vertical line is at a radial distance equivalent to R_g. The $\frac{C}{H}$ ratio of the strip and the particle are the same along the red horizontal line.

The changes in chemical properties inside the incipient soot particles can also be observed

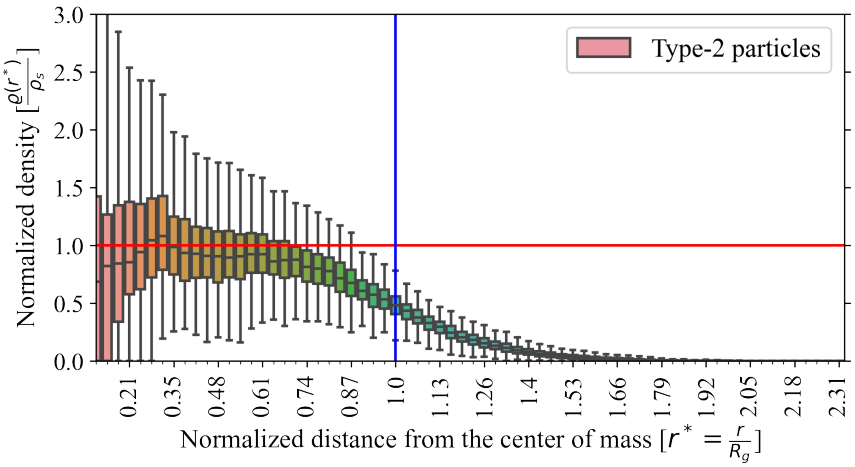
in the radial distribution of carbon to hydrogen ratio. The normalized local C/H ratio for type 1 (Fig. 7(a)) and type 2 (Fig. 7(b)) particles are plotted as a function of the normalized radial distance from the center of mass in Fig. 7 using box and whisker plots. The blue vertical line depicts the location where the radial distance becomes equal to the radius of gyration (R_g) of individual particles. The red horizontal line indicates where the local C/H ratio equals the particle C/H ratio.

The normalized local C/H ratio increases up to a certain distance from the center of mass and then starts to drop. For type 1 particles, the increase in the local C/H ratio take longer distance from the center of mass, and the median value gradually reaches a peak value slightly higher than the particle C/H ratio (about 1.3 times). After that, the local C/H ratio starts to drop, and the median value reaches a value close to the particle C/H ratio at around 65% of the radius of gyration. The value stays close to the particle C/H ratio up to the radius of gyration and then slowly drops to zero. This indicates a very small or no dense core region in type 1 particles. The distinction between the core and shell regions is not clear in type 1 particles because of the absence of a pronounced core as the local C/H ratio remains close to the particle C/H ratio. For type 2 particles, however, the local C/H ratio increases rapidly in the central region, and the median value reaches a peak value of about 1.7 times the particle C/H ratio. After that, the local C/H ratio starts to drop and reaches a value equal to the global C/H ratio at the radius of gyration. Unlike type 1 particles, the region where the median of the normalized local C/H ratio is close to unity is very narrow in type 2 particles. This indicates that the dense core region of type 2 particle is well-developed, and the separation between the core and shell regions is more pronounced than in type 1 particles.

The radial distribution of the normalized density inside type 1 (Fig. 8(a)) and type 2 (Fig. 8(b)) incipient soot particles is presented in Fig. 8 as a function of the normalized radial distance from the center of mass of soot particles (r^*) as box and whisker plots. The normalized C/H ratio shown in Fig. 7 and the normalized local density in Fig. 8



(a)



(b)

Figure 8: Radial distribution of normalized local density ($\rho^*(r^*) = \frac{\varrho(r^*)}{\rho_s}$) inside incipient particles as a function of normalized radial distance ($r^* = \frac{r}{R_g}$) from the center of mass. The blue vertical line is at a radial distance equivalent to R_g. The strip and particle density are the same along the red horizontal line.

represent fundamentally different aspects of particle structure and composition. Specifically, the normalized C/H ratio is a measure of the chemical composition within a specific region, independent of the actual mass density in that region. Hence, both Figs. 7 and 8 convey important and complementary pieces of information.

For type 1 particles, Fig. 8 shows a tiny dense central region extending up to about 40% of the radius of gyration. The local density of type 1 particles is maximum near the center

of mass and drops gradually as the distance from the center of mass increases. For type 2 particles, on the other hand, the dense core region is more prominent and extends up to about 50-60% of the radius of gyration. The local density of type 2 particles stays close to the particle density up to around 60% of the radius of gyration, and then it quickly drops.

3.5 The boundary between the core and shell

The results discussed so far indicate the presence of a core and shell structure in incipient soot particles. This is further examined by looking at the median values of normalized local density (ρ_s^*), normalized C/H ratio ($\theta_{C/H}^*$), and radial distribution of carbon atoms ($N''_{©}$ and $N''_{¢}$) in Fig. 9 for both type 1 (in red) and type 2 (in blue) incipient soot particles. No significant trend is noticed from the radial distribution of non-cyclic carbon per unit area (Fig. 9(d)). However, the normalized density (Fig. 9(a)), C/H ratio (Fig. 9(b)) and the radial distribution of cyclic carbon atoms per unit area (Fig. 9(c)) demonstrate unique common trends in type 2 particles, which is not observed in type 1 particles. All of the quantities

1. reach a maximum value near the center of mass and decrease gradually to a local minimum at a distance of about 50% of the radius of gyration,
2. show the presence of a plateau region between 50% and 60% of the radius of gyration,
3. and drop monotonously after approximately 60% of the radius of gyration.

The first region or the central region can be identified as the *core* of the soot, as it is near the center of mass, it is denser and contains more rings than other regions (as seen in Fig. 5(b) and 5(d)). The local minima marks the beginning of the boundary between the *core* and the *shell*. The narrow plateau region can be thought of as the boundary region between *core* and *shell* regions. And finally, the gradual descent of these quantities indicates the *shell* region. In the type 1 particles, we can only see the gradual descent stage, indicating that the core-shell differentiation is not yet developed, i.e., no developed core. The reason for

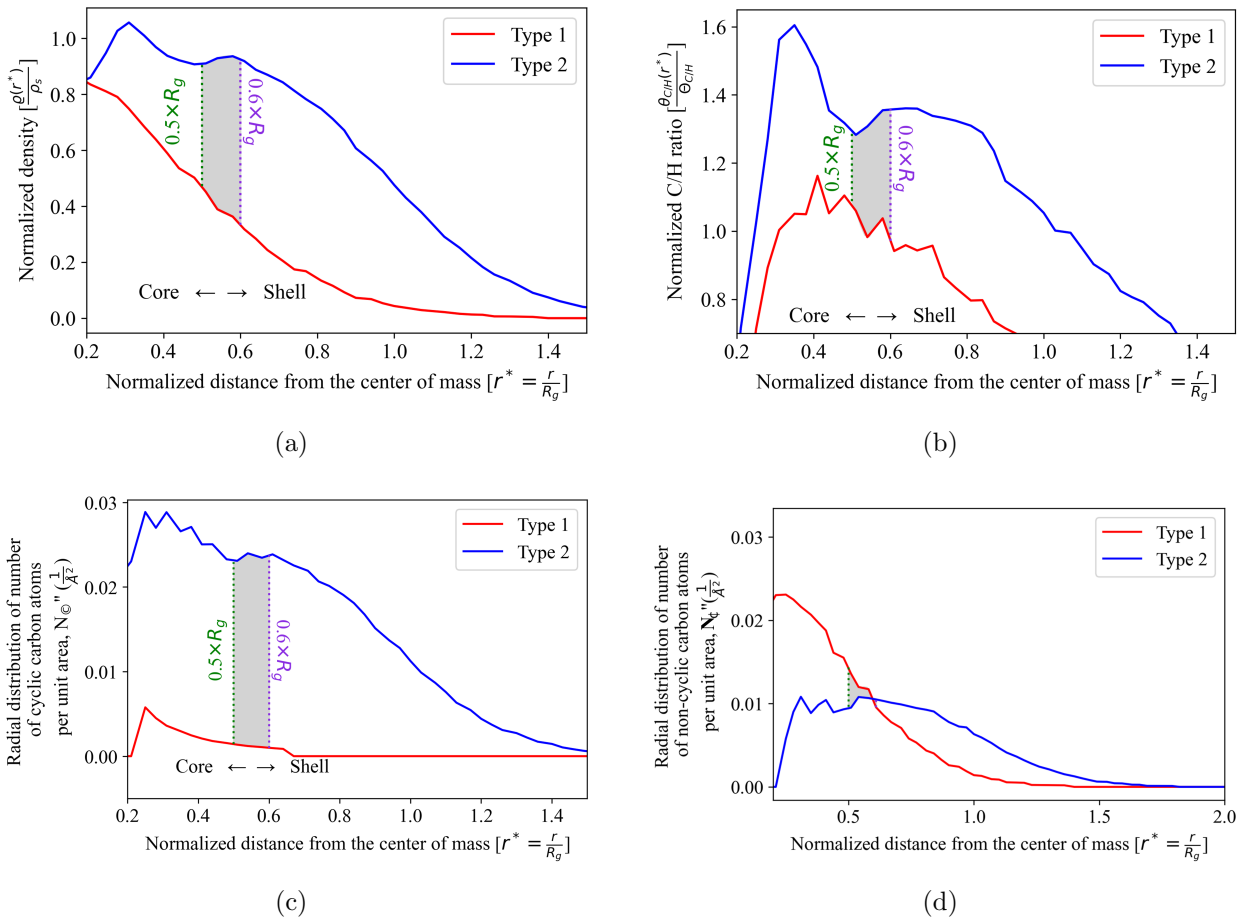


Figure 9: Identification of core and shell based on the radial distribution of normalized medians (a) density, (b) C/H ratio and radial distribution of the number of (c) cyclic and (d) non-cyclic carbon atoms per unit area in incipient particles as a function of normalized radial distance ($r^* = \frac{r}{R_g}$) from the center of mass.

such a trend in density, C/H ratio, and ring structures can be explained by the nature of the stacking of cyclic molecules (disordered and ordered), as shown in Fig. 10. As observed in the schematic in Fig. 10, the core region comprises an interconnected cross-linked network of cyclic molecules, while the shell region contains a sheet-like organization of cyclic molecules. Such structural differences in core and shell are also supported by results presented by Pascazio et al.[40] The core size of $0.5 - 0.60R_g$ obtained here is also consistent with the core size of $0.5 - 0.75R_g$ suggested by lattice Monte Carlo simulations[67], as well as with the core size of soot produced in diesel engines[41] and diffusion flames[20].

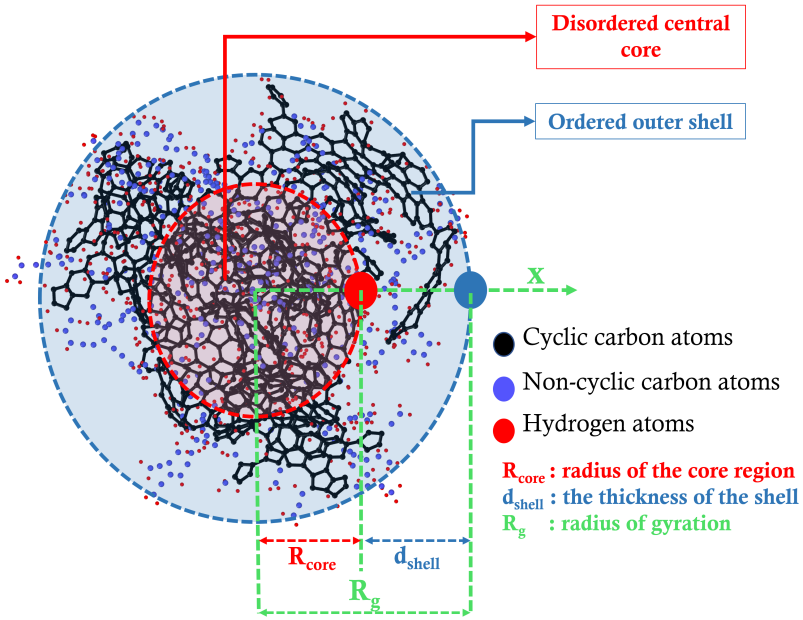


Figure 10: Schematic of core and shell structure of incipient particles.

4 Conclusion

A series of reactive molecular dynamics (RMD) simulations were performed to study the internal structure of incipient soot particles during acetylene pyrolysis at four different temperatures. A total of 3324 incipient soot particles were obtained at different stages of their evolution from these simulations. The mass, volume, surface area, radius of gyration, density, C/H ratio, and number of cyclic structures were calculated for each particle. It was found that all incipient soot particles contained both cyclic and non-cyclic moieties.

Using unsupervised machine-learning techniques, it was found that incipient soot particles can be classified into two types based on their morphological and chemical features. The particle size serves as a good approximate predictor of the class of the incipient particle; smaller particles tend to belong to type 1 and larger ones particles tend to belong to type 2 category. This indicated that the two types correspond to early and late stages of incipient soot.

The internal structure of the incipient particles was investigated using the radial distribution of cyclic and non-cyclic carbon-based moieties, C/H ratio, and density. We did

not observe any direct or obvious sensitivity of the internal structure of the particles to temperature for the temperature rang e of 1350 to 1800 K employed here.

The internal structures of the two types of incipient soot particles, however, showed distinctly different characteristics. The center of type 1 particle was primarily made of non-cyclic structures whereas the center of the type 2 particles is dominated by cyclic structures. Type 1 particles did not show any clear core-shell structure. The radial distribution of the particle density showed a clear existence of a dense "core" at the center and a less dense "shell" at the periphery of type 2 particles. The core in type 2 particles extends from the center of the particle up to approximately 0.5–0.6R_g. This core region were made of an interconnected corss-linked network of cyclic structures while the shell contained sheet-like organizations of cyclic structures.

This study presents a detailed exploration of internal structure of incipient soot particles obtained from acetylene pyrolysis. Acetylene can serve as a good model fuel for exploring soot inception. However, further investigations are needed to confirm whether the results obtained from acetylene pyrolysis can be generalized to other fuels and combustion conditions.

Acknowledgement

The research benefited from computational resources provided through the NCMAS, supported by the Australian Government, The University of Melbourne's Research Computing Services, and the Petascale Campus Initiative. K.M.M. and S.P.R. acknowledge funding support from the National Science Foundation as some of this material is based upon work supported by the National Science Foundation under Grant No. 2144290.

References

(1) Jacobsen, N. R.; Pojana, G.; White, P.; Møller, P.; Cohn, C. A.; Korsholm, K. S.; Vogel, U.; Marcomini, A.; Loft, S.; Wallin, H. Genotoxicity, cytotoxicity, and reactive

441 oxygen species induced by single-walled carbon nanotubes and C60 fullerenes in the
442 FE1-Muta™Mouse lung epithelial cells. *Environ. Mol. Mutagen.* **2008**, *49*, 476–487.

443 (2) Jerrett, M.; Burnett, R. T.; Beckerman, B. S.; Turner, M. C.; Krewski, D.; Thurston, G.;
444 Martin, R. V.; van Donkelaar, A.; Hughes, E.; Shi, Y. et al. Spatial Analysis of Air
445 Pollution and Mortality in California. *Am. J. Respir. Crit. Care Med.* **2013**,

446 (3) Hansen, J.; Nazarenko, L. Soot climate forcing via snow and ice albedos. *Proc. Natl.*
447 *Acad. Sci. U.S.A.* **2004**, *101*, 423–428.

448 (4) Bond, T. C.; Doherty, S. J.; Fahey, D. W.; Forster, P. M.; Berntsen, T.; DeAngelo, B. J.;
449 Flanner, M. G.; Ghan, S.; Kärcher, B.; Koch, D. et al. Bounding the role of black
450 carbon in the climate system: A scientific assessment. *Journal of Geophysical Research:*
451 *Atmospheres* **2013**, *118*, 5380–5552.

452 (5) Irimiea, C.; Faccinetto, A.; Mercier, X.; Ortega, I.-K.; Nuns, N.; Therssen, E.; Des-
453 groux, P.; Focsa, C. Unveiling trends in soot nucleation and growth: When secondary
454 ion mass spectrometry meets statistical analysis. *Carbon* **2019**, *144*, 815–830.

455 (6) Appel, J.; Bockhorn, H.; Frenklach, M. Kinetic modeling of soot formation with detailed
456 chemistry and physics: laminar premixed flames of C2 hydrocarbons. *Combust. Flame*
457 **2000**, *121*, 122–136.

458 (7) Wang, H. Formation of nascent soot and other condensed-phase materials in flames.
459 *Proc. Combust. Inst.* **2011**, *33*, 41–67.

460 (8) Rigopoulos, S. Modelling of Soot Aerosol Dynamics in Turbulent Flow. *Flow, Turbu-*
461 *lence and Combustion* **2019**, *103*, 565–604.

462 (9) Li, D. D.; Wang, C.; Chan, Q. N.; Yeoh, G. H. Soot: A review of computational models
463 at different length scales. *Experimental and Computational Multiphase Flow* **2023**, *5*,
464 1–14.

(10) Dobbins, R. A.; Subramaniasivam, H. *Soot Formation in Combustion*; Springer: Berlin, Germany, 1994; pp 290–301.

(11) Balthasar, M.; Kraft, M. A stochastic approach to calculate the particle size distribution function of soot particles in laminar premixed flames. *Combust. Flame* **2003**, *133*, 289–298.

(12) Wang, B.; Mosbach, S.; Schmutzhard, S.; Shuai, S.; Huang, Y.; Kraft, M. Modelling soot formation from wall films in a gasoline direct injection engine using a detailed population balance model. *Appl. Energy* **2016**, *163*, 154–166.

(13) Michelsen, H. A.; Colket, M. B.; Bengtsson, P.-E.; D'Anna, A.; Desgroux, P.; Haynes, B. S.; Miller, J. H.; Nathan, G. J.; Pitsch, H.; Wang, H. A Review of Terminology Used to Describe Soot Formation and Evolution under Combustion and Pyrolytic Conditions. *ACS Nano* **2020**, *14*, 12470–12490.

(14) Vander Wal, R. L. Soot precursor carbonization: Visualization using LIF and LII and comparison using bright and dark field TEM. *Combust. Flame* **1998**, *112*, 607–616.

(15) Zhao, B.; Uchikawa, K.; Wang, H. A comparative study of nanoparticles in premixed flames by scanning mobility particle sizer, small angle neutron scattering, and transmission electron microscopy. *Proc. Combust. Inst.* **2007**, *31*, 851–860.

(16) Mao, Q.; van Duin, A. C. T.; Luo, K. H. Formation of incipient soot particles from polycyclic aromatic hydrocarbons: A ReaxFF molecular dynamics study. *Carbon* **2017**, *121*, 380–388.

(17) Chen, M.; Li, W.; Zhang, H.; Liu, M.; Zhang, J.; Li, X.; Han, Y. Recent ReaxFF MD studies on pyrolysis and combustion mechanisms of aviation/aerospace fuels and energetic additives. *Energy Advances* **2023**, *2*, 54–72.

(18) Sabbah, H.; Biennier, L.; Klippenstein, S. J.; Sims, I. R.; Rowe, B. R. Exploring the Role of PAHs in the Formation of Soot: Pyrene Dimerization. *J. Phys. Chem. Lett.* **2010**, *1*, 2962–2967.

(19) Sánchez, N. E.; Callejas, A.; Millera, Á.; Bilbao, R.; Alzueta, M. U. Polycyclic Aromatic Hydrocarbon (PAH) and Soot Formation in the Pyrolysis of Acetylene and Ethylene: Effect of the Reaction Temperature. *Energy Fuels* **2012**, *26*, 4823–4829.

(20) Botero, M. L.; Sheng, Y.; Akroyd, J.; Martin, J.; Dreyer, J. A. H.; Yang, W.; Kraft, M. Internal structure of soot particles in a diffusion flame. *Carbon* **2019**, *141*, 635–642.

(21) Chang, Q.; Gao, R.; Gao, M.; Yu, G.; Wang, F. The structural evolution and fragmentation of coal-derived soot and carbon black during high-temperature air oxidation. *Combust. Flame* **2020**, *216*, 111–125.

(22) Morajkar, P. P.; Abdrabou, M. K.; Raj, A.; Elkadi, M.; Stephen, S.; Ibrahim Ali, M. Transmission of trace metals from fuels to soot particles: An ICP-MS and soot nanostructural disorder study using diesel and diesel/Karanja biodiesel blend. *Fuel* **2020**, *280*, 118631.

(23) Gleason, K.; Carbone, F.; Sumner, A. J.; Drollette, B. D.; Plata, D. L.; Gomez, A. Small aromatic hydrocarbons control the onset of soot nucleation. *Combust. Flame* **2021**, *223*, 398–406.

(24) Carbone, F.; Gleason, K.; Gomez, A. Probing gas-to-particle transition in a moderately sooting atmospheric pressure ethylene/air laminar premixed flame. Part I: gas phase and soot ensemble characterization. *Combust. Flame* **2017**, *181*, 315–328.

(25) Jacobson, R. S.; Korte, A. R.; Vertes, A.; Miller, J. H. The Molecular Composition of Soot. *Angew. Chem. Int. Ed.* **2020**, *59*, 4484–4490.

(26) Barone, A. C.; D'Alessio, A.; D'Anna, A. Morphological characterization of the early process of soot formation by atomic force microscopy. *Combust. Flame* **2003**, *132*, 181–187.

(27) Schulz, F.; Commodo, M.; Kaiser, K.; De Falco, G.; Minutolo, P.; Meyer, G.; D`anna, A.; Gross, L. Insights into incipient soot formation by atomic force microscopy. *Proc. Combust. Inst.* **2019**, *37*, 885–892.

(28) Commodo, M.; D'Anna, A.; De Falco, G.; Larciprete, R.; Minutolo, P. Illuminating the earliest stages of the soot formation by photoemission and Raman spectroscopy. *Combust. Flame* **2017**, *181*, 188–197.

(29) Commodo, M.; Kaiser, K.; De Falco, G.; Minutolo, P.; Schulz, F.; D'Anna, A.; Gross, L. On the early stages of soot formation: Molecular structure elucidation by high-resolution atomic force microscopy. *Combust. Flame* **2019**, *205*, 154–164.

(30) Parker, D. S. N.; Zhang, F.; Kim, Y. S.; Kaiser, Ralf. I.; Mebel, A. M. On the Formation of Resonantly Stabilized C5H3 Radicals—A Crossed Beam and Ab Initio Study of the Reaction of Ground State Carbon Atoms with Vinylacetylene. *J. Phys. Chem. A* **2011**, *115*, 593–601.

(31) Johansson, K. O.; Head-Gordon, M. P.; Schrader, P. E.; Wilson, K. R.; Michelsen, H. A. Resonance-stabilized hydrocarbon-radical chain reactions may explain soot inception and growth. *Science* **2018**, *361*, 997–1000.

(32) Gentile, F. S.; Picca, F.; De Falco, G.; Commodo, M.; Minutolo, P.; Causà, M.; D'Anna, A. Soot inception: A DFT study of σ and π dimerization of resonantly stabilized aromatic radicals. *Fuel* **2020**, *279*, 118491.

(33) Rundel, J. A.; Thomas, C. M.; Schrader, P. E.; Wilson, K. R.; Johansson, K. O.; Bambha, R. P.; Michelsen, H. A. Promotion of particle formation by resonance-stabilized radicals during hydrocarbon pyrolysis. *Combust. Flame* **2022**, *243*, 111942.

(34) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

(35) Ashraf, C.; van Duin, A. C. T. Extension of the ReaxFF Combustion Force Field toward Syngas Combustion and Initial Oxidation Kinetics. *J. Phys. Chem. A* **2017**, *121*, 1051–1068.

(36) Chenoweth, K.; van Duin, A. C. T.; Goddard, W. A. ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *J. Phys. Chem. A* **2008**, *112*, 1040–1053.

(37) Schuetz, C. A.; Frenklach, M. Nucleation of soot: Molecular dynamics simulations of pyrene dimerization. *Proc. Combust. Inst.* **2002**, *29*, 2307–2314.

(38) Han, S.; Li, X.; Nie, F.; Zheng, M.; Liu, X.; Guo, L. Revealing the Initial Chemistry of Soot Nanoparticle Formation by ReaxFF Molecular Dynamics Simulations. *Energy Fuels* **2017**, *31*, 8434–8444.

(39) Chen, C.; Jiang, X. Molecular dynamics simulation of soot formation during diesel combustion with oxygenated fuel addition. *Phys. Chem. Chem. Phys.* **2020**, *22*, 20829–20836.

(40) Pascazio, L.; Martin, J. W.; Bowal, K.; Akroyd, J.; Kraft, M. Exploring the internal structure of soot particles using nanoindentation: A reactive molecular dynamics study. *Combust. Flame* **2020**, *219*, 45–56.

(41) Ishiguro, T.; Takatori, Y.; Akihama, K. Microstructure of diesel soot particles probed by electron microscopy: First observation of inner core and outer shell. *Combust. Flame* **1997**, *108*, 231–234.

(42) Pathak, A. D.; Potphode, D.; Sharma, C. S. Graphitization induced structural transfor-

mation of candle soot carbon into carbon nano-onion as a functional anode for metal-ion batteries. *Mater. Adv.* **2022**, *3*, 3610–3619.

(43) Liu, Y.; Lv, G.; Fan, C.; Li, N.; Wang, X. Surface Functional Groups and Graphitization Degree of Soot in the Sooting History of Methane Premixed Flame. WCX™ 17: SAE World Congress Experience. 2017.

(44) Ruiz, M.; Villoria, R.; Millera, A.; Alzueta, M.; Bilbao, R. Influence of the temperature on the properties of the soot formed from C2H2 pyrolysis. *Chemical Engineering Journal* **2007**, *127*, 1–9.

(45) Frenklach, M.; Clary, D. W.; Gardiner, W. C.; Stein, S. E. Detailed kinetic modeling of soot formation in shock-tube pyrolysis of acetylene. *Symp. Combust.* **1985**, *20*, 887–901.

(46) Violi, A.; D'Anna, A.; D'Alessio, A. Modeling of particulate formation in combustion and pyrolysis. *Chem. Eng. Sci.* **1999**, *54*, 3433–3442.

(47) Vlasov, P. A.; Warnatz, J. Detailed kinetic modeling of soot formation in hydrocarbon pyrolysis behind shock waves. *Proc. Combust. Inst.* **2002**, *29*, 2335–2341.

(48) D'Anna, A.; Violi, A. Detailed Modeling of the Molecular Growth Process in Aromatic and Aliphatic Premixed Flames. *Energy Fuels* **2005**, *19*, 79–86.

(49) Violi, A. Modeling of soot particle inception in aromatic and aliphatic premixed flames. *Combust. Flame* **2004**, *139*, 279–287.

(50) Sharma, A.; Mukut, K. M.; Roy, S. P.; Goudeli, E. The coalescence of incipient soot clusters. *Carbon* **2021**, *180*, 215–225.

(51) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; In 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D. et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(52) Castro-Marcano, F.; Kamat, A. M.; Russo, M. F.; van Duin, A. C. T.; Mathews, J. P. Combustion of an Illinois No. 6 coal char simulated using an atomistic char representation and the ReaxFF reactive force field. *Combust. Flame* **2012**, *159*, 1272–1285.

(53) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637–649.

(54) Evans, D. J.; Holian, B. L. The Nose–Hoover thermostat. *J. Chem. Phys.* **1985**, *83*, 4069–4074.

(55) Mukut, K. M.; Sharma, A.; Goudeli, E.; Roy, S. P. A Closer Look into the Formation of Soot Particles: A Molecular Dynamics Study. 12th US National Combustion Meeting, College Station, TX, USA. 2021.

(56) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305–320.

(57) Mukut, K. M.; Roy, S.; Goudeli, E. Molecular arrangement and fringe identification and analysis from molecular dynamics (MAFIA-MD): A tool for analyzing the molecular structures formed during reactive molecular dynamics simulation of hydrocarbons. *Comput. Phys. Commun.* **2022**, *276*, 108325.

(58) Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* **2009**, *18*, 015012.

(59) Zhang, C.; Zhang, C.; Ma, Y.; Xue, X. Imaging the C black formation by acetylene pyrolysis with molecular reactive force field simulations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 11469–11480.

(60) Siegmann, K.; Sattler, K. Formation mechanism for polycyclic aromatic hydrocarbons in methane flames. *J. Chem. Phys.* **2000**, *112*, 698–709.

(61) Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137.

(62) van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.

(63) Mukut, K. M.; Ganguly, A.; Goudeli, E.; Kelesidis, G. A.; Roy, S. P. Physical, chemical and morphological evolution of incipient soot obtained from molecular dynamics simulation of acetylene pyrolysis. arXiv:2402.06460v1 (physics.atm-clus), submitted on 02/09/2024, 2024; `https://doi.org/10.48550/arXiv.2402.06460`, accessed 2024-04-19.

(64) Johansson, K. O.; El Gabaly, F.; Schrader, P. E.; Campbell, M. F.; Michelsen, H. A. Evolution of maturity levels of the particle surface and bulk during soot growth and oxidation in a flame. *Aerosol Sci. Technol.* **2017**, *51*, 1333–1344.

(65) Michelsen, H. A.; Campbell, M. F.; Johansson, K. O.; Tran, I. C.; Schrader, P. E.; Bambha, R. P.; Cenker, E.; Hammons, J. A.; Zhu, C.; Schaible, E. et al. Soot-particle core-shell and fractal structures from small-angle X-ray scattering measurements in a flame. *Carbon* **2022**, *196*, 440–456.

(66) Kholghy, M. R.; Veshkini, A.; Thomson, M. J. The core–shell internal nanostructure of soot – A criterion to model soot maturity. *Carbon* **2016**, *100*, 508–536.

(67) Kelesidis, G. A.; Crepaldi, P.; Pratsinis, S. E. Oxidation dynamics of soot or carbon black accounting for its core-shell structure and pore network. *Carbon* **2023**, 118764.

(68) DeCarlo, P. F.; Slowik, J. G.; Worsnop, D. R.; Davidovits, P.; Jimenez, J. L. Particle Morphology and Density Characterization by Combined Mobility and Aerodynamic Diameter Measurements. Part 1: Theory. *Aerosol Sci. Technol.* **2004**, *38*, 1185–1205.

(69) Theiler, J. Estimating fractal dimension. *J. Opt. Soc. Am. A, JOSAA* **1990**, *7*, 1055–1073.

(70) Forrest, S. R.; Witten, T. A., Jr. Long-range correlations in smoke-particle aggregates. *J. Phys. A: Math. Gen.* **1979**, *12*, L109.

A Symbols and nomenclature

a: A parameter in Eqn B.4

A: Surface area of a particle (Å^2)

c: A parameter in Eqn B.4

D_f: Atomic fractal dimension of a particle

$m_{p,i}$: Mass of i^{th} particle (kg)

M: Molar mass of a particle (kg/kmol)

M_p: Mass of a particle (kg)

$n_{\copyright}$: Number of cyclic carbon atoms at a specific location

$n_{¢}$: Number of non-cyclic carbon atoms at a specific location

N: Total number of atoms in the entire particle

$N_{\bigcirc}$: Number of rings in the entire particle

N_5: Number of 5-membered rings in the entire particle

N_6: Number of 6-membered rings in the entire particle

N_7: Number of 7-membered rings in the entire particle

N_C: Number of carbon atoms in the entire particle

N_H: Number of hydrogen atoms in the entire particle

$N_{\copyright}$: Number of cyclic carbon atoms in the entire particle

$N_{¢}$ Number of non-cyclic carbon atoms in the entire particle

r: Local radius (Å)

R_{eq}: Volume equivalent radius (Å)

R_g: Radius of gyration (Å)

ρ_s: Simulated density of a particle (kg/m^3)

ρ_e: Empirical density of a particle (kg/m^3)

ϱ: Local (actual) density (kg/m^3)

T: Temperature (K)

$\theta_{C/H}$: Local C/H ratio

$\Theta_{C/H}$: C/H ratio of the entire particle

V: Volume of a particle (Å^3)

w_C: Mass of a carbon atom (kg)

w_H: Mass of a hydrogen atom (kg)

Superscripts:

$*$: Denotes normalized value

$''$: Denotes per unit area value

B Expressions for physical properties of soot particles

The trajectory files obtained from RMD simulations contain coordinates of each atom with reference to a global reference frame. This coordinate information, along with the mass of each atom, is used to calculate the coordinates of the center of mass of each particle. The mass of a particle M_p is calculated by summing up the mass of the atoms in the cluster. The volume (V) is computed using MSMS[56] with a pore size of 1.5 Å. The volume equivalent radius of a particle with volume V is calculated via Eqn. B.1.

$$R_{eq} = \left(\frac{3V}{4\pi}\right)^{1/3} \tag{B.1}$$

The radius of gyration (R_g) is calculated following the standard definition using Eqn. B.2.

$$R_g = \sqrt{\frac{\sum_{i=1}^{N} m_{p,i} r_i^2}{\sum_{i=1}^{N} m_{p,i}}}, \tag{B.2}$$

where r_i is the distance of the i^{th} atom from the center of mass, $m_{p,i}$ is the mass of individual atoms, and N is the total number of atoms in the cluster.

The simulated density (ρ_s) is calculated using the particle mass (M_p) and volume (V) of the incipient particle using Eqn. B.3.

$$\rho_s = \frac{M_p}{V} \tag{B.3}$$

Empirical (bulk) density[64,68] of an incipient particle is calculated using Eqn. B.4.

$$\rho_e \quad = (0.260884 a^2 c)^{-1} \left(\frac{w_C \Theta_{C/H} + w_H}{\Theta_{C/H} + 1} \right), \tag{B.4}$$

where w_C and w_H are the molar masses of carbon and hydrogen atoms, a is the length of the graphite unite cell in the basal plane, c is the interlayer spacing in Angstroms, and $\Theta_{C/H}$ represents the carbon to hydrogen ratio of the cluster. More details can be found in.[64,68]

The atomic fractal dimension (D_f), following the approach used in,[50] is calculated using the sandbox method[69,70] using Eqn. B.5.

$$D_f = \frac{\log M_p(r)}{\log r}, \tag{B.5}$$

where $M_p(r)$ is the mass of atoms in the cluster as a function of radial distance from the center of mass. Please note that this "atomic" fractal dimension is for a single incipient particle and differs from the traditional fractal dimension used in aggregate characterization.[50]

The molar mass (M) is calculated by summing the atomic masses of each element in a compound, multiplied by the number of atoms of that element in the molecular formula.

For an incipient particle with the molecular formula C_mH_n, where m and n represent the number of carbon and hydrogen atoms respectively, the molar mass (M) can be calculated using the atomic masses of carbon (12 kg/kmol) and hydrogen (1 kg/kmol). Therefore, the molar mass (M) of the particle is given by Eqn. B.6.

$$M = 12m + n \tag{B.6}$$

C Physicochemical data used and analyzed in this study

C.1 Feature set

The feature set used in this study for each particle includes the following

1. Temperature (T)
2. Number of carbon atoms (N_C)
3. Number of hydrogen atoms (N_H)
4. Number of atoms (N)
5. Molar mass (M)
6. C/H ratio ($\Theta_{C/H}$)
7. Radius of gyration (R_g)
8. Atomic fractal dimension (D_f)
9. Simulated density (ρ_s)
10. Empirical density (ρ_e) (also referred to as the bulk density in literature)

11. Total number of cyclic structures ($N_{\bigcirc}$)

12. Fraction of cyclic carbon atoms ($^{N_{\textcircled{c}}}/_{N_C}$)

13. Fraction of 5-member rings ($^{N_5}/_{N_{\bigcirc}}$)

14. Fraction of 6-member rings ($^{N_6}/_{N_{\bigcirc}}$)

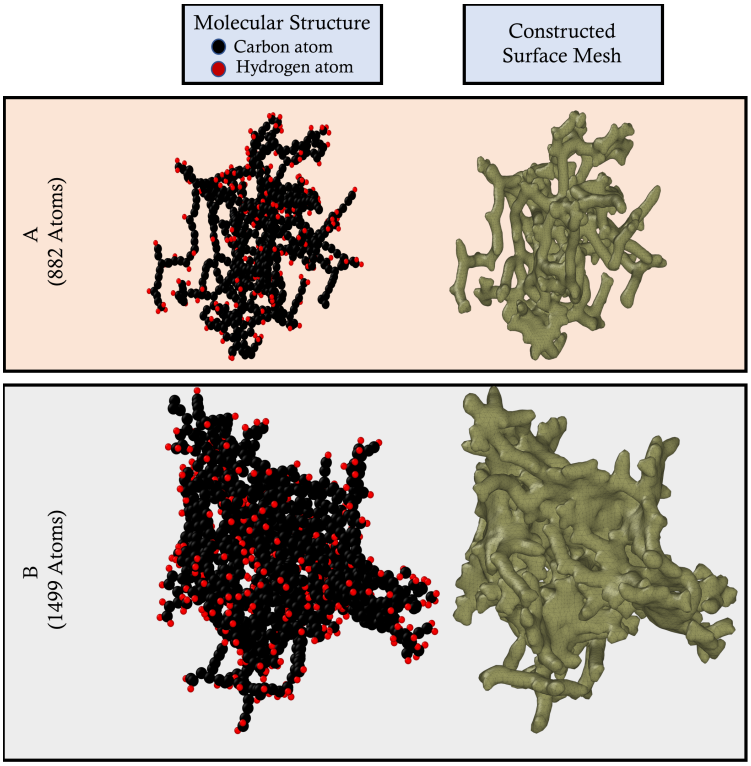
15. Fraction of 7-member rings ($^{N_7}/_{N_{\bigcirc}}$)

16. Surface area (A)

17. Volume (V)

18. Area to volume ratio ($^{A}/_{V}$)

C.2 Sample data



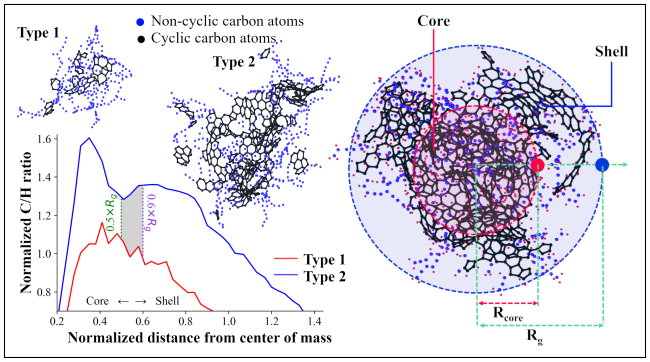
A sample set of the attributes of nascent soot particles investigated in the present study

Identifier	A	B
Temperature (K)	1500	1500
Atoms	882	1499
Number of carbon	659	1083
Number of hydrogen	223	416
Molar mass (kg/kmol)	8139.8	13426.9
C/H ratio	2.955	2.603
Radius of gyration (Å)	13.473	15.076
Atomic fractal dimension	2.440	2.367
Simulated density (g/cm^3)	1.380	1.548
Empirical density (g/cm^3)	1.670	1.621
Fraction of cyclic carbon	0.114	0.554
Fraction of non-cyclic carbon	0.886	0.446
Fraction of 5-membered ring	0.529	0.195
Fraction of 6-membered ring	0.235	0.569
Fraction of 7-membered ring	0.235	0.236
Total number of cyclic structures	17	174
Surface area (Å^2)	5273.3	6754.7
Volume (Å^3)	9791.6	14403.0
Area to Volume Ratio (1/Å)	0.539	0.469

Figure 11: Two sample soot particles and their attributes investigated in this study

Figure 11 shows two sample soot clusters and their properties as examples. These clusters, labeled as `A` and `B`, were extracted from a simulation at 1500 K at two different times. The left side shows the molecular structure of the particle and the three-dimensional volumetric representation by constructing a surface mesh using `OVITO`[58] (this is what the incipient particles would actually look like). The physicochemical properties of these particles, as analyzed in this work for classification via machine-learning, are tabulated on the right side of the figure.

TOC Graphic


Type 1
Non-cyclic carbon atoms
Cyclic carbon atoms
Type 2
Core
Shell
Normalized C/H ratio
0.8
1.0
1.2
1.4
1.6
0.3×R_g
0.6×R_g
Core ← → Shell
Type 1
Type 2
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Normalized distance from center of mass
R_{core}
R_g