HOMO–LUMO Gaps and Molecular Structures of Polycyclic Aromatic Hydrocarbons in Soot Formation

Yabei Xu[1], Qingzhao Chu[1], Dongping Chen[1]* and Andrés Fuentes[2]

[1]*State Key Lab of Explosion Science and Technology, Beijing Institute of Technology, Beijing, China,* [2]*Departamento de Industrias, Universidad Técnica Federico Santa María, Valparaíso, Chile*

A large number of PAH molecules is collected from recent literature. The HOMO-LUMO gap value of PAHs was computed at the level of B3LYP/6-311+G (d,p). The gap values lie in the range of 0.64–6.59 eV. It is found that the gap values of all PAH molecules exhibit a size dependency to some extent. However, the gap values may show a big variation even at the same size due to the complexity in the molecular structure. All collected PAHs are further classified into seven groups according to features in the structures, including the types of functional groups and the molecular planarity. The impact of functional groups, including –OH, –CHO, –COOH, =O, –O– and –C_nH_m on the bandgap is discussed in detail. The substitution of ketone group has the greatest reduction on the HOMO-LUMO gap of PAH molecules. Besides functional groups, we found that both local structure and the position of five-member rings make critical impacts on the bandgap *via* a detailed analysis of featured PAHs with unexpected low and high gap values. Among all these factors, the five-member rings forming nonplanar PAHs impact the gap most. Furthermore, we developed a machine learning model to predict the HOMO-LUMO gaps of PAHs, and the average absolute error is only 0.19 eV compared with the DFT calculations. The excellent performance of the machine learning model provides us an accurate and efficient way to explore the band information of PAHs in soot formation.

Keywords: PAH, HOMO-LUMO gap, functional group, five-member ring, machine learning

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) generated from the incomplete combustion of hydrocarbon fuels are accepted as the precursors of soot. The identification of PAHs and their structures is critical to interpret their growth mechanism, which is the basis for the reduction of the soot emission (Wang and Chung, 2019). With the application of novel measurement methods, recent researchers have made important progress in the investigation of the key process in soot formation by identifying the potential intermediates. Johansson and his coworkers proposed a radical-assisted PAH growth mechanism supported by the aerosol mass spectrometry measurements (Johansson et al., 2018). Schulz et al. (Schulz et al., 2019) used the state-of-art Atomic Force Microscopy (AFM) to identify the detailed configuration of large PAHs (>300 amu). Although the abundance and relevance of these identified PAHs to soot formation is unknown, it was the first time that the configurations of large PAHs are confirmed in measurements. Commodo (Commodo et al., 2019) further studied the early formation stage of different soot samples by AFM/SEM, providing the direct evidence of the formation of cross-linked structures. Adamson (Adamson et al., 2018) detected

aliphatic bridged multi-core PAHs by atmospheric sampling high-resolution tandem mass spectrometry, revealing the presence of alkylated aromatic compounds. It is found that the main component of soot, *e.g.*, PAHs, are more complicated than we expected and they are affected by many factors, such as size, functional group, cross-linking, and aliphatic chains (Commodo et al., 2019; Schulz et al., 2019; Gentile et al., 2020; Wang et al., 2021). In the formation process of soot, PAHs containing 30–40 carbon atoms are the main components of soot (Elvati and Violi, 2013; Adkins and Miller, 2015; Johansson et al., 2016; Adkins et al., 2017; Kholghy et al., 2018). Lots of aliphatic, aromatic and oxygenated functionalities are substituted on the surface of PAHs, among which the common oxygenated functional groups include hydroxyl (–OH), formyl (–CHO) and carboxyl (–COOH) (Öktem et al., 2005; Cain et al., 2010). Some special PAHs also contain different types of functional groups (–O–, =O) and five-membered rings. Note that the structures of PAHs containing five-membered rings are mostly nonplanar (Wang et al., 2021).

PAH molecules and soot particles have unique optical band gap (OBG) and electronic properties (Chen and Wang, 2019a; Li et al., 2020). It is feasible, in theory, to correlate PAH configurations with OBGs. The studies of OBG of soot starts from the pioneering work of Minutolo (Minutolo et al., 1996). The OBG of soot samples was found to be proportional to the H/C ratio. In later works, Adkins and Miller analyzed the optical band gap of soot in diffusion flames by laser extinction measurements combined with density function theory (DFT) calculations (Adkins and Miller, 2015). They determined that the size of polycyclic aromatic hydrocarbons is 10–20 aromatic rings. As the molecular PAH always has a well-defined band structure, it is straightforward to estimate the gap value using computational methods (Adkins and Miller, 2015; Chen and Wang, 2019a). Thus, it is possible to approximate the PAH size in soot particle based on the values of the optical band gap. A successful example can be seen in a recent work of Menon et al. (Menon et al., 2019). They found that the computed gap values from DFT results can well reproduce the experimental OBGs measured by UV-Visible absorption spectrum. Not only the PAH size but also the particle size of soot impacts the magnitude of band gaps. In another recent work, soot particles were found to exhibit quantum dot behaviors that the OBG is inversely proportional to the particle size. (Liu et al., 2019). This indicates that the impact of particle size on the OBG must be excluded when interpreting experimental measurements for PAH identification. The DFT calculations were further conducted to evaluate the impact of surface functionalization on soot particle. Chen and Wang explored the impact of different types of functional groups (–OH, –CH₃, –CHO, –COOH) on the computed HOMO-LUMO gaps of PAH monomers, showing that –CHO group substitution have the largest influence factor on gap values. The PAH clusters with surface formyl groups were further built to evaluate the impact of surface functionalization together with the effect of particle size (Chen and Wang, 2019a; Li et al., 2020). Previous works had made significant contribution to understand the impact of particle size, PAH size and even surface

functionalization on the OBG, however, it is still lack of a systematic work to study the band structures and the corresponding gaps of all the important PAHs in soot formations, especially those large PAHs extracted from AFM measurements (Commodo et al., 2019; Schulz et al., 2019).

In this work, we collected hundreds of reliable PAHs relevant with the formation of soot from recent literature. The selected PAHs were classified into seven groups according to the features in the structures. The optimized structures and bandgaps of PAHs were obtained from DFT calculations. The effect of functional groups and structure planarity on the bandgaps were discussed from the analysis of the orbital structures in a set of unique PAHs. We further developed a machine learning model to evaluate the bandgaps from PAH structures directly. The accuracy in the predictions was examined by comparing with DFT calculations.

COMPUTATIONAL METHOD

To build a comprehensive data set of PAHs, we exanimated a large number of publications about soot particles mostly in the last ten years, but potential structures with S and N elements are excluded to simplify our analysis in the bandgap. In this work, we selected 323 PAHs ranging from 6 up to 96 carbon numbers (Lafleur et al., 1993; Elvati and Violi, 2013; Kislov et al., 2013; Lowe et al., 2015; Johansson et al., 2016; Zhang et al., 2016; Johansson et al., 2017; Adamson et al., 2018; Kholghy et al., 2018; Li et al., 2018; Commodo et al., 2019; Elvati et al., 2019; Giaccai and Miller, 2019; Kozliak et al., 2019; Schulz et al., 2019; Wang et al., 2019; Zhang, 2019; Frenklach and Mebel, 2020; Gavilan Marin et al., 2020; Gentile et al., 2020; Leon et al., 2020; Michelsen, 2020; Pascazio et al., 2020; Saldinger et al., 2020; Zhao et al., 2020; Chen et al., 2021; Shi et al., 2021; Wang et al., 2021). The molecular structures are all included in **Supplementary Table S1** (see **Supplementary Material**). We performed DFT calculations to optimize the geometry of each molecule using the B3LYP method with the 6-311+G (d,p) basis set. An empirical dispersion correction was also included. All DFT calculations were carried out using Gaussian 09 (Frisch et al., 2009). The HOMO-LUMO gap value is computed from the energy difference between HOMO and LUMO-Kohn-Sham (KS) orbitals. The predicted HOMO-LUMO gap using B3LYP/6-311+G (d,p) is consistent with those from the B3LYP/6-31G(d) method (**Supplementary Table S2**), which is commonly used in previous works (Li et al., 2020). We built a structure and bandgap database including all the collected PAHs from literature. The selected PAHs are divided into seven groups according to the substituted groups and molecular structures. The detail of each group is discussed below. Based on the database, we further developed a machine learning model that can predict the HOMO-LUMO gap value of PAHs using molecular structures. The coordinates are preprocessed by Smooth Overlap of Atomic Positions (SOAP) method (De et al., 2016) to guarantee invariance of geometry under translation, rotation, and permutation among identical particles. The SOAP method provides a robust descriptor that encodes regions of atomic

FIGURE 1 | (A) HOMO energies of all PAH molecules. **(B)** LUMO energies of all PAH molecules. **(C)** HOMO-LUMO gap of all PAH molecules. The line is fitted using $E_{H-L} = E_{H-L}^{\infty} + am^{-2/3}$, where E_{H-L}^{∞} is the gap value for a bulk counterpart, m refer to the PAHs mass and a is a fitting parameter. Each color represents the gap information of a specific group including "Planar", "Nonplanar", "–OH, –CHO and –COOH", "=O", "–O–", "–C$_n$H$_m$" and "Linear".

geometries by using a local expansion of a Gaussian smeared atomic density with orthonormal functions based on spherical harmonics and radial basis functions.

RESULTS AND DISCUSSION

The relationship between the number of carbon atoms in PAHs and HOMO-LUMO gap values is presented in **Figure 1**. **Figures 1A,B** include the HOMO and LUMO energy as a function of PAH size, respectively. The HOMO energy increases as the increase of PAH size, while the LUMO energy shows an opposite trend. The overall effect reduces the gap values with an increase in the PAH size (**Figure 1C**). This finding is consistent with previous works that a bigger PAH always has a lower gap value (Miller et al., 2013; Adkins et al., 2017; Chen and Wang, 2019a). The gap values lie in the range of 0.64–6.59 eV. The gap values cluster in the range of 20–50 carbon atoms, and the variation of the same sized PAHs in the gap values can be as large as ~2 eV. Such big variation is correlated with the complicity in the PAH structures, which will be explored in later contents.

The number of PAHs larger than fifty carbons is rather limited compared to its smaller counterpart, and this can be attributed to the technical difficulty in the sampling method. But the recent advances in AFM allows researchers to identify large PAHs (Commodo et al., 2019; Chen et al., 2021; Wang et al., 2021). In a previous work (Chen and Wang, 2019b), we know that the bandgap of PAHs follows a dependence of $m^{-2/3}$, where m is the PAH mass, due to the quantum confinement effect. All data in **Figure 1C** is fitted with the same correlation (black line). The fitted line cannot capture all the data due to the complicity in the selected PAH structures. Some PAHs show unexpected low and high gap values, and the underlying reason will be examined below.

We group PAHs into seven groups in order to better analyze the gap values of PAHs. According to the types of functional groups substituted on the surface of PAHs, four groups are included as "–OH, –CHO, –COOH", "=O", "–O–" and "–C$_n$H$_m$". The groups of "Planar" and "Nonplanar" refer to planar and nonplanar PAH with only five- and six-member rings, respectively. Besides above groups, the PAHs with straight carbon rings are regarded as the group of "Linear".

TABLE 1 | Key statistics of PAHs in different groups.

Group	Planar	Nonplanar	–OH, –CHO, –COOH[a]	=O	–O–	–C_nH_m	Linear	All data
Bandgap[b]	0.64–6.59	0.75–5.1	2.07–4.42	2.07–3.68	2.38–4.7	0.67–6.28	1.45–5.97	0.64–6.59
Carbon number[c]	6–96	18–64	12–36	13–33	10–50	6–71	10–26	6–96
Number of Data[d]	145	14	15	9	17	121	9	323

[a]In this group, five PAHs molecule contain a "–O–" group, and two PAH molecules contain a "=O" group.
[b]The value range of HOMO-LUMO gap.
[c]The number of carbon atoms in PAHs.
[d]The number of PAH molecules in each group.
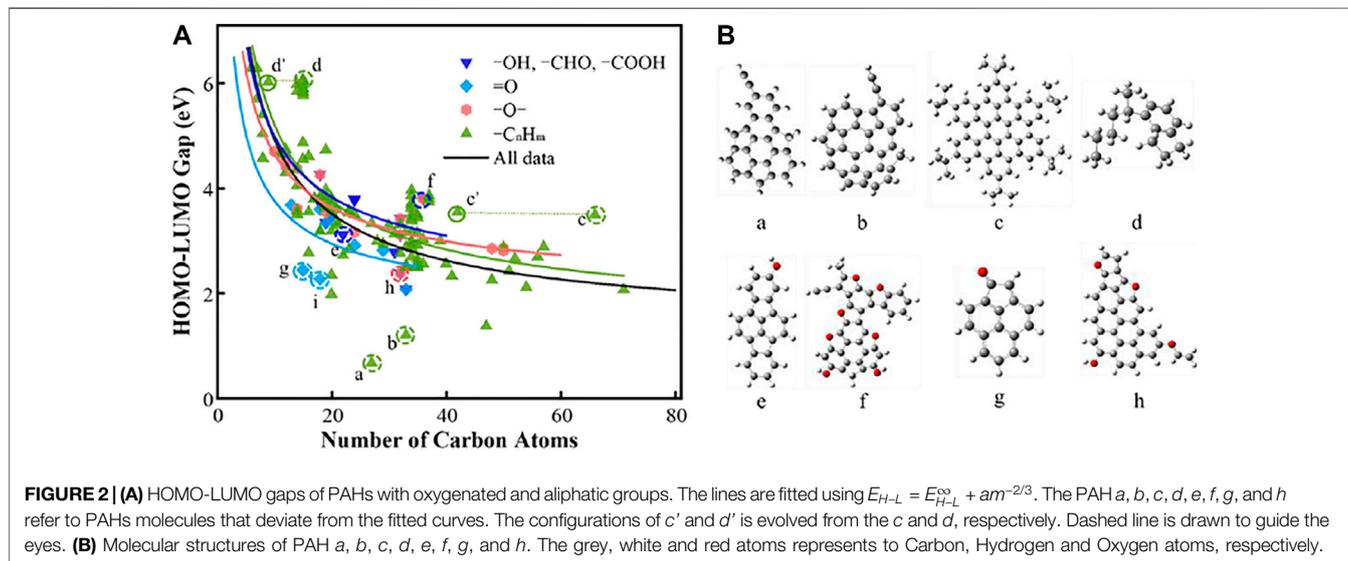
FIGURE 2 | (A) HOMO-LUMO gaps of PAHs with oxygenated and aliphatic groups. The lines are fitted using $E_{H-L} = E_{H-L}^{\infty} + am^{-2/3}$. The PAH a, b, c, d, e, f, g, and h refer to PAHs molecules that deviate from the fitted curves. The configurations of c' and d' is evolved from the c and d, respectively. Dashed line is drawn to guide the eyes. **(B)** Molecular structures of PAH a, b, c, d, e, f, g, and h. The grey, white and red atoms represents to Carbon, Hydrogen and Oxygen atoms, respectively.

The relevant statistics of each PAH group are shown in **Table 1**. It should be noted that the PAHs with –OH, –CHO, and –COOH groups are classified into one group due to the limited number of samples in all data. Some PAH molecules are substituted by multiple functional groups; for example, $C_{32}H_{14}O_4$ (**Supplementary Table S1**) contains –OH, –C_2H_3 and three –O– groups (Elvati et al., 2019), and this PAH is classified into two groups of "–OH, –CHO, –COOH" and "–O–".

We first analyze the PAH molecules substituted with functional groups including –OH, –CHO, –COOH, =O, –O– and –C_nH_m (**Figure 2A**). From previous works (Li et al., 2020) the optical band gap of soot exhibits quantum confinement effects (Chen and Wang, 2019a), and the gap value is related to R^{-2} or $m^{-2/3}$, where R and m are the radius and mass of the soot particles, respectively. In this work, we adapt the formula $E_{H-L} = E_{H-L}^{\infty} + am^{-2/3}$ to fit the gap values of PAHs, where a, E_{H-L} and E_{H-L}^{∞} are the fitting coefficient, the gap value of PAHs with mass m, and the bulk value, respectively. The range of gap values in the group of "–OH, –CHO, –COOH", "–O–" and "–C_nH_m" is 2.07–4.42, 2.38–4.7, and 0.67–6.28 eV, respectively. The range of gap values in the first two groups is close, while the PAHs in the group of "–C_nH_m" exhibits a large variation. The minimum values in the group of "–C_nH_m" reach 0.67 eV, and this is attributed to the unique configurations (a), which will be

analyzed in the later section. The fitted curves of these three groups locate all above the fitting curve of all PAHs, suggesting that the PAHs in these groups follow the overall trend. The bulk gap values (e.g., E_{H-L}^{∞}) of the "–OH, –CHO, –COOH", "–O–" and "–C_nH_m" group are 1.84, 1.89, and 1.26 eV, respectively. The corresponding value is 1.1 eV considering all selected PAHs. The PAH molecules substituted with ketone groups show smaller gap values comparing with other three groups. The gap values are comprised in the range of 2.07–3.68 eV. However, the scattering data in **Figure 2A** indicates that unique structures in each group have gap values far away from the others cases. To address this issue, we further identify eight PAHs as the featured species to highlight the abnormal gap values (**Figure 2B**).

Prior to analyze the featured species, we shall explore the impact of the substitution for the groups of "–OH, –CHO, –COOH", "O–", "=O" and "–C_nH_m". A four-ring PAH (e.g., benz(a)anthracene) (Giaccai and Miller, 2019) is selected as a base configuration to illustrate the impact of different groups. Note that all the substituted PAHs presented in **Figure 3** are included in the data of **Figure 1**. In **Figure 3A**, we can find that all substitution lowers the gap values for all groups; the impact of –CH_3, –OH, –COOH, –O–, –CHO, =O groups on the HOMO-LUMO gap values are 0.001, 0.03, 0.04, 0.13, 0.39, 1.46 eV, respectively. Among all the functional groups, the ketone
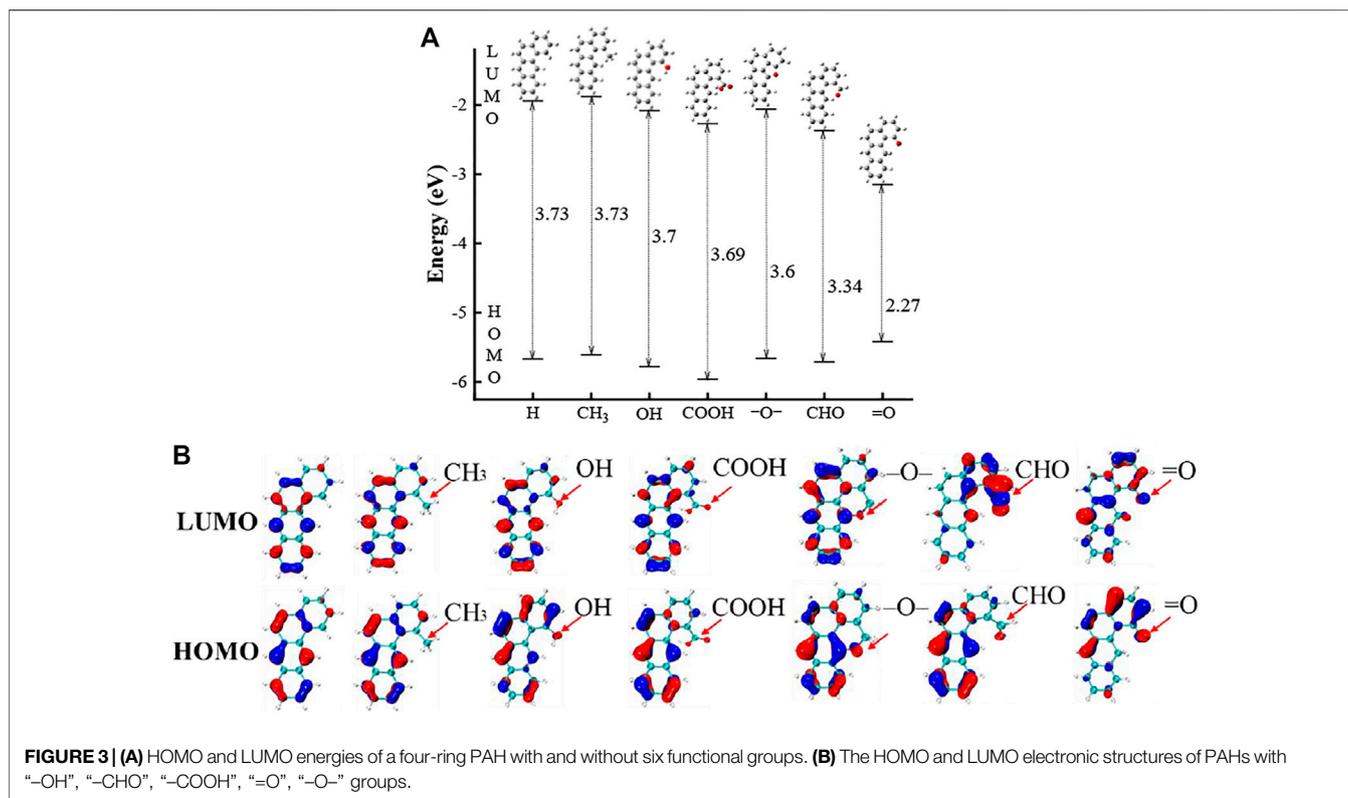
FIGURE 3 | **(A)** HOMO and LUMO energies of a four-ring PAH with and without six functional groups. **(B)** The HOMO and LUMO electronic structures of PAHs with "–OH", "–CHO", "–COOH", "=O", "–O–" groups.
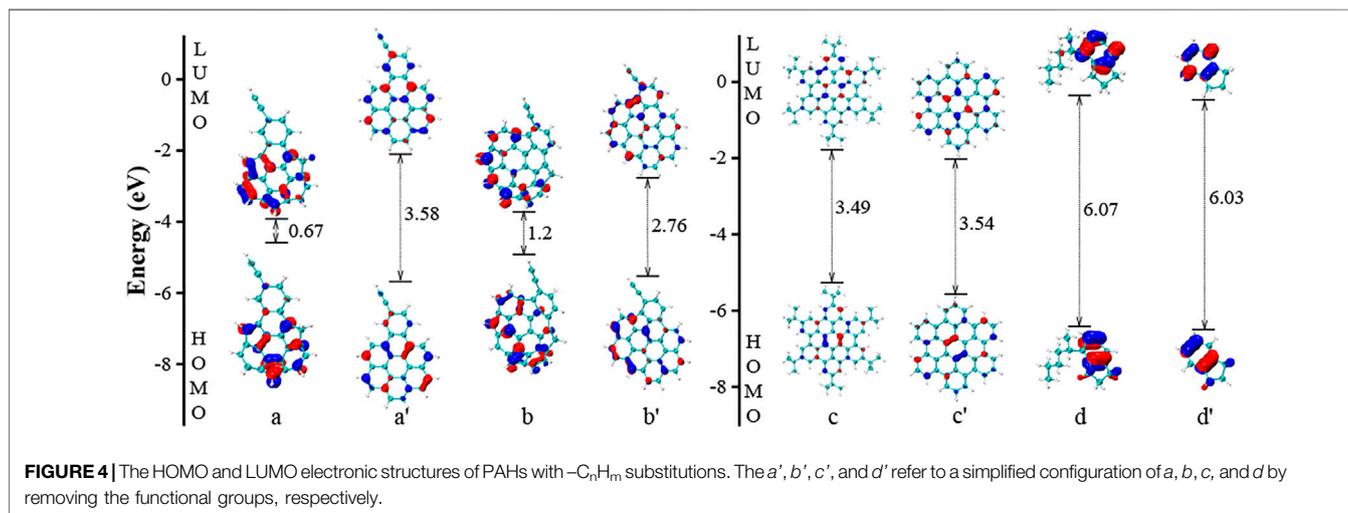
FIGURE 4 | The HOMO and LUMO electronic structures of PAHs with $-C_nH_m$ substitutions. The a', b', c', and d' refer to a simplified configuration of a, b, c, and d by removing the functional groups, respectively.

group substitution has the greatest reduction on the HOMO-LUMO gap value, while the –CH₃, –OH and –COOH substitution has a relatively small influence on the gap value. The finding here is consistent with the previous works (Giaccai and Miller, 2019). From the electronic structure diagrams in **Figure 3B**, the substitution of –CH₃, –OH and –COOH has very limited influence on the electronic structures of HOMO and LUMO. The –CHO group has an obvious bonding effect on the LUMO contributed by the substituted carbon atom and

neighbors, but the impact on the HOMO structure is limited. The ketone group induces a clear impact on the closest aromatic rings lowering the LUMO energy significantly. We can also identify its impact on the HOMO orbital, but the change in the HOMO energy is minor. The overall effect of "=O" substitution on the LUMO and HOMO orbitals causes a large reduction in the HOMO-LUMO gap.

Now, we shall examine the featured species in **Figure 2B** (a, b, c, d, e, f, g, and h). Here, simplified configurations are built by

FIGURE 5 | The HOMO and LUMO electronic structures of PAHs with –OH substitutions. The e' and e'' are two simplified configurations of e by removing five-member rings and –OH functional group, respectively. The furan rings in f is replaced by six-member ring, and this configuration is termed as f'. Further replacing five-member ring in f', we have the configuration of f''.
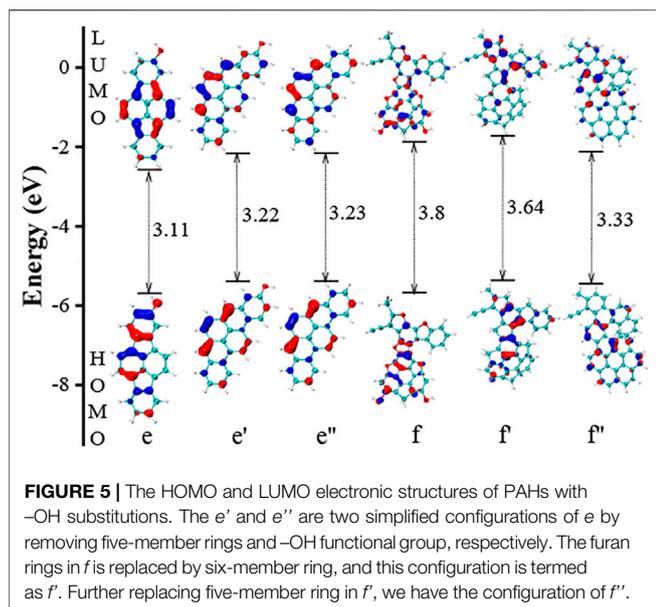
removing the features in the eight species. This strategy allows us to directly evaluate the key factors in each case. The featured species a, b, c, and d are taken from the group of "–C_nH_m". In **Figure 4**, the two PAHs molecules, e.g., a and b, have aliphatic substitution (–C_2H group) and five-member rings. Both structures are bent into a "bowl" shape due to the five-member rings on one side, and the HOMO-LUMO gap values are 0.67 and 1.2 eV, respectively. We learn from **Figure 3** that aliphatic substitution has a very weak effect on the electronic structures of HOMO and LUMO. Thus, we only consider the influence of the five-member ring here for a and b. Two simplified PAH molecules a' and b' are built by replacing five-member rings with six-member rings in a and b, respectively and the two molecules a' and b' are both planar structures. The gap value of a' is much larger than that of a. The same trend can be found in the comparison between b and b'. Clearly, the five-member ring has an obvious influence on the electronic structures of HOMO and LUMO, and the overall effect on the HOMO-LUMO gap values are 2.91 and 1.56 eV, respectively. Such effect can be also viewed as the effect of non-planarity in the PAH structures. The molecules c and d have multiple or a long –C_nH_m aliphatic groups. After removing the aliphatic groups, the molecules are marked as c' and d', which have almost the same gap values as the counterparts, because the aromatic planes dominate the HOMO and LUMO energies. However, due to the reduction of carbon atoms in the molecular structure, the c' and d' approach to the fitting curve, as highlighted by the dotted line in **Figure 2A**.

In **Figure 5**, the two five-member rings in molecule e is replaced with six-carbon rings, labeled as e'. Such modification in the structure causes an obvious impact on the electronic structures of HOMO and LUMO, increasing the HOMO-LUMO gap value by 0.11 eV. Further removing the group of "–OH", the electronic structures of HOMO and LUMO of e'' is almost unchanged. Therefore, the key factor in molecule e lies in the five-member rings. Unlike molecule a and b, the two five-

member rings do not bend the PAH into a nonplanar one, and its impact on gap values is very limited. In the case of f, we built a simplified configuration of f' by removing two –OH groups and replacing five oxygen atoms with carbon atoms. The reduction of gap values is 0.16 eV comparing f and f'. Further replacing the five-member rings with six-member rings, we build a molecule f'', and the reduction in gap values is increased to 0.47 eV. This explains the unexpected large gap value in f.

In **Figure 6**, the ketone group in molecule g is replaced by a six-member rings, and we term the configuration as g'. Comparing the LUMO molecular orbital, the bonding effect can be identified from g, which is consistent with the finding in **Figure 3B**. Such effect causes a big reduction in the LUMO energy, and in turn, the gap value is reduced by 1.37 eV accordingly. All the functional groups in h are removed, and we view this configuration as h'. A moderate increase in the gap value is observed.

The relationship between the HOMO-LUMO gaps and PAH planarity is shown in **Figure 7**. The planar and nonplanar PAHs discussed in this section excludes those with any functional groups discussed above. The fitting curve of planar PAHs is very close to the overall fitting curve, and the range of HOMO-LUMO gap value is 0.64–6.59 eV. However, the nonplanar PAH shows very different trend compared to the planar ones. In the selected PAHs, we have 14 ones classified as the "nonplanar" ones. More importantly, the gap values scatter in **Figure 7A**, indicating that the large variation exist in the nonplanar PAHs. Again, we selected four PAHs molecules j, k, l, and m marked in the figure as the target to explain the results behind the unexpected low gap values. The detailed structures are shown in **Figure 7B**.

The selected HOMO and LUMO electronic orbitals of j and k are shown in **Figure 8**. These two molecules are nonplanar ones. The molecule j contains five five-member rings, and the overall structure is a "bowl" shape. The band gap is 1.08 eV. The

FIGURE 6 | The HOMO and LUMO electronic structures of PAHs with ketone and furan groups. The g' is the simplified PAH of g without the ketone group. The h' is made from the configuration of h by replacing the furan ring with six-member rings and removing the –OH group.

FIGURE 7 | **(A)** HOMO-LUMO gaps of PAH molecules with planar and nonplanar structures. The curves presented as solid line are obtained fitting with $E_{H-L} = E_{H-L}^{\infty} + am^{-2/3}$, where m refer to the PAHs mass and a is a fitting parameter. Molecules j, k, l, and m indicates a few representative PAH molecules away from the fitted curves. **(B)** Molecular structures of j, k, l, and m.

FIGURE 8 | The HOMO and LUMO electronic structures of selected nonplanar j and k. The j' and k' is made from replacing five-member rings in j and k by six-member rings, respectively.

FIGURE 9 | The HOMO and LUMO electronic structures of two planar PAHs (l and m). The l' is a coronene having two more hydrogens compared to l. The m' is made from replacing two five-member rings with six-member rings in m. The m'' is made from replacing all five-member rings with six-member rings in m.

molecule j' is made by replacing all five-member ring in j with six-member rings, and it becomes a planar PAH. Similar as the molecule a and b in **Figure 4**, the non-planarity significantly impacts the gap values, changing from 1.08 to 2.15 eV in **Figure 8**. The non-planarity in molecule k also lower the band gap by 2.25 eV comparing with the molecule k'. **Figure 9** presents the HOMO and LUMO electronic structures of two selected planar PAHs (l and m). The molecule l has a unique feature that two of its edge hydrogens are missing. This feature leaves two reactive sites impacting HOMO and LUMO orbitals, and the overall band gap is only 0.64. Compared to l', the reduction caused by the two missing hydrogens is 3.36 eV. The m' is made from replacing two five-member rings with six-member rings in m, and the

corresponding change in the gap value is 0.23 eV. However, further replacing the other two five-member rings in m, the HOMO-LUMO gap value is reduced by 0.95 eV, which is much larger than the change from m to m'. This finding highlights that the effect of five-member rings depends on their position. Overall, the factors such as functional groups, local structure, and the position of five-member rings all contribute to the band gap to some extent. Among all these factors, the five-member rings leading non-planar PAHs impact the gap most.

In this work, two types of PAH fall into the group of "Linear"; the ones with benzene rings connected by each other (n, q, and r) or by a single C-C bond (o and p). **Figure 10** illustrates the

FIGURE 10 | HOMO-LUMO gaps of linear PAH molecules. The lines are fitted using $E_{H-L} = E_{H-L}^{\infty} + am^{-2/3}$, where m refer to the PAHs mass and a is a fitting parameter. The molecule n, o, p, q, and r indicate the representative PAHs in the group of "Linear".
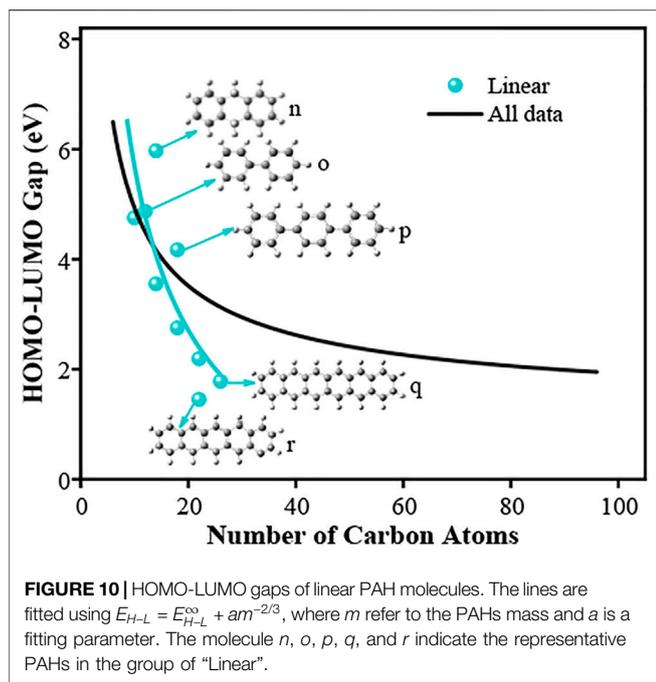
relationship between the HOMO-LUMO gap and the carbon number of linear PAHs. The overall trend follows the same as other cases that the band gap values decreases rapidly with the increase in the carbon number. However, the reduction is more significant compared to other cases. The bulk gap value (E_{H-L}^{∞}) extracted from the fitted expression is a negative value as -2.36 eV, which have no physical meaning.

The above analysis of each group proves that the band gap is a unique molecular property depending on its local structures. We further develop a machine learning model using the above data to predict the gap information from an arbitrary structure.

The atomic coordinates are chosen as the features to train the machine learning model. The coordinates are preprocessed by Smooth Overlap of Atomic Positions (SOAP) method (De et al., 2016) to guarantee the invariance of geometry under translation, rotation, and permutation among identical particles. The SOAP method could not only describe the numbers and types of atoms in PAH molecules, but also records the connectivity of atoms.

In the training section, we use the 10-fold cross-validation linear regression, where the full datasets are divided into 10 equal sized subsets. One of the subsets is treated as the test set for model validation, and the remaining 9 subsets are used as training sets to train the model. As **Figure 11A** shows, the predictions from the model agrees well with the DFT data; the mean absolute error (MAE) is 0.19 eV and the coefficient of determination values (R^2) is 0.96. The insert in **Figure 11A** also shows that the MSE evolution of training set and test set; the learning curve decline gradually with the increase of the data size, supporting that the problem of overfitting does not occur. We further evaluate the quality of the prediction using the machine learning model on the selected PAHs including planar, nonplanar, and linear ones (**Supplementary Table S3**). In **Figure 11B**, the results show that the model gives a very good prediction for planar and linear ones with a MAE smaller than 0.01 eV. In the case of nonplanar PAHs, the MAE is a bit higher as 0.11 eV. Considering the complicity in the molecular structures of nonplanar PAHs, the quality of the prediction is excellent. These results suggest that the current machine learning model is quite accurate and efficient. Also, the computational cost of our machine learning model is several order of magnitudes lower than the DFT calculations. Therefore, the present machine-learning model provide a good tool to predict the HOMO-LUMO gaps of PAH molecules with good accuracy and high computational efficiency.

FIGURE 11 | (A) HOMO-LUMO gaps of 323 PAHs predicted by the machine learning model and DFT calculations. The inset indicates the convergence of learning curve for the 10-fold cross-validation split of the data. **(B)** The average gap values of planar, nonplanar and linear PAHs calculated from DFT method and the machine learning model. The detail of PAH is included in **Supplementary Table S3–S5**.

CONCLUSION

In this work, we used DFT method to calculate the HOMO-LUMO gaps of PAHs in a comprehensive data set collected from recent publications. The range of gap values lies in 0.64–6.59 eV for 323 PAHs in the data set. The HOMO energy increases with the increase in the PAH size, but the LUMO energy shows an opposite trend. The overall effect reduces the gap values. However, gap values may vary several eV at the same size due to the differences in the molecular structure. To address this issue, seven groups are identified from the PAHs, including "–OH, –CHO, –COOH", "=O", "–O–", "–C_nH_m", "Planar", "Nonplanar" and "Linear". We explore the impact of –CH_3, –OH, –COOH, –O–, –CHO, =O groups on the HOMO-LUMO gap values are 0.001, 0.03, 0.04, 0.13, 0.39, 1.46 eV, respectively, and the ketone group substitution has the greatest reduction on the HOMO-LUMO gap value. By applying a detailed analysis of PAHs with unexpected low and high gap values, we found that the factors such as functional groups, local structure, and the number of five-member rings all contribute to the band gap to some extent. Among all these factors, the aliphatic substitution has a very weak effect on the electronic structures of HOMO and LUMO, the five-member rings leading non-planar PAHs impact the gap most. We developed a machine learning model to predict the HOMO-LUMO gaps of PAHs, which is in good agreement with the DFT data set, and the average absolute error is only 0.19 eV. This model can quickly and accurately predict the HOMO-LUMO gaps of PAHs.

REFERENCES

Adamson, B. D., Skeen, S. A., Ahmed, M., and Hansen, N. (2018). Detection of Aliphatically Bridged Multi-Core Polycyclic Aromatic Hydrocarbons in Sooting Flames with Atmospheric-Sampling High-Resolution Tandem Mass Spectrometry. *J. Phys. Chem. A.* 122 (48), 9338–9349. doi:10.1021/acs.jpca.8b08947

Adkins, E. M., Giaccai, J. A., and Miller, J. H. (2017). Computed Electronic Structure of Polynuclear Aromatic Hydrocarbon Agglomerates. *Proc. Combust. Inst.* 36 (1), 957–964. doi:10.1016/j.proci.2016.06.186

Adkins, E. M., and Miller, J. H. (2015). Extinction Measurements for Optical Band Gap Determination of Soot in a Series of Nitrogen-Diluted Ethylene/Air Non-premixed Flames. *Phys. Chem. Chem. Phys.* 17 (4), 2686–2695. doi:10.1039/c4cp04452e

Cain, J. P., Gassman, P. L., Wang, H., and Laskin, A. (2010). Micro-FTIR Study of Soot Chemical Composition-Evidence of Aliphatic Hydrocarbons on Nascent Soot Surfaces. *Phys. Chem. Chem. Phys.* 12 (20), 5206–5218. doi:10.1039/b924344e

Chen, D., and Wang, H. (2019a). HOMO-LUMO Energy Splitting in Polycyclic Aromatic Hydrocarbons and Their Derivatives. *Proc. Combust. Inst.* 37 (1), 953–959. doi:10.1016/j.proci.2018.06.120

Chen, D., and Wang, H. (2019b). HOMO-LUMO Gaps of Homogeneous Polycyclic Aromatic Hydrocarbon Clusters. *J. Phys. Chem. C* 123 (45), 27785–27793. doi:10.1021/acs.jpcc.9b08300

Chen, P., Fatayer, S., Schuler, B., Metz, J. N., Gross, L., Yao, N., et al. (2021). The Role of Methyl Groups in the Early Stage of Thermal Polymerization of Polycyclic Aromatic Hydrocarbons Revealed by Molecular Imaging. *Energy Fuels* 35 (3), 2224–2233. doi:10.1021/ACS.ENERGYFUELS.0C04016

Commodo, M., Kaiser, K., De Falco, G., Minutolo, P., Schulz, F., D'Anna, A., et al. (2019). On the Early Stages of Soot Formation: Molecular Structure Elucidation

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YX prepared the dataset and analyzed the data. QC made the machine learning model. YX, QC, DC, and AF all wrote the paper, drew conclusion and edited the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmech.2021.744001/full#supplementary-material

by High-Resolution Atomic Force Microscopy. *Combust. Flame* 205, 154–164. doi:10.1016/j.combustflame.2019.03.042

De, S., Bartók, A. P., Csányi, G., and Ceriotti, M. (2016). Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* 18 (20), 13754–13769. doi:10.1039/c6cp00415f

Elvati, P., Turrentine, K., and Violi, A. (2019). The Role of Molecular Properties on the Dimerization of Aromatic Compounds. *Proc. Combust. Inst.* 37 (1), 1099–1105. doi:10.1016/j.proci.2018.05.065

Elvati, P., and Violi, A. (2013). Thermodynamics of Poly-Aromatic Hydrocarbon Clustering and the Effects of Substituted Aliphatic Chains. *Proc. Combust. Inst.* 34 (1), 1837–1843. doi:10.1016/j.proci.2012.07.030

Frenklach, M., and Mebel, A. M. (2020). On the Mechanism of Soot Nucleation. *Phys. Chem. Chem. Phys.* 22 (9), 5314–5331. doi:10.1039/d0cp00116c

Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., et al. (2009). *Gaussian 09, Revision D. 01*. Wallingford CT: Gaussian, Inc.

Gavilan Marin, L., Bejaoui, S., Haggmark, M., Svadlenak, N., de Vries, M., Sciamma-O'Brien, E., et al. (2020). Low-Temperature Formation of Carbonaceous Dust Grains from PAHs. *ApJ* 889 (2), 101. doi:10.3847/1538-4357/ab62b7

Gentile, F. S., Picca, F., De Falco, G., Commodo, M., Minutolo, P., Causà, M., et al. (2020). Soot Inception: A DFT Study of σ and π Dimerization of Resonantly Stabilized Aromatic Radicals. *Fuel* 279, 118491. doi:10.1016/j.fuel.2020.118491

Giaccai, J. A., and Miller, J. H. (2019). Examination of the Electronic Structure of Oxygen-Containing PAH Dimers and Trimers. *Proc. Combust. Inst.* 37 (1), 903–910. doi:10.1016/j.proci.2018.05.057

Johansson, K. O., Dillstrom, T., Elvati, P., Campbell, M. F., Schrader, P. E., Popolan-Vaida, D. M., et al. (2017). Radical-radical Reactions, Pyrene Nucleation, and Incipient Soot Formation in Combustion. *Proc. Combust. Inst.* 36 (1), 799–806. doi:10.1016/j.proci.2016.07.130

Johansson, K. O., Dillstrom, T., Monti, M., El Gabaly, F., Campbell, M. F., Schrader, P. E., et al. (2016). Formation and Emission of Large Furans and Oxygenated

Hydrocarbons from Flames. *Proc. Natl. Acad. Sci. USA* 113 (30), 8374–8379. doi:10.1073/pnas.1604772113

Johansson, K. O., Head-Gordon, M. P., Schrader, P. E., Wilson, K. R., and Michelsen, H. A. (2018). Resonance-Stabilized Hydrocarbon-Radical Chain Reactions May Explain Soot Inception and Growth. *Science* 361 (6406), 997–1000. doi:10.1126/science.aat3417

Kholghy, M. R., Kelesidis, G. A., and Pratsinis, S. E. (2018). Reactive Polycyclic Aromatic Hydrocarbon Dimerization Drives Soot Nucleation. *Phys. Chem. Chem. Phys.* 20 (16), 10926–10938. doi:10.1039/c7cp07803j

Kislov, V. V., Sadovnikov, A. I., and Mebel, A. M. (2013). Formation Mechanism of Polycyclic Aromatic Hydrocarbons beyond the Second Aromatic Ring. *J. Phys. Chem. A.* 117 (23), 4794–4816. doi:10.1021/jp402481y

Kozliak, E., Sulkes, M., Alhroub, I., Kubátová, A., Andrianova, A., and Seames, W. (2019). Influence of Early Stages of Triglyceride Pyrolysis on the Formation of PAHs as Coke Precursors. *Phys. Chem. Chem. Phys.* 21 (36), 20189–20203. doi:10.1039/c9cp02025j

Lafleur, A. L., Howard, J. B., Marr, J. A., and Yadav, T. (1993). Proposed Fullerene Precursor Corannulene Identified in Flames Both in the Presence and Absence of Fullerene Production. *J. Phys. Chem.* 97 (51), 13539–13543. doi:10.1021/j100153a020

Leon, G., Menon, A., Pascazio, L., Bringley, E. J., Akroyd, J., and Kraft, M. (2020). Kinetic Monte Carlo Statistics of Curvature Integration by HACA Growth and Bay Closure Reactions for PAH Growth in a Counterflow Diffusion Flame. *Proc. Combust. Inst.* 38, 1449–1457. doi:10.1016/j.proci.2020.06.352

Li, H., Wang, H., Chen, D., and Kang, Z. (2020). Revealing the Optical Properties of Polycyclic Aromatic Hydrocarbon Clusters with Surface Formyl Groups. *Proc. Combust. Inst.* 38, 1207–1215. doi:10.1016/j.proci.2020.07.126

Li, Y., Tan, H., Wang, X., Bai, S., Mei, J., You, X., et al. (2018). Characteristics and Mechanism of Soot Formation during the Fast Pyrolysis of Biomass in an Entrained Flow Reactor. *Energy Fuels* 32 (11), 11477–11488. doi:10.1021/acs.energyfuels.8b00752

Liu, C., Singh, A. V., Saggese, C., Tang, Q., Chen, D., Wan, K., et al. (2019). Flame-Formed Carbon Nanoparticles Exhibit Quantum Dot Behaviors. *Proc. Natl. Acad. Sci. USA* 116 (26), 12692–12697. doi:10.1073/pnas.1900205116

Lowe, J. S., Lai, J. Y. W., Elvati, P., and Violi, A. (2015). Towards a Predictive Model for Polycyclic Aromatic Hydrocarbon Dimerization Propensity. *Proc. Combust. Inst.* 35 (2), 1827–1832. doi:10.1016/j.proci.2014.06.142

Menon, A., Dreyer, J. A. H., Martin, J. W., Akroyd, J., Robertson, J., and Kraft, M. (2019). Optical Band Gap of Cross-Linked, Curved, and Radical Polyaromatic Hydrocarbons. *Phys. Chem. Chem. Phys.* 21 (29), 16240–16251. doi:10.1039/c9cp02363a

Michelsen, H. A. (2020). Effects of Maturity and Temperature on Soot Density and Specific Heat. *Proc. Combust. Inst.* 38, 1197–1205. doi:10.1016/j.proci.2020.06.383

Miller, J. H., Herdman, J. D., Green, C. D. O., and Webster, E. M. (2013). Experimental and Computational Determinations of Optical Band Gaps for PAH and Soot in a N2-Diluted, Ethylene/Air Non-premixed Flame. *Proc. Combust. Inst.* 34 (2), 3669–3675. doi:10.1016/j.proci.2012.05.054

Minutolo, P., Gambi, G., and D'Alessio, A. (1996). The Optical Band Gap Model in the Interpretation of the UV-Visible Absorption Spectra of Rich Premixed Flames. *Symp. Combust.* 26 (1), 951–957. doi:10.1016/S0082-0784(96)80307-9

Öktem, B., Tolocka, M. P., Zhao, B., Wang, H., and Johnston, M. V. (2005). Chemical Species Associated with the Early Stage of Soot Growth in a Laminar

Premixed Ethylene-Oxygen-Argon Flame. *Combust. Flame* 142 (4), 364–373. doi:10.1016/j.combustflame.2005.03.016

Pascazio, L., Martin, J. W., Menon, A., Hou, D., You, X., and Kraft, M. (2020). Aromatic Penta-Linked Hydrocarbons in Soot Nanoparticle Formation. *Proc. Combust. Inst.* 38, 1525–1532. doi:10.1016/j.proci.2020.09.029

Saldinger, J. C., Wang, Q., Elvati, P., and Violi, A. (2020). Characterizing the Diversity of Aromatics in A Coflow Diffusion Jet A-1 Surrogate Flame. *Fuel* 268, 117198. doi:10.1016/j.fuel.2020.117198

Schulz, F., Commodo, M., Kaiser, K., De Falco, G., Minutolo, P., Meyer, G., et al. (2019). Insights into Incipient Soot Formation by Atomic Force Microscopy. *Proc. Combust. Inst.* 37 (1), 885–892. doi:10.1016/j.proci.2018.06.100

Shi, X., Wang, Q., and Violi, A. (2021). Reaction Pathways for the Formation of Five-Membered Rings onto Polyaromatic Hydrocarbon Framework. *Fuel* 283, 119023. doi:10.1016/j.fuel.2020.119023

Wang, Q., Elvati, P., Kim, D., Johansson, K. O., Schrader, P. E., Michelsen, H. A., et al. (2019). Spatial Dependence of the Growth of Polycyclic Aromatic Compounds in an Ethylene Counterflow Flame. *Carbon* 149, 328–335. doi:10.1016/j.carbon.2019.03.017

Wang, Q., Saldinger, J. C., Elvati, P., and Violi, A. (2021). Molecular Structures in Flames: A Comparison between SNapS2 and Recent AFM Results. *Proc. Combust. Inst.* 38 (1), 1133–1141. doi:10.1016/j.proci.2020.06.250

Wang, Y., and Chung, S. H. (2019). Soot Formation in Laminar Counterflow Flames. *Prog. Energ. Combust. Sci.* 74, 152–238. doi:10.1016/j.pecs.2019.05.003

Zhang, H.-B., Hou, D., Law, C. K., and You, X. (2016). Role of Carbon-Addition and Hydrogen-Migration Reactions in Soot Surface Growth. *J. Phys. Chem. A.* 120 (5), 683–689. doi:10.1021/acs.jpca.5b10306

Zhang, Y. (2019). "Similarities in Diverse Polycyclic Aromatic Hydrocarbons of Asphaltenes and Heavy Oils Revealed by Noncontact Atomic Force Microscopy: Aromaticity, Bonding, and Implications for Reactivity," in *Chemistry Solutions to Challenges in the Petroleum Industry* (Washington, DC: ACS Symposium Series; American Chemical Society), 1320, 39–65. doi:10.1021/bk-2019-1320.ch003

Zhao, J., Lin, Y., Huang, K., Gu, M., Lu, K., Chen, P., et al. (2020). Study on Soot Evolution under Different Hydrogen Addition Conditions at High Temperature by ReaxFF Molecular Dynamics. *Fuel* 262, 116677. doi:10.1016/j.fuel.2019.116677