**HOMO–LUMO gaps and molecular structures of polycyclic aromatic hydrocarbons in soot formation**

Yabei Xu1, Qingzhao Chu1, Dongping Chen1\*, Andres Fuentes2

*1State Key Lab of Explosion Science and Technology, Beijing Institute of Technology, Beijing, 100081, China*

*2Departamento de Industrias, Universidad Técnica Federico Santa María, Av. España 1680, Casilla 110-V, Valparaíso, Chile*

\*Corresponding Author Email: [dc516@bit.edu.cn](mailto:dc516@bit.edu.cn)

1. The data set of PAHs

To build a more comprehensive data set of polycyclic aromatic hydrocarbons (PAHs), we screened a large number of publications about soot particles mostly in the last ten years. 323 reliable PAHs relevant with the formation of soot were collected. We computed the HOMO-LUMO gap for PAHs using the DFT method. The features, including functional groups, number of carbon rings, HOMO energy, LUMO energy, gap values and the resource, are included in Table S1. The sequence of the PAHs follow the number of carbon atoms in PAHs.

**Table S1**.Molecular structures and featured properties of PAHs.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Molecular Formula | Rings1 | Function groups | Structure | HOMO (eV) | LUMO  (eV) | *EH-L*  (eV) | References |
| C6H6 | 1 | - |  | -7.08 | -0.48 | 6.59 | (Kholghy et al., 2018; Li et al., 2018; Martin et al., 2019) |
| C6H8 | 1 | -CnHm |  | -6.74 | -0.46 | 6.28 | (Kholghy et al., 2018) |
| C7H7 | 1 | -CnHm |  | -6.2  -6.89 | -0.51  -2.03 | 5.7  4.86 | (Kozliak et al., 2019) |
| C7H8 | 1 | -CnHm |  | -6.74 | -0.46 | 6.28 | (Li et al., 2018; Commodo et al., 2019; Schulz et al., 2019) |
| C8H6 | 1 | -CnHm |  | -6.69 | -1.29 | 5.40 | (Kholghy et al., 2018; Li et al., 2018) |
| C8H8 | 1 | -CnHm |  | -6.40 | -1.36 | 5.04 | (Li et al., 2018) (Commodo et al., 2019) |
| C8H8 | 1 | -CnHm |  | -5.64 | -1.87 | 3.77 | (Commodo et al., 2019) |
| C8H10 | 1 | -CnHm |  | -5.73 | -1.16 | 4.57 | (Commodo et al., 2019) |
| C9H8 | 2 | - |  | -6.12 | -0.93 | 5.19 | (Li et al., 2018) |
| C9H8 | 2 | - |  | -5.30 | -1.98 | 3.31 | (Martin et al., 2019) |
| C10H8 | 2 | - |  | -6.15 | -1.40 | 4.75 | (Kholghy et al., 2018; Li et al., 2018; Martin et al., 2019; Michelsen, 2020) |
| C10H9O | 2 | -CnHm  -O- |  | -4.04  -6.55 | -0.54  -1.85 | 3.49  4.7 | (Commodo et al., 2019) |
| C11H12O | 1 | -CnHm  -O- |  | -5.91 | -1.24 | 4.66 | (Johansson et al., 2016) |
| C12H8 | 3 | - |  | -6.17 | -2.29 | 3.89 | (Adamson et al., 2018; Kholghy et al., 2018; Martin et al., 2019; Michelsen, 2020) |
| C12H8 | 2 | -CnHm |  | -6.12 | -1.82 | 4.30 | (Kholghy et al., 2018) |
| C12H8 | 2 | - |  | -5.72 | -1.65 | 4.06 | (Li et al., 2018) |
| C12H8O | 2 | -O-  -CnHm |  | -5.91 | -1.40 | 4.50 | (Johansson et al., 2016) |
| C12H8O4 | 2 | -OH(3)  -O- |  | -5.84 | -1.42 | 4.42 | (Wang et al., 2019) |
| C12H10 | 2 | - |  | -6.26 | -1.39 | 4.87 | (Li et al., 2018; Michelsen, 2020) |
| C12H10 | 3 | - |  | -5.79 | -1.18 | 4.61 | (Kislov et al., 2013; Martin et al., 2019; Michelsen, 2020; Shi et al., 2021) |
| C12H12 | 2 | -CnHm |  | -5.93 | -1.20 | 4.74 | (Michelsen, 2020) |
| C12H12 | 2 | -CnHm |  | -5.89 | -1.24 | 4.65 | (Michelsen, 2020) |
| C13H10 | 3 | - |  | -6.10 | -1.18 | 4.92 | (Kislov et al., 2013; Li et al., 2018; Michelsen, 2020; Shi et al., 2021) |
| C13H10 | 3 | - |  | -5.57 | -1.57 | 4.00 | (Li et al., 2018; Martin et al., 2019) |
| C13H10O2 | 4 | =O |  | -6.28 | -2.60 | 3.68 | (Wang et al., 2020) |
| C14H8 | 3 | -CnHm |  | -6.07 | -2.57 | 3.50 | (Kholghy et al., 2018) |
| C14H8O | 4 | -O- |  | -5.54 | -1.93 | 3.61 | (Saldinger et al., 2020) |
| C14H10 | 3 | - |  | -5.58 | -2.02 | 3.55 | (Kholghy et al., 2018; Li et al., 2018; Michelsen, 2020) |
| C14H10 | 3 | - |  | -6.09 | -1.41 | 4.68 | (Adamson et al., 2018; Kholghy et al., 2018; Li et al., 2018; Kozliak et al., 2019; Martin et al., 2019; Michelsen, 2020) |
| C14H10O | 3 | -O- |  | -5.80 | -2.23 | 3.57 | (Johansson et al., 2016) |
| C14H10O | 2 | -O-  -CnHm |  | -6.16 | -1.79 | 4.37 | (Johansson et al., 2016) |
| C14H12 | 2 | -CnHm |  | -5.85 | -1.80 | 4.05 | (Kozliak et al., 2019; Michelsen, 2020) |
| C14H12 | 3 | - |  | -6.65 | -0.68 | 5.97 | (Michelsen, 2020) |
| C14H14 | 2 | -CnHm |  | -6.36 | -0.38 | 5.98 | (Commodo et al., 2019) |
| C14H20 | 2 | -CnHm |  | -6.16 | -0.32 | 5.85 | (Adamson et al., 2018) |
| C15H8O | 4 | =O |  | -5.81 | -3.37 | 2.44 | (Zhao et al., 2020) |
| C15H10 | 4 | - |  | -6.01 | -1.41 | 4.60 | (Li et al., 2018; Shi et al., 2021) |
| C15H12 | 3 | -CnHm |  | -5.92 | -1.37 | 4.55 | (Wang et al., 2020) |
| C15H14 | 3 | -CnHm |  | -5.87 | -1.00 | 4.87 | (Zhang, 2019) |
| C15H16 | 2 | -CnHm |  | -6.45 | -0.56 | 5.90 | (Schulz et al., 2019) |
| C15H16 |  | -CnHm |  | -6.25 | -0.36 | 5.90 | (Commodo et al., 2019) |
| C15H22 | 2 | -CnHm |  | -6.42 | -0.38 | 6.04 | (Adamson et al., 2018) |
| C15H22 | 2 | -CnHm |  | -6.20 | -0.38 | 5.82 | (Adamson et al., 2018) |
| C15H22 | 2 | -CnHm |  | -6.17 | -0.33 | 5.84 | (Adamson et al., 2018) |
| C15H22 | 2 | -CnHm |  | -6.41 | -0.35 | 6.07 | (Adamson et al., 2018) |
| C15H22 | 2 | -CnHm |  | -6.10 | -0.33 | 5.77 | (Adamson et al., 2018) |
| C15H22 | 2 | -CnHm |  | -6.20 | -0.22 | 5.98 | (Adamson et al., 2018) |
| C15H22 | 2 | -CnHm |  | -6.41 | -0.37 | 6.04 | (Adamson et al., 2018) |
| C15H22 | 2 | -CnHm |  | -6.18 | -0.28 | 5.90 | (Adamson et al., 2018) |
| C16H8 | 4 | - |  | -5.44 | -3.69 | 1.74 | (Mao et al., 2018) |
| C16H9 | 4 | - |  | -5.78  -5.8 | -1.99  -2.57 | 3.79  3.23 | (Mao et al., 2018) |
| C16H10 | 4 |  |  | -5.67 | -1.86 | 3.81 | (Elvati and Violi, 2013; Pascazio et al., 2019),(Lafleur et al., 1993; Kholghy et al., 2018; Li et al., 2018; Michelsen, 2020; Chen et al., 2021),(Mao et al., 2018) |
| C16H10 | 4 | - |  | -6.11 | -2.15 | 3.96 | (Michelsen, 2020) |
| C16H10 | 4 | - |  | -5.71 | -2.63 | 3.08 | (Kholghy et al., 2018) |
| C16H10 | 4 | - |  | -6.11 | -2.24 | 3.86 | (Lafleur et al., 1993; Kholghy et al., 2018) |
| C16H10 | 4 | -CnHm |  | -5.67 | -2.90 | 2.77 | (Johansson et al., 2017) |
| C16H10 | 3 | -CnHm |  | -6.14 | -2.60 | 3.55 | (Johansson et al., 2017) |
| C16H10 | 4 | - |  | -6.11 | -2.15 | 3.96 | (Lafleur et al., 1993; Li et al., 2018; Gavilan Marin et al., 2020) |
| C16H14 | 3 | -CnHm |  | -5.97 | -1.36 | 4.61 | (Gavilan Marin et al., 2020) |
| C17H12 | 4 | -CnHm |  | -5.62 | -1.81 | 3.80 | (Wang et al., 2020; Chen et al., 2021) |
| C17H12 | 4 | - |  | -5.65 | -1.34 | 4.30 | (Commodo et al., 2019) |
| C17H12 | 4 | -CnHm |  | -5.62 | -1.82 | 3.80 | (Chen et al., 2021) |
| C18H10 | 5 | - |  | -5.32 | -2.97 | 2.36 | (Martin et al., 2019) |
| C18H10 | 5 | - |  | -5.75 | -2.49 | 3.26 | (Kholghy et al., 2018) |
| C18H10 | 5 | - |  | -6.22 | -2.23 | 3.99 | (Lafleur et al., 1993; Li et al., 2018; Zhao et al., 2020) |
| C18H10 | 5 | - |  | -6.23 | -2.73 | 3.51 | (Lafleur et al., 1993) |
| C18H10 | 5 | - |  | -5.75 | -2.49 | 3.26 | (Lafleur et al., 1993; Adamson et al., 2018) |
| C18H10 | 4 | -CnHm |  | -5.82 | -1.99 | 3.83 | (Zhao et al., 2020) |
| C18H10O | 4 | -O- | C:\Users\QQ\AppData\Local\Temp\1617696272(1).jpg | -5.66 | -2.06 | 3.60 | (Giaccai and Miller, 2019) |
| C18H10O3 | 3 | -O-(2)  -OH |  | -5.64 | -1.38 | 4.26 | (Wang et al., 2019) |
| C18H12 | 5 | - |  | -5.35 | -1.92 | 3.43 | (Li et al., 2018) |
| C18H12 | 4 | - |  | -5.98 | -1.76 | 4.21 | (Li et al., 2018) |
| C18H12 | 4 | -CnHm |  | -5.55 | -2.38 | 3.18 | (Zhao et al., 2020) |
| C18H12 | 4 | - |  | -5.20 | -2.45 | 2.75 | (Michelsen, 2020) |
| C18H12 | 4 | - |  | -5.67 | -1.94 | 3.72 | (Michelsen, 2020) |
| C18H12 | 4 | - |  | -5.87 | -1.65 | 4.22 | (Michelsen, 2020) |
| C18H12 | 4 | -CnHm | C:\Users\QQ\AppData\Local\Temp\1617711032(1).jpg | -5.71 | -1.91 | 3.81 | (Elvati and Violi, 2013) |
| C18H12O | 4 | –OH | C:\Users\QQ\AppData\Local\Temp\1617667391(1).jpg | -5.78 | -2.08 | 3.70 | (Giaccai and Miller, 2019) |
| C18H12O | 4 | =O | C:\Users\QQ\AppData\Local\Temp\1617695211(1).jpg | -5.42 | -3.15 | 2.27 | (Giaccai and Miller, 2019) |
| C18H14 | 4 | -CnHm |  | -5.57 | -1.78 | 3.79 | (Chen et al., 2021) |
| C18H14 | 4 | -CnHm |  | -5.50 | -1.79 | 3. 72 | (Chen et al., 2021) |
| C18H14 | 4 | --CnHm |  | -5.57 | -1.77 | 3.79 | (Chen et al., 2021) |
| C18H14 | 4 | -CnHm |  | -5.57 | -1.77 | 3.79 | (Chen et al., 2021) |
| C18H14 | 4 | - |  | -5.64 | -1.62 | 4.02 | (Wang et al., 2020) |
| C18H14 | 3 | - |  | -5.93 | -1.76 | 4.17 | (Michelsen, 2020) |
| C18H14 | 3 | - |  | -6.25 | -1.15 | 5.10 | (Michelsen, 2020) |
| C18H14 | 4 | -CnHm | C:\Users\QQ\AppData\Local\Temp\1617710178(1).jpg | -5.60 | -1.79 | 3.81 | (Elvati and Violi, 2013) |
| C18H16 | 4 | -CnHm |  | -5.34 | -2.16 | 3.18 | (Commodo et al., 2019) |
| C19H12 | 5 | -CnHm |  | -6.01 | -2.04 | 3.97 | (Commodo et al., 2019) |
| C19H12O | 4 | –CHO | C:\Users\QQ\AppData\Local\Temp\1617670438(1).jpg | -5.71 | -2.37 | 3.34 | (Giaccai and Miller, 2019) |
| C19H12O | 4 | -O- |  | -5.41 | -1.87 | 3.54 | (Giaccai and Miller, 2019) |
| C19H12O2 | 4 | –COOH | C:\Users\QQ\AppData\Local\Temp\1617671052(1).jpg | -5.96 | -2.27 | 3.69 | (Giaccai and Miller, 2019) |
| C19H14 | 4 | -CnHm |  | -5.56 | -1.89 | 3.66 | (Michelsen, 2020) |
| C19H14 | 4 | -CnHm |  | -5.56 | -1.93 | 3.63 | (Commodo et al., 2019; Michelsen, 2020) |
| C19H14 | 4 | -CnHm |  | -6.06 | -1.33 | 4.73 | (Zhang, 2019) |
| C19H16 | 4 | -CnHm |  | -5.45 | -1.75 | 3.70 | (Chen et al., 2021) |
| C19H16 | 4 | -CnHm) |  | -5.46 | -1.75 | 3.71 | (Chen et al., 2021) |
| C19H16 | 4 | -CnHm |  | -5.52 | -1.73 | 3.78 | (Chen et al., 2021) |
| C19H16 | 4 | -CnHm | C:\Users\QQ\AppData\Local\Temp\1617710611(1).jpg | -5.59 | -1.79 | 3.81 | (Elvati and Violi, 2013) |
| C20H9O4 | 6 | =O  =O  -COOH  -O- |  | -6.63  -6.55 | -3.16  -4.41 | 3.47  2.14(Wang et al., 2020) | (Wang et al., 2020) |
| C20H10 | 6 | - |  | -5.65 | -2.84 | 2.81 | (Commodo et al., 2019) |
| C20H10 | 6 | - |  | -6.25 | -2.08 | 4.16 | (Martin et al., 2019; Leon et al., 2020) |
| C20H12 | 5 | - |  | -4.75 | -3.30 | 1.44 | (Zhao et al., 2020) |
| C20H12 | 5 | -CnHm |  | -4.97 | -2.99 | 1.98 | (Zhao et al., 2020) |
| C20H12 | 4 | -CnHm |  | -5.16 | -2.81 | 2.35 | (Zhao et al., 2020) |
| C20H12 | 5 | - |  | -5.45 | -2.10 | 3.35 | (Elvati and Violi, 2013; Lowe et al., 2015; Michelsen, 2020; Zhao et al., 2020) |
| C20H12 | 5 | - |  | -5.76 | -1.79 | 3.96 | (Adamson et al., 2018; Gavilan Marin et al., 2020; Michelsen, 2020; Wang et al., 2020) |
| C20H12 | 5 | - |  | -5.29 | -2.31 | 2.98 | (Li et al., 2018; Michelsen, 2020) |
| C20H13 | 5 | -CnHm |  | -4.48  -5.98 | -1.3  -2.79 | 3.18  3.19 | (Commodo et al., 2019) |
| C20H14 | 4 | --CnHm | C:\Users\QQ\AppData\Local\Temp\1617696962(1).jpg | -5.73 | -1.95 | 3.78 | (Elvati and Violi, 2013) |
| C20H14 | 5 | - |  | -4.83 | -2.18 | 2.65 | (Li et al., 2018) |
| C20H16 | 4 | -CnHm |  | -5.45 | -1.92 | 3.54 | (Michelsen, 2020) |
| C20H18 | 4 | -CnHm |  | -5.59 | -1.78 | 3.81 | (Elvati and Violi, 2013) |
| C20H18 | 4 | -CnHm |  | -5.41 | -1.71 | 3.70 | (Chen et al., 2021) |
| C21H14 | 5 | -CnHm |  | -5.37 | -2.08 | 3.29 | (Michelsen, 2020) |
| C21H14 | 5 | -CnHm |  | -5.40 | -2.05 | 3.35 | (Michelsen, 2020) |
| C21H16 | 5 | - |  | -5.51 | -1.84 | 3.67 | (Michelsen, 2020) |
| C21H16 | 5 | -CnHm |  | -5.34 | -1.75 | 3.59 | (Michelsen, 2020) |
| C22H10 | 5 | -CnHm |  | -5.81 | -3.09 | 2.72 | (Wang et al., 2019) |
| C22H10 | 7 | - |  | -5.82 | -3.27 | 2.55 | (Lowe et al., 2015) |
| C22H12 | 6 | - |  | -6.05 | -2.21 | 3.84 | (Kozliak et al., 2019) |
| C22H12 | 6 | - |  | -5.54 | -2.05 | 3.48 | (Lowe et al., 2015) |
| C22H12O | 6 | -OH |  | -5.69 | -2.58 | 3.11 | (Saldinger et al., 2020) |
| C22H14 | 6 | - |  | -5.90 | -1.46 | 4.44 | (Frenklach and Mebel, 2020) |
| C22H14 | 5 | - |  | -4.65 | -3.20 | 1.45 | (Commodo et al., 2019) |
| C22H14 | 5 | - |  | -4.94 | -2.75 | 2.19 | (Michelsen, 2020) |
| C22H14 | 5 | - |  | -5.72 | -1.87 | 3.85 | (Michelsen, 2020) |
| C22H16 | 5 | - |  | -6.01 | -1.44 | 4.57 | (Michelsen, 2020) |
| C23H14 | 7 | - |  | -5.30 | -1.97 | 3.33 | (Commodo et al., 2019) |
| C23H14 | 6 | -CnHm |  | -5.42 | -1.89 | 3.53 | (Commodo et al., 2019) |
| C23H14 | 6 | -CnHm |  | -5.41 | -1.85 | 3.57 | (Commodo et al., 2019) |
| C23H14 | 6 | -CnHm |  | -5.16 | -2.33 | 2.83 | (Zhang, 2019) |
| C24H10 | 8 | - |  | -5.89 | -3.57 | 2.33 | (Lowe et al., 2015) |
| C24H10 | 8 |  |  | -5.53 | -3.52 | 2.01 | (Wang et al., 2020) |
| C24H10 | 7 | - |  | -5.17 | -4.53 | 0.64 | (Mao et al., 2018) |
| C24H10O | 7 | -O- |  | -5.72 | -2.56 | 3.16 | (Wang et al., 2020) |
| C24H11 | 7 | - |  | -5.86  -5.8 | -1.87  -2.67 | 3.98  3.14 | (Mao et al., 2018) |
| C24H12 | 7 | - |  | -5.78 | -1.78 | 4.00 | (Elvati and Violi, 2013; Lowe et al., 2015; Adamson et al., 2018; Mao et al., 2018; Martin et al., 2019; Pascazio et al., 2019; Michelsen, 2020) |
| C24H12 | 7 | - |  | -5.21 | -2.68 | 2.53 | (Commodo et al., 2019) |
| C24H12 | 7 | - |  | -5.67 | -2.93 | 2.74 | (Commodo et al., 2019) |
| C24H12 | 7 | - |  | -5.54 | -2.32 | 3.22 | (Commodo et al., 2019) |
| C24H12 | 7 | - |  | -5.63 | -2.76 | 2.87 | (Commodo et al., 2019) |
| C24H12 | 7 | - |  | -5.56 | -2.91 | 2.65 | (Commodo et al., 2019) |
| C24H12 | 7 | - |  | -5.71 | -2.86 | 2.85 | (Martin et al., 2019) |
| C24H12 | 7 | - |  | -5.34 | -2.46 | 2.88 | (Leon et al., 2020) |
| C24H12O | 7 | -OH |  | -5.63 | 1.84 | 7.48 | (Mao et al., 2018) |
| C24H14 | 5 | - |  | -5.54 | -2.03 | 3.51 | (Adamson et al., 2018) |
| C24H14 | 7 | -CnHm |  | -5.32 | -1.82 | 3.50 | (Commodo et al., 2019) |
| C24H14 | 6 | - |  | -5.51 | -2.11 | 3.40 | (Michelsen, 2020) |
| C24H14 | 6 | - |  | -5.48 | -2.67 | 2.81 | (Pascazio et al., 2020) |
| C24H14O | 6 | -CnHm  =O |  | -5.81 | -2.89 | 2.91 | (Zhang, 2019) |
| C24H14O |  | -OH |  | -5.73 | -1.96 | 3.77 | (Saldinger et al., 2020) |
| C24H16 | 6 | - |  | -5.38 | -2.08 | 3.29 | (Pascazio et al., 2020) |
| C24H16 | 7 | - |  | -5.84 | -1.56 | 4.28 | (Frenklach and Mebel, 2020) |
| C24H18 | 4 | - |  | -6.18 | -1.61 | 4.57 | (Michelsen, 2020) |
| C25H14 | 7 | - |  | -5.65 | -1.84 | 3.81 | (Leon et al., 2020) |
| C26H12 | 8 | - |  | -5.63 | -2.47 | 3.16 | (Commodo et al., 2019) |
| C26H12 | 8 | - |  | -5.76 | -2.43 | 3.34 | (Martin et al., 2019) |
| C26H12 | 8 | - |  | -5.29 | -2.97 | 2.32 | (Leon et al., 2020) |
| C26H14 | 7 | - |  | -5.20 | -2.38 | 2.82 | (Zhang, 2019) |
| C26H14 | 8 | - |  | -5.57 | -1.72 | 3.85 | (Martin et al., 2019) |
| C26H16 | 6 | - |  | -4.75 | -2.96 | 1.78 | (Michelsen, 2020) |
| C26H16 | 8 | - |  | -5.18 | -2.57 | 2.61 | (Commodo et al., 2019) |
| C27H12 |  | -CnHm |  | -4.59 | -3.92 | 0.67 | (Wang et al., 2020) |
| C27H14 | 7 | -CnHm |  | -5.70 | -2.37 | 3.33 | (Wang et al., 2020) |
| C28H12 | 9 | - |  | -5.89 | -2.58 | 3.32 | (Lowe et al., 2015; Commodo et al., 2019) |
| C28H13 | 9 | - |  | -4.75 | -2.29 | 2.46 | (Commodo et al., 2019) |
| C28H14 | 8 | - |  | -5.61 | -1.96 | 3.65 | (Lowe et al., 2015) |
| C28H14 | 9 | - |  | -5.58 | -2.36 | 3.22 | (Gentile et al., 2020) |
| C28H14 | 8 | - |  | -5.47 | -2.48 | 2.99 | (Commodo et al., 2019) |
| C28H16 | 7 | - |  | -5.34 | -2.26 | 3.08 | (Kozliak et al., 2019) |
| C28H16 | 7 | - |  | -5.19 | -2.53 | 2.66 | (Wang et al., 2020) |
| C28H20 | 6 | -CnHm |  | -5.21 | -2.22 | 2.99 | (Zhang, 2019) |
| C29H19 | 8 | -CnHm |  | -4.29  -5.78 | -1.37  -2.84 | 2.91  2.94 | (Commodo et al., 2019) |
| C29H12O | 9 | =O |  | -5.56 | -2.74 | 2.82 | (Commodo et al., 2019) |
| C30H12 | 10 | - |  | -5.25 | -3.63 | 1.62 | (Lowe et al., 2015) |
| C30H14 | 9 | - |  | -5.40 | -2.20 | 3.20 | (Lowe et al., 2015; Adamson et al., 2018; Commodo et al., 2019) |
| C30H14 | 9 | - |  | -5.31 | -2.67 | 2.64 | (Commodo et al., 2019) |
| C30H14 | 9 | - |  | -5.60 | -2.43 | 3.18 | (Commodo et al., 2019) |
| C30H14 | 9 | - |  | -5.23 | -2.61 | 2.62 | (Commodo et al., 2019) |
| C30H14 | 9 | - |  | -5.32 | -2.11 | 3.22 | (Commodo et al., 2019) |
| C30H14 | 9 | - |  | -5.42 | -2.46 | 2.96 | (Leon et al., 2020) |
| C30H14 | 9 | - |  | -5.41 | -2.64 | 2.77 | (Leon et al., 2020) |
| C30H16 | 9 | - |  | -4.99 | -2.20 | 2.79 | (Commodo et al., 2019) |
| C30H17 | 9 | - |  | -4.48  -5.55 | -2.39  -2.83 | 2.09  2.72 | (Commodo et al., 2019) |
| C30H18 | 9 | - |  | -4.97 | -2.75 | 2.23 | (Gentile et al., 2020) |
| C30H22 | 7 | -CnHm |  | -5.36 | -2.07 | 3.29 | (Zhang, 2019) |
| C31H16 | 9 | -CnHm |  | -5.31 | -2.18 | 3.13 | (Commodo et al., 2019; Schulz et al., 2019) |
| C31H16O | 9 | -OH |  | -4.90 | -2.11 | 2.78 | (Commodo et al., 2019) |
| C31H18 | 9 | - |  | -5.50 | -1.91 | 3.59 | (Martin et al., 2019) |
| C32H12 | 11 |  |  | -5.23 | -3.81 | 1.42 | (Wang et al., 2020) |
| C32H12O4 | 11 | -OH  -O-(3) |  | -5.23 | -2.12 | 3.11 | (Elvati et al., 2019) |
| C32H14 | 10 | - |  | -5.27 | -2.35 | 2.92 | (Lowe et al., 2015; Commodo et al., 2019; Pascazio et al., 2019) |
| C32H14 | 10 | - |  | -5.74 | -3.35 | 2.38 | (Lowe et al., 2015) |
| C32H14 | 10 | - |  | -5.26 | -3.51 | 1.75 | (Lowe et al., 2015) |
| C32H14 | 10 | - |  | -5.76 | -2.72 | 3.03 | (Commodo et al., 2019) |
| C32H14 | 10 | - |  | -5.27 | -2.35 | 2.92 | (Michelsen, 2020) |
| C32H14O4 | 10 | -OH  -O-(3)  -CnHm |  | -5.26 | -2.88 | 2.38 | (Elvati et al., 2019) |
| C32H14O4 | 10 | -O-(3)  -OH  -CnHm |  | -5.34 | -1.91 | 3.42 | (Elvati et al., 2019) |
| C32H18 | 8 | - |  | -5.71 | -1.90 | 3.81 | (Elvati et al., 2019) |
| C32H20 | 8 | -CnHm |  | -5.38 | -2.07 | 3.30 | (Chen et al., 2021) |
| C32H20 | 8 | -CnHm |  | -5.39 | -2.07 | 3.32 | (Chen et al., 2021) |
| C32H20 | 8 | -CnHm |  | -5.10 | -2.34 | 2.76 | (Chen et al., 2021) |
| C33H12 | 11 | -CnHm |  | -4.92 | -3.72 | 1.20 | (Wang et al., 2020) |
| C33H15O2 | 10 | =O  -OH |  | -5.31  -5.38 | -3.23  -3.5 | 2.07  1.88(Wang et al., 2020) | (Wang et al., 2020) |
| C33H16 | 10 | - |  | -5.21 | -2.05 | 3.16 | (Commodo et al., 2019) |
| C33H16 | 10 | -CnHm |  | -5.12 | -2.24 | 2.88 | (Commodo et al., 2019) |
| C33H18 | 9 | -CnHm |  | -5.01 | -2.51 | 2.50 | (Chen et al., 2021) |
| C33H18 | 9 | -CnHm |  | -5.02 | -2.54 | 2.49 | (Chen et al., 2021) |
| C33H18 | 9 | -CnHm |  | -5.05 | -2.51 | 2.53 | (Chen et al., 2021) |
| C33H18 | 10 | -CnHm |  | -5.16 | -1.98 | 3.17 | (Commodo et al., 2019) |
| C33H18 | 10 | - |  | -5.07 | -1.87 | 3.20 | (Commodo et al., 2019) |
| C34H16 | 9 | -CnHm |  | -5.67 | -2.76 | 2.91 | (Wang et al., 2019) |
| C34H16 | 11 | - |  | -5.23 | -2.27 | 2.96 | (Commodo et al., 2019) |
| C34H16 | 11 | - |  | -5.25 | -2.31 | 2.93 | (Commodo et al., 2019) |
| C34H18 | 9 | - |  | -5.43 | -2.25 | 3.18 | (Adamson et al., 2018) |
| C34H18 | 9 | - |  | -5.36 | -2.60 | 2.76 | (Frenklach and Mebel, 2020) |
| C34H18 | 10 | - |  | -5.54 | -1.92 | 3.62 | (Frenklach and Mebel, 2020) |
| C34H18 | 9 | - |  | -5.34 | -2.35 | 2.99 | (Michelsen, 2020) |
| C34H18 | 11 | - |  | -4.89 | -2.31 | 2.58 | (Martin et al., 2019) |
| C34H20 | 8 | -CnHm |  | -5.61 | -1.99 | 3.62 | (Frenklach and Mebel, 2020) |
| C34H20 | 8 | -CnHm |  | -5.31 | -2.45 | 2.86 | (Frenklach and Mebel, 2020) |
| C34H20 | 9 | -CnHm |  | -5.43 | -2.44 | 2.99 | (Chen et al., 2021) |
| C34H24 | 8 | -CnHm |  | -4.95 | -2.26 | 2.69 | (Chen et al., 2021) |
| C34H26 | 7 | -CnHm |  | -5.55 | -1.84 | 3.72 | (Adamson et al., 2018) |
| C34H26 | 7 | -CnHm |  | -5.53 | -2.10 | 3.43 | (Adamson et al., 2018) |
| C34H26 | 8 | -CnHm |  | -5.63 | -2.47 | 3.16 | (Adamson et al., 2018) |
| C34H26 | 7 | -CnHm |  | -5.50 | -2.12 | 3.38 | (Adamson et al., 2018) |
| C34H26 | 7 | -CnHm |  | -5.70 | -1.75 | 3.96 | (Adamson et al., 2018) |
| C34H26 | 7 | -CnHm |  | -5.48 | -1.98 | 3.50 | (Adamson et al., 2018) |
| C34H26 | 7 | -CnHm |  | -6.03 | -2.17 | 3.86 | (Adamson et al., 2018) |
| C34H26 | 8 | -CnHm |  | -5.94 | -2.26 | 3.68 | (Adamson et al., 2018) |
| C35H12 | 13 |  |  | -4.93 | -3.85 | 1.08 | (Wang et al., 2020) |
| C35H18 | 11 | -CnHm |  | -5.42 | -2.49 | 2.92 | (Commodo et al., 2019) |
| C35H18 | 10 | -CnHm |  | -5.56 | -2.41 | 3.14 | (Commodo et al., 2019) |
| C35H20 | 10 | -CnHm |  | -4.95 | -2.39 | 2.56 | (Chen et al., 2021) |
| C35H21 | 11 | -CnHm |  | -4.17  -5.34 | -1.53  -2.74 | 2.64  2.6 | (Commodo et al., 2019) |
| C35H22 | 9 | -CnHm |  | -4.97 | -2.46 | 2.51 | (Chen et al., 2021) |
| C35H22 | 9 | -CnHm |  | -5.00 | -2.44 | 2.56 | (Chen et al., 2021) |
| C35H22 | 9 | -CnHm |  | -4.97 | -2.47 | 2.49 | (Chen et al., 2021) |
| C35H22 | 9 | -CnHm |  | -4.99 | -2.45 | 2.54 | (Chen et al., 2021) |
| C35H22 | 9 | -CnHm |  | -5.42 | -1.98 | 3.44 | (Chen et al., 2021) |
| C35H24 | 8 | -CnHm |  | -5.41 | -1.92 | 3.49 | (Chen et al., 2021) |
| C36H12 | 12 | - |  | -5.23 | -4.31 | 0.92 | (Kozliak et al., 2019) |
| C36H14O7 | 12 | -OH(2)  -O-(5)  -CnHm |  | -5.67 | -1.87 | 3.80 | (Saldinger et al., 2020) |
| C36H16 | 11 | - |  | -5.31 | -2.31 | 2.99 | (Adamson et al., 2018) |
| C36H16 | 11 | - |  | -5.10 | -2.36 | 2.74 | (Commodo et al., 2019) |
| C36H16 | 11 | - |  | -5.47 | -2.54 | 2.94 | (Commodo et al., 2019) |
| C36H16 | 11 | - |  | -5.08 | -2.39 | 2.69 | (Commodo et al., 2019) |
| C36H17 | 11 | -CnHm |  | -4.35  -5.6 | -1.79  -3.07 | 2.56  2.53 | (Commodo et al., 2019) |
| C36H18 | 10 | - |  | -5.62 | -2.59 | 3.04 | (Pascazio et al., 2020) |
| C36H18 | 11 | - |  | -4.67 | -3.26 | 1.41 | (Pascazio et al., 2020) |
| C36H20 | 10 | - |  | -5.30 | -2.34 | 2.96 | (Pascazio et al., 2020) |
| C36H20 | 9 | - |  | -5.72 | -1.90 | 3.81 | (Elvati et al., 2019) |
| C36H22 | 9 | -CnHm |  | -5.28 | -2.25 | 3.03 | (Chen et al., 2021) |
| C36H24 | 7 | - |  | -6.00 | -1.63 | 4.37 | (Michelsen, 2020) |
| C37H16 | 12 | - |  | -5.18 | -2.14 | 3.04 | (Commodo et al., 2019) |
| C37H20 | 12 | - |  | -4.31 | -3.16 | 1.14 | (Commodo et al., 2019) |
| C37H28 | 8 | -CnHm |  | -5.54 | -1.80 | 3.74 | (Chen et al., 2021) |
| C37H30 | 9 | -CnHm |  | -5.54 | -1.73 | 3.80 | (Schulz et al., 2019) |
| C37H30 | 9 | -CnHm |  | -5.46 | -1.60 | 3.86 | (Commodo et al., 2019) |
| C38H16 | 12 | - |  | -4.93 | -3.24 | 1.69 | (Commodo et al., 2019) |
| C38H16 | 12 | - |  | -5.00 | -3.33 | 1.68 | (Martin et al., 2019) |
| C38H16 | 12 | - |  | -5.15 | -2.46 | 2.69 | (Lowe et al., 2015) |
| C38H18 | 12 |  |  | -5.17 | -3.43 | 1.73 | (Wang et al., 2020) |
| C39H28 | 9 | -CnHm |  | -5.19 | -2.18 | 3.00 | (Chen et al., 2021) |
| C40H16 | 13 | - |  | -4.74 | -3.02 | 1.71 | (Commodo et al., 2019) |
| C40H18 | 13 |  |  | -5.54 | -3.12 | 2.42 | (Wang et al., 2020) |
| C40H21 | 12 | -CnHm |  | -4.75  -5.56 | -2.19  -3.44 | 2.56  2.12 | (Commodo et al., 2019) |
| C40H22 | 10 | - |  | -5.79 | -1.84 | 3.96 | (Elvati et al., 2019) |
| C41H18 | 12 | -CnHm |  | -5.46 | -3.13 | 2.33 | (Wang et al., 2020) |
| C42H14 | 14 |  |  | -4.86 | -4.03 | 0.83 | (Wang et al., 2020) |
| C42H16 | 14 | - |  | -5.08 | -2.57 | 2.52 | (Lowe et al., 2015; Pascazio et al., 2019) |
| C42H18 | 13 | - |  | -5.57 | -2.02 | 3.55 | (Michelsen, 2020) |
| C44H20 | 15 | - |  | -4.72 | -2.61 | 2.11 | (Schulz et al., 2019) |
| C44H20 | 15 | - |  | -4.67 | -2.55 | 2.12 | (Commodo et al., 2019) |
| C47H18 | 15 |  |  | -5.14 | -3.67 | 1.47 | (Wang et al., 2020) |
| C47H18 | 15 | -CnHm |  | -5.22 | -4.09 | 1.13 | (Wang et al., 2020) |
| C48H16O4 | 17 | -O-(4) |  | -5.07 | -2.22 | 2.85 | (Elvati et al., 2019) |
| C48H18 | 16 | -CnHm |  | -5.42 | -3.17 | 2.25 | (Wang et al., 2020) |
| C48H20 | 15 | - |  | -5.18 | -2.48 | 2.70 | (Michelsen, 2020) |
| C49H16 | 8 |  |  | -4.65 | -3.90 | 0.75 | (Wang et al., 2020) |
| C49H22 | 16 |  |  | -5.65 | -2.87 | 2.79 | (Wang et al., 2020) |
| C50H16O5 | 18 | -O-(5) |  | -5.02 | -2.21 | 2.81 | (Elvati et al., 2019) |
| C50H24 | 15 | -CnHm |  | -5.24 | -2.36 | 2.87 | (Commodo et al., 2019) |
| C50H24O | 15 | -O- |  | -5.54 | -2.72 | 2.86 | (Wang et al., 2020) |
| C51H18 | 15 | -CnHm |  | -5.83 | -3.41 | 2.42 | (Wang et al., 2020) |
| C52H22 | 16 | - |  | -5.38 | -2.89 | 2.49 | (Commodo et al., 2019) |
| C52H26 | 16 | - |  | -4.96 | -2.70 | 2.26 | (Schulz et al., 2019; Pascazio et al., 2020) |
| C52H26 | 16 | - |  | -4.89 | -2.63 | 2.26 | (Commodo et al., 2019) |
| C52H26 | 15 | -CnHm |  | -5.11 | -2.45 | 2.66 | (Chen et al., 2021) |
| C54H18 |  | - |  | -5.25 | -2.44 | 2.81 | (Lowe et al., 2015) |
| C54H20 | 18 | - |  | -5.08 | -2.57 | 2.51 | (Lowe et al., 2015) |
| C54H28 | 16 | -CnHm |  | -4.77 | -2.66 | 2.11 | (Chen et al., 2021) |
| C56H20 |  |  |  | -4.95 | -3.83 | 1.12 | (Wang et al., 2020) |
| C56H36 | 14 | -CnHm |  | -5.04 | -2.35 | 2.69 | (Chen et al., 2021) |
| C57H34 | 15 | -CnHm |  | -5.20 | -2.32 | 2.88 | (Chen et al., 2021) |
| C64H30 | 19 |  |  | -5.41 | -2.94 | 2.47 | (Wang et al., 2020) |
| C66H20 | 24 | - |  | -4.90 | -2.84 | 2.06 | (Lowe et al., 2015) |
| C66H30 | 19 | - |  | -5.10 | -2.58 | 2.52 | (Chen et al., 2021) |
| C66H66 | 13 | -CnHm |  | -5.27 | -1.78 | 3.49 | (Lowe et al., 2015) |
| C71H40 | 19 | -CnHm |  | -4.97 | -2.89 | 2.07 | (Chen et al., 2021) |
| C74H22 | 27 | - |  | -4.75 | -2.98 | 1.78 | (Lowe et al., 2015) |
| C80H22 | 30 | - |  | -4.80 | -2.97 | 1.84 | (Lowe et al., 2015) |
| C90H24 | 34 | - |  | -4.63 | -3.16 | 1.48 | (Lowe et al., 2015) |
| C96H24 | 37 | - |  | -4.95 | -2.84 | 2.11 | (Lowe et al., 2015) |

1 Including five-member ring and six-member ring.

**2. Computational methods**

In this work, we evaluate the accuracy of our computational method by comparing the computed HOMO-LUMO gap of PAHs with the measured optical bandgap from experiments (Parac and Grimme, 2003; Menon et al., 2019a), as shown in Fig. S1. The TDDFT predictions (Menon et al., 2019b) are also included for comparison. For the PAHs considered, the average error between the calculated values in this work and the experimental ones is 0.28 eV. The error value decreases with the increase of the number of carbon atoms. Compared to the TDDFT predictions, our method overestimates the gap value by 0.2 eV on average. Overall, our results are consistent with the experimental and computed gap values reported in previous works.



Figure. S1. Experimental and computed gap values of nine PAHs. The black points represent the measured values (Parac and Grimme, 2003; Menon et al., 2019a). The red and blue points refer to the computed gap values using TDDFT method (Menon et al., 2019) and the DFT method at the level of B3LYP/6-311+G(d,p). Lines are drawn to guide the eyes.

The impact of basis set is further explored by comparing our results with a smaller basis set as B3LYP/6-31G(d) method (Li et al., 2020). We selected several PAHs molecules in the Table S1 as the target to evaluate the calculation results behind the two methods. The predicted HOMO-LUMO gap are almost consistent with each other; the maximum error is 0.09 eV, and the average error is 0.04 eV (Table S2).

**Table S2.** HOMO-LUMO gap values from two computational methods.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | B3LYP/6-311+G(d,p) | | | B3LYP/6-31G(d) | | |  |
| Molecular formula | HOMO | LUMO | *EL-H* | HOMO | LUMO | *EL-H* | △*E*(eV)1 |
| C20H12 | -5.45 | -2.1 | 3.35 | -5.1 | -1.72 | 3.38 | 0.03 |
| C13H10O2 | -6.17 | -2.43 | 3.74 | -5.92 | -2.16 | 3.76 | 0.03 |
| C8H10 | -5.72 | -1.75 | 3.97 | -5.44 | -1.45 | 3.99 | 0.02 |
| C12H8O4 | -5.63 | -1.18 | 4.46 | -5.39 | -0.86 | 4.53 | 0.07 |
| C14H12 | -5.76 | -1.65 | 4.11 | -5.51 | -1.34 | 4.17 | 0.06 |
| C24H18 | -6.10 | -1.45 | 4.65 | -5.85 | -1.11 | 4.74 | 0.09 |
| C14H10O | -5.69 | -2.10 | 3.59 | -5.43 | -1.82 | 3.61 | 0.01 |
| C18H10O3 | -5.50 | -1.21 | 4.29 | -5.26 | -0.93 | 4.33 | 0.04 |
| C54H18 | -5.19 | -2.37 | 2.82 | -4.93 | -2.10 | 2.83 | 0.01 |

1 The difference between two methods

**3. The date sets of Planar, nonplanar and linear PAHs**

To evaluate the prediction quality from the machine learning model, some PAHs from three groups of PAHs (Planar, Nonplanar and Linear) were selected as the target. The detail of PAH is included in Table S3-S5.

**Table S3.** Structure and band information of PAHs with planar structures.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Molecular Formula | Structure | HOMO | LUMO | *EH-L*(eV) |
| C20H12 |  | -5.45 | -2.10 | 3.35 |
| C22H10 |  | -5.82 | -3.27 | 2.55 |
| C22H12 |  | -5.54 | -2.05 | 3.48 |
| C24H10 |  | -5.89 | -3.57 | 2.33 |
| C24H12 |  | -5.78 | -1.78 | 4.00 |
| C28H12 |  | -5.89 | -2.58 | 3.32 |
| C28H14 |  | -5.61 | -1.96 | 3.65 |
| C30H12 |  | -5.25 | -3.63 | 1.62 |
| C30H14 |  | -5.40 | -2.20 | 3.20 |
| C32H14 |  | -5.27 | -2.35 | 2.92 |
| C32H14 |  | -5.26 | -3.51 | 1.75 |
| C38H16 |  | -5.15 | -2.46 | 2.69 |
| C42H16 |  | -5.08 | -2.57 | 2.52 |
| C54H18 |  | -5.25 | -2.44 | 2.81 |
| C54H20 |  | -5.08 | -2.57 | 2.51 |
| C66H20 |  | -4.90 | -2.84 | 2.06 |
| C74H22 |  | -4.75 | -2.98 | 1.78 |
| C80H22 |  | -4.80 | -2.97 | 1.84 |
| C90H24 |  | -4.63 | -3.16 | 1.48 |
| C96H24 |  | -4.95 | -2.84 | 2.11 |

**Table S4.** Structure and band information of PAHs with nonplanar structures.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Molecular Formula | Structure | HOMO | LUMO | *EH-L*(eV) |
| C47H18 |  | -5.14 | -3.67 | 1.47 |
| C56H20 |  | -4.95 | -3.83 | 1.12 |
| C40H18 |  | -5.54 | -3.12 | 2.42 |
| C64H30 |  | -5.41 | -2.94 | 2.47 |
| C48H18 |  | -5.42 | -3.17 | 2.25 |
| C49H22 |  | -5.65 | -2.87 | 2.79 |
| C34H18 |  | -5.36 | -2.60 | 2.76 |
| C18H14 |  | -6.25 | -1.15 | 5.10 |
| C36H24 |  | -6.00 | -1.63 | 4.37 |
| C32H18 |  | -5.71 | -1.90 | 3.81 |
| C36H20 |  | -5.72 | -1.90 | 3.81 |

**Table S5.** Structure and band information of linear PAHs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Molecular Formula | Structure | HOMO | LUMO | *EH-L*(eV) |
| C14H12 |  | -6.65 | -0.68 | 5.97 |
| C12H10 |  | -6.26 | -1.39 | 4.87 |
| C14H10 |  | -5.58 | -2.02 | 3.55 |
| C10H8 |  | -6.15 | -1.40 | 4.75 |
| C26H16 |  | -4.75 | -2.96 | 1.78 |
| C22H14 |  | -4.94 | -2.75 | 2.19 |
| C18H14 |  | -5.93 | -1.76 | 4.17 |
| C18H12 |  | -5.20 | -2.45 | 2.75 |

**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. *Journal of Physical Chemistry A* 122(48)**,** 9338-9349. doi: 10.1021/acs.jpca.8b08947.

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 : an American Chemical Society journal*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 by high-resolution atomic force microscopy. *Combustion and Flame* 205**,** 154-164. doi: 10.1016/j.combustflame.2019.03.042.

Elvati, P., Turrentine, K., and Violi, A. (2019). The Role of Molecular Properties on the Dimerization of Aromatic Compounds. *Proceedings of the Combustion Institute* 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. *Proceedings of the Combustion Institute* 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. *Physical Chemistry Chemical Physics* 22(9)**,** 5314-5331. doi: 10.1039/d0cp00116c.

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. *Astrophysical Journal* 889(2). 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. 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. *Proceedings of the Combustion Institute* 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. *Proceedings of the Combustion Institute* 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. *Proceedings of the National Academy of Sciences of the United States of America* 113(30)**,** 8374-8379. doi: 10.1073/pnas.1604772113.

Kholghy, M.R., Kelesidis, G.A., and Pratsinis, S.E. (2018). Reactive Polycyclic Aromatic Hydrocarbon Dimerization Drives Soot Nucleation. *Physical Chemistry Chemical Physics* 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. *The Journal of Physical Chemistry A* 117(23)**,** 4794-4816. doi: 10.1021/jp402481y.

Kozliak, E., Sulkes, M., Alhroub, I., Kubatova, A., Andrianova, A., and Seames, W. (2019). Influence of Early Stages of Triglyceride Pyrolysis on the Formation of PAHs as Coke Precursors. *Physical Chemistry Chemical Physics* 21(36)**,** 20189-20203. doi: 10.1039/c9cp02025j.

Lafleur, A.L., Howard, a.B., Man, J.A., and Yadav, T. (1993). Proposed Fullerene Precursor Corannulene Identified in Flames Both in the Presence and Absence of Fullerene Production. *The Journal of Physical Chemistry A* 97（51）**,** 13539-13543.

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. *Proceedings of the Combustion Institute*. 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. *Proceedings of the Combustion Institute*. 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.

Lowe, J.S., Lai, J.Y.W., Elvati, P., and Violi, A. (2015). Towards a Predictive Model for Polycyclic Aromatic Hydrocarbon Dimerization Propensity. *Proceedings of the Combustion Institute* 35(2)**,** 1827-1832. doi: 10.1016/j.proci.2014.06.142.

Mao, Q., Hou, D., Luo, K.H., and You, X. (2018). Dimerization of Polycyclic Aromatic Hydrocarbon Molecules and Radicals under Flame Conditions. *J Phys Chem A* 122(44)**,** 8701-8708. doi: 10.1021/acs.jpca.8b07102.

Martin, J.W., Hou, D., Menon, A., Pascazio, L., Akroyd, J., You, X., et al. (2019). Reactivity of Polycyclic Aromatic Hydrocarbon Soot Precursors: Implications of Localized π-Radicals on Rim-Based Pentagonal Rings. *The Journal of Physical Chemistry C* 123(43)**,** 26673-26682. doi: 10.1021/acs.jpcc.9b07558.

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

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

Michelsen, H.A. (2020). Effects of Maturity and Temperature on Soot Density and Specific Heat. *Proceedings of the Combustion Institute*. doi: 10.1016/j.proci.2020.06.383.

Parac, M., and Grimme, S. (2003). A TDDFT Study of the Lowest Excitation Energies of Polycyclic Aromatic Hydrocarbons. *Chemical Physics* 292(1)**,** 11-21. doi: 10.1016/s0301-0104(03)00250-7.

Pascazio, L., Martin, J.W., Botero, M.L., Sirignano, M., D’Anna, A., and Kraft, M. (2019). Mechanical Properties of Soot Particles: The Impact of Crosslinked Polycyclic Aromatic Hydrocarbons. *Combusion Science and Technology* 1-21, doi:10.1080/00102202.2019.1668380 doi: 10.1080/00102202.2019.1668380.

Pascazio, L., Martin, J.W., Menon, A., Hou, D., You, X., and Kraft, M. (2020). Aromatic Penta-Linked Hydrocarbons in Soot Nanoparticle Formation. *Proceedings of the Combustion Institute*. 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. 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. *Proceedings of the Combustion Institute* 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. 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. (2020). Molecular structures in flames: A comparison between SNapS2 and recent AFM results. *Proceedings of the Combustion Institute* 000**,** 1-9. doi: 10.1016/j.proci.2020.06.250.

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. *ACS Symposium Series; American Chemical Society: Washington, DC* 1320(3)**,** 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 Nnder Different Hydrogen Addition Conditions at High Temperature by ReaxFF Molecular Dynamics. *Fuel* 262. doi: 10.1016/j.fuel.2019.116677.