Fuel 338 (2023) 127304

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Full Length Article

Application of principal-component analysis to the interpretation of coal tar physico-chemical properties

Man Zhang [a,b], Xining Zang [a,b,*]

[a] Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China
[b] Key Laboratory for Advanced Materials Processing Technology, Ministry of Education, Beijing, Tsinghua University, 100084, China

ARTICLE INFO

Keywords:
Principal-component analysis
Density functional theory
Coal tar
Polyaromatic hydrocarbons

ABSTRACT

Coal tars, mixtures of heavy polyaromatic hydrocarbons (PAHs) by-produced in the coking processes of coals, are used as resources for further refining, burnt for carbon black, or just disposed in landfills. Direct application of rich PAHs in coal tars in renewable applications such as photovoltaic devices will not only reduce the carbon emission and waste production, but also upgrade the tar-to-product chain with higher additive value. Decoding the component of coal tars and their optical properties is the basis to apply these complex PAHs in optical devices. Herein, we perform density functional theory (DFT) simulations and time-dependent DFT (TDDFT) simulations of all organic molecules in coal tar, calculating their molecular dipole moment, optical gap, fundamental gap, Urbach tail energy, vibrational spectrum, etc. Using Principal-component analysis (PCA) with maximum variance criterion, we decouple three principal components associated with structural, electronic, and vibrational properties. Through comparison between simulation results and experiments, an electronic-scale comprehension for the correlation of properties can be convinced, contributing to the interpretation of coal tar.

1. Introduction

Coal tars are mixtures of polyaromatic hydrocarbons (PAHs) condensed in the pipelines during the coal coking processes, which are typically used as resources for further refining, burnt for carbon black, or just disposed in landfills [1,2]. It can be further distilled to produce light oil, carbolic oil, naphthalene oil, washing oil, anthracene oil, coal asphalt, and other products by steam cracking, composed of 25 kinds of organic molecules (excluding isomers) and bitumen (50 %–60 %) [3]. The extensive structural and chemical richness guarantees tunable electricity and physical properties, making it economically and environmental-friendly appealing for electric and optoelectronic applications [4]. The complexity of coal tars' chemistry, on the other hand, brings difficulties to simulate the physical properties and establish correlations between the characteristic and their molecular level structures. In this paper, 15 factors of chemical, electrical, optical, and spectral data of 45 molecules (with 11 types of isomers and PAHs in bitumen) in coal tar (Fig. 1) are generated using density functional theory (DFT) and time-dependent DFT (TDDFT). With the abundance of characterization data, principal component analysis (PCA) is performed to reduce the dimensions. The three principal components PC1, PC2, PC3 are linear combinations of C:H ratio, sp^3/sp^2 ratio, internal energy (E), enthalpy (H), Gibbs free energy (G), single-point energy (S), molecular mass, dipole moment, Urbach tail energy (E_u), optical gap ($E_{optical}$), fundamental gap ($E_{fundamental}$), optical band gap (E_g), and vibration characteristics. Experimental data show that the coal tar is mainly composed of PAHs with 4–6 rings (molecular mass in between 217 and 300). And the simulated optical band gap ($3 \sim 4$ eV) is wider than the measured value (2.857 eV), which can be explained by the solid-state effect and impurity influence. This study provides new insights into an electronic-scale understanding for the coal tar chemical properties, which can be expanded to other heavy hydrocarbon interpretation.

2. Methodology

2.1. Organic molecules simulation.

In DFT formalism, the energy of an electronic system can be regarded as a function of electron density, which in turn is a function of spatial coordinates. For a system with n electrons, the energy E relates to electron density with three coordinates, rather than wavefunctions with 3n or 4n (with spin) coordinates [5,6], which can be given by

* Corresponding author at: Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China.
E-mail address: xzang@mail.tsinghua.edu.cn (X. Zang).

https://doi.org/10.1016/j.fuel.2022.127304
Received 17 November 2022; Received in revised form 19 December 2022; Accepted 24 December 2022
Available online 31 December 2022

$$E(n) = T_s(n) + V(n) + U_H(n) + E_{xc} \tag{1}$$

In equation (1), $T_s(n)$, $V(n)$, and $U_H(n)$ are the kinetic energy term that ignores the interaction between electrons, the external potential energy term which is obtained by Born Oppenheimer approximation, and the energy term about the interaction of electrons obtained by Thomas Fermi model, respectively. The last term is the error between the approximation energy and the exact value, which is called the exchange–correlation energy. For E_{xc} calculation, the generalized gradient approximation method (GGA) is designed for spatially inhomogeneous systems, where E_{xc} is dependent on both density and the gradient of the density. On that basis, hybrid functionals combing GGA with a percentage Hartree-Fock exchange is the most widely used approach for atomization energy simulation.

Becke's three-parameter hybrid exchange functional [7,8,9] (B3) using the correlational functional of Lee, Yang, and Parr [10] (LYP) along with 6-31G(d) pople split valence basis sets, have been proved to be favorable for accurate calculation of organic molecules with a compromise of expense and accuracy [11]. Molecular orbitals, internal energy, enthalpy, Gibbs free energy, single-point energy, molecular mass, and dipole moment can be obtained by single-point energy calculation performed on the fully optimized molecular geometry. And on that basis, the excited state simulation can be executed to model the molecular vibration characteristics and ultraviolet–visible (UV–vis) absorption spectrum, where Urbach tail energy and the optical gap can be acquired. DFT and TDDFT were performed at B3lyp/6-31G(d) level using Gaussian 16 package with software copyright [12].

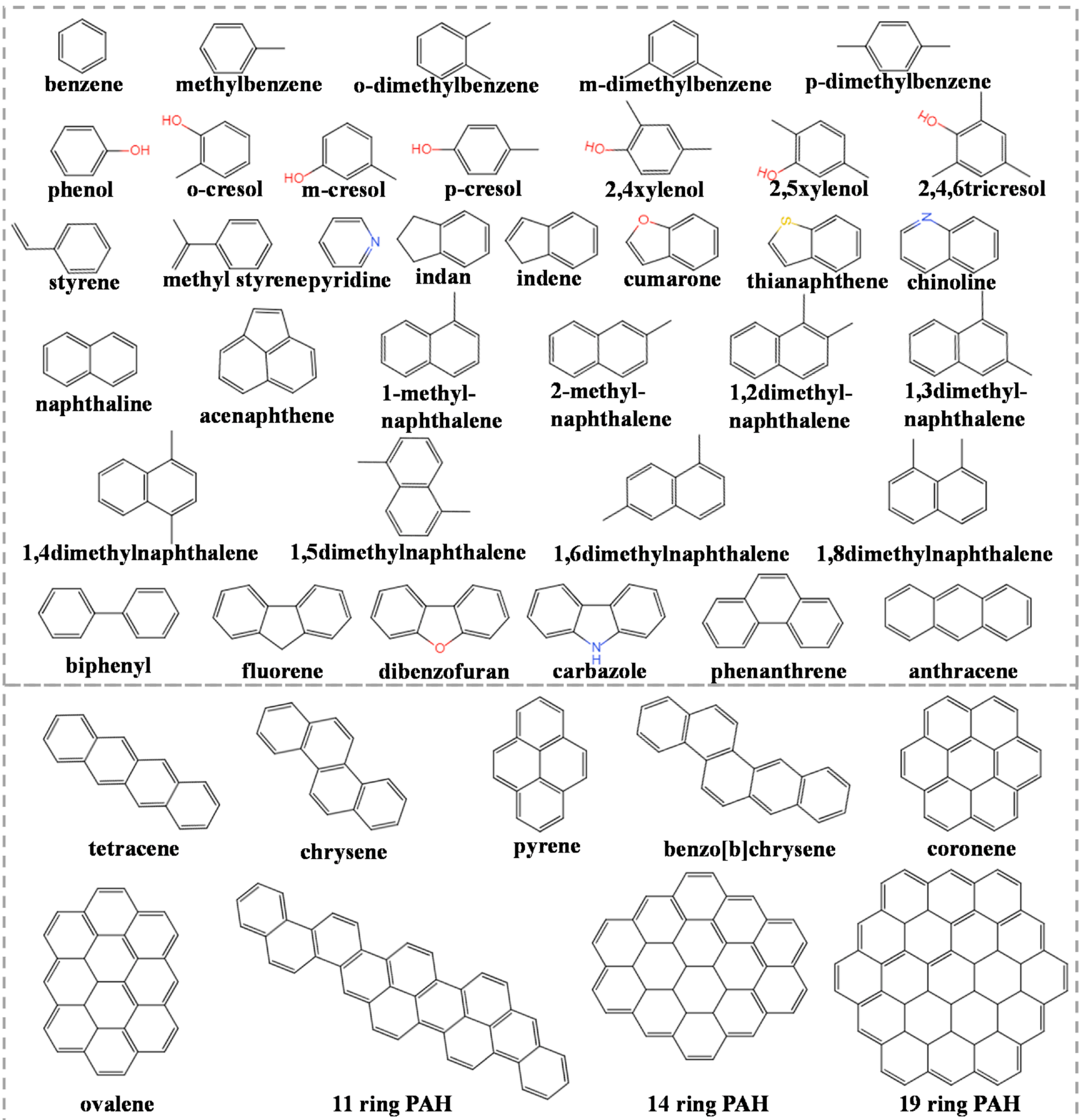

Fig. 1. Molecular structures of the investigated small organic molecules and PAHs.

2.2. Data processing.

For 45 kinds of small organic molecules and PAHs, C:H ratio and sp^3/sp^2 ratio were counted carefully, and internal energy (E), enthalpy (H), Gibbs free energy (G), single-point energy (S), molecular mass, and dipole moment were calculated via Gaussian software. For confusing single point energy and internal energy, the former refers to the energy of a molecule with a given geometric configuration in 273.15 K considered as a point on the potential surface and the latter refers to energy at 298.15 K containing translational, rotational, and vibrational energy. And vibration characteristics were also identified, including the alkyl C—H stretch region near 3000 cm^{-1}, aromatic skeleton (1450–1600 cm^{-1}), and hydroxyl (~3736 cm^{-1}) vibrations in simulated Infrared spectroscopy. Furthermore, for these neutral molecules, molecular orbitals can be acquired, of which the highest occupied and lowest unoccupied orbitals (HOMO & LUMO) gap was singled out. In the DFT calculation method following Koopmans' theorem, the energy of the HOMO level is a good approximation to the (minus) vertical ionization potential (IP), while the LUMO level corresponds to the (minus) vertical electron affinity (EA) [13]. Therefore, the deviation of HOMO and LUMO can represent the fundamental gap, which is defined as IP-EA, as exhibited in Fig. 2(a).

Another kind of gap in the isolated molecular group, optical gap, can be determined by UV–vis absorption spectrum through TDDFT. The energy of the optical gap is defined as the excitation energy corresponding to the energy of the lowest excited state by absorbing photons. There is still electrostatic bounding between electron and hole in the excited state, therefore, the optical gap would be narrower than the fundamental gap as depicted in Fig. 2(a). Ten excited states were set in the simulation to ensure encompassing all the excitation conditions of target molecules. As listed in Table 1, all the simulated optical gaps meet the law, where the deviation is the electron-hole pair binding energy.

In addition, the DFT method can also be employed to investigate the optical band gap, where the intermolecular distance is long enough to ignore the interaction in the organic solution. The optical band gap for 45 kinds of organic liquids can be obtained with [14]:

$$(\alpha h\upsilon)^{1/n} = K\left(h\nu - E_g\right) \tag{2}$$

In equation (2), α is the absorption, hυ is the photon energy, K is a constant, and E_g represents the optical band gap. For the direct band gap semiconductor, n = 0.5, and for the indirect one, n = 2. For organic molecules, the carriers are more localized for intermolecular force, so it is prone to display the characteristics of a direct band gap. As shown in Fig. 2(b), E_g can be represented by the intercept on the x-coordinate for comparison.

In addition to the energy gap, Urbach energy is another important character of organic or amorphous semiconductors which is related to the density of tail states in semiconductors. As shown in Fig. 2(c), due to the structural disorder, carrier concentration, and the interaction between carriers and impurities, the band tail state composed of localized states is generated near the bottom of the conduction band and the top of the valence band, resulting in the absorption of photons with energy less than the width of the band gap [15,16,17]. The state density decays exponentially as expressed in equation (3) [18]:

$$\alpha = \alpha_0 exp\left(\left(h\nu - E_g\right)/E_u\right) \tag{3}$$

Where α_0 is a constant, E_u and E_g represent Urbach tail energy and band gap, respectively. As presented in Fig. 2(d), with logarithmic transformation, the Urbach energy value can be compared by the reciprocal of the decay slope.

Based on the multiple data simulated, principal-component analysis can be applied as a powerful multivariate technique for dimension reduction [19,20]. With a concentration on variances, variables with inter-correlation can be transformed into a group of orthogonal components [21] reserving at least 80 % of the information [22], named as principal components. With the more comprehensive and representative components, the quantitative evaluation and cluster analysis of the organic molecules in coal tar can be achieved.

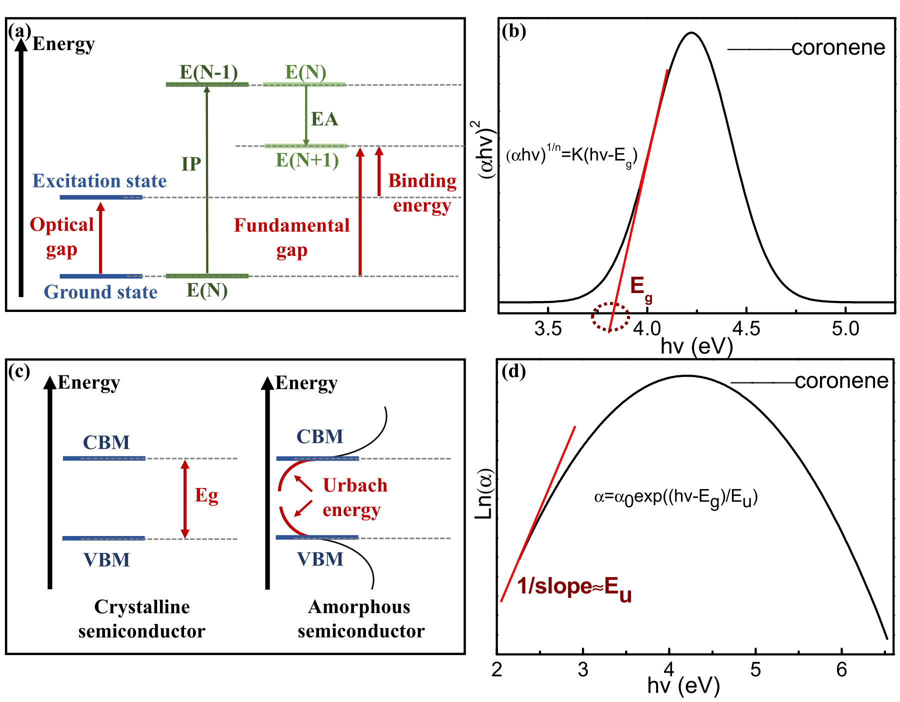

Fig. 2. (a) Gap identification for molecules and semiconductors (E(N) represents the electron energy of N electrons; Binding energy represents the electron-hole pair binding energy); (b) Fitting optical band gap; (c)E_u illustration; (d) Fitting E_u. Using coronene as an example.

Table 1

The 15 properties of 45 kinds of small organic molecules and PAHs calculated based on DFT and TDDFT methods.

Small molecule & PAHs	CHr	Sp.r	E (Hartree)	H (Hartree)	G (Hartree)	S (Hartree)	Mass (amu.)	Dip.M (Debye.)	E_u (eV)	E_o (eV)	E_f (eV)	E_g (eV)	IR_a (D)	IR_O (D)	IR_C (D)
benzene	1	0	−232.14	−232.14	−232.18	−232.25	78.047	0.0001	0.0417	5.545	6.801	6.984	37.78	0	0
Methyl-benzene	0.875	0.167	−271.43	−271.43	−271.47	−271.57	92.063	0.320	0.0776	5.436	6.549	6.773	35.49	0	99.15
o-dimethyl-benzene	0.8	0.333	−310.72	−310.72	−310.76	−310.88	106.08	0.546	0.0810	5.390	6.448	6.680	96.28	0	182.4
m-dimethyl-benzene	0.8	0.333	−310.72	−310.72	−310.76	−310.89	106.08	0.296	0.0819	5.347	6.411	6.587	58.82	0	131.1
p-dimethyl-benzene	0.8	0.333	−310.72	−310.72	−310.76	−310.89	106.08	0.0724	0.0822	5.324	6.326	6.626	83.41	0	233.2
phenol	1	0	−307.35	−307.35	−307.38	−307.46	94.042	1.60	0.0756	5.574	6.479	6.804	151.2	145.26	0
o-cresol	0.875	0.167	−346.64	−346.64	−346.68	−346.78	108.06	1.67	0.0630	5.184	5.985	6.584	232.7	141.84	120.4
m-cresol	0.875	0.167	−346.64	−346.64	−346.68	−346.78	108.06	1.09	0.0617	5.172	5.987	6.536	190.5	119.40	76.63
p-cresol	0.875	0.167	−346.64	−346.64	−346.68	−346.78	108.06	1.33	0.0685	5.092	5.812	6.652	296.0	135.97	110.4
2,4-xylenol	0.8	0.333	−385.93	−385.93	−385.98	−386.10	122.07	1.07	0.0646	5.083	5.837	6.499	140.2	116.65	151.6
2,5-xylenol	0.8	0.333	−385.93	−385.93	−385.98	−386.10	122.07	1.38	0.0642	5.117	5.936	6.410	210.2	125.62	144.4
2,4,6-tri-methylphenol	0.75	0.5	−425.22	−425.22	−425.27	−425.42	136.09	1.38	0.0656	5.051	5.827	6.352	190.0	90.634	137.1
styrene	1	0	−309.51	−309.51	−309.55	−309.65	104.06	0.191	0.0697	4.929	5.201	4.721	30.27	0	0
methyl styrene	0.9	0.125	−348.80	−348.80	−348.84	−348.97	118.08	0.413	0.0768	4.879	5.114	4.609	41.10	0	107.7
pyridine	1	0	−248.19	−248.19	−248.22	−248.29	79.042	2.19	0.0736	4.913	6.261	7.188	53.01	0	0
indan	0.9	0.5	−348.82	−348.82	−348.85	−348.99	118.08	0.505	0.0879	5.310	6.283	6.539	35.89	0	201.4
indene	1.125	0.125	−347.62	−347.62	−347.65	−347.76	116.06	0.663	0.0734	4.866	5.280	4.580	32.89	0	57.03
cumarone	1.333	0	−383.55	−383.55	−383.59	−383.67	118.04	0.743	0.0665	5.083	5.527	4.817	109.6	0	0
benzo-thiophene	1.333	0	−706.53	−706.53	−706.57	−706.65	134.02	0.738	0.0936	4.723	5.238	5.373	57.08	0	0
quinoline	1.286	0	−401.79	−401.79	−401.83	−401.93	129.06	2.02	0.0784	4.290	4.907	5.671	69.29	0	0
naphthalene	1.25	0	−385.74	−385.74	−385.78	−385.89	128.06	0	0.0755	4.461	4.829	5.731	24.26	0	0
1-Methyl- naphthalene	1.1	0.1	−425.03	−425.03	−425.07	−425.21	142.08	0.295	0.0788	4.380	4.751	5.616	23.29	0	100.0
2-Methyl- naphthalene	1.1	0.1	−425.03	−425.03	−425.07	−425.21	142.08	0.417	0.0781	4.421	4.835	5.599	19.69	0	109.1
1,2-dimethyl-naphthalene	1	0.2	−464.31	−464.31	−464.36	−464.52	156.09	0.593	0.0862	4.285	4.659	5.481	37.28	0	166.5
1,3-dimethyl-naphthalene	1	0.2	−464.32	−464.31	−464.36	−464.53	156.09	0.454	0.0842	4.365	4.761	5.503	35.93	0	213.6
1,4-dimethyl-naphthalene	1	0.2	−464.31	−464.31	−464.36	−464.53	156.09	0.0283	0.0898	4.293	4.671	5.476	41.52	0	227.4
1,5-dimethyl-naphthalene	1	0.2	−464.31	−464.31	−464.36	−464.53	156.09	0	0.0849	4.311	4.679	5.530	59.85	0	209.4
1,6-dimethyl-naphthalene	1	0.2	−464.31	−464.31	−464.36	−464.53	156.09	0.434	0.0856	4.356	4.760	5.493	37.71	0	210.5
1,8-dimethyl-naphthalene	1	0.2	−464.30	−464.30	−464.35	−464.52	156.09	0.537	0.0855	4.336	4.705	5.486	64.27	0	158.1
acenaphthene	1.2	0.2	−463.12	−463.12	−463.17	−463.32	154.08	0.818	0.0834	4.338	4.711	5.569	6.793	0	181.8
biphenyl	1.2	0	−463.12	−463.11	−463.16	−463.31	154.08	0.0001	0.0691	5.025	5.372	4.641	84.53	0	0
fluorene	1.3	0.083	−501.23	−501.23	−501.27	−501.42	166.08	0.482	0.0883	4.677	5.042	4.335	62.37	0	47.12
dibenzofuran	1.5	0	−537.16	−537.16	−537.20	−537.33	168.06	0.714	0.0690	4.533	5.087	4.301	131.7	0	0
carbazole	1.333	0	−517.29	−517.28	−517.33	−517.47	167.07	1.65	0.105	4.151	4.799	4.97	119.5	162.19	0
phenanthrene	1.4	0	−539.33	−539.33	−539.37	−539.53	178.08	0	0.103	3.276	3.594	4.901	14.40	0	0
anthracene	1.4	0	−539.33	−539.33	−539.38	−539.54	178.08	0.0417	0.0715	4.026	4.739	4.656	35.68	0	0
tetracene	1.5	0	−692.91	−692.91	−692.96	−693.17	228.09	0.0012	0.129	2.344	2.777	4.469	2.595	0	0
chrysene	1.5	0	−692.93	−692.93	−692.98	−693.18	228.09	0.0002	0.0881	3.873	4.245	4.450	7.457	0	0
pyrene	1.6	0	−615.56	−615.55	−615.60	−615.77	202.08	0	0.0899	3.543	3.843	3.162	11.09	0	0
Benzo[b]- chrysene	1.57	0	−846.52	−846.52	−846.58	−846.82	278.11	0.0303	0.0921	3.135	3.499	4.052	5.595	0	0
coronene	2	0	−921.60	−921.60	−921.66	−921.90	300.09	0	0.0492	3.297	4.037	3.820	1.539	0	0
ovalene	2.29	0	−1227.6	−1227.6	−1227.7	−1228.0	398.11	0.0003	0.0847	2.658	2.934	3.410	5.321	0	0
11 ring	1.909	0	−1613.3	−1613.3	−1613.4	−1613.9	526.17	0.0175	0.106	2.330	2.392	1.958	8.447	0	0
14 ring	2.625	0	−1609.9	−1609.9	−1610.0	−1610.4	520.13	0.0004	0.0998	2.361	2.524	1.992	6.156	0	0
19 ring	3	0	−2068.4	−2068.4	−2068.5	−2069.0	666.14	0.0002	0.0851	2.217	2.824	2.519	6.135	0	0

[a]CHr, C:H ratio; Sp.r, sp^3/sp^2 ratio; E, Internal energy; H, Enthalpy; G, Gibbs free energy; S, Single-point energy; Dip.M, Dipole Moment; E_o, optical gap; E_f, fundamental gap; IR_a, $IR_{aromatic}$; IR_O, IR_{OH}; IR_C, $IR_{alkyl\ C—H\ stretch}$; D, $10^{-40}esu^2cm^2$.

3. Results and discussion

3.1. PCA computation.

For the simulated 15 factors, PCA is computed in SPSS (Statistical Product Service Solutions) software, and three principal-components were figured out as the lithotripsy map depicted in Fig. 3(a). As listed in Table 2, the first principal component (PC1) was mainly composed of C:H ratio, E, H, G, single-point energy, molecular mass, fundamental gap, optical gap, optical band gap, and Urbach tail energy, which are all related to band structure of the molecule. The second component contained dipole moment, the character of the aromatic skeleton, and hydroxyl vibrations. The third principal component consisted of the sp^3/sp^2 ratio and the character of alkyl C—H stretch vibration, which can be concluded as aliphatic hydrocarbon characteristics. In the plot of the three principal component weights (Fig. 3(b)), the correlations among the 15 factors can be represented by the angle, where a small angle means a strong positive correlation, an angle near 180° means a negative correlation and unrelated features would manifest orthogonal lines [23]. For organic molecules in coal tar, the fundamental gap, optical gap, optical band gap, E, H, G, and single-point energy are positively correlated with each other, where the E, H, G, and single-point energy overlap completely in the plot, and negatively correlated with C:H ratio, molecular mass, and Urbach tail energy. And the three principal components are orthogonal to each other, with a clear exhibition in the plot. The correlation analysis could serve as a reference for a fast and convincing physical property analysis of heavy hydrocarbon substances.

The component matrix was applied to calculate the spatial coordinates of each component, with three principal components as coordinate axes. Based on the component matrix A_i (Table 1) and the eigenvalues λ_i of the three principal components in and total variance table (Table S1), the principal component load matrix U_i can be obtained:

$$U_i = A_i/\sqrt{\lambda_i} \tag{4}$$

The three-dimensional scatter diagram was drawn by multiplying the U_i with the standard values of 45 variables as shown in Fig. 4(a). The isomers are clustered in the diagram with similar properties, and the three principal component line charts are depicted in Fig. S1. The comprehensive score can be calculated as follows:

$$PCc = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}PC1 + \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}PC2 + \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}PC3 \tag{5}$$

For PAHs with ring number greater than 4 in coal tar (in bitumen), there is a declining trend as revealed in Fig. 4(a) and in the comprehensive score line chart in Fig. 4(b). The essential characteristics of organic molecules in coal tar can be clearly recognized with clustering

Table 2
The principal component matrix.

Factors	Principal Components		
	1	2	3
C:H ratio	−0.9482	0.1533	−0.1151
sp^3/sp^2 ratio	0.5439	−0.0151	0.7904
E	0.9448	−0.2320	−0.1750
H	0.9448	−0.2320	−0.1750
G	0.9448	−0.2320	−0.1750
S	0.9448	−0.2320	−0.1751
Mass	−0.9454	0.2062	0.1961
Dip.M	0.5281	0.6359	−0.1006
E_u	−0.4914	−0.2519	0.2396
E_o	0.9350	0.0225	−0.0444
E_f	0.9160	0.0779	−0.0361
E_g	0.9172	0.0342	0.0893
IR_a	0.5838	0.7260	0.0976
IR_O	0.3940	0.8212	0.0669
IR_C	0.4846	−0.2173	0.8007

analysis based on the comprehensive scores, as shown in Fig. 4(c). With the Ward method, the molecules are divided into four small groups: single benzene derivatives, naphthalene or biphenyl and their derivatives, PAHs with 4–7 aromatic rings, and PAHs with fused rings, respectively. Since tar is a mixture of refinery by-product, the complexity and variation of tar compositions from batch to batch increase the information redundancy to inventor them. Through PCA, we reduce the dimensions of tar component properties matrix that is essential for data storage, data processing and further screening of tar for target properties.

3.2. Experimental results.

The characterization experiments were performed to explore the coal tar molecular composition and verify the simulation results (details are provided in the supporting information). The distribution of molecular weight fragments in coal tar was obtained by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. As depicted in Fig. 5(a), the coal tar is composed of molecules with mass-to-charge ratios between 100 and 500, corresponding to the ingredient list provided by the manufacturer (Table S2). And the molecular fragments between 217 and 300 are dominant, proving that the PAHs with 4–6 rings in coal tar are more abundant based on the calculation results in Table 1. Therefore, with a combination of calculated optical band gaps, the simulated value for coal tar should be around 3 ~ 4 eV. The infrared spectrum is shown in Fig. 5(b), and the ratio of absorbance at peak 2 (2855–2950 cm^{-1}) to peak 1 (1450–1600 cm^{-1}) and peak 3 (~3736 cm^{-1}) to peak 1 are calculated for the convenience of comparison as listed in Table 3. Furthermore, for determining the value of

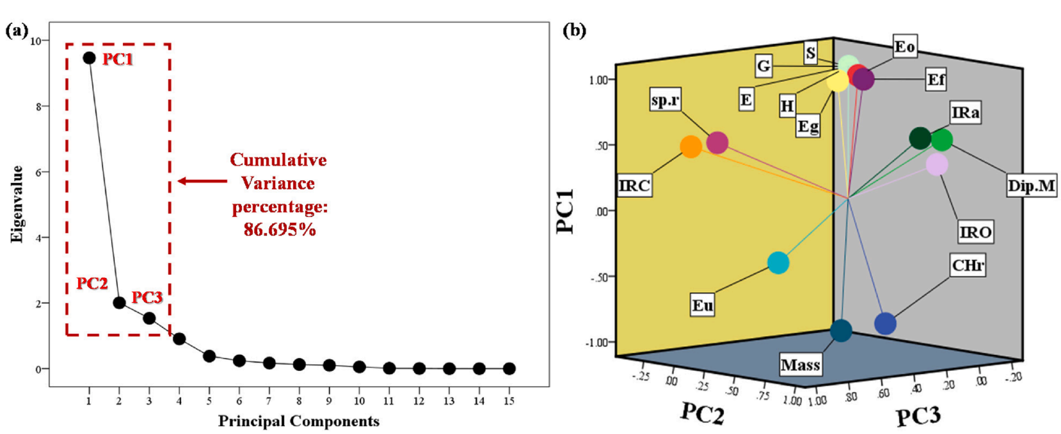

Fig. 3. (a) Lithotripsy map of influencing factors; (b) The three principal component weights.

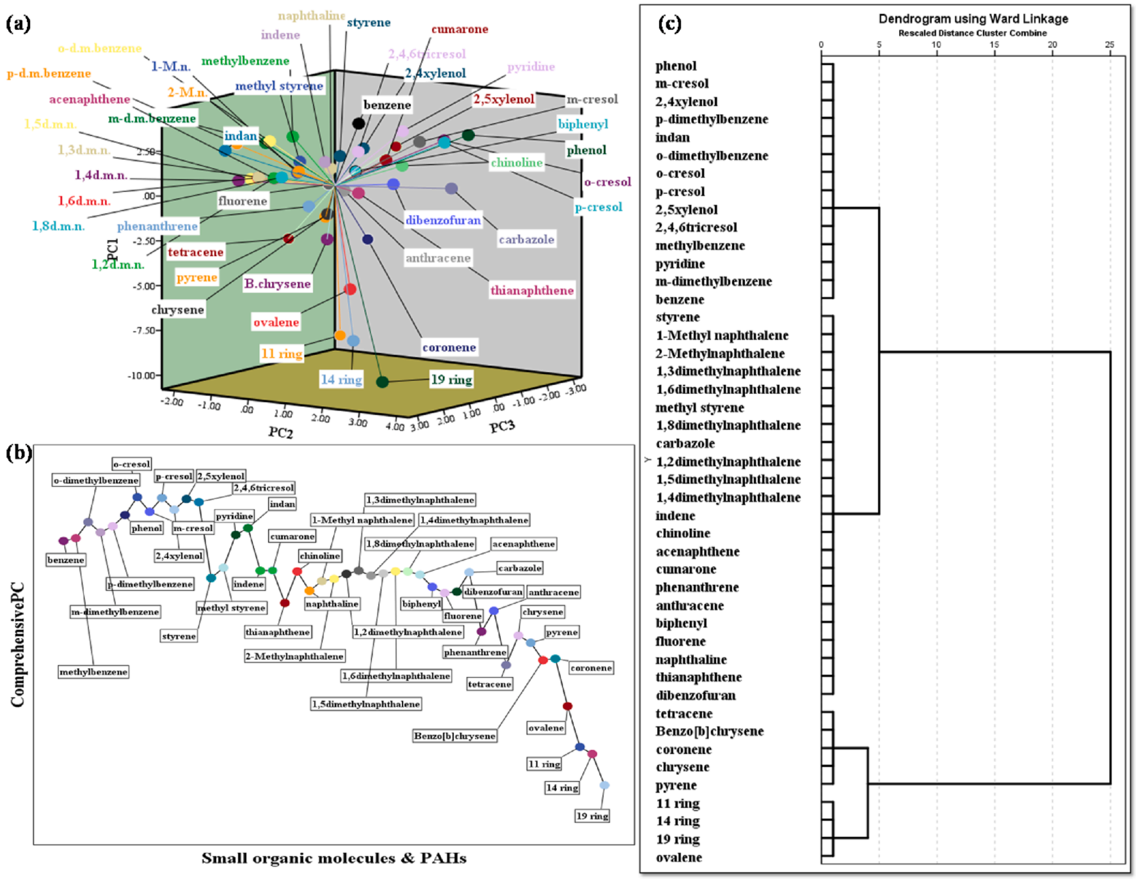

Fig. 4. (a) Three-dimensional scatter diagram of 45 variables; (b) The comprehensive score line chart; (c) The cluster analysis diagram based on the comprehensive score.

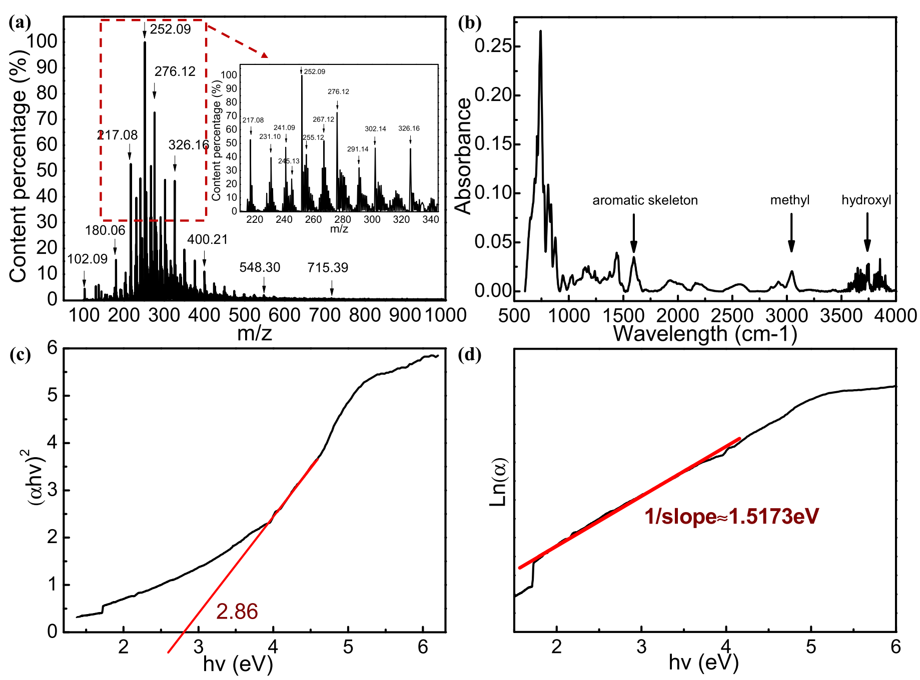

Fig. 5. (a) The distribution of molecular weight fragments measured by MALDI-TOF MS; (b) The infrared spectrum of coal tar; (c) The optical band gap value measured by UV–visible absorption spectrum; (d) Calculated E_u based on the measured UV–visible absorption spectrum.

Table 3
The comparison of simulated and characterization data.

Small molecules & PAHs	$IR_{OH}/IR_{aromatic}$ (calculated)	$IR_{CH3}/IR_{aromatic}$ (calculated)	E_g(c) (calculated) (eV)
benzene	0	0	6.984
methylbenzene	0	2.794	6.773
o-dimethylbenzene	0	1.894	6.680
m-dimethylbenzene	0	2.228	6.587
p-dimethylbenzene	0	2.795	6.626
phenol	0.9606	0	6.804
o-cresol	0.6095	0.5173	6.584
m-cresol	0.6267	0.4022	6.536
p-cresol	0.4594	0.3731	6.652
2,4-xylenol	0.8323	1.081	6.499
2,5-xylenol	0.5977	0.6868	6.410
2,4,6-trimethylphenol	0.4770	0.7216	6.352
styrene	0	0	4.721
methyl styrene	0	2.620	4.609
pyridine	0	0	7.188
indan	0	5.611	6.539
indene	0	1.734	4.580
cumarone	0	0	4.817
benzothiophene	0	0	5.373
quinoline	0	0	5.671
naphthalene	0	0	5.731
1-Methyl naphthalene	0	4.296	5.616
2-Methylnaphthalene	0	5.541	5.599
1,2-dimethylnaphthalene	0	4.466	5.481
1,3-dimethylnaphthalene	0	5.946	5.503
1,4-dimethylnaphthalene	0	5.476	5.476
1,5-dimethylnaphthalene	0	3.498	5.530
1,6-dimethylnaphthalene	0	5.582	5.493
1,8-dimethylnaphthalene	0	2.460	5.486
acenaphthene	0	26.76	5.569
biphenyl	0	0	4.641
fluorene	0	0.7555	4.335
dibenzofuran	0	0	4.301
carbazole	1.357	0	4.968
phenanthrene	0	0	4.901
anthracene	0	0	4.656
tetracene	0	0	4.469
chrysene	0	0	4.450
pyrene	0	0	3.162
Benzo[*b*]chrysene	0	0	4.052
coronene	0	0	3.820
ovalene	0	0	3.410
11 ring	0	0	1.958
14 ring	0	0	1.992
19 ring	0	0	2.519
Sample	**$IR_{OH}/IR_{aromatic}$ (measured)**	**$IR_{CH3}/IR_{aromatic}$ (measured)**	**E_g(c) (measured) (eV)**
Coal tar	0.5759	0.2761	2.857

the optical band gap, the absorption spectrum was computed on the powder of coal tar with naturally drying and grinding. As marked in Fig. 5(c), with the same processing method as Fig. 2a, the optical band gap value can be calculated as 2.857 eV. The deviation between the experimental value and the simulated result (3 ~ 4 eV) is mainly due to the stronger intermolecular interactions in solids and more ordered phase structures leading to narrower band gaps [24,25,26]. Furthermore, as illustrated in Fig. 2(c), the impurities of the real sample in the industrial grade would generate a stronger Urbach band tail as shown in Fig. 5(d), which also brings a smaller optical gap.

However, there is a broad distribution of energy levels and a shortage of long-range C=C conjunction as depicted in DFT simulation, leading to poor electrical conductivity in tar. Furthermore, with an abundance of the aromatic ring structure, there are a great deal of delocalized large π bonds and conjugate structures, resulting in the tendency for gaining and losing electrons. Therefore, the solar cell can also be a promising application, where both conductive electrode and semiconducting acceptor/donor can be processed from tar.

4. Conclusions

In conclusion, a non-destructive identification of 15 factors for 45 types of organic molecules in coal tar was computed with simulation, relating to element composition, semiconducting properties, electronic energy, and structure. DFT and TDDFT using the B3lyp/6-31G(d) basis set was computed with Gaussian 16 program to optimize molecule structure, calculate electronic energy, and simulate visible spectrum and pattern vibration characteristics. For avoiding time-consuming and laborious analysis, PCA was implemented to demonstrate the correlations between the factors with covariance. And in the maximum variance criteria, three principal components can be constructed representing structural, electronic, and vibrational properties, respectively.

CRediT authorship contribution statement

Man Zhang: Conceptualization, Methodology, Software, Writing – original draft. **Xining Zang:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

Prof. Xining Zang acknowledges the National Natural Science Foundation of China (number 52105459), Tsinghua University Initiative Scientific Research Program, and the Tsinghua University Start-up Funding.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fuel.2022.127304.

References

[1] Małachowska A, Mastalerz M, Hampton L, Hupka J, Drobniak A. Origin of bitumen fractions in the Jurassic-early Cretaceous Vaca Muerta Formation in Argentina: Insights from organic petrography and geochemical techniques. Int J Coal Geol 2019;205:155–65.
[2] Zang X, Dong Y, Jian C, Ferralis N, Grossman JC. Upgrading carbonaceous materials: Coal, tar, pitch, and beyond. Matter 2022;5(2):430–47.
[3] Lunge G. Coal-tar and ammonia. 5th and enl. ed. New York: D. Van Nostrand; 1916.
[4] Zang X, Jian C, Ingersoll S, Li H, Adams JJ, Lu Z, et al. Laser-engineered heavy hydrocarbons: Old materials with new opportunities. Sci Adv 2020;6(17):z5231.
[5] Sousa SF, Fernandes PA, Ramos MJ. General performance of density functionals. Chem A Eur J 2007;111(42):10439–52.
[6] Kostova I, Peica N, Kiefer W. Theoretical and spectroscopic studies of lanthanum (III) complex of 5-aminoorotic acid. Chem Phys 2006;327(2–3):494–505.
[7] Becke AD. Density-functional thermochemistry. 1. the effect of the exchange-only gradient correction. J Chem Phys 1992;96(3):2155–60.
[8] Becke AD. Density-functional thermochemistry. 2. the effect of the perdew-wang generalized-gradient correlation correction. J Chem Phys 1992;97(12):9173–7.
[9] Gelfand LS, Iaconianni FJ, Pytlewski LL, Speca AN, Mikulski CM, Karayannis NM. Nicotinic and isonicotinic-acid n-oxide interactions with 3d metal perchlorates. J Inorg Nucl Chem 1980;42(3):377–85.
[10] Lee CT, Yang WT, Parr RG. Development of the colle-salvetti correlation-energy formula into a functional of the electron-density. Phys Rev B 1988;37(2):785–9.
[11] Santhanamoorthi N, Lo C, Jiang J. Molecular design of porphyrins for Dye-Sensitized solar cells: A DFT/TDDFT study. J Phys Chem Lett 2013;4(3):524–30.
[12] Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.
[13] Bredas JL. Mind the gap! Mater Horiz 2014;1(1):17–9.
[14] Kilic G, Issa SAM, Ilik E, Kilicoglu O, Tekin HO. A journey for exploration of Eu_2O_3 reinforcement effect on zinc-borate glasses: Synthesis, optical, physical and nuclear radiation shielding properties. Ceram Int 2021;47(2):2572–83.
[15] Seyidov MY, Suleymanov RA, şale Y.. Unusual Urbach tail in $TlGaSe_2$ ferroelectric-semiconductor with incommensurate phase. J Appl Phys 2012;112(10):103106.
[16] Nishiwaki M, Nagaya K, Kato M, Fujimoto S, Tampo H, Miyadera T, et al. Tail state formation in solar cell materials: First principles analyses of zincblende, chalcopyrite, kesterite, and hybrid perovskite crystals. Phys Rev Mater 2018;2(8).
[17] Hebig J, Kühn I, Flohre J, Kirchartz T. Optoelectronic properties of $(CH_3NH_3)_3sb_2i_9$ thin films for photovoltaic applications. ACS Energy Lett 2016;1(1):309–14.
[18] Mishra V, Warshi MK, Sati A, Kumar A, Mishra V, Kumar R, et al. Investigation of temperature-dependent optical properties of TiO_2 using diffuse reflectance spectroscopy. SN Appl Sci 2019;1(3).
[19] Evans GT. Transformation of factor matrices to achieve congruence. British Journal of Mathematical & Statistical Psychology 1971; 24 (MAY):22-&.
[20] Krzanowski WJ. Between-groups comparison of principal components. J Am Stat Assoc 1979;74(367):703–7.
[21] Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev Comput Stat 2010;2(4):433–59.
[22] Adler N, Yazhemsky E. Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction. Eur J Oper Res 2010;202(1):273–84.
[23] Tesch S, Otto M. Application of principal-component analysis to the interpretation of brown-coal properties. Fuel 1995;74(7):978–82.
[24] Bera TK, Song JH, Freeman AJ, Jang JI, Ketterson JB, Kanatzidis MG. Soluble direct-band-gap semiconductors $LiAsS_2$ and $NaAsS_2$: Large electronic structure effects from weak As center dot center dot center dot S interactions and strong nonlinear optical response. Angew Chem Int Ed 2008;47(41):7828–32.
[25] Guo Z, Lee D, Schaller RD, Zuo X, Lee B, Luo T, et al. Relationship between Interchain Interaction, Exciton Delocalization, and Charge Separation in Low-Bandgap Copolymer Blends. J Am Chem Soc 2014;136(28):10024–32.
[26] Dkhissi A, Ouhib F, Chaalane A, Hiorns RC, Dagron-Lartigau C, Iratcabal P, et al. Theoretical and experimental study of low band gap polymers for organic solar cells. PCCP 2012;14(16):5613–9.