POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Computational study of the energy gap of benzenoid hydrocarbons through topological descriptors

Master's Degree in Materials Engineering and Nanotechnology

Author: Noemi Carol D'Ausilio
Supervisor: Professor Matteo Maria Saverio Tommasini
Co-supervisor: Gianluca Serra
Academic year: 2022-2023

---

1. Introduction

The comprehension of the link between the topological properties and the electronic properties of nanostructured materials is of primary interest. In this context, my investigation focuses on the energy separation between frontier orbitals, commonly referred to as the HOMO-LUMO gap. The energy gap is an electronic property of molecules relevant to the determination of their reactivity and their optical characteristics.

I considered benzenoid hydrocarbons (BHs). BHs are a class of organic molecules characterized by fused benzene rings, and they belong to the family of graphene molecules. For BHs, crucial properties such as aromaticity are tied to the topological arrangement of the fused rings [1].

Therefore, for these compounds, it holds significance investigating the energy gap through topological descriptors within the framework of chemical graph theory.

2. Methods

The molecules under analysis can be represented in a simplified manner through chemical graph theory. In this framework, each molecule can be depicted as a graph, where every carbon atom corresponds to a vertex and each carbon-carbon bond corresponds to an edge. Additionally, through the planar graphs of graph theory, it is possible to define polyhexes, which are polygonal shapes consisting of fused hexagonal rings. Consequently, each BH molecule can be associated to a polyhex [2]. In particular, the molecules that I have studied are associated to n-hexes, namely BHs with a fixed number n of rings, as shown in Figure 1. Then I considered highly symmetric BHs associated to: oblate rectangles $O(p, q)$ with $1 \leq p, q \leq 6$; parallelograms $L(p, q)$ with $2 \leq p, q \leq 7$; prolate rectangles $P(p,q)$ with $2 \leq p,q \leq 6$; hexagons $H(3,3,r)$ with $2 \leq r \leq 7$, $H(4,q,r)$ with $3 \leq q \leq 4$ and $3 \leq r \leq 6$, $H(5,5,r)$ with $5 \leq r \leq 6$, as shown in Figure 2.

Within the graph theory, it is also possible to define graph invariants to describe the properties of the molecules. Among these invariants, I focused on β representing the ratio of carbon-carbon bonds (M) to the number of carbon atoms (N), and on the number of Kekulé structures K. Kekulé structures have both chemical and mathematical significance, and correspond to the possible distributions of π-electrons along the molecule. They also serve as a chemical graph that represents the possible configurations of non-adjacent edges (associated to C=C bonds) covering all vertices. The number of Kekulé structures associated to the analyzed molecules is tabulated or determined through combinatorial formulas [3]. Moreover, for each molecular graph, I defined its dualist graph, that indicates certain peripheral conformations of BHs, such as fjords, and coves, leading to non-planar configurations due to steric hindrance, and bays that result in a planar configuration of the molecule [1], as shown in Figure 3.

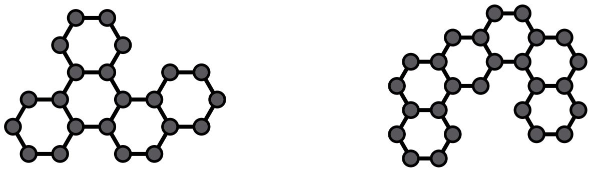
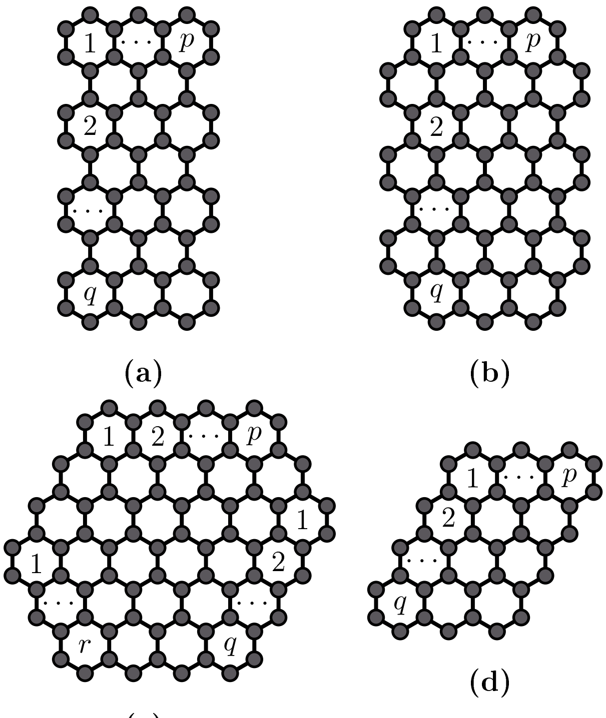
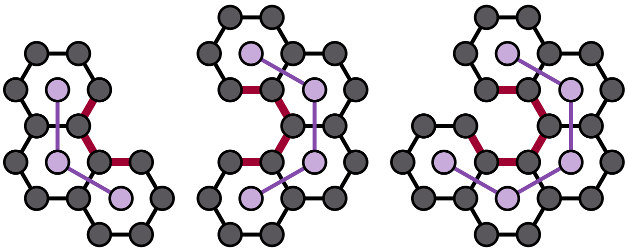


Figure 1: Molecular graphs of BHs associated to n-hexes, with $n = 5, 6$. Grey circles represent the vertices of the molecular graph; black segments are the edges of the molecular graph.



(a) **(b)** **(c)** **(d)**

Figure 2: Molecular graphs of BHs associated to prolate rectangles $P(p,q)$ (Panel **a**), parallelograms $L(p,q)$ (Panel **b**), oblate rectangles $O(p,q)$ (Panel **c**), hexagons $H(p,q,r)$ (Panel **d**). Grey circles represent the vertices of the molecular graph; black segments are the edges of the molecular graph.



(a) Bay **(b)** Cove **(c)** Fjord

Figure 3: Particular perimeter configuration in BHs highlighted by the thick red segments. Purple circles represent the vertices of the dual graph and purple segments are the edges of the dual graph.

I study the electronic many-body problem of a molecule within the Born-Oppenheimer approximation and the Hohenberg-Sham theorems of Density Functional Theory (DFT). The HOMO-LUMO gaps of specific classes of BHs have been computed using both Hückel molecular orbital theory and DFT. Hückel theory, a simplified model, facilitates the study of π-conjugated systems, while the DFT method, grounded in the Hohenberg-Sham theorem, strives to approximate the actual electron density and the proper geometry of the molecule. The basis set that I used for studying the HOMO-LUMO gap of all the molecules is Pople's 6-311G(d,p).

3. Energy gap of BHs

In the '80s, Ciosłowski proposed a formula to estimate the HOMO-LUMO gap calculated using Hückel theory (χ_{HL}), for n-hexes with $1 \leq n \leq 8$ [4]. The formula employs two topological parameters, $(2M/N)^{1/2}$ and $K^{2/N}$, where M is the number of C-C bonds, N is the number of carbon atoms, and K is the number of Kekulé structures.

$$\chi_{HL} = 2(\ 2.90611\ (2M/N)^{1/2} \\ +\ 3.91744\ K^{2/N})$$

I assessed the formula's applicability to the DFT-calculated gap of molecules linked to n-hexes molecules with $1 \leq n \leq 6$. Through the square least method, I determined the coefficients, obtaining a refined version of the Ciosłowski formula with a correlation coefficient $r = 0.9540$, as shown in Figure 4.

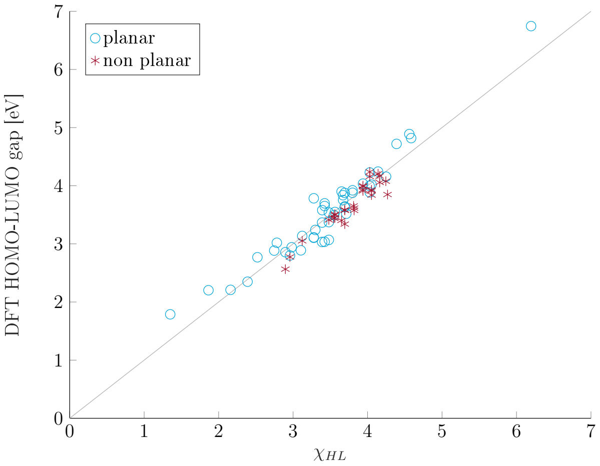


Figure 4: The scatter plot of DFT HOMO-LUMO gap over χ_{HL} for planar (blue circles) and non-planar molecules (red stars).

From Figure 4, I observed that the difference between the DFT-calculated HOMO-LUMO gap and the estimated gap χ_{HL} is comparable for both planar and non-planar molecules. This suggests that the HOMO-LUMO gap of molecules associated to n-hexes with $1 \leq n \leq 6$ depends mainly on topology rather than the molecule being planar (i.e., the perimeter classification of Figure 3).

I then shifted my focus on testing the validity of the Ciosłowski formula for an expanded set of molecules, including those associated to prolate rectangles, oblate rectangles, parallelograms, and hexagons. By analyzing the spatial distribution of the HOMO and LUMO orbitals, I observed that in smaller molecules the frontier orbitals span the entire structure, while in larger molecules, they localize along the zig-zag edges. Given this observation, I hypothesized that the topological parameters effectively describe the HOMO-LUMO gap for BHs with more than 8 rings but less than 20 rings. However, I noticed that for some classes of molecules, for example the ones associated to prolate rectangles, $K^{2/N}$ is not an effective topological parameter since $K^{2/N} = (p+1)^{\frac{1}{p+4}}$ only increases with p, not q. Through another best fitting of the coefficients again, I derived a new formula that predicts the DFT-calculated HOMO-LUMO gap of molecules associated to n-hexes with $1 \leq n \leq 6$, oblate rectangles, parallelograms and hexagons. In Figure 5, the modified Ciosłowski formula demonstrates good predictive power, with relative errors within 0.20 for parallelograms, 0.25 for hexagons, and 0.20 for oblates. Having initially studied molecules with one to six rings, I extended my investigation to larger molecules. To explore the formula's limits, I examined graphene, an infinite two-dimensional layer, and find that the Ciosłowski formula may predict the zero gap of graphene. Acknowledging the main limitations of the formula, I sought new indices that are easier to calculate than the number of Kekulé structures but still exhibit a high correlation coefficient.

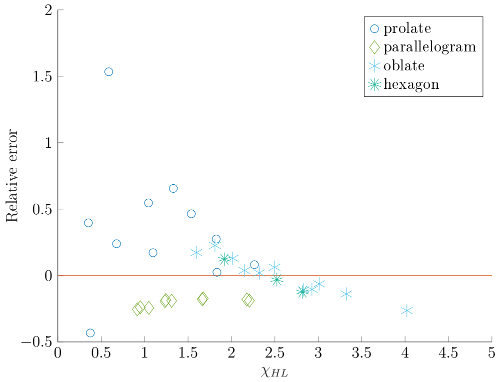


Figure 5: The relative error over χ_{HL} of the BHs associated to parallelograms, oblate rectangles, prolate rectangles and hexagons.

I considered four parameters. Two of them, $1/P$ and $1/N$, where P is the number of external edges, are connected to the size of the molecules. The other two parameters explored are P/M_{int} and P/N, and they reflect the relative weight of peripherical region with respect to the overall connectivity. M_{int} is defined as the number of C-C internal bonds.

	$1/P$	$1/N$	P/M_{int}	P/N
prolate rectangles	0.931	0.925	0.884	0.828
oblate rectangles	0.794	0.694	0.657	0.551
parallelograms	0.994	0.996	0.989	0.967
hexagons	0.642	0.753	0.500	0.486

Table 1: Correlation coefficients of the HOMO-LUMO gap calculated with DFT for molecules with more than 20 rings.

As shown in Table 1, for molecules with more than 20 rings, I observed that $1/P$ performs well as a correlation coefficient for those characterized by both armchair and zig-zag profiles. For molecules with only zig-zag edges, $1/N$ yields the best correlation coefficients. P/M_{int} and P/N exhibit lower performances. Thus, I infer that, for molecules with more than 20 rings, topological descriptors strictly increasing with the dimension correlate better with the DFT-calculated HOMO-LUMO gap than descriptors reflecting the general connectivity of the molecule.

4. Bond length alternation

I investigated bond length alternation within all analyzed classes of BHs, revealing an increase in average bond length from 1.400 Å to 1.425 Å. I noticed that the bond length alternation is more pronounced in the peripheral region than in the core region across all the molecules. Furthermore, I identified specific trends in different classes of BHs. For molecules associated to prolate and oblate rectangles, the armchair edges show a bond length alternation more marked than zig-zag edges. In the case of parallelograms, this alternation is more pronounced on some edges than on others.

In my analysis, the scatter plots representing the Hückel theory-calculated HOMO-LUMO gap against the DFT-calculated HOMO-LUMO reveal a general linear correlation, except for parallelograms. This deviation is linked to the instability of parallelograms, primarily characterized by the presence of just zig-zag edges, which leads to an open-shell configuration, marked by unpaired electrons.

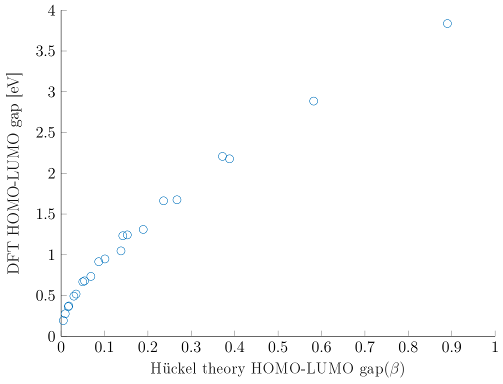


Figure 6: In blue circles, the scatter plot showing the relation between the HOMO-LUMO gap evaluated with the Hückel theory and the HOMO-LUMO gap calculated with DFT for molecules associated to parallelograms.

5. Isoarithmicity

I finally considered selected unbranched catacondensed BHs, where "catacondensed" denotes that each carbon atom is shared with no more than two benzenoid rings (i.e. ,no internal carbon atoms), and "unbranched" implies the absence of ramifications in the dualist graph. The ring concatenation in these BHs is denoted by a sequence of L (representing a ring attached to the adjacent one forming no angle in the dualist graph) or A (when the ring is linked to the adjacent one forming an angle in the dualist graph). Balaban introduced the concept of isoarithmicity [5], which identifies BHs characterized by the same LA sequence but different chemical structures. An example is shown in Figure 7. It has been shown that isoarithmic molecules, sharing identical LA sequences, exhibit the same HOMO-LUMO gap as calculated by Hückel theory.

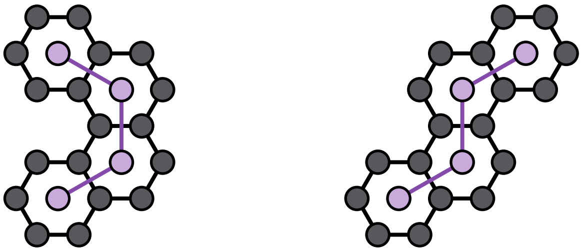


Figure 7: The molecular graphs of two isoarithmic molecules, characterized by the same sequence $LAAL$. The two unbranched catacondensed BHs are associated to the same dual graphs and the same LA sequences but their dualist graphs(in violet) are different.

My aim is to validate this statement for the DFT-calculated HOMO-LUMO gap of isoarithmic n-hexes with $1 \leq n \leq 6$. The results demonstrate that within each isoarithmic class, the deviation from the average DFT-calculated HOMO-LUMO gap is small. Hence, it is reasonable to assert that the HOMO-LUMO gap of isoarithmic BHs calculated by DFT does not depend on the periphery or planarity of the molecule but rather on its topology. The parameter that contains the definition of isoarithmicity is K since isoarithmic molecules have the same number of topological structures [5].

6. Conclusions

My investigation reveals that the DFT-calculated HOMO-LUMO gap for small n-hexes can be accurately predicted using parameters introduced by Ciosłowski. My data also show that the HOMO-LUMO gap is independent of the planarity of the molecule. Moreover, the extension of the Ciosłowski formula to BHs with up to 20 rings demonstrates its efficacy in describing the HOMO-LUMO gap of small BHs. However, limitations arise as the Ciosłowski formula fails to predict the zero gap of graphene and the HOMO-LUMO gap for certain classes of BHs. To address this, I explored alternative topological parameters. For molecules with more than 20 rings, my analysis identifies $1/P$ (inverse perimeter) and $1/N$ (inverse number of carbons) as the most effective correlation coefficients across all classes, showing the relation between molecular size and the HOMO-LUMO gap. Furthermore, my study establishes a robust correlation between the DFT-calculated HOMO-LUMO gap and the Hückel theory-calculated counterpart, also supported by the analysis of carbon-carbon bond lengths. Lastly, my investigation of isoarithmic BHs reveals that these molecules exhibit (as expected [5]) nearly identical DFT-calculated HOMO-LUMO gaps, emphasizing that the gap's dependence lies not in planarity but solely in the molecular topology.

References

[1] Yosadara Ruiz-Morales. Homo- lumo gap as an index of molecular size and structure for polycyclic aromatic hydrocarbons (pahs) and asphaltenes: A theoretical study. i. *The Journal of Physical Chemistry A*, 106(46):11283–11308, 2002.

[2] Ivan Gutman and Sven J Cyvin. *Introduction to the theory of benzenoid hydrocarbons*. Springer Science & Business Media, 2012.

[3] Sven J Cyvin and Ivan Gutman. *Kekulé structures in benzenoid hydrocarbons*, volume 46. Springer Science & Business Media, 2013.

[4] J Cioslowski. A unified theory of the stability of benzenoid hydrocarbons. *International journal of quantum chemistry*, 31(4):581–590, 1987.

[5] Alexandru T Balaban. Applications of graph theory in chemistry. *Journal of chemical information and computer sciences*, 25(3):334–343, 1985.