



# Doping of Two-Dimensional Structures of Carbon, Nitrogen, and Boron: A Study with Genetic Algorithms

**Keywords:** nanotechnology, electronic structures, genetic algorithms

Julio Smanioto Garcia, IFGW, UNICAMP

Prof. Dr. Douglas Soares Galvão, IFGW, UNICAMP

## 1 Introduction

Materials science is a rapidly evolving field that involves the study of the properties and behavior of various materials that have numerous applications in modern technology. The research in this field involves the development of new materials with improved properties and the discovery of new applications for existing materials. The *designs* of materials for specific applications [1] is a crucial focus area that has the potential to revolutionize industries and applications ranging from electronics to medicine and energy.

As an important class of materials, semiconductors gained its relevance in the late 1940s with the invention of the first transistor [2]. Since then, the ability to engineer the electronic and optical properties of semiconductors through doping, hetero structuring, and other techniques has also opened up new avenues for research and development in the field.

In this project, we intend to investigate how the type of dopants (C, B and N), their number and spatial distribution pattern affect the electronic structure of some two-dimensional benzenoid structures: graphene [3, 4], and polyanilines [5, 6, 7]. In particular, we intend to design some structures for responses within some ranges (gap and electronic localization). For this study we intend to use genetic algorithm techniques [8, 9, 10] coupled with the simple Hückel tight-binding Hamiltonian [11, 12, 13].

## 2 Methodology

### 2.1 Hückel Molecular Orbital

For the electronic part, we used the simple Hückel molecular orbital theory (HMO) [11]. The tight-binding Hamiltonian is defined as follows:

$$\mathbf{H} = \sum_i \alpha_i |i\rangle\langle i| + \sum_{\substack{i,j \\ i \neq j}} \beta_{i,j} |i\rangle\langle j|$$

where  $i$  and  $j$  are the atomic site indices,  $\alpha_i$  and  $\beta_{ij}$  are the usual Coulomb and hopping integrals, respectively. The values for these integrals in the case of carbon, nitrogen and boron are well parameterized in the [11, 14] literature.

This method is based on LCAOs (linear combination of atomic orbitals) approximation:

$$\psi_{\pi_j} = \sum_r^n c_{jr} \phi_r$$

and the degree of location can be easily estimated by calculating the IPN (inverse participation number):

$$I_j = \frac{\sum_r |c_{jr}|^4}{(\sum_r |c_{jr}|^2)^2}$$

the IPN can assume values from zero (maximum delocalization) to one (located on a single orbital). Despite its simplicity, Hückel's method is capable of producing reliable qualitative results for planar systems [15, 16].

After solving the eigensystem of the Hamiltonian, the *gap* is found by counting the number of  $\pi$ -electrons given off by each atom, then filling in the molecular orbitals and identifying the HOMO (*Highest Occupied Molecular Orbital*) and then subtract its energy value with the value for the energy of the LUMO (*Lowest Unoccupied Molecular Orbital*).

For the one-dimensional PANIs, Hamiltonian matrices were generated similarly to the benzene case[11], by counting the atomic sites along a single line that runs through the entire structure and visits each site in order. This counting logic produces block diagonal Hamiltonian matrices using the simple Hückel method, where each block represents the monomeric units of the PANIs family Figure 1.

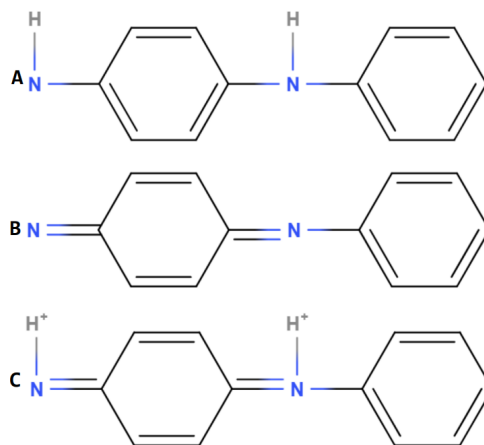


Figure 1: The monomeric units of the PANIs family. A - leucoemeraldine, B - emeraldine, C - diprotonated emeraldine

However, this same counting logic does not produce block diagonal matrices for two-dimensional structures like graphene. Instead, the structure was divided into multiple lines running through each "row" of the space lattice, as shown in Figure 2.

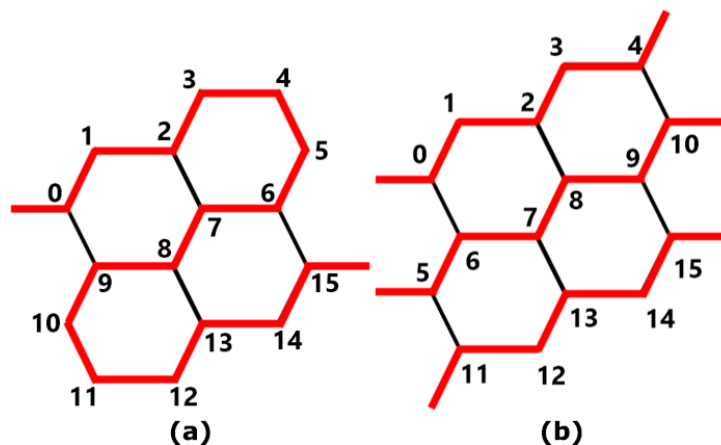


Figure 2: Exemple of the counting methods tested in the study. (a) a single line running through all the atomic sites of the molecule. (b) multiple lines running through each row of the molecule.

## 2.2 Genetic Algorithms

Chromosomes in the GA population are typically in the form of a bit string, so that each position on the chromosome has two possible alleles, 0 and 1 (e.g. [0, 1, 1, 0, ..., 0, 0, 1]). Each chromosome represents

a point in the hyperspace of possible solutions to the problem at hand. The GA then iterates through the population of chromosomes, calculating each individual performance (fitness) value. Then, the code proceeds to reproduce the population by random selecting the breeders within the desired performance range. Therefore, as the generations goes on, the algorithm replaces the less fit individuals by more fit ones.

Most GAs need this evaluator function that assigns a numerical value (performance) to each chromosome in the population, this function is interpreted as a hyperparameter of the algorithm. The simplest forms of GA involve, at least, this three operators [8, 9, 10]: **selection**, **reproduction/crossover** and **mutation**.

Each iteration of this process is called a generation. The entire set of generations is what we called a round or a run. At the end of each run, there are chromosomes that are distinctly more fit than others. As the process is intended to be stochastic, two runs using different seeds for the random number generator will generally produce different results in the details. We present a flow diagram of the algorithm in Figure 3 [15].

We defined the performance function  $f(x)$  in absolute units for the project as:

$$f(x) = 70\% \times gap + 30\% \times IPN \quad (1)$$

wherein  $gap = HOMO - LUMO$ , and is calculated for the highest occupied orbital (HOMO). The weights used in the function are not particularly significant, except for the normalization and the requirement that the  $gap$  should be the primary factor. We understand that if the  $gap$  is large, the delocalization of the orbital loses its significance.

### 3 Results and Discussion

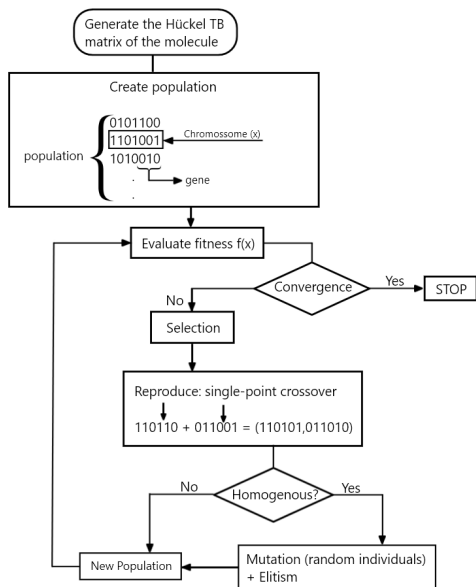


Figure 3: Flow chart of the genetic algorithm.

The algorithm was first developed to work with polyanilines (PANIs) and then adapted to work with the 2D-structures of a honeycomb lattice. We treated graphene with its edges in zigzag form (Figure 2), since the method only takes into account the first neighbors.

We present the results of the genetic search for the PANIs polymers with 200, 300 and 500 rings in Table 1. Only some of the best candidates are shown for the  $A_x B_y C_z$  composition. Also, the first solutions of the graphene  $BN$  search can be found in Table 2 for different sizes.

The code presented results that are in agreement with the literature for known molecules, such as benzene, pyridine [11, 12, 17], the PANIs polymer family [7, 15] and pure graphene sheets. This validates the implementation of the simple Hückel method; once the entry of the secular determinant of the molecule is correct, the output will yield sat-

isfactory results for electronic  $gap$  (in hopping units for carbon bonding,  $\beta \approx 2.5 V$  [11]) and IPNs. Most of the data and its analysis do not fit into this document.

Size	Composition	$gap [\beta]$	$IPN$	$f(x)$
200 rings	$A_{48}B_{44}C_8$	0.00018	0.00175	0.00065
300 rings	$A_{18}B_{127}C_5$	0.00016	0.00139	0.00053
500 rings	$A_{28}B_{212}C_{10}$	0.00003	0.00082	0.00027

Table 1: Some of the best candidates of the genetic search for the PANIs polymers.

Solutions	$f(x)$	$gap [\beta]$	$IPN$
$C_{22}N_5B_3$	0.02976	0.0151	0.06396
$C_{39}N_7B_2$	0.01714	0.00351	0.04895
$C_{66}N_4B_0$	0.01472	0.00291	0.04227

Table 2: First results of the  $BN$  search for graphene square sheets with 30, 48, and 70 atoms, respectively.

## 4 Conclusion

As said before, the Hückel method results are in agreement with the literature for known molecules. The genetic algorithm was developed and was able to reproduce the chromosomes through several generations, yielding an enormous variety of results. The code can be adapted to work with other bidimensional carbon structures, taking into account the changes in the parameters of the Hamiltonian. Most of the data still needs to be presented and discussed into future works

## References

- [1] P. V. Balachandran, J. Young, T. Lookman, and J. M. Rondinelli, “Learning from data to design functional materials without inversion symmetry,” *Nature communications*, vol. 8, no. 1, pp. 1–13, 2017.
- [2] A. C. . J. Ouellette, “November 17 - december 23, 1947: Invention of the first transistor,” November 2000 (Volume 9, Number 10).
- [3] A. K. Geim and K. S. Novoselov, “The rise of graphene,” *Nature materials*, vol. 6, no. 3, pp. 183–191, 2007.
- [4] A. C. Neto, F. Guinea, N. M. Peres, K. S. Novoselov, and A. K. Geim, “The electronic properties of graphene,” *Reviews of modern physics*, vol. 81, no. 1, p. 109, 2009.
- [5] E. Genies, A. Boyle, M. Lapkowski, and C. Tsintavis, “Polyaniline: A historical survey,” *Synthetic metals*, vol. 36, no. 2, pp. 139–182, 1990.
- [6] S. Manohar, A. MacDiarmid, and A. Epstein, “Polyaniline: Pernigranile, an isolable intermediate in teh conventional chemical synthesis of emeraldine,” *Synthetic metals*, vol. 41, no. 1-2, pp. 711–714, 1991.
- [7] D. S. Galvão, D. A. Dos Santos, B. Laks, C. P. De Melo, and M. J. Caldas, “Role of disorder in the conduction mechanism in polyanilines,” *Physical review letters*, vol. 63, no. 7, p. 786, 1989.
- [8] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.
- [9] M. M. Mitchell, *An introduction to genetic algorithms*. Cambridge. MA: MIT Press, 1996.

- [10] D. E. Golberg, *Genetic algorithms in search, optimization, and machine learning*. Addison Wesley, 1989.
- [11] A. Streitwieser, *Molecular orbital theory for organic chemists*. John Wiley & Sons, 1962.
- [12] C. A. Coulson, B. O’Leary, and R. B. Mallion, *Hückel theory for organic chemists*. Academic Press, 1978.
- [13] W. L. Jorgensen and L. Salem, *The Organic Chemist’s Book of Orbitals*. Academic Press, 1973.
- [14] N. Baird and M. Whitehead, “Coulomb parameters in simple huckel and omega technique molecular orbital methods,” *Canadian Journal of Chemistry*, vol. 44, no. 16, pp. 1933–1943, 1966.
- [15] R. Giro, M. Cyrillo, and D. S. Galvao, “Designing conducting polymers using genetic algorithms,” *Chemical Physics Letters*, vol. 366, no. 1-2, pp. 170–175, 2002.
- [16] F. C. Lavarda, M. C. Dos Santos, D. S. Galvao, and B. Laks, “Near resonant scattering from nonsymmetric dimers: Applications to substituted polyanilines,” *Physical review letters*, vol. 73, no. 9, p. 1267, 1994.
- [17] P. Peter Atkins and J. De Paula, *Atkins’ physical chemistry*. OUP Oxford, 2014.