DNA as a semiconductor: Analysis of charge localization

L. G. Mesa¹, A.C. García-Castro¹ and C. J. Páez¹
¹Escuela de Física, Universidad Industrial de Santander, Ciudad Universitaria, Bucaramanga 680002, Colombia
E-mail: cjpaezg@uis.edu.co

Abstract.
In this work, we analyze the charge localization in DNA molecules using an effective tight-binding approach that includes the backbone onsite energies. The localization length and participation number are examined as a function of energy dependence. We see that for specific energy ranges, the electronic states spread out into all sites, while in other energy ranges, the probability density is highly concentrated on either the backbone sites or the nucleotide like sites.

1. Introduction
The DNA can be considered as a conductor[1, 2, 3, 4] or a semiconductor[5, 6, 7, 8, 9, 10, 11] nano-material; this possibility made the DNA an attractive candidate as a template in nanoelectronics. These controversial results suggested a complex scenario not yet completely understood. Several theoretical models, from tight-binding models [12, 13, 14] until ab-initio calculus [15, 16], have been focused on understanding the details of the DNA electronic structures, the nature of electronic states and how the geometrical factors affect the conduction properties of the DNA in the nano-electronics devices [17, 18, 19, 20].

Another important point has proved to be relevant: the charge localization properties in these macromolecules. This involved question has to handle several problems, like the environment, possible intra-chain correlations, contacts, and molecular structure. The concept of localization has been initially defined for the thermodynamic limit, but the advent of mesoscopic systems challenged the understanding of the localization properties. Considering an increasing degree of disorder in a mesoscopic system, the transport can be tuned from ballistic down to the localized regime, with a diffusive transport window in between. The localization length $LL$ is a relevant length scale defining the ballistic and the localized regimes [21]. In particular, one has a localized regime when $L$, the length of the system is much larger than the localization length: $L \gg LL$. The diffusive regime is partially characterized by $L \ll LL$. Furthermore, the transition from localized to diffusive transport is instead a wide crossover in the range of $L \approx LL$. Hence, the degree of localization for mesoscopic systems has to be carefully investigated.

In previous work, this analysis has been undertaken for a double-strand ladder model [22], while here we investigate a double-strand ladder model with backbones, which modify the electronic structure drastically, as already mentioned. The same qualitative behavior is observed here, with the presence of backbones, only at specific energy ranges: different definitions of the
degree of localization converge to the same value. We focus on the role of charge localization as an essential parameter for studying the electronic transport through the DNA.

2. Numerical model
Many mechanisms have been suggested to explain the charge transport along DNA, where generally \( \pi \)-pathway transfer due to the overlap of molecular orbitals of the stacked aromatic bases of DNA can lead to charge propagation even at long distances [23]. In the present work we will consider heuristic DNA double chain with backbones[24] were the base pair sequences are considered completely random. This heuristic model capture the essential electronic properties replaced the complex molecular units (base nucleotides as well as the phosphates and sugars in the backbones) by effective orbitals. The electronic properties of the structure depicted in Figure 1 are described by the tight-binding Hamiltonian:

\[
H_{DNA} = \sum_{i=1}^{N_L} \left[ \sum_{j=1}^{2} \epsilon_{ij} |i,j><i,j| + t_{i,i+1} |i,j><i+1,j| + \sum_{q=\downarrow,\uparrow} \epsilon_{q} |i,q><i,q| + t_{q} |i,j><i,q| + t_{12} |i,1><i,2| \right] + h.c. 
\] (1)

The different terms are better appreciated if referred to Figure 1. \( \epsilon_{i,j} \) are the effective orbital energies of the base nucleotides, while \( \epsilon_{q} \) stand for the backbone molecules. The two DNA strands are connected by the \( t_{1,2} \) hopping parameters. Intra strand and backbone to nucleotide couplings are given by \( t_{i,i+1} \) and \( t_{q} \), respectively. In order to reduce the number of model parameters and to simplify our computation, we have adopted a simple parameterization taking the intrachain hopping parameters and interchain coupling as \( t_{i,i+1} = 1.0eV \) and \( t_{12} = 0.5eV \), respectively. This set of parameters has been used in several ladder models for DNA [25], respecting the expected condition of \( t_{12} < t_{i,i+1} \). The additional hopping onto the backbone is \( t_{q}=0.7 \) eV, and the backbone onsite energy is taken to be \( \epsilon_{q} = 11 \) eV. The molecule-electrode coupling is given by \( \tau = 0.35eV \). It is well known that the quality of the contacts strongly affects the transport properties of molecular devices [26].

The choice of the tight-binding parameters is far from uniquely determined, being rather a controversial issue since several parameter sets have been proposed in the literature [27]. The four different values for the effective nucleotide orbital energies, \( \epsilon_A = 8.24eV, \epsilon_T = 9.14eV, \epsilon_C = 8.87eV \) and \( \epsilon_G = 7.75eV \) are randomly assigned in one of the strands, with the same probability, while the sites of the second strand are set to follow the DNA pairing (A-T and C-G). In the present phenomenological description, we adopt an early choice [28], which has been recently rather revalidated [26].

The localization length (LL) for the DNA-like ladder model is computed from the exponential decrease in the transmission probability,[29, 30], so, it can be defined as

\[
LL^{-1}(E) = -\lim_{N_L \to \infty} \frac{1}{2N_L} < T(E) >, 
\] (2)

Where \( < \cdots > \) means an average of transmission probability \( T(E) \) over one hundred different chain configurations. This is done to avoid spurious effects due to a particular configuration. Here \( N_L \) is the length of the system given by base pairs number. Hence, for the sake of completeness in the forthcoming discussion, the total number of effective sites in the system is \( 4N_L \). The transmission probability \( T(E) \) between the electrodes can be evaluated by

\[
T(E) = Tr \left[ \Gamma_S G^r \Gamma_D (G^r)^\dagger \right], 
\] (3)

where \( G^r \) is the retarded Green’s function of the system which can be found from [31]
Figure 1. (color online). Schematic representation of The ladder model for electronic transport along DNA attached to two semi-infinite electrodes left (L) and right (R). A nucleobases-pairs sequence is given as red circles representing the four effective nucleotides: A, T, C, and G. Sugar-phosphate (backbone) are given as blue circles, and electronic pathways are shown as lines. Throughout this work we will consider the limiting case of a completely random sequencing along the central chains, maintaining the correlation between the base pairs (A-T and C-G).

\[ G^r = [E - H_{DNA} - \Sigma_S - \Sigma_D]^{-1}. \]  

(4)

Alternatively, another way to define the localization degree of an electronic state is obtained directly from the wave function, namely, the participation ratio \( PR \) is defined by [32]:

\[ PR(E) = \frac{1}{4N_L \sum_{i=1}^{N_L} \sum_{j=1}^{4} |\Psi_{ij}|^4}, \]

(5)

where \( \Psi_{jj}^2 = \rho_j \) is the wave-function amplitude in the \( j \) site, the local density of states \( (\rho_j) \) at a given site \( j \) can be straightforwardly extracted from \( \rho = -\frac{1}{\pi} \text{Im} \text{G}_{j,j} \), where \( \text{G}_{j,j} \) is the total green’s function at site \( j \). For localized states, \( PR \to 0 \) in the limit \( NL \to \infty \), while truly localized states - extended periodically modulated states - exhibit \( PR = \frac{2}{3} \) in one dimensional systems [32]. A quantity related to the \( PR \) is participation number \( (PN) \):

\[ PN = N_L \cdot PR, \]

where \( PN \) would be a measure of the actual number of base pairs having appreciable wave-function amplitudes at a given energy.

3. Results and Discussion

The transmission probabilities as a function of energy are shown in Figure 2, for two DNA models: a base pairs chain without backbones, Figure 2(a), and a base pairs chain with backbones Figure 2(b). We show that the backbone could strongly affect charge transport along the overlapping orbitals of the base pairs. Indeed, In the Figure 2(b), the backbones open a semiconducting gap, absent in the simple double chain model. It should be noticed that the ordered backbones considered in Figure 2(b) define the sharp states at the top(bottom) of the valence(conduction) band. Including disorder in the backbones representing environmental effects, like dryness - will smooth out these threshold states.[6].

A further insight is achieved by inspecting directly the probability density associated with selected electronic states of the ladder model with backbones attached to leads in Figure 3, considering a short system 50 bps long. In Figure 3(a) we compare the \( LL \) obtained from the
transmission probability, Figure 2(b), employing Eq. (2), having in mind that here we are far from the limit $N_L \to \infty$, with the participation number, obtained as described above, starting from the participation ratio, Eq. (5). Both quantities show the same qualitative behavior, as a function of energy, but do not agree quantitatively through the entire energy range. This quantitative agreement should only be expected in the limit of $NL \to \infty$.

In Figure 3(b-g), one observes that the distribution of the probability densities for randomly chosen disorder realization at certain energies, varies from being concentrated on the base pairs ($E = 6.42eV$ and $E = 10.95eV$), on the backbones (sharp peaks at the edges of the gap: $E = 8.49eV$ and $E = 8.92eV$) or on the entire system ($E = 8.05eV$ and $E = 9.28eV$).

Notice that the states are shown in Figure 3(b) are spread out along the entire length of the system, consistent with an effective delocalization picture, hence $LL > L$. Although relevant in defining the transport regime, the degree of localization, together with the system length, is not sufficient to infer the actual transport properties, even within the framework of the present effective model, because of the unequal distribution of the probability densities, on either backbone or central ladder sites.

The presence of backbones leads to systems where two qualitatively different chains occur, the backbone themselves and the base pairs double chain, promoting a transversal modulation in the $PR$. Therefore, the inclusion of backbones, leading a ladder model to be an effective four chain-wide model, introduces an extra component to the results expected for the $PN$. We see that for certain energy ranges, as clearly seen in Figure 3 (shorter systems), the electronic states spread out into all sites, while in other energy ranges, the probability density is highly concentrated on either the backbone sites or the nucleotide like sites. Hence, an effective modulation perpendicular to the system seems to become relevant. The inspection of the wave

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig2}
\caption{a) Average Transmission as a function of energy for a double-strand DNA of 50bps of length, a) without and b) with backbone.}
\end{figure}
Figure 3. (color online): a) Average Participation Number, $PN$, (red) and localization length, $LL$, (blue) as a function of energy for a double-strand DNA with backbones 50 bps long. Local probability densities for different energies of the spectrum shown in (a) along the four chains of the system: b) 6.42 eV, c) 8.05 eV, d) 8.49 eV, e) 8.92 eV, f) 9.28 eV and g) 10.95 eV. External ones are the backbones and the internal ones are for the base pairs ladder.

function, in Figure 3, reveals exponentially decaying states from the contacts into the DNA-like double chain with attached backbones. The resonances are disorder realization dependent and may be washed out if averages over many configurations are undertaken, as in the cases of $LL$ and $PN$ calculations.

4. Final Remarks
Finally, having in mind that experiments suggest that random DNA is not a good conductor [33] over average distances of 300 nucleotides, our results indicate that, indeed, a few hundred base pairs is the transition from diffusive to localized transport regimes. However, DNA samples, a few tenths of base pairs long, may also show a diffusive transport regime, but at such small
lengths, the strength of the contacts to the leads may severely influence the results. On the 
other hand, discrepancies between experimental results for more extended systems should 
have an origin in other environmental aspects than the contact to the leads. The present work 
addresses the question of characterizing the degree of charge localization in finite disordered 
systems, considering a more realistic, although the heuristic model for finite DNA-like double 
strands, by including the backbones, absent in previous work.

The degree of charge localization of electronic states in small systems is relevant in the sense 
that one could define an effective delocalization, i.e., $LL \gg N_L$. Therefore, a given system could 
show the transport properties of a localized regime if long enough, while a smaller piece could 
fall within a diffusive regime. One definition of the degree of localization, $PN$, does not depend 
on the system size, presenting a well-behaved evolution: increases with increasing system length, 
characteristic of a diffusive regime, saturating at long enough systems, the fingerprint of a truly 
localized system.

References

[1] Fink H W and Schonenberger C 1999 Nature 398 407
[2] Fathizadeh S and Behnia S 2020 Journal of the Physical Society of Japan 89 024004
[3] Rakitin A, Aich P, Papadopoulos C, Kobzar Y, Venedeev A S, Lee J S and Xu J M 2001 
Phys. Rev. Lett. 86 3670
[4] Schaff J F, Akdeniz Z and Vignolo P 2010 Physical Review A 81 41604
[5] Porath D, Bezyradin A, de Vries S and Dekker C 2000 Nature 403 635
[6] Cuniberti G, Craco L, Porath D and Dekker C 2002 Physical Review B 65 241314
[7] González J, Cruz-Irisson M, Sánchez V and Wang C 2020 Journal of Physics and Chemistry 
of Solids 136 109136
[8] Yoo K H, Ha D H, Lee J O, Park J W, Kim J, Kim J J, Lee H Y, Kawai T and Choi H Y 
2001 Phys. Rev. Lett. 87 198102
[9] Zhang W and Ullon S E 2004 Phys. Rev. B 69 153203
[10] Feng J F and Xiong S J 2002 Phys. Rev. E 66 21908
[11] Maia F F, Freire V N, Caetano E W S, Azevedo D L, Sales F A M and Albuquerque E L 
2011 The Journal of Chemical Physics 134 175101
[12] Mehrrez H and Anantram M P 2005 Phys. Rev. B 71 115405
[13] Bagci V M K and Krokhin A A 2007 Physical Review B 76 134202
[14] Maciá E 2007 Physical Review B 76 245123
[15] Pemmaraju C D, Rungger I, Chen X, Rocha A R and Sanvito S 2010 Physical Review B 
82 125426
[16] Mallajosyula S S, Lin J C, Cox D L, Pati S K and Singh R R P 2008 Physical Review 
Letters 101 176805
[17] Chang P C, Chang C Y, Jian W B, Yuan C J, Chen Y C and Chang C C 2019 Nano 
Research 12 1293–1300
[18] Paez C and Schulz P 2008 Microelectronics Journal 39
[19] Páez C, Rey-González R and Schulz P 2010 Physical Review B 81
[20] Páez C and Schulz P 2013 European Physical Journal B 86
[21] Guhr T, Müller-Groeling A and Weidenmüller H A 1998 Phys. Rep. 299 189
[22] Carrillo-Nuñez H and Schulz P A 2008 Physical Review B 78 235404
[23] Endres R G, Cox D L and Singh R R P 2004 Rev. Mod. Phys. 76 195
[24] Klotsa D K, Römer R A and Turner M S 2005 *Biophys. J.* **89** 2187
[25] Díaz E, Sedrakyan A, Sedrakyan D and Domínguez-Adame F 2007 *Phys. Rev. B* **75** 14201
[26] Waleed Shinwari M, Jamal Deen M, Starikov E B and Cuniberti G 2010 *Advanced Functional Materials* **20** 1865
[27] Hawke L G D, Kalosakas G and Simserides C 2010 *The European Physical Journal E* **32** 291
[28] Roche S, Bicout D, Macia E and Kats E 2003 *Phys. Rev. Lett.* **91** 228101
[29] Kissel G J 1991 *Physical Review A* **44** 1008
[30] Johnston R and Kunz H 1983 *Journal of Physics C: Solid State Physics* **16** 3895
[31] Páez C J, Bahamon D A, Pereira A L C and Schulz P A 2016 *Beilstein J. Nanotechnol.* **7**
[32] Edwards J T and Thouless D J 1972 *J. Phys. C* **5** 807
[33] de Pablo P J, Moreno-Herrero F, Colchero J, Gómez Herrero J, Herrero P, Baró A M, Ordejón P, Soler J M and Artacho E 2000 *Physical Review Letters* **85** 4992