

# Spectral, localization and charge transport properties of periodic, aperiodic and random binary sequences

K. Lambropoulos<sup>1,\*</sup> and C. Simserides<sup>1,†</sup>

<sup>1</sup>*National and Kapodistrian University of Athens, Department of Physics, Panepistimiopolis, 15784 Zografos, Athens, Greece*

(Dated: August 15, 2018)

We study periodic, deterministic aperiodic (Thue-Morse, Fibonacci, Period Doubling, Rudin-Shapiro, Cantor, generalized Cantor, Kolakoski) and random binary sequences within the framework of the Tight-Binding wire model. We use B-DNA as our prototype system, where all sequences have purines on the same strand and start with guanine. We investigate the interplay between structure, magnitude of parameters, and spectral, localization and charge transport properties. To this end, we comparatively study relevant quantities such as autocorrelation functions, integrated density of states, Lyapunov exponents, transmission coefficients and current-voltage curves, using the Landauer-Büttiker formalism. Our results reveal that there is a correspondence between the degree of structural complexity and the presence of correlations in the sequences and the aforementioned properties. Apart from periodic segments, which display enhanced transport properties, there are several cases of deterministic aperiodic sequences that can support significant currents, depending on the Fermi level of the leads. Random segments represent the less efficient category. Finally, we discuss factors that influence the results and examine homogeneous sequences, for which transport efficiency is maximal.

## I. INTRODUCTION

We focus on periodic, aperiodic and random binary sequences, i.e., sequences based on a binary alphabet, like  $\{0, 1\}$ . We use B-DNA as a prototype system and investigate sequences based on the couple  $\{G, A\}$ . This means that in one strand of double helix B-DNA we have either Guanine (G) or Adenine (A), and of course, in the complementary strand we have Cytosine (C) and Thymine (T), respectively. The persistence length  $\ell_p$  of a polymer somehow quantifies its stiffness, in the sense that pieces shorter than  $\ell_p$  behave rather like a flexible elastic beam, while much longer pieces are more likely to bend. DNA is among the stiffest of known polymers with  $\ell_p \approx 50$  nm or 150 base pairs<sup>1</sup>. This is one of the reasons we chose B-DNA as our prototype system, along with its biological and nanoscientific importance. On the other hand, if we stretch and join the DNA of all chromosomes of a single cell, that would give us a length of the order of a meter and would consist of billions of base pairs.

DNA is fundamental to living organisms because the sequence of its bases (adenine, guanine, thymine, cytosine) carries their genetic code. Its remarkable properties have drawn the interest of a broad interdisciplinary scientific community, beyond molecular biology and genetics. From a physics point of view, its electronic structure and its charge transfer and transport properties are studied with the aim to understand its biological functions and their potential applications in nanotechnology (e.g., nanocircuits, molecular wires)<sup>2,3</sup>. The base-pair stack of the double-helix DNA structure creates a nearly one-dimensional  $\pi$ -pathway that favors charge transfer and transport. The term *transfer* means that a carrier, created (e.g., by oxidation or reduction) or injected at a specific place, moves to a more favorable location, while the term *transport* implies the use of electrodes between

which electric voltage is applied.

Recent research has shown that carrier movement through DNA can be manipulated. For example, the carrier transfer rate through DNA can be tuned by chemical modification, e.g. using various natural and artificial nucleobases with different highest occupied molecular orbital (HOMO) levels<sup>4</sup>. Transfer rates can be increased by many orders of magnitude with appropriate sequence choice<sup>18,19</sup>. Furthermore, dynamical fluctuations, arising from either solvent fluctuations or base-pair vibrations can gate charge transport, counteracting the intrinsic disordered potential profile of the sequence<sup>6</sup>.

Many external factors (such as aqueousness, counterions, extraction process, electrodes, purity, substrate), influence carrier motion along DNA<sup>7</sup>. Hence, the need for a better understanding of the intrinsic factors that affect charge transfer and transport, such as geometry and base-pair sequence, arises. *Ab initio* calculations<sup>8-16</sup> and model Hamiltonians<sup>17-31</sup> have been used to explore the variety of experimental results and the underlying mechanisms. The former are currently limited to short segments for computational reasons, while the latter allow to address systems of realistic length. Here we study rather long sequences, hence we adopt the latter approach. The aim of this work is a comparative examination of the influence of base-pair sequence on charge transport.

Several works have been devoted to the study of transfer and transport in specific DNA structures (periodic<sup>18,19,32,33</sup>, quasiperiodic<sup>34-36</sup>, random and natural<sup>21,22,37-39</sup>) using variants of the Tight-Binding (TB) method. Here, we employ the TB wire model, with the sites of the chain being the base pairs, to study the spectral, localization and charge transport properties of periodic, deterministic aperiodic [Thue-Morse ( $\mathcal{TM}$ ), Fibonacci ( $\mathcal{F}$ ), Period Doubling ( $\mathcal{PD}$ ), Rudin-Shapiro ( $\mathcal{RS}$ ), Cantor set ( $\mathcal{CS}$ ), generalized Cantor set ( $\mathcal{GCS}$ ),

Kolakoski ( $\mathcal{KOL}$ ) and random DNA segments.

We use a parametrization that allows for different hopping parameters. This leads to quantitative and qualitative consequences. It gives a clearer picture than considering all hopping parameters equal, as in the bulk of the literature. We calculate quantities relevant to the aforementioned properties, such as autocorrelation functions, integrated density of states, Lyapunov exponents, transmission coefficients and current-voltage (I-V) curves.

The rest of the paper is organized as follows: In Sec. II we outline our notation and the transfer matrix method (TMM) theoretical framework. In Sec. III we focus on structural properties of the categories of studied segments. In Sec. IV, we discuss spectral properties in terms of the integrated density of states (IDOS). In Sec. V we present Lyapunov exponents, which characterize the localization length of eigenstates. In Sec. VI we discuss zero-bias transmission coefficients. In Sec. VII, we study I-V characteristics using the Landauer-Büttiker formalism. In Sec. VIII we state some remarks on the effect the parameters have on the results. Finally, in Sec. IX, we state our conclusions.

## II. NOTATION AND THEORETICAL FRAMEWORK

In the present work, we focus on periodic, deterministic aperiodic and random DNA segments consisting of different base pairs with their purines (A and G) on the 5'-3' strand. We will use this strand to denote the segments. For example, the notation GGAG means that we have the GGAG bases in the 5'-3' strand and the complementary ones, CCTC, in the 3'-5' strand. All studied sequences start with G.

The TB system of equations for a DNA segment in the Wire Model<sup>19,40</sup> reads

$$E\psi_n = E_n\psi_n + t_{n-1}\psi_{n-1} + t_n\psi_{n+1}, \quad (1)$$

$\forall n = 1, 2, \dots, N$ , where  $E$  is the eigenenergy,  $E_n$  is the on-site energy of base pair  $n$ ,  $|\psi_n|^2$  is the relevant occupation probability, and  $t_\ell$  is the hopping integral between base pairs  $l$  and  $l+1$ . The on-site energies are taken  $E_{A-T} = -8.3$  eV for the A-T base pair and  $E_{G-C} = -8.0$  eV for the G-C base pair.<sup>17-19,41-43</sup> The hopping integrals between successive base pairs that are involved in the segments studied here are shown in Table I.<sup>17-19,41-43</sup> The values of the parameters correspond to the HOMO of the base pairs and are discussed in Ref. 17.

TABLE I: HOMO Hopping integrals between successive base pairs involved in the segments studied in this work.  $r(c)$  stands for the base pair in the row (column) of the table.

$t_{rc}^{53}$ (eV)	G	A
G	-0.100	-0.110
A	-0.030	-0.020

Eq. (1) can equivalently be solved using the TMM, by rewriting it in the matrix form

$$\begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix} = P_n(E) \begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix}, \quad (2)$$

where

$$P_n(E) = \begin{pmatrix} \frac{E-E_n}{t_n} & -\frac{t_{n-1}}{t_n} \\ 1 & 0 \end{pmatrix} \quad (3)$$

is the Transfer Matrix (TM) of base pair  $n$ . The product

$$M_N(E) = \prod_{n=N}^1 P_n(E) \quad (4)$$

defines the Global Transfer Matrix (GTM) of the segment, containing all the information about its energetics. The elements of the GTM are recurrently given by

$$M_N^{11(12)} = \frac{E - E_N}{t_N} M_{N-1}^{11(12)} - \frac{t_{N-1}}{t_N} M_{N-2}^{11(12)} \quad (5a)$$

$$M_N^{21(22)} = M_{N-1}^{11(12)} \quad (5b)$$

with initial conditions  $M_1^{11} = (E - E_1)/t_1$ ,  $M_1^{12} = -t_N/t_1$ ,  $M_0^{21} = 1$ ,  $M_0^{22} = 0$ .  $M^{ij}$  is the element  $ij$  of matrix  $M$ . If we cyclically bound the segment, the GTM is a symplectic matrix, hence it is always unimodular ( $\det(M_N) = 1$ ).

We denote periodic segments by  $(XY\dots Z)_m$ , where  $m$  is the total number of repetition units. Details about the studied deterministic aperiodic segments and our notation to describe them can be found in Appendix A.

## III. STRUCTURAL PROPERTIES

To obtain a clear picture of the interplay between structure and energy profile of the segments, as well as its effect on localization and transport properties, we present some details on the structural characteristics of each category of sequences.

We deal with binary sequences, that is sequences based on a binary alphabet, like  $\{0, 1\}$  or  $\{G, A\}$  in our case. Therefore, a useful classification of their structural properties can be done through the study of the different base-pair triplets that are found in each category<sup>44</sup>. A triplet is made of a base pair and its next and previous neighbors. Since in a realistic treatment we need to simultaneously consider the difference in the on-site energies and the hopping integrals (as done here), the total number of possible triplets ( $2^3$  for a binary sequence) corresponds to the total number of different TMs that can be found in the GTM; cf. Eq. (3). The number of triplets in each category of DNA segments as well as the occurrence percentage of each triplet (for large  $N$ ) are depicted in Fig. 1. Finally, we notice, it has been claimed that the on-site

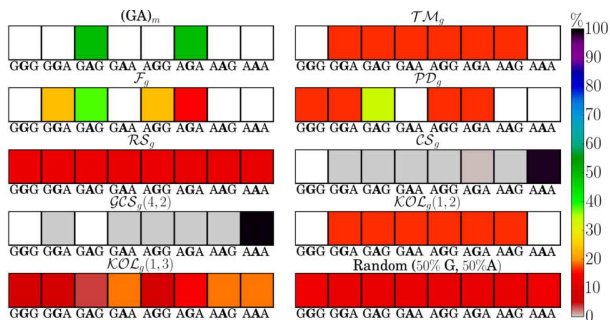


FIG. 1: Classification of the DNA segments studied in this work based on the number and occurrence percentage of base-pair triplets. The boxes correspond to each of the 8 possible triplets. For each segment, white boxes correspond to forbidden triplets, and the color of the rest corresponds to their occurrence percentage (calculated for large  $N$ ).

energy of a base depends on its flanking bases, an idea beyond the scope of our present calculations<sup>45</sup>.

From Fig. 1 it is obvious that the periodic  $(GA)_m$  segment represents the most ordered case (2 triplets with equal occurrence percentages).  $\mathcal{F}$  and  $\mathcal{PD}$  segments possess 4 and 5 different triplets, respectively, and have one dominant GAG triplet.  $\mathcal{TM}$  and  $\mathcal{KOL}(1,2)$  segments possess 6 equidistributed triplets.  $\mathcal{RS}$ , random and  $\mathcal{KOL}(1,3)$  segments possess all possible triplets; in the first two cases they are equidistributed; in the latter there are some predominant triplets. Finally, the Cantor Set family segments possess many of the possible triplets (7 for  $\mathcal{CS}$ , 6 for  $\mathcal{GCS}(4,2)$ ). However, the AAA triplets are predominant, asymptotically reaching 100% occurrence percentage as  $N$  increases.

The structural complexity determines the total number of TB parameters and the occurrence percentage of each inside a given segment. In Appendix B (Fig. B.1) we present the scaling of each TB parameter occurrence percentage for all the categories of studied segments. Among other things, we observe: The occurrence percentage of  $t_{GA}$  is always equal to that of  $t_{AG}$ . In all deterministic aperiodic cases, the occurrence percentages reach specific values as the generation,  $g$ , increases. Comparing  $\mathcal{F}$  and  $\mathcal{PD}$  sequences, although the former is structurally simpler (cf. Fig. 1), it has the same total number of TB parameters with the latter, since it has the additional triplet GGG.

Having obtained an estimate of the structural complexity of the sequences, we move to the estimation of the correlations of their energy landscape. We will do this by calculating the autocorrelation function (ACF)<sup>46</sup> for the quantities  $\frac{E_n}{t_n}$ ,  $n = 1, \dots, N$ . This ratio is used to fully capture the energy intricacy of the sequences. The lag- $j$  normalized ACF,  $ACF(j)$ , of  $\frac{E_j}{t_j}$ ,  $j = 1, 2, \dots, N-1$ , expresses the degree the base pairs are correlated with their  $j$ -th neighbors. Using the notation  $y_k = \frac{E_k}{t_k}$ , it is given by the expression

$$ACF(j) = \frac{\sum_{k=1}^{N-j} (y_k - \bar{y})(y_{j+k} - \bar{y})}{\sum_{k=1}^N (y_k - \bar{y})^2}, \quad (6)$$

where,  $\bar{y}$  is the mean value of  $y_{\{j\}}$ .

In Fig. 2, we present the ACF all the categories of studied segments, for three different lengths for each. The horizontal axes are normalized over the total number of neighbors ( $N-1$ ), thus corresponding to the *relative* neighbor distances. We notice that the ACF of each category has a characteristic shape. Furthermore, from the inspection of Fig. 2, we observe that there is a correspondence between the degree of structural complexity of the segments and the strength of correlations. Random and  $\mathcal{RS}$  sequences, which possess 8 equidistributed triplets, display weak correlations.  $\mathcal{KOL}(1,2)$  and  $\mathcal{TM}$  sequences, which possess 6 equidistributed triplets, display somehow stronger correlations. Then follow  $\mathcal{KOL}(1,3)$ ,  $\mathcal{CS}$ , and  $\mathcal{GCS}(4,2)$  sequences, which possess predominant triplets. The fractal sequences of the Cantor Set family possess strong correlations in the regions where G is present, interrupted by long, largely homogeneous, regions where it is not present. Deterministic aperiodic segments with the least possible triplets ( $\mathcal{F}$  and  $\mathcal{PD}$ , with 4 and 5 triplets, respectively) display strong correlations, and the periodic case is the dominant one.

Finally, we mention that by comparing the ACF of each category for different chain sizes, we can come to conclusions about their inflation/deflation symmetry. Sequences that possess this symmetry have similar autocorrelations at similar relative neighbor distances. This is the case for all aperiodic sequences studied here, apart from  $\mathcal{KOL}(1,2)$  and the random ones and [cf. Fig. 2(h) and (j), respectively]. As far as the  $\mathcal{KOL}(p,q)$  family segments are concerned, we have checked no inflation/deflation symmetry exists when  $|p-q| = 2\nu + 1$ ,  $\nu \in \mathbb{N}^*$ , in contrast with the cases  $|p-q| = 2\nu$ , such as  $\mathcal{KOL}(1,3)$ , shown in Fig. 2(i).

#### IV. SPECTRAL PROPERTIES

For fixed boundary conditions ( $\psi_{N+1} = \psi_0 = 0$ ), the eigenspectrum, i.e. the eigenenergies  $E_j$ ,  $j = 1, 2, \dots, N$  of a segment, can be given by the roots of the polynomial  $M_N^{11}(E)$ <sup>47,48</sup>. For periodic segments, the eigenspectrum can be recursively obtained with the help of the Chebyshev polynomials of the second kind<sup>48</sup>. Here, the eigenspectra of the segments have been calculated by numerical diagonalization of the Hamiltonian matrix, which is generally a real tridiagonal symmetric matrix. In the periodic case, the matrix is  $u$ -Toeplitz, where  $u$  is the size of the repetition unit.

The Density of States (DOS) can be obtained by

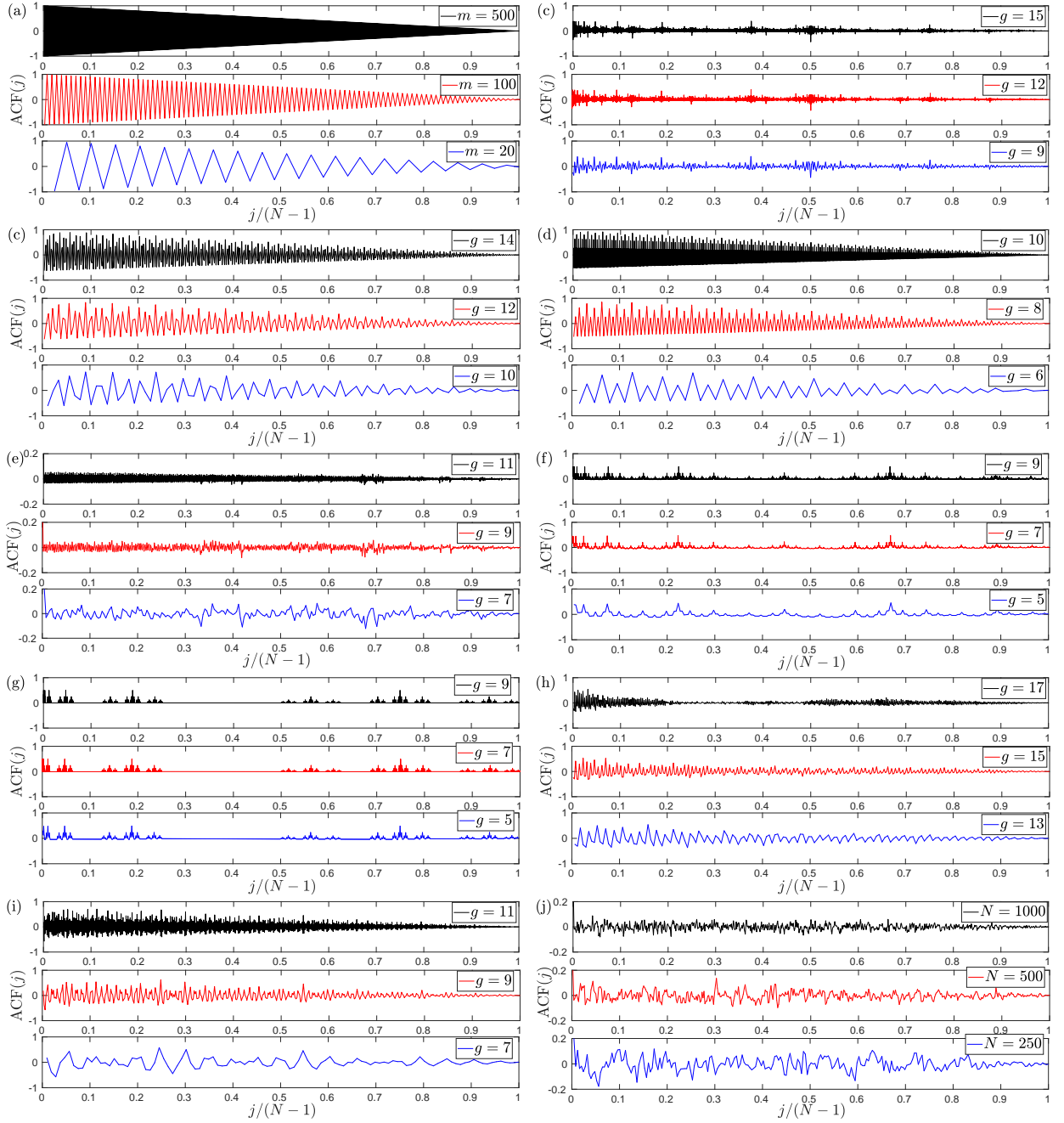


FIG. 2: Scaling of the autocorrelation function of various categories of DNA segments. (a) Periodic  $(GA)_m$ . (b)  $\mathcal{TM}_g$ . (c)  $\mathcal{F}_g$ . (d)  $\mathcal{PD}_g$  (e)  $\mathcal{RS}_g$ . (f)  $\mathcal{CS}_g$ . (g)  $\mathcal{GCS}_g(4, 2)$ . (h)  $\mathcal{KOL}_g(1, 2)$ . (i)  $\mathcal{KOL}_g(1, 3)$ . (j) Random (50% G content, 50% A content).

$$g(E) = \frac{N}{\pi} \frac{d}{dE} \left| \text{acos} \left( \frac{\text{Tr}(M_N(E))}{2} \right) \right|. \quad (7)$$

The Integrated DOS (IDOS) is given by the expression

$$IDOS(E) = \int_{E_1}^{E_2} g(E) dE. \quad (8)$$

The eigenspectra and the corresponding DOS for all the categories of DNA segments studied in this work are pre-

sented in Appendix C (Fig. C.1). We notice that for all studied deterministic aperiodic sequences, the allowed energies do not exceed the energy interval defined by the eigenspectrum of the random sequence. We have verified that this also holds for periodic polymers with only G and A in the 5'-3' strand, as their repetition unit increases. Hence, the above mentioned interval of the random sequence represents a limit. Two subsets of the aforementioned interval gather around the on-site energies of G and A, so will be henceforth referred to as G and A en-



ergy regions.

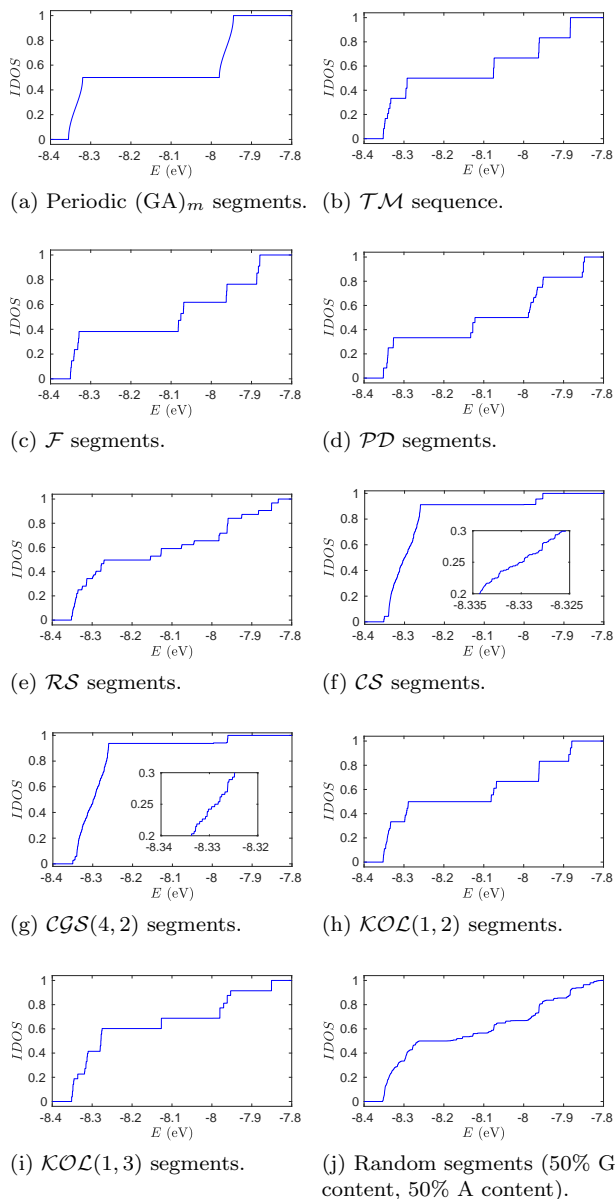


FIG. 3: Normalized IDOS of various categories of DNA segments.

The normalized IDOS for all categories of DNA segments, for large  $N$ , is presented in Fig. 3. In each panel, the largest energy gap, which is the region between two consecutive discontinuities of the DOS, corresponds to the separation between the upper limit of the allowed energies in the A region and the lower limit of the allowed energies in the G region. The value of the normalized IDOS in this gap corresponds to the relative number of A inside the sequence. Periodic  $(GA)_m$  segments possess two narrow, continuous bands, which can be recursively obtained; also, an analytical expression for the DOS exists<sup>48</sup>.  $TM$ ,  $F$ ,  $PD$ ,  $RS$ , and  $KOL$  family sequences possess step-like IDOS, which indicates that the eigen-

ergies concentrate at specific energy regimes, separated by small gaps. Cantor set family sequences have allowed energies predominantly in the A region. Although at first glance, the IDOS in this region may seem rather homogeneous, it can be seen from the insets of Fig. 3(f)-(g), that the spectrum is very rough. The random sequence IDOS has a shape that resembles to that of the  $RS$  sequence, although it is much more disrupted.

## V. LOCALIZATION

For the GTM of a given segment,  $M_N(E)$ , there exists a limiting matrix  $L(E)$  such that

$$L(E) = \lim_{N \rightarrow \infty} [M_N(E)^T M_N(E)]^{\frac{1}{2N}}. \quad (9)$$

The existence of  $L(E)$  is guaranteed by the Oseledec multiplicative ergodic theorem<sup>49</sup>. The Lyapunov Exponents of the segment are connected with the  $\nu$ -th eigenvalue of  $L(E)$ ,  $L_\nu(E)$ , through

$$\gamma_\nu(E) = \lim_{N \rightarrow \infty} \frac{1}{2N} \ln[L_\nu(E)]. \quad (10)$$

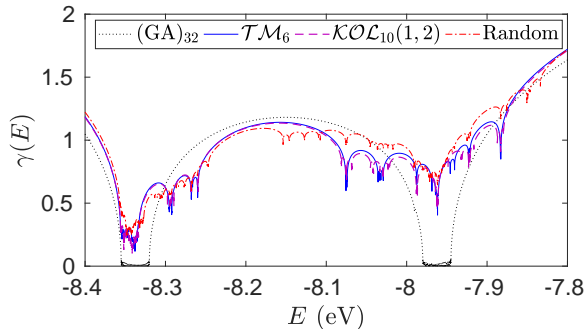
If the GTM is a  $2d \times 2d$  symplectic matrix, as in our case ( $d = 1$ ), the Lyapunov exponents are distinct and have the property  $-\gamma_1 < -\gamma_2 < \dots < -\gamma_d < \gamma_d < \dots < \gamma_2 < \gamma_1$ , hence  $\sum_{\nu=1}^{2d} \gamma_\nu = 0$ <sup>50,51</sup>. Since the Lyapunov exponents control the growth/decay rate of the solutions of Eq. (1), they are associated with the system's inverse localization length. In the case of symplectic GTMs, the localization length is given by the inverse of the smallest positive Lyapunov exponent,  $\gamma_d(E)$ <sup>51</sup>.

Since we deal with finite segments, the numerical Lyapunov exponents presented below correspond to finite values of  $N$ , hence the limit is dropped. To avoid numerical overflows when the matrix product is constructed, we use a QR decomposition scheme: We start with the initial matrix  $M_N(E)^T M_N(E) = P_1^T P_2^T \dots P_N^T P_N \dots P_2 P_1$ . We perform a QR decomposition of  $P_1$ , i.e.  $P_1 = Q_1^{(1)} R_1^{(1)}$ , so that  $M_N(E)^T M_N(E) = P_1^T P_2^T \dots P_N^T P_N \dots (P_2 Q_1^{(1)}) R_1^{(1)}$ . By consecutively performing QR decompositions at  $P_j Q_{j-1}^{(1)}$ , we arrive at  $M_N(E)^T M_N(E) = Q_{2N}^{(1)} \prod_{j=2N}^1 R_j^{(1)} := Q^{(1)} R^{(1)}$ . Hence, the matrix  $R^{(1)} Q^{(1)}$  and the initial matrix are similar, i.e., they have the same eigenvalues. By iterating this procedure, we arrive at a form  $R^{(k)} Q^{(k)}$ , where  $Q^{(k)}$  converges to a unit matrix and  $R^{(k)} = \prod_{j=2N}^1 R_j^{(k)}$ , i.e., a product of upper triangular matrices with positive diagonal entries in descending order. Hence, the eigenvalue  $L_\nu(E)$  is given by the  $\frac{1}{2N}$ -th power of the diagonal elements of

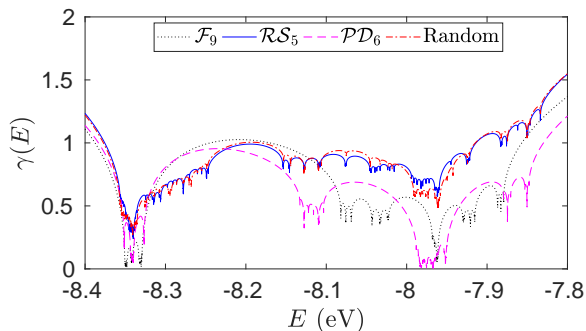
$R^{(k)}$ ,  $R^{(k)\nu\nu}$ . The Lyapunov exponents are thus

$$\gamma_\nu(E) = \frac{1}{2N} \sum_{j=1}^{2N} \ln[R_j^{(k)\nu\nu}]. \quad (11)$$

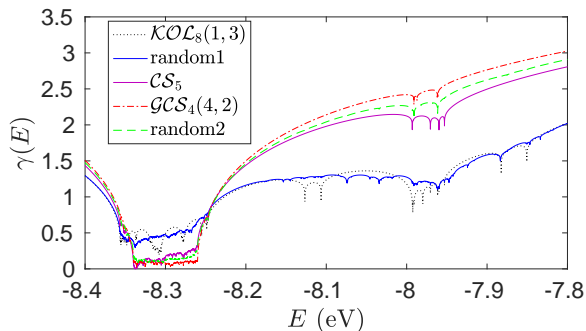
In our case, where  $d = 1$ , the only exponent to be determined is  $\gamma_1(E)$ . The index 1 will be dropped below.



(a) Periodic  $(GA)_m$  (black/dotted),  $\mathcal{TM}$  (blue/filled),  $\mathcal{KOL}(1,2)$  (magenta/dashed) and random (red/dashed-dotted) segments. All segments have 50% G content.



(b)  $\mathcal{F}$  (61.82%, black/dotted),  $\mathcal{RS}$  (56.25%, blue/filled),  $\mathcal{PD}$  (67.19%, magenta/dashed) and random (56.25%, red/dashed-dotted) segments. Percentages in parentheses denote G content.



(c)  $\mathcal{KOL}(1,3)$  (40.00%, black/dotted),  $\mathcal{CS}$  (13.17%, magenta),  $\mathcal{GCS}(4,2)$  (6.25%, red/dashed-dotted), and two random sequences (40.00% blue, 10.00% green/dashed). Percentages in parentheses denote G content.

FIG. 4: Lyapunov exponents of various categories of DNA segments.

The Lyapunov exponents of all categories of periodic

and deterministic aperiodic DNA segments, for large  $N$ , are presented in Fig. 4, together with some sequences with randomly rearranged base pairs. We have grouped together the segments according to the percentages of G and A they possess. Cases with similar G and A content are depicted in Fig. 4(a), with dominant G content in Fig. 4(b) and with dominant A content in Fig. 4(c). Segments grouped together have similar sizes where possible.

Starting with Fig. 4(a), we notice that the Lyapunov exponents follow the trend of the autocorrelation functions; stronger correlations lead generally to less localized states. Periodic  $(GA)_m$  segments have vanishing exponents inside their bands; this is a signature of the Bloch character of the wavefunctions.  $\mathcal{TM}$ , and  $\mathcal{KOL}(1,2)$  sequences have non-vanishing exponents of similar magnitude. This similarity is direct consequence of the similar base-pair triplet distribution those two categories possess (cf. Fig. 1). The random sequence has generally much more localized states. As a general remark, we notice that the Lyapunov exponents in the A energy region are rather smaller than the ones in the G energy region.

The conclusion that segments with stronger correlations possess less localized states is also evident from Fig. 4(b). Furthermore, the Lyapunov exponents of  $\mathcal{F}$  and  $\mathcal{PD}$  segments reach very small values in both base-pair energy regions, while those of  $\mathcal{RS}$  and random segments do not.  $\mathcal{F}$  ( $\mathcal{PD}$ ) segments possess larger energy intervals of less localized states in the A (G) region than  $\mathcal{PD}$  ( $\mathcal{F}$ ), while for  $\mathcal{RS}$  and random segments the exponents follow resembling trends. The dominance of smaller exponents in  $\mathcal{PD}$  segments over  $\mathcal{F}$  segments in the G region can be explained by the enhanced presence of  $t_{GG}$  (which are of large magnitude) in the former, induced by the occurrence of GGG triplets (cf. Fig. 1).

In segments with dominant A content, which are depicted in Fig. 4(c), the Lyapunov exponents in the A energy region are much smaller than those in the G region.  $\mathcal{KOL}(1,3)$  segments possess less localized states than random ones with similar G content in their common allowed energy intervals. The more dominant A becomes, the less (more) localized are the states in the A (G) region; this is the case for segments  $\mathcal{CS}$ ,  $\mathcal{GCS}(4,2)$  and random sequences with similar G content. In these cases, there are large A-rich regions within the segments, interrupted by Gs, which act like a disorder. The more homogeneous regions the segments possess, the less localized their eigenstates will be in the A energy region. Comparing these segments in Fig. 4(c), we can see that, generally, as the percentage of G decreases, the exponents become smaller in the A region; however, there are always energies at which the fractal sequences, which possess stronger correlations, are more delocalized than the random one. The very small percentage of G leads to highly localized states in the corresponding energy interval.

## VI. TRANSMISSION COEFFICIENT

The transmission coefficient describes the probability of an incident wave to be transmitted through a specific segment. We connect the segment to semi-infinite homogeneous metallic leads, which act as carrier baths. The leads' energy spectrum is given by the dispersion relation  $E = E_M + 2t_M \cos(qa)$ , where  $E_M$  is the on-site energy of the leads and  $t_M$  is the hopping integral between the leads sites. The coupling between the segment and the left (right) lead is described by the effective parameters  $t_{cL(R)}$ . Assuming incident waves from the left, we have

$$\psi_{\{n\} \leq 1} = e^{iqna} + re^{-iqna}, \quad \psi_{\{n\} \geq N} = te^{iqna}. \quad (12)$$

The transmission coefficient is defined as  $T(E) = |t|^2$ . The GTM of the scattering region obeys the equation

$$\begin{pmatrix} \psi_{N+1} \\ \psi_N \end{pmatrix} = P_R M_N P_L \begin{pmatrix} \psi_1 \\ \psi_0 \end{pmatrix}. \quad (13)$$

$$P_R = \begin{pmatrix} \frac{t_N}{t_{cR}} & 0 \\ 0 & \frac{t_{cR}}{t_M} \end{pmatrix}, \quad P_L = \begin{pmatrix} \frac{t_M}{t_{cL}} & 0 \\ 0 & \frac{t_{cL}}{t_N} \end{pmatrix} \quad (14)$$

are the matrices that describe the coupling of the three subsystems. After some manipulations, we arrive at the following expression for the transmission coefficient

$$T(E) = \frac{1}{1 + \Lambda(E)}, \quad (15)$$

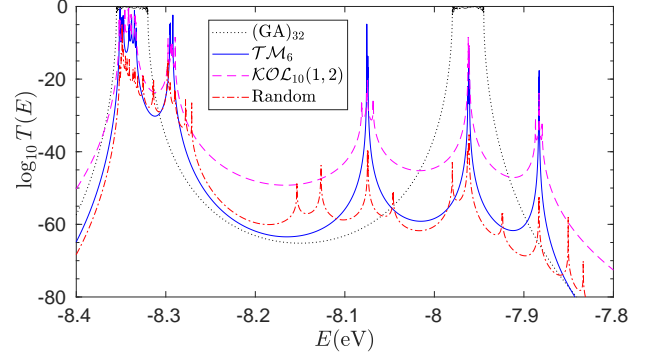
$$\Lambda(E) = \frac{[W_N(E) + X_N^+(E) \cos(qa)]^2}{4 \sin^2(qa)} + \frac{X_N^-(E)^2}{4}. \quad (16)$$

$$W_N(E) = M_N^{11} \omega - M_N^{22} \omega^{-1}, \quad (17a)$$

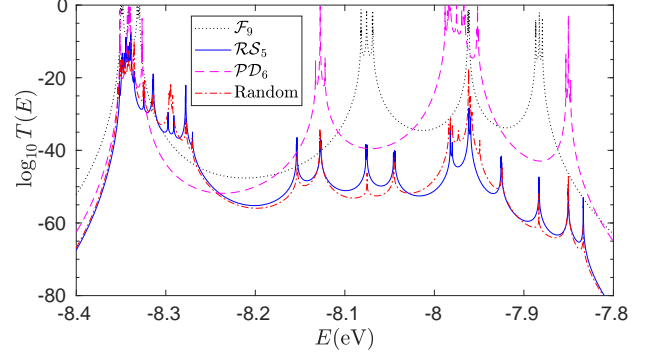
$$X_N^\pm(E) = M_N^{12} \chi \pm M_N^{21} \chi^{-1}, \quad (17b)$$

$$\omega = \frac{t_M t_N}{t_{cR} t_{cL}}, \quad \chi = \frac{t_{cL}}{t_{cR}}. \quad (17c)$$

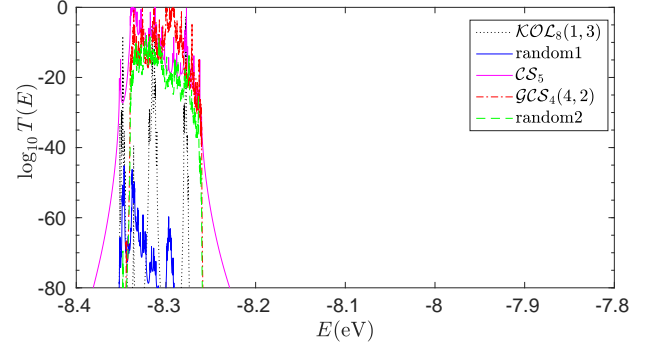
$\omega$ , included only in  $W_N(E)$ , expresses the deviation of the coupling of the system to the leads from the ideal case in which they are interconnected as if they were connected to themselves; hence  $\omega$  is a coupling strength factor.  $\chi$ , included only in  $X_N^\pm(E)$ , expresses the difference of the coupling between the leads and each end of the system; hence,  $\chi$  is a coupling asymmetry factor. In Ref. <sup>48</sup> we discuss the effects of  $\omega$  and  $\chi$ , as well as of the leads properties, to the transmission profiles of periodic segments. In the following, we choose the coupling parameters to satisfy the ideal and symmetric coupling conditions,  $|\omega| = |\chi| = 1$ . These have been shown to be the optimal coupling conditions for periodic segments<sup>48</sup>. We choose  $E_M = \frac{(E_{A-T} + E_{G-C})}{2} = -8.15$  eV and  $t_M = -0.25$



(a) Periodic  $(GA)_m$  (black/dotted),  $\mathcal{TM}_6$  (blue/filled),  $\mathcal{KOL}_{10}(1,2)$  (magenta/dashed) and random (red/dashed-dotted) segments. All segments have 50% G content.



(b)  $\mathcal{F}_9$  (61.82%, black/dotted),  $\mathcal{RS}_5$  (56.25%, blue/filled),  $\mathcal{PD}_6$  (67.19%, magenta/dashed) and random (56.25%, red/dashed-dotted) segments. Percentages in parentheses denote G content.



(c)  $\mathcal{KOL}_s(1,3)$  (40.00%, black/dotted),  $\mathcal{CS}_5$  (13.17%, magenta),  $\mathcal{GCS}_4(4,2)$  (6.25%, red/dashed-dotted), and two random sequences (40.00% blue, 10.00% green/dashed). Percentages in parentheses denote G content.

FIG. 5: Transmission coefficients of various categories of DNA segments.

eV, so that all eigenstates of the systems under examination are contained within the leads' bandwidth.

In Fig. 5 we present the transmission coefficients. At first glance, the transmission coefficients qualitatively follow the trend of the Lyapunov exponents (cf. Fig. 4). The less localized the eigenstates are, the more transpar-

ent the segments are to the incident waves at their energy region. Periodic  $(GA)_m$  segments display the most enhanced transmission, and reach the full transmission condition at specific energies<sup>48</sup>; this does not hold in general for deterministic aperiodic and random segments. Furthermore, apart from periodic  $(GA)_m$ ,  $\mathcal{F}$ , and  $\mathcal{PD}$  segments, transmission in the G energy region is from very small to negligible. These categories, together with the Cantor Set family ones, display the most enhanced transmission.  $\mathcal{TM}$  and  $\mathcal{KOL}(1,2)$  sequences display some energies at which transmission is rather significant. Deterministic aperiodic segments are more transparent than random ones with similar base-pair content, with the exception of  $\mathcal{RS}$ , that generally follows the trend of its randomly redistributed counterpart. Finally, we notice that the sequences shown in Fig. 5(c) have negligible transmission in the G energy region. This is due to the small role  $t_{GG}$  plays, since it rarely occurs within the segments.

## VII. CURRENT-VOLTAGE CURVES

We apply a constant bias voltage  $V_b$  between the leads, so that their chemical potential takes the form  $\mu_L^R = E_M \pm \frac{V_b}{2}$ . Then, a linear voltage drop within the DNA segment is induced and the transmission coefficient becomes bias-dependent. The energy regime between the leads' chemical potentials defines the conductance channel. The electrical current at zero temperature can be computed using the Landauer-Büttiker formalism<sup>52-56</sup> as

$$I(V) = \frac{2e}{h} \int_{E_M - \frac{V_b}{2}}^{E_M + \frac{V_b}{2}} T(E, V_b) dE, \quad (18)$$

since the Fermi-Dirac distributions,  $f(E_M \pm \frac{V_b}{2})$ , are Heaviside step-functions. The factor 2 in Eq. (18) comes from the double spin-degeneracy of each electronic level.

Again, we choose the coupling parameters to satisfy the ideal and symmetric coupling conditions,  $|\omega| = |\chi| = 1$ . We set the leads hopping integral  $t_M = -0.5$  eV to ensure that the leads' bands are wide enough to capture the whole picture. The choice of the leads Fermi level,  $E_M$ , plays a major role in both the shape of the I-V curves and the magnitude of the currents. This is demonstrated in Fig. 6, where the I-V curve of a periodic  $(GA)_{16}$  segment is determined as a function of  $E_M$ . It is evident that larger currents ( $\sim 0.1 \mu\text{A}$ ) occur at small biases when  $E_M$  lies within the bands of the segment. When this is not the case, voltage thresholds appear, and the (smaller in magnitude) turn-on currents emerge at biases that increase in a linear fashion with changing  $E_M$ . The magnitude of the currents becomes gradually smaller as  $E_M$  moves further away from the segments' bands, and is negligible when  $E_M$  lies well outside the bands. Finally, we should mention that the I-V curves are symmetric with respect to the difference between  $E_M$  and  $\frac{(E_{A-T} + E_{G-C})}{2}$ . The

above mentioned conclusions hold also qualitatively for segments consisting of identical monomers with crosswise purines, such as  $(GC)_m$ , where only one on-site energy ( $E_{G-C}$ ) is involved, with the difference that the curves are symmetric with respect to the difference between  $E_M$  and  $E_{G-C}$ .

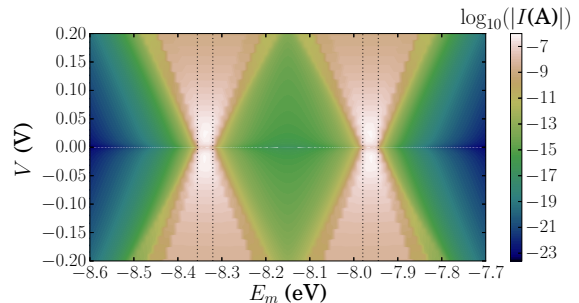


FIG. 6: The role of the leads' Fermi level,  $E_M$ , to the I-V curve of a  $(GA)_{16}$  segment. The vertical dotted lines encompass the bands of the segment.

Given the previous discussion and Fig. 6, we chose to study the I-V curves of all segments for two values of  $E_M$ , specifically  $-7.95$  eV and  $-8.35$  eV (i.e. at the center of the periodic segment's bands), to capture both G and A energy regions. In the following, we will only present curves the currents of which reach the pA regime. Our results are depicted in Figs. 7 and 8, for  $E_M = -8.35$  eV and  $E_M = -7.95$  eV, respectively.

From Fig 7(a), it is evident that periodic segments can carry significantly larger currents ( $\sim 0.1 \mu\text{A}$ ) than other categories. The deterministic aperiodic  $\mathcal{TM}$  and  $\mathcal{KOL}(1,2)$  segments display quite smaller currents than the periodic ones, of similar magnitude ( $\sim 1$  nA), but with clearly distinct shapes. The similarity of current magnitudes between  $\mathcal{TM}$  and  $\mathcal{KOL}(1,2)$  segments is in accordance with the similarity in the values of the Lyapunov exponents and zero-bias transmission coefficient for these cases, cf. Figs. 4(a) and 5(a), respectively. The random segment displays significantly smaller currents compared to the rest categories, reaching  $\sim 10$  pA.

As far as segments with dominant G content are concerned, we can see in Fig. 7(b) that  $\mathcal{F}$  and  $\mathcal{PD}$  segments can carry significantly larger currents than the  $\mathcal{RS}$  and random ones. This is again in accordance with the magnitude of the Lyapunov exponents and the transmission coefficients for these cases, cf. Figs. 4(b) and 5(b). In the A energy region, there is a larger energy range in which  $\mathcal{F}$  segments display less localized states and higher transmission than  $\mathcal{PD}$  ones. This fact is reflected on the magnitude of the currents ( $\sim 1$  nA for  $\mathcal{F}$ ,  $\sim 0.1$  nA for  $\mathcal{PD}$ ).  $\mathcal{RS}$  and random segments display currents in the  $\sim 10$  pA regime, but their curves have different shapes.

Sequences with dominant A content are depicted in Fig. 7(c).  $\mathcal{KOL}(1,3)$  sequences display rather small currents, that hardly reach 10 pA, due to the fact that the hopping integral with the largest occurrence percentage,



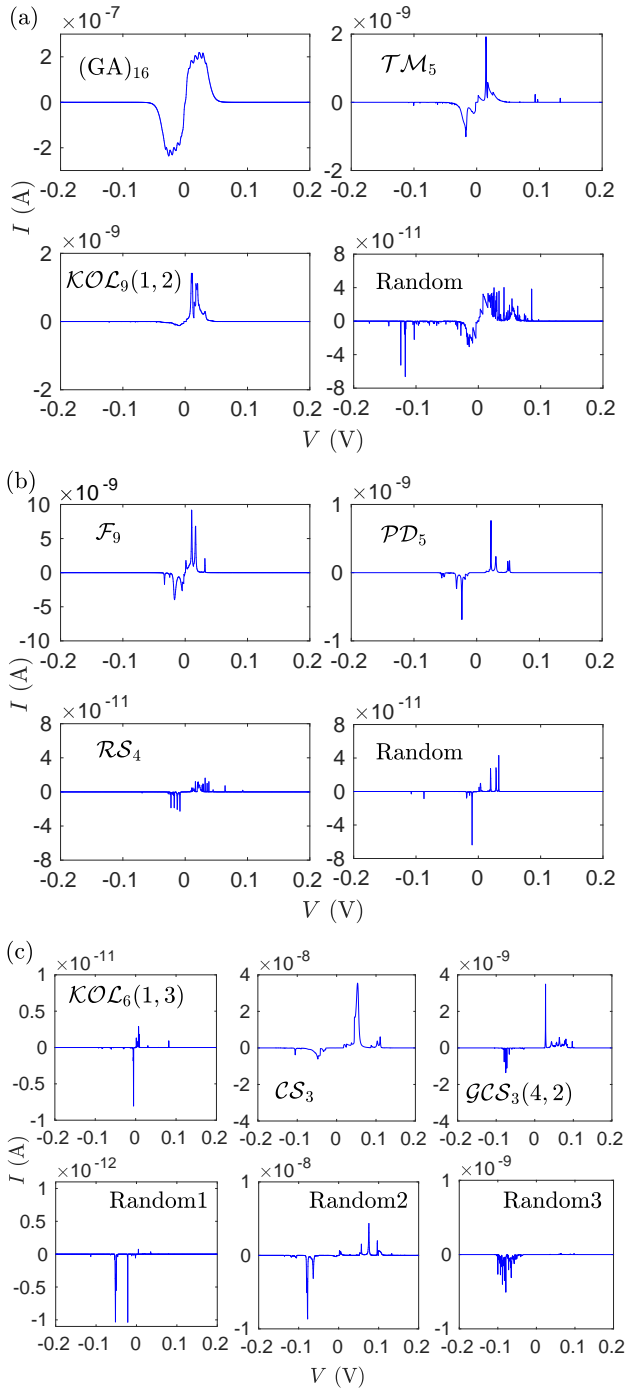


FIG. 7: I-V curves of various categories of DNA segments for  $E_M = -8.35$  eV. Categories as in Figs. 4 and 5. (a) Periodic  $(GA)_m$ ,  $TM$ ,  $KOL(1,2)$  segments and a random segment with similar G content. (b)  $\mathcal{F}$ ,  $\mathcal{PD}$ ,  $\mathcal{RS}$  segments, and a random segment with similar G content. (c) (top)  $KOL(1,3)$ ,  $\mathcal{CS}$ ,  $\mathcal{GCS}(4,2)$  segments. (Bottom) Random rearrangements of  $KOL(1,3)$ ,  $\mathcal{CS}$ ,  $\mathcal{GCS}(4,2)$  segments, respectively.

i.e.  $t_{AA}$ , is of rather small value. Albeit their small magnitude, the currents of  $KOL(1,3)$  sequences are larger than of their random rearrangement, which hardly reach

1 pA. In Cantor set family sequences, A content is much larger than G content, leading to large parts of the segment being essentially homogeneous. Hence, although  $t_{AA}$  has a small value, rather large currents occur ( $\sim 10$  nA for  $\mathcal{CS}$ ,  $\sim 1$  nA for  $\mathcal{GCS}(4,2)$ ). In this class of sequences, G, which, due to its small presence acts as a disorder in an otherwise homogeneous segment, is gathered in specific regions. Therefore, the currents they display are about one order of magnitude larger than their random rearrangements ( $\sim 1$  nA and  $\sim 10$  nA, respectively).

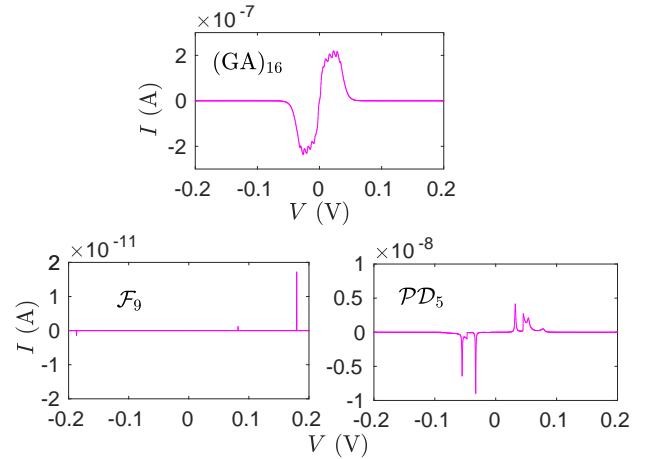


FIG. 8: I-V curves of various categories of DNA segments for  $E_M = -7.95$  eV.

As discussed in previous sections, in the G energy region the eigenstates of most segment categories are highly localized and display very small or negligible transmission. This, for  $E_M = -7.95$  eV, leads to currents that lie well below the pA regime. The only cases that do not follow this trend are the periodic,  $\mathcal{F}$ , and  $\mathcal{PD}$  segments, the I-V curves of which are depicted in Fig. 8. The periodic segments curve in this case is identical to the one for  $E_M = -8.35$  eV, due to the symmetry of the I-V curves with respect to the difference between  $E_M$  and  $\frac{(E_{A-T} + E_{G-C})}{2}$ , cf. Fig. 6. The rest two cases display energy intervals in the G region for which less localized states and enhanced transmission occur, as shown in previous sections. Close to  $E_M$ , the interval for  $\mathcal{F}$  segments is much smaller than the one for  $\mathcal{PD}$  segments, leading to a great difference in the current magnitudes between the two cases: a single spike of  $\sim 100$  pA for  $\mathcal{F}$  segments, currents in the  $\sim 10$  nA regime for  $\mathcal{PD}$  segments. This is due to the presence of GGG triplets in  $\mathcal{PD}$  segments, which leads to enhanced presence of  $t_{GG}$  (the magnitude of which is large), compared to  $\mathcal{F}$  segments, cf. Fig B.1(c)-(d).

### VIII. EFFECT OF PARAMETERS

It is common in the literature that all hopping parameters between different moieties are considered equal, for simplicity. Let us provide some example results occurring for identical hopping parameters, with reference to the Lyapunov exponents: In this case,  $\mathcal{F}$  segments possess more delocalized states in the G region (results not presented here), in contrast with the discussion of Fig. 4(b). Additionally, for all studied sequences, if we take equal hopping parameters, the act of substituting G with A and vice versa leads to a mere reflection of  $\gamma(E)$  relative to the mean value of the on-site energies,  $\frac{(E_{A-T} + E_{G-C})}{2}$  (results not presented here). This is not the case when different hopping parameters are considered. Their relative presence and magnitude can lead to significant differences in the electronic properties. Another example is the  $\mathcal{TM}$  sequence. If we equalize all hopping parameters, the Lyapunov exponent is also symmetric relative to  $\frac{(E_{A-T} + E_{G-C})}{2}$  (results not presented here), a scenario that does not hold for different hopping parameters, cf. Fig. 4(a). Of course, the inclusion of different hopping parameters plays a significant role not only in the Lyapunov exponents, but also in all properties that are determined by the electronic structure, such as the transmission coefficient and the I-V curves. To conclude, besides the fact that, in terms of chemical complexity, taking identical hopping parameters is unrealistic, our treatment reveals that considering different hopping parameters leads to a better understanding of the interplay between structural and transport properties, both quantitatively and qualitatively.

Furthermore, as far as transport properties are concerned, different results occur for different parameter values. For example, we have been able to reproduce the results reported for the transmission coefficients in Refs. [22,32,57](#), and for the I-V curves in Ref. [32](#), using the corresponding parametrizations, which are different from the one used here (all with equal hopping integrals). Different shapes as well as current-voltage regimes can be obtained, if the parameters are modified. For example, in Ref. [58](#) where microRNA chains are studied, taking different hopping integrals between nucleotides but of significantly larger magnitude than the ones used here, the authors report currents in the nA regime for voltages up to 16 V. These curves have been reproduced as well. The difference in the current-voltage regimes can also be seen by comparing the I-V curves of the homogeneous  $(G)_m$  and  $(A)_m$  segments (Fig. 9), which, due to their structural simplicity, represent the most efficient cases for charge transport. The curves have been calculated for  $E_M = E_{G-C}$  ( $E_{A-T}$ ) for the former (latter) case, i.e., in the center of the bands, with  $t_M = -0.5$  eV, and ideal and symmetric coupling conditions. Since the leads are aligned with the band centers, the only defining factor of the current-voltage regime is the value of the hopping parameter  $t_{GG}$  ( $t_{AA}$ ). Since  $t_{GG} > t_{AA}$ ,  $(G)_m$  segments

display greater currents than  $(A)_m$  segments ( $\sim 10 \mu\text{A}$  vs.  $\sim 1 \mu\text{A}$ ) and lie in a larger bias regime. Generally, increasing the value of the hopping parameter results in an increase of both the current magnitude and the voltage regime, until the states of the segment reach the bandwidth of the leads. For both I-V curves, the conductance at zero bias is equal to the quantum of conductance, i.e.,  $\frac{\partial I}{\partial V}|_{V=0} = G_0 = \frac{2e^2}{h} \approx 7.748 \times 10^{-5}$  S.

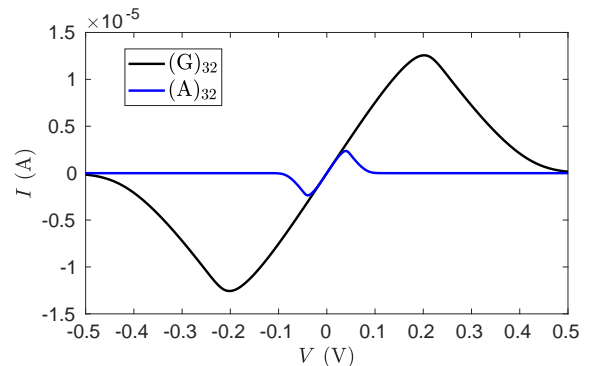


FIG. 9: I-V curves of  $(G)_{32}$  and  $(A)_{32}$  segments.

As discussed in Sec. VII (cf. Fig. 6), the occurrence of voltage gaps in the I-V curves depends on the relative position of the Fermi level of the leads and the eigenenergies of the segments. For example, a typical semiconducting I-V curve occurs for  $(G)_{30}$  segments, if we set  $E_M - E_{G-C} = 0.3$  eV (i.e. for  $E_M$  lying outside the band of the segment), with a voltage gap of  $\approx 0.7$  V and currents  $\sim 1$  nA. This is in accordance with the experimental I-V curves reported for the same system in Ref. [59](#), where the authors also attribute the voltage gap to the offset between the Fermi level of the electrode and the energy levels of the  $(G)_{30}$  segment.

### IX. CONCLUSION

We comparatively studied periodic, deterministic aperiodic (Thue-Morse, Fibonacci, Period Doubling, Rudin-Shapiro, Cantor set, generalized Cantor set, Kolakoski) and random binary sequences within the framework of the Tight-Binding wire model. We used B-DNA and the couple  $\{G,A\}$  as a prototype system. All segments had their purines on the same strand and started with guanine. Our aim was to gain a better understanding of the interplay between the structure of the segments and their spectral, localization and charge transport properties. To this end, we took differences in hopping parameters between successive base pairs into account. This led to a more realistic evaluation of the role the structure of the segments plays in the aforementioned properties.

We determined the number and occurrence percentage of all possible base-pair triplets that can be found within these segments, as well as their autocorrelation functions. Our results showed that there is a correspondence be-

tween the number of possible triplets, the existence of dominant triplets and the strength of correlations within the segments.

We calculated the density of states of the segments, and presented the integrated density of states, a measure of their energy structure. The allowed energies of all studied deterministic aperiodic segments lie within the interval defined by the eigenspectrum of random segments, and their IDOS has rough shapes.

Furthermore, we calculated the Lyapunov exponents of the segments, and showed that the structure, the relative presence of each base pair, and the values of the parameters play a major role in the degree of eigenstates localization. Generally, segments with strong correlations possess less localized states.

Next, we connected the segments to semi-infinite homogeneous leads and studied the zero-bias transmission coefficients, reaching similar conclusions regarding their transparency to incident carriers.

We also studied the current-voltage characteristics of the segments, using the Landauer-Büttiker formalism. We showed that the shape of the curves and the magnitude of the currents strongly depends on the leads' on-site energy (Fermi level). The current-voltage characteristics were calculated for two values of the latter, corresponding to positions that catch the energy regions of interest. For

the parametrization used, we found that periodic binary segments can carry currents in the  $\mu\text{A}$  regime. Several deterministic aperiodic segments (specifically, Fibonacci, Period-doubling, Cantor set and generalized Cantor set) can also display rather large currents, namely in the nA regime, depending on the Fermi level of the leads. Random sequences hold the smallest currents, in accordance with the weak correlations they possess.

Finally, the I-V curves of the homogeneous  $(\text{G})_m$  and  $(\text{A})_m$  segments, due to their structural simplicity, represent the most efficient cases for charge transport with conductance at zero bias equal to the quantum of conductance. Typical semiconducting I-V curves occur for these segments when there is a mismatch between their eigenstates and Fermi level of the leads, in accordance with experimental results.

### Acknowledgements

K. Lambropoulos wishes to thank the General Secretariat of Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI) for a PhD research scholarship.

---

### Related work at

[http://users.uoa.gr/~csimseri/physics\\_of\\_nanostructures\\_and\\_biomaterials.html](http://users.uoa.gr/~csimseri/physics_of_nanostructures_and_biomaterials.html)

---

## Appendix A: Deterministic Aperiodic segments

### 1. Fibonacci sequence

The Fibonacci sequence, named after the Italian mathematician Leonardo Pisano (Fibonacci) who introduced it in his 1212 book *Liber Abaci*, in a study of the population growth of rabbits<sup>60</sup>, is a number sequence the terms of which are generated by the addition of the two previous terms, with given initial conditions. However, this sequence appears many centuries before in Indian mathematics, in connection with Sanskrit prosody. For example, the possible ways to arrange short (S) and double, long (L) syllables with given total duration measured as  $g$  S syllables is the Fibonacci number of the  $g + 1$  generation. If  $\mathcal{N}_g$  is the Fibonacci number of generation  $g$ , and we set  $\mathcal{N}_0 = \mathcal{N}_1 = 1$ , the recurrence relation  $\mathcal{N}_g = \mathcal{N}_{g-1} + \mathcal{N}_{g-2}$  produces the number sequence 1, 1, 2, 3, 5, 8, 13, 21, 34, ... Using the two-letter alphabet  $\{\text{G}, \text{A}\}$ , we can define the Fibonacci word  $\mathcal{F}_g$  by the substitution rules  $\text{A} \rightarrow \text{G}$ ,  $\text{G} \rightarrow \text{GA}$ , starting with  $\mathcal{F}_0 = \text{A}$ . Hence,  $\mathcal{F}_1 = \text{G}$ ,  $\mathcal{F}_2 = \text{GA}$ ,  $\mathcal{F}_3 = \text{GAG}$ ,  $\mathcal{F}_4 = \text{GAGGA}$ , etc. Obviously, the length of the word  $\mathcal{F}_g$  is  $\mathcal{N}_g$ .

### 2. Thue-Morse sequence

The Thue-Morse (TM) sequence (aka Prouhet-Thue-Morse sequence) was first studied by Eugene Prouhet in the field of number theory<sup>61</sup>, defined by Alex Thue in the field of combinatorics<sup>62</sup>, and rediscovered by Marston Morse in the context of differential geometry<sup>63</sup>. It is a binary sequence of 0s and 1s, starting with 0, with its  $g^{\text{th}}$  generation constructed by appending the Boolean complement of the previous generation to the sequence. Using the two-letter alphabet  $\{\text{G}, \text{A}\}$ , we can define the TM word  $\mathcal{TM}_g$  by the substitution rules  $\text{G} \rightarrow \text{GA}$ ,  $\text{A} \rightarrow \text{AG}$ , starting with  $\mathcal{TM}_0 = \text{G}$ . Hence,  $\mathcal{TM}_1 = \text{GA}$ ,  $\mathcal{TM}_2 = \text{GAAG}$ ,  $\mathcal{TM}_3 = \text{GAAGAGGA}$ , etc. The length of the word  $\mathcal{TM}_g$  is  $2^g$ .

### 3. Period-Doubling sequence

The Period-Doubling (PD) sequence is closely connected with the TM sequence. Specifically, its elements are given by the first differences of the elements of the TM binary sequence modulo 2. Using the two-letter alphabet  $\{\text{G}, \text{A}\}$ , we can define the PD word  $\mathcal{PD}_g$  by the substitution rules  $\text{G} \rightarrow \text{GA}$ ,  $\text{A} \rightarrow \text{GG}$ , starting with

$\mathcal{PD}_0 = G$ . Hence,  $\mathcal{PD}_1 = GA$ ,  $\mathcal{PD}_2 = GAGG$ ,  $\mathcal{PD}_3 = GAGGGAGA$ , etc. The length of the word  $\mathcal{PD}_g$  is  $2^g$ .

**4. Rudin-Shapiro sequence**

The Rudin-Shapiro (RS, aka Golay-Rudin-Shapiro) sequence is the sequence of the appended coefficients of the RS polynomials<sup>64,65</sup>. It contains only  $\pm 1$ s and is generated by starting with  $+1, +1$  and employing the rules

$$\begin{aligned} +1, +1 &\rightarrow +1, +1, +1, -1 \\ +1, -1 &\rightarrow +1, +1, -1, +1 \\ -1, +1 &\rightarrow -1, -1, +1, -1 \\ -1, -1 &\rightarrow -1, -1, -1, +1. \end{aligned}$$

Using the two-letter alphabet  $\{G, A\}$ , we can define the RS word  $\mathcal{RS}_g$  by the substitution rules  $GG \rightarrow GGGA$ ,  $GA \rightarrow GGAG$ ,  $AG \rightarrow AAGA$ ,  $AA \rightarrow AAAG$ , starting with  $\mathcal{RS}_0 = GG$ . Hence,  $\mathcal{RS}_1 = GGGGA$ ,  $\mathcal{RS}_2 = GGGAGGAG$ , etc. The length of the word  $\mathcal{RS}_g$  is  $2^{g+1}$ .

**5. Cantor Set sequence**

The Cantor Set (CS), named after mathematician Georg Cantor who introduced it<sup>66</sup>, is one of the most well-known deterministic fractals. It is obtained as follows: given the continuous interval  $[0, 1]$ , the middle third,  $(\frac{1}{3}, \frac{2}{3})$  is deleted, resulting in the union  $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ . Then, the open middle third of each remaining interval is deleted, and the process is repeated *ad infinitum*. Using the two-letter alphabet  $\{G, A\}$ , we can define the CS word  $\mathcal{CS}_g$  by the substitution rules  $G \rightarrow GAG$ ,  $A \rightarrow AAA$ , starting with  $\mathcal{CS}_0 = G$ . Hence,  $\mathcal{CS}_1 = GAG$ ,  $\mathcal{CS}_2 = GAGAAAAGAG$ , etc. All generations are palindromic words. The length of the word  $\mathcal{CS}_g$  is  $3^g$ .

**6. Generalized Cantor Set sequences**

In accordance with the rationale described above, one can imagine the construction of a generalized CS word,  $\mathcal{GCS}_g(s, d)$ , produced by the two-letter alphabet  $\{G, A\}$ , where  $s$  is the total number of letters substituting each letter of the sequence in the next generation and  $d$  is the number of letters that correspond to the “deleted” middle segment ( $s > d$ ).  $s$  and  $d$  are mutually odd or even, to preserve the palindromicity of the words. For example, the generalized word  $\mathcal{GCS}_g(4, 2)$  is given by the rules  $G \rightarrow GAAG$ ,  $A \rightarrow AAAAA$ , starting with  $\mathcal{GCS}_0(4, 2) = G$ . The length of the word  $\mathcal{GCS}_g(s, d)$  is  $s^g$ .

**7. Kolakoski sequences**

The Kolakoski  $\{p, q\}$  sequences are a family of sequences of the integers  $p \neq q$  that are their own run

length encodings (a *run* is defined here as the maximal subsequence of identical numbers). The classic and most well known sequence of this class, Kolakoski $\{1, 2\}$ <sup>67</sup>, also referred to as Oldenburger-Kolakoski sequence, was popularized by recreational mathematician William Kolakoski<sup>68</sup>, but it was independently introduced by Rufus Oldenburger<sup>69</sup>. This family of sequences possesses different properties in different cases. For example, for specific values of  $p$  and  $q$ , they may show pure-point or continuous diffraction spectra<sup>70</sup>. Each generation,  $\text{Kol}_g(p, q)$ , of the sequences can be seen as the run length encoding of the next generation, starting with  $\text{Kol}_0(p, q) = q^p$  and following the substitution rules

$$\begin{aligned} q &\rightarrow p^q && \text{if } q \text{ was at odd } n, \\ q &\rightarrow q^q && \text{if } q \text{ was at even } n, \\ p &\rightarrow p^p && \text{if } p \text{ was at odd } n, \\ p &\rightarrow q^p && \text{if } p \text{ was at even } n. \end{aligned}$$

For example  $\mathcal{KOL}_0(1, 2) = 2$ ,  $\mathcal{KOL}_1(1, 2) = 11$ ,  $\mathcal{KOL}_2(1, 2) = 12$ ,  $\mathcal{KOL}_3(1, 2) = 122$ ,  $\mathcal{KOL}_4(1, 2) = 12211$ ,  $\mathcal{KOL}_5(1, 2) = 1221121$ , etc. Accordingly, using the two-letter alphabet  $\{G, A\}$ , we can define the  $\mathcal{KOL}(p, q)$  word  $\mathcal{KOL}_g(p, q)$  by assigning  $G$  to  $p$  and  $A$  to  $q$ . Thus, e.g.,  $\mathcal{KOL}_5(1, 2) = GAAGGAG$ . The length of  $\mathcal{KOL}(1, 2)$  as the generation increases is given by the OEIS sequence A001083<sup>71</sup>. Generally, the length of the word  $\mathcal{KOL}_g(p, q)$  is equal to the sum of the terms of  $\mathcal{KOL}_{g-1}(p, q)$ .



## Appendix B: TB parameters percentages

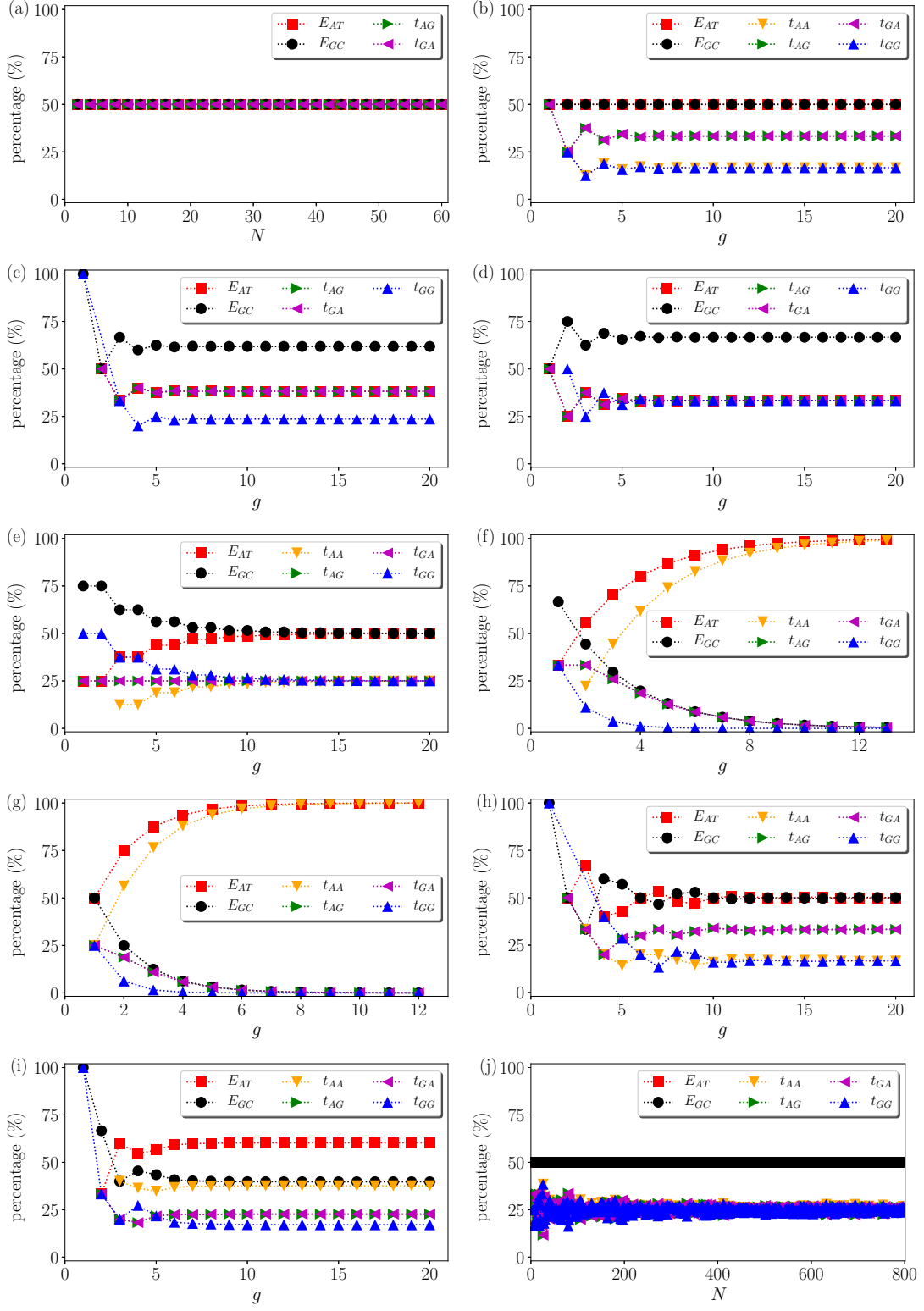
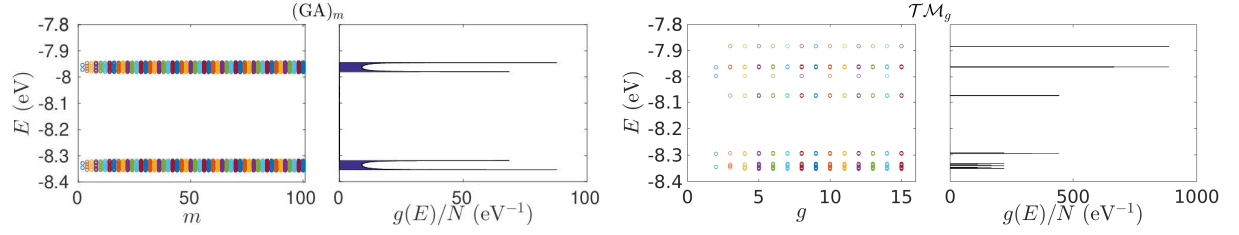


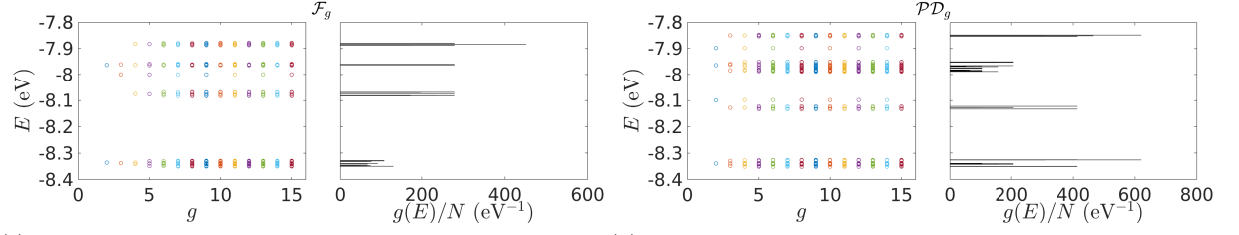
FIG. B.1: Scaling of the occurrence percentage of each TB parameter in various categories of DNA segments. (a)  $(GA)_m$ . (b)  $TM_g$ . (c)  $F_g$ . (d)  $PD_g$ . (e)  $RS_g$ . (f)  $CS_g$ . (g)  $GCS_g$ . (h)  $KOL_g(1,2)$ . (i)  $KOL_g(1,3)$ . (j) Random (50% G, 50% A).

## Appendix C: Eigenspectra and DOS



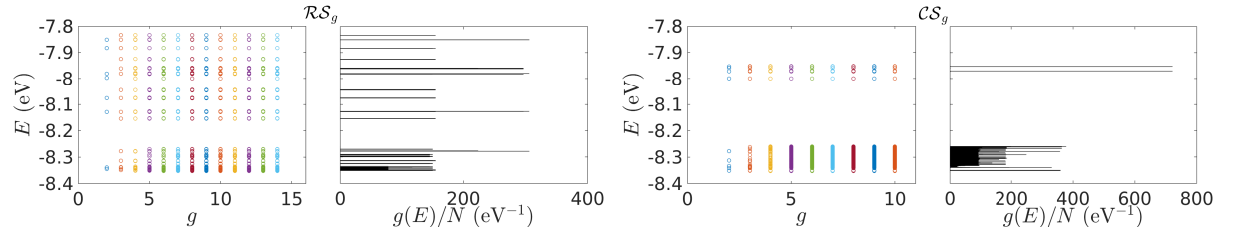
(a) Periodic sequence.

(b) Thue-Morse sequence.



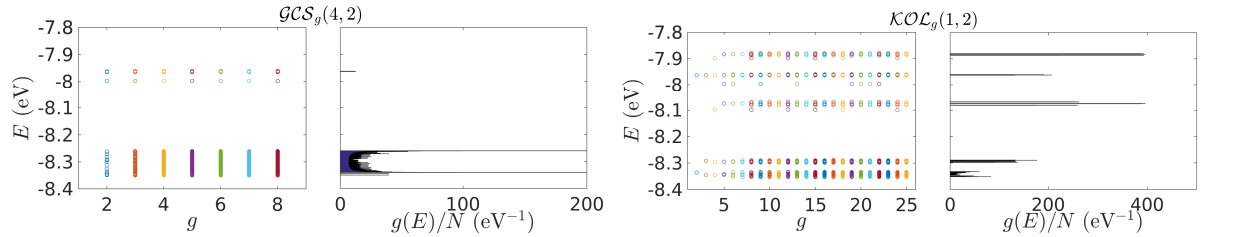
(c) Fibonacci sequence.

(d) Period Doubling sequence.



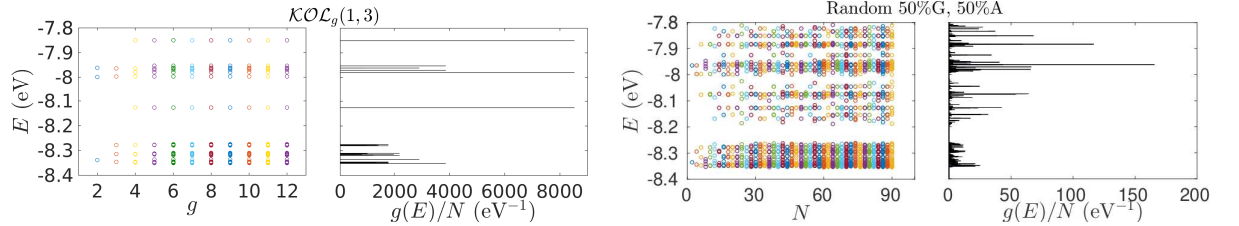
(e) Rudin-Shapiro sequence.

(f) Cantor set sequence.



(g) Generalized (4,2) Cantor set sequence.

(h) Kolakoski (1,2) sequence.



(i) Kolakoski (1,3) sequence.

(j) Random sequence (50% G content, 50% A content).

FIG. C.1: Eigenspectra and DOS of various categories of DNA segments.

- \* Electronic address: [klambro@phys.uoa.gr](mailto:klambro@phys.uoa.gr)
- † Electronic address: [csimseri@phys.uoa.gr](mailto:csimseri@phys.uoa.gr)
- <sup>1</sup> G. S. Manning, The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force, *Biophys J.* **91**, 3607 (2006).
  - <sup>2</sup> C. H. Wohlgamuth, M. A. McWilliams, and J. D. Slinker, DNA as a molecular wire: Distance and sequence dependence, *Anal. Chem.* **85**, 8634 (2013).
  - <sup>3</sup> F. D. Lewis and M. R. Wasielewski, Dynamics and efficiency of photoinduced charge transport in DNA: Toward the elusive molecular wire, *Pure Appl. Chem.* **85**, 1379 (2013).
  - <sup>4</sup> K. Kawai and T. Majima, Chapter 44, Increasing the hole transfer rate through DNA by chemical modification, in T. Akasaka, A. O. S. Fukuzumi, H. Kandori, and Y. Aso (eds.), *Chemical Science of  $\pi$ -Electron Systems*, Springer Tokyo Heidelberg New York Dordrecht London (2015). ISBN: [978-4-431-55357-1](#)
  - <sup>5</sup> K. Lambropoulos, C. Vantaraki, P. Mpilia, M. Mantela, and C. Simserides, Periodic polymers with increasing repetition unit: Energy structure and carrier transfer, submitted.
  - <sup>6</sup> R. Gutiérrez, R. Caetano, P. B. Woiczikowski, T. Kubař, M. Elstner, and G. Cuniberti, Structural fluctuations and quantum transport through DNA molecular wires: a combined molecular dynamics and model Hamiltonian approach, *New J. Phys.* **12**, 023022 (2010).
  - <sup>7</sup> E. Maciá, *Aperiodic Structures in Condensed Matter: Fundamentals and Applications*, Taylor & Francis CRC, Boca Raton FL, USA (2009). ISBN: [978-1-420-06827-6](#).
  - <sup>8</sup> Y.-J. Ye and L.-L. Shen, DFT Approach to calculate electronic transfer through a segment of DNA double helix, *J. Comput. Chem.* **21**, 1109 (2000).
  - <sup>9</sup> Y.-J. Ye and Y. Jiang, Electronic structures and long-range electron transfer through DNA molecules, *Int. J. Quantum Chem.* **78**, 112 (2000).
  - <sup>10</sup> R. N. Barnett, C. L. Cleveland, U. Landman, E. Boone, S. Kanvah, and G.B. Schuster, Effect of base sequence and hydration on the electronic and hole transport properties of duplex DNA: Theory and experiment, *J. Phys. Chem. A* **107**, 3525 (2003).
  - <sup>11</sup> E. Artacho, M. Machado, D. Sánchez-Portal, P. Ordejón, and J. M. Soler, Electrons in dry DNA from density functional calculations, *Mol. Phys.* **101**, 1587 (2003).
  - <sup>12</sup> C. Adessi, S. Walch, and M. P. Anantram, Environment and structure influence on DNA conduction, *Phys. Rev. B* **67**, 081405(R) (2003).
  - <sup>13</sup> H. Mehrez and M. P. Anantram, Interbase electronic coupling for transport through DNA, *Phys. Rev. B* **71**, 115405 (2005).
  - <sup>14</sup> A. A. Voityuk, Electronic couplings and on-site energies for hole transfer in DNA: Systematic quantum mechanical molecular dynamic study, *J. Chem. Phys.* **128**, 115101 (2008).
  - <sup>15</sup> T. Kubař, P. B. Woiczikowski, G. Cuniberti, and M. Elstner, Efficient calculation of charge-transfer matrix elements for hole transfer in DNA, *J. Phys. Chem. B* **112**, 7937 (2008).
  - <sup>16</sup> M. Tassi, A. Morphis, K. Lambropoulos, and C. Simserides, RT-TDDFT study of hole oscillations in B-DNA monomers and dimers, *Cogent Physics* **4**, 1361077 (2017).
  - <sup>17</sup> C. Simserides, A systematic study of electron or hole transfer along DNA dimers, trimers and polymers, *Chem. Phys.* **440**, 31 (2014).
  - <sup>18</sup> K. Lambropoulos, M. Chatzieleftheriou, A. Morphis, K. Kaklamanis, M. Theodorakou, and C. Simserides, Unbiased charge oscillations in B-DNA: Monomer polymers and dimer polymers, *Phys. Rev. E* **92**, 032725 (2015).
  - <sup>19</sup> K. Lambropoulos, M. Chatzieleftheriou, A. Morphis, K. Kaklamanis, R. Lopp, M. Theodorakou, M. Tassi, and C. Simserides, Electronic structure and carrier transfer in B-DNA monomer polymers and dimer polymers: Stationary and time-dependent aspects of wire model vs. extended ladder model, *Phys. Rev. E* **94**, 062403 (2016).
  - <sup>20</sup> G. Cuniberti, L. Craco, D. Porath, and C. Dekker, Backbone-induced semiconducting behavior in short DNA wires, *Phys. Rev. B* **65**, 241314(R) (2002).
  - <sup>21</sup> S. Roche, D. Bicout, E. Maciá, and E. Kats, Long range correlations in DNA: Scaling properties and charge transfer efficiency, *Phys. Rev. Lett.* **91**, 228101 (2003).
  - <sup>22</sup> S. Roche, Sequence dependent DNA-mediated conduction, *Phys. Rev. Lett* **91**, 108101 (2003).
  - <sup>23</sup> F. Palmero, J. F. R. Archilla, D. Hennig, and F. R. Romero, Effect of base-pair inhomogeneities on charge transport along the DNA molecule, mediated by twist and radial polarons, *New J. Phys.* **6**, 13 (2004).
  - <sup>24</sup> H. Yamada, Localization of electronic states in chain models based on real DNA sequence, *Phys. Lett. A* **332**, 65 (2004).
  - <sup>25</sup> V. M. Apalkov and T. Chakraborty, Electron dynamics in a DNA molecule, *Phys. Rev. B* **71**, 033102 (2005).
  - <sup>26</sup> D. K. Klotsa, R. A. Römer, and M. S. Turner, Electronic transport in DNA, *Biophys. J.* **89**, 2187 (2005).
  - <sup>27</sup> C. T. Shih, S. Roche, and R. A. Römer, Point-mutation effects on charge-transport properties of the tumor-suppressor gene p53, *Phys. Rev. Lett.* **100**, 018105 (2008).
  - <sup>28</sup> Y. S. Joe, S. H. Lee, and E. R. Hedin, Electron transport through asymmetric DNA molecules, *Phys. Lett. A* **374**, 2367 (2010).
  - <sup>29</sup> J. Yi, Conduction of DNA molecules: A charge-ladder model, *Phys. Rev. B* **68**, 193103 (2003).
  - <sup>30</sup> R. A. Caetano and P. A. Schulz, Sequencing-independent delocalization in a DNA-like double chain with base pairing, *Phys. Rev. Lett.* **95**, 126601 (2005).
  - <sup>31</sup> X. F. Wang and T. Chakraborty, Charge transfer via a two-strand superexchange bridge in DNA, *Phys. Rev. Lett.* **97**, 106602 (2006).
  - <sup>32</sup> E. Maciá, F. Triozon, and S. Roche, Contact-dependent effects and tunneling currents in DNA molecules, *Phys. Rev. B* **71**, 113106 (2005).
  - <sup>33</sup> R. G. Sarmento, E. L. Albuquerque, P. D. Sesion Jr., U. L. Fulco, B. P. W. de Oliveira, Electronic transport in double-strand poly(dG)poly(dC) DNA segments, *Phys. Lett. A* **373**, 1486 (2009).
  - <sup>34</sup> R. G. Sarmento, G. A. Mendes, E. L. Albuquerque, U. L. Fulco, M. S. Vasconcelos, O. Ujsághy, V. N. Freire, and E. W. S. Caetano, The DNA electronic specific heat at low temperature: The role of aperiodicity, *Phys. Lett. A* **376**, 2413 (2012).
  - <sup>35</sup> E. L. Albuquerque, U. L. Fulco, V. N. Freire, E. W. S. Caetano, M. L. Lyra, F. A. B. F. de Moura, DNA-based

- nanobiostuctured devices: The role of quasiperiodicity and correlation effects, *Phys. Rep.* **535**, 139 (2014).
- <sup>36</sup> E. Maciá, Electronic structure and transport properties of double-stranded Fibonacci DNA, *Phys. Rev. B* **74**, 245105 (2006).
- <sup>37</sup> C. J. Páez, P. A. Schulz, N. R. Wilson and R. A. Römer, Robust signatures in the current-voltage characteristics of DNA molecules oriented between two graphene nanoribbon electrodes, *New J. Phys.* **14**, 093049 (2012).
- <sup>38</sup> S. Kundu and S. N. Karmakar, Electronic specific heat of DNA: Effects of backbones and disorder, *Phys. Lett. A* **379**, 1377 (2015).
- <sup>39</sup> S. Fathizadeh, S. Behnia, and J. Ziaei, Engineering DNA Molecule Bridge between Metal Electrodes for High-Performance Molecular Transistor: An Environmental Dependent Approach, *J. Phys. Chem. B* **122**, 2487 (2018).
- <sup>40</sup> G. Cuniberti, E. Maciá, A. Rodríguez, and R. A. Römer, Chapter 1: Tight-Binding Modeling of Charge Migration in DNA Devices, pp. 1-20 in T. Chakraborty (ed.) *Charge Migration in DNA Perspectives from Physics, Chemistry, and Biology*, Springer-Verlag Berlin Heidelberg (2007). ISBN: **978-3-540-72494-0**.
- <sup>41</sup> L. G. D. Hawke, G. Kalosakas, and C. Simserides, Electronic parameters for charge transfer along DNA, *Eur. Phys. J. E* **32**, 291 (2010); *ibid.* Erratum to: Electronic parameters for charge transfer along DNA **34**, 118 (2011).
- <sup>42</sup> K. Lambropoulos, K. Kaklamanis, G. Georgiadis, and C. Simserides, THz and above THz electron or hole oscillations in DNA dimers and trimers, *Ann. Phys. (Berlin)* **526**, 249 (2014).
- <sup>43</sup> K. Lambropoulos, K. Kaklamanis, A. Morphis, M. Tassi, R. Lopp, G. Georgiadis, M. Theodorakou, M. Chatzieleftheriou, and C. Simserides, Wire and extended ladder model predict THz oscillations in DNA monomers, dimers and trimers *J. Phys.: Condens. Matter* **28**, 495101 (2016).
- <sup>44</sup> E. Maciá, Spectral classification of one-dimensional binary aperiodic crystals: An algebraic approach, *Ann. Phys. (Berl.)* **529**, 1700079 (2017).
- <sup>45</sup> K. Senthilkumar, F. C. Grozema, C. Fonseca Guerra, F. M. Bickelhaupt, F. D. Lewis, Y. A. Berlin, M. A. Ratner, L. D. A. Siebbeles, Absolute Rates of Hole Transfer in DNA, *J. Am. Chem. Soc.* **127** (2005) 14894.
- <sup>46</sup> E. L. Albuquerque, M. S. Vasconcelos, M. L. Lyra, and F. A. B. F. de Moura, Nucleotide correlations and electronic transport of DNA sequences, *Phys. Rev. E* **71**, 021910 (2005).
- <sup>47</sup> L. Molinari, Transfer matrices and tridiagonal-block Hamiltonians with periodic and scattering boundary conditions, *J. Phys. A: Math. Gen.* **30**, 983 (1997).
- <sup>48</sup> K. Lambropoulos, and C. Simserides, Spectral and transmission properties of periodic 1D tight-binding lattices with a generic unit cell: an analysis within the transfer matrix approach, *J. Phys. Commun.* **2**, 035013 (2018).
- <sup>49</sup> V. I. Oseledec, A multiplicative ergodic theorem. Characteristic Lyapunov exponents of dynamical systems, *Trans. Moscow Math. Soc.* **19**, 197 (1968).
- <sup>50</sup> A. Crisanti, G. Paladin, A. Vulpiani, *Products of random matrices in statistical physics*, Springer (1993). ISBN: **978-3-642-84942-8**.
- <sup>51</sup> J. A. Scales, and E. S. Van Vleck, Lyapunov exponents and localization in randomly layered media, *J. Comput. Phys.* **133**, 27 (1997).
- <sup>52</sup> R. Landauer, Spatial variation of currents and fields due to localized scatterers in metallic conduction, *IBM J. Res. Dev.* **1**, 223 (1957).
- <sup>53</sup> M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Generalized many-channel conductance formula with application to small rings, *Phys. Rev. B* **31**, 6207 (1985).
- <sup>54</sup> M. Büttiker, Symmetry of electrical conduction, *IBM J. Res. Dev.* **32**, 317 (1988).
- <sup>55</sup> G. C. Liang, A. W. Ghosh, M. Paulsson, and S. Datta, Electrostatic potential profiles of molecular conductors, *Phys. Rev. B* **69**, 115302 (2004).
- <sup>56</sup> S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, Cambridge (1995). ISBN: **978-0-511-80577-6**.
- <sup>57</sup> A.-M. Guo, Long-range correlation and charge transfer efficiency in substitutional sequences of DNA molecules, *Phys. Rev. E* **75** 061915 (2007).
- <sup>58</sup> J. I. N. Oliveira, E. L. Albuquerque, U. L. Fulco, P. W. Mauriz, R. G. Sarmento, E. W. S. Caetano, and V. N. Freire, Conductance of single microRNAs chains related to the autism spectrum disorder, *EPL* **107** 68006 (2014).
- <sup>59</sup> D. Porath, A. Bezryadin, S. de Vries and C. Dekker, Direct measurement of electrical transport through DNA molecules, *Nature* **403** 635 (2000).
- <sup>60</sup> L. E. Sigler (trans.), *Fibonacci's Liber Abaci: A translation into modern English of Leonardo Pisano's Book of Calculation*, Springer, 2002, ISBN: **978-0-387-40737-1**.
- <sup>61</sup> E. Prouhet, Mémoire sur les relations entre les puissances des nombres, *Comptes rendus de l'Académie des Sciences Paris* **33**, 225 (1851)(in French).
- <sup>62</sup> T. Nagell, A. Selberg, S. Selberg, K. Thalberg (eds.), *Selected mathematical papers of Axel Thue*, Universitetsforlaget, Oslo (1977).
- <sup>63</sup> M. Morse, Recurrent geodesics on a surface of negative curvature, *Trans. Amer. Math. Soc.* **22**, 84 (1921).
- <sup>64</sup> H. S. Shapiro, Extremal problems for polynomials and power series, Master's thesis, Massachusetts Institute of Technology (1951).
- <sup>65</sup> W. Rudin, Some theorems on Fourier coefficients, *Proc. Amer. Math. Soc.* **10**, 855 (1959).
- <sup>66</sup> G. Cantor, Über unendliche, lineare Punktmannigfaltigkeiten V, *Mathematische Annalen*, **21**, 545 (1883) (in German).
- <sup>67</sup> *The On-Line Encyclopedia of Integer Sequences*, <https://oeis.org>, 2010, Sequence **A000002**.
- <sup>68</sup> W. Kolakoski, Advanced Problems: Problem 5304, *Am. Math. Mon.* **72**, 674 (1965).
- <sup>69</sup> R. Oldenburger, Exponent trajectories in symbolic dynamics, *Trans. Amer. Math. Soc.* **49**, 453 (1939).
- <sup>70</sup> B. Sing, Kolakoski sequences - an example of aperiodic order, *J. Non. Cryst. Solids* **334-335**, 100 (2004).
- <sup>71</sup> *The On-Line Encyclopedia of Integer Sequences*, <https://oeis.org>, 2010, Sequence **A001083**.