Codon thermoelectric signature in molecular junctions

Enrique Maciá
Dpto. Física de Materiales, Facultad CC. Físicas, Universidad Complutense de Madrid, Madrid E-28040, Spain
(Received 25 May 2010; revised manuscript received 13 July 2010; published 30 July 2010)

The thermoelectric power of trimer oligonucleotides connected in between metallic contacts at different temperatures is theoretically studied. The obtained analytical expressions reveal the existence of important resonance effects leading to a significant thermopower enhancement for certain characteristic energies which depend on the specific electronic structure of considered codons. This result suggests the existence of a thermoelectric signature for different triplet associations of biological interest.

DOI: 10.1103/PhysRevB.82.045431 PACS number(s): 85.65.+h, 84.60.Rb, 87.10.Vg, 87.14.gk

I. INTRODUCTION

The proliferation of large-scale DNA-sequencing projects for applications in clinical medicine, health care, and criminal research has driven the quest for alternative approaches to the commonly used Sanger sequencing method in order to reduce time, error rate, and cost.[1] Since Sanger's method relies on chemistry to read the bases guanine (G), adenine (A), cytosine (C), and thymine (T) in DNA,[2] this quest has spurred new perspectives in nanotechnology looking for sequencing methods entirely based on physical differences between the bases for noninvasive detection of nucleotides along the DNA strands.[3] Measuring transverse (i.e., perpendicular to the helix axis) tunnel currents through single-stranded DNA as it translocates through a nanopore has been proposed as a suitable physical method enabling single-base resolution,[4,5] and it has experimentally been shown that the four bases provide a distinguishable transverse electronic signature when measured with a scanning tunnel microscope which directly detects the molecular levels of single DNA bases.[1,6] On the other hand, improvements of current nanotechnology allow us to confidently measure thermoelectricity at the molecular level as well. Thus, the thermoelectric properties of molecular junctions containing different benzene-related moieties chemically bond to gold nanocontacts has been investigated in a series of experiments with a suitably modified scanning tunneling microscope.[7,8] Positive values of the Seebeck coefficient [indicating that the Fermi level is closer to the highest occupied molecular orbital (HOMO) level] were obtained for all considered molecules when contacted through thiol groups, indicating that the charge carriers are holes in this case. On the contrary, a negative value is obtained for a benzene molecule contacted to gold electrodes with cyanide groups. Thus, end groups are key to control the very nature of charge carriers and by properly varying end groups and molecular junction constituents, one can engineer metal-molecule heterostructures with targeted thermoelectric properties.

Motivated by these experimental results in this paper, we will extend previous works proposing the possibility of sequencing short DNA fragments by employing thermoelectric measurements.[9,10] The working hypothesis inspiring our proposal is the following. The basic unit of information in the genetic code is the so-called codon. A codon is an ordered sequence of three consecutive nucleotides that specifies a particular amino acid in a protein or initiation (stops) sites where translation into protein synthesis begins (ends). From the viewpoint of condensed-matter physics, each codon is characterized by its electronic structure, stemming from orbital overlapping among neighboring nucleobases. Therefore, the resulting electronic structure provides a characteristic spectral portrait, also determining the codon transport properties. Then, one may regard charge carrier propagating through the oligonucleotide as a physical probe sensing its electronic structure *as a whole*. In this way, rather than a one-by-one nucleotide reading, typical of chemistry-based techniques, we will be able to *directly sensing* triplet nucleobases associations (including codons in codifying regions) at once. In order to substantiate this approach, we shall analyze the thermoelectric spectral curves of codons trapped between appropriate contacts at different temperatures, determining their characteristic *thermoelectric signature*. Due to the extreme sensitivity of thermopower to finer details in the codon-electrode electronic structure, the thermoelectric response of trimer nucleobases forming a molecular junction exhibits several narrow resonant features where the Seebeck coefficient attains very large values (200–2000 μV K^{-1} at room temperature). The position of these peaks sensitively depends on the characteristic electronic structure of the considered trimer, hence providing a very accurate method to properly identify different codons of biological interest.

II. MODEL DESCRIPTION

As a first approximation, a DNA-based molecular junction (Fig. 1 inset) can be described in terms of three noninteracting subsystems according to the tight-binding Hamiltonian,[11]

$$\mathcal{H} = \left(\sum_{n=1}^{N} \varepsilon_n c_n^\dagger c_n - \sum_{n=1}^{N-1} t_{n,n+1} c_n^\dagger c_{n+1} + \text{H.c.} \right) - \tau (c_0^\dagger c_1 + c_N^\dagger c_{N+1} + \text{H.c.}) + \sum_{l=0}^{-\infty} (\varepsilon_M c_l^\dagger c_l - t_M c_{l-1}^\dagger c_l + \text{H.c.}) + \sum_{l=N+1}^{+\infty} (\varepsilon_M c_l^\dagger c_l - t_M c_l^\dagger c_{l+1} + \text{H.c.}). \quad (1)$$

The first term in Eq. (1) describes DNA in terms of an effective linear chain with an orbital per site, the second term

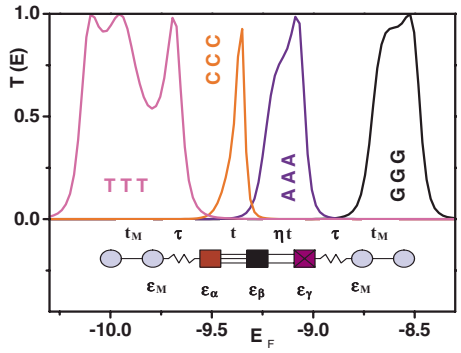

FIG. 1. (Color online) Transmission coefficient as a function of the Fermi energy for a TTT, CCC, AAA, and GGG codons with τ =0.5 eV, η=1, t_{TT}=0.158 eV, t_{GG}=0.084 eV, t_{AA}=0.070 eV, and t_{CC}=0.041 eV (Refs. 14 and 15). Full transmission energies (in eV) are given by E_{TTT}=−10.092, −9.953, −9.696 (triplet); E_{CCC} =−9.357; E_{AAA}=−9.080; and E_{GGG}=−8.526. (inset) Tight-binding codon-based molecular junction model considered in this work.

describes the DNA-metal contacts, and the last two terms describe the contacts at both sides, modeled as semi-infinite one-dimensional chains of atoms with one orbital per site, where N is the number of bases, $c_j^\dagger$ (c_j) is the creation (annihilation) operator for a charge at the jth site in the chain, ε_n are the on-site energies of the bases, $t_{n,n+1}$ is the hopping term between them, τ measures the coupling strength between the leads and the end nucleotides, ε_M is the leads on-site energy, and $t_M(>\tau)$ is their hopping term. Within the transfer-matrix framework, and considering nearest-neighbors interactions only, the Schrödinger equation corresponding to the Hamiltonian (1) can be expressed in the form

$$\begin{pmatrix} \psi_{N+1} \\ \psi_N \end{pmatrix} = \mathbf{T}_{N+1}\mathbf{T}_N \cdots \mathbf{T}_1\mathbf{T}_0 \begin{pmatrix} \psi_0 \\ \psi_{-1} \end{pmatrix}, \tag{2}$$

where ψ_n is the wave-function amplitude for the energy E at site n and

$$\mathbf{T}_n(E) = \begin{pmatrix} \dfrac{E-\varepsilon_n}{t_{n,n+1}} & -\dfrac{t_{n,n-1}}{t_{n,n+1}} \\ 1 & 0 \end{pmatrix} \tag{3}$$

is the local transfer matrix. The zero-bias transmission coefficient, $\mathcal{T}_N(E)$, describing the fraction of charge carriers transmitted through a chain of length N in the absence of any applied voltage, can be obtained from the knowledge of the leads dispersion relation, $E=\varepsilon_M+2t_M \cos k$, and the matrix elements of the global transfer matrix $\mathbf{M}(E)\equiv\Pi_{n=N+1}^{0}\mathbf{T}_n(E)$, by means of the relationship $\mathcal{T}_N(E)=4 \sin^2 k/\mathcal{D}_N(E)$, where $\mathcal{D}_N(E)=[M_{12}-M_{21}+(M_{11}-M_{22})\cos k]^2+(M_{11}+M_{22})^2\sin^2 k$. Experimental measurements of the Seebeck coefficient of molecular tunnel junctions have been conducted at zero bias.[7,8] In that case, the system can be described by only a single Fermi level and the Seebeck coefficient can be expressed in terms of $\mathcal{T}_N(E)$ by means of the expression,[12]

$$S_N(T) = -\frac{\pi^2 k_B^2}{3|e|}\left(\frac{\partial \ln \mathcal{T}_N(E)}{\partial E}\right)_{E_F} T, \tag{4}$$

where e is the electron charge, k_B is the Boltzmann constant, E_F is the Fermi energy, and T is the mean temperature of the contacts.

In the case of a codon (N=3, Fig. 1 inset), we have three nucleobases of energies ε_α, ε_β, and ε_γ, coupled with hopping terms $t_{\alpha\beta}\equiv t$ and $t_{\beta\gamma}\equiv \eta t$, and the transmission coefficient takes the form

$$\mathcal{T}_{\alpha\beta\gamma}(E) = [1 + 4t_M^2 q_{\alpha\beta}^2 W^{-1}(E) R_{\alpha\beta\gamma}^2(E)]^{-1}, \tag{5}$$

where $q_{\alpha\beta}\equiv(\lambda\mu_{\alpha\beta})^{-2}\eta^{-1}$, $\lambda\equiv\tau/t_M$ measures the coupling strength between the nucleobase and the leads (in units of the lead bandwidth $4t_M$), $\mu_{\alpha\beta}\equiv t_{\alpha\beta}/t_M$ measures the coupling strength between the bases, $W\equiv(E-E_-)(E_+-E)$, with $E_\pm \equiv \varepsilon_M \pm 2t_M$, define the allowed spectral window and

$$R_{\alpha\beta\gamma} \equiv 4x_\alpha x_\beta x_\gamma + \lambda^2(\eta^2\mu_{\alpha\beta}^2 - 4x_\beta x_\gamma)\cos k - \mu_{\alpha\beta}^2(x_\gamma + \eta^2 x_\alpha) + \lambda^4 x_\beta, \tag{6}$$

where $2x_v\equiv(E-\varepsilon_v)/t_M$. The roots of polynomial $R_{\alpha\beta\gamma}(E)$ determine the full transmission ($\mathcal{T}_{\alpha\beta\gamma}$=1) peaks energy values as well as the thermopower crossing points [see Eq. (4)] where the Seebeck coefficient change its sign from p type to n type (and vice versa). By plugging Eq. (5) into Eq. (4), we obtain

$$S_{\alpha\beta\gamma}(E_F,T) = S_0(T)[1-\mathcal{T}_{\alpha\beta\gamma}(E_F)] \times \left\{B(E_F) + \left(\frac{\partial \ln R_{\alpha\beta\gamma}}{\partial E}\right)_{E_F}\right\}, \tag{7}$$

where $S_0(T)=2|e|L_0T$, $L_0=\pi^2k_B^2/(3e^2)$ is the Lorenz number, and $B(E_F)\equiv(E_F-\varepsilon_M)/W(E_F)$. The Seebeck coefficient is then expressed as a product involving three contributions. The factor S_0 sets the thermovoltage scale (in μV K^{-1} eV units) at a given temperature. The factor $1-\mathcal{T}_{\alpha\beta\gamma}$ links the thermopower magnitude to the conductance properties of the chain so that the Seebeck coefficient progressively decreases (increases) as the conductance increases (decreases), vanishing when $\mathcal{T}_{\alpha\beta\gamma}$=1. The last factor in Eq. (7) depends on two additive contributions in turn. The value of $B(E_F)$ depends on the relative position of the Fermi level with respect to both the band center, ε_M, and the band edges of the contacts, $E_\pm$. Thus, its contribution vanishes when $E_F\to\varepsilon_M$, whereas B (and consequently $S_{\alpha\beta\gamma}$) asymptotically diverges as the Fermi level approaches the spectral window edges (i.e., $E_F\to E_\pm$). Finally, the logarithmic derivative term in Eq. (7) determines the overall behavior of the $S_{\alpha\beta\gamma}$ curve.

III. TRANSMISSION AND THERMOPOWER CURVES

In order to provide a realistic description of codon triplets, we have used the *ab initio* derived HOMO energy levels of G, A, C, and T nucleotides (consisting of a nucleobase, a ribose sugar, and a phosphate group) in the presence of water given by ε_G=−8.58 eV, ε_A=−9.12 eV, ε_C=−9.37 eV, and ε_T=−9.86 eV.[13] Depending on the DNA sequence composi-

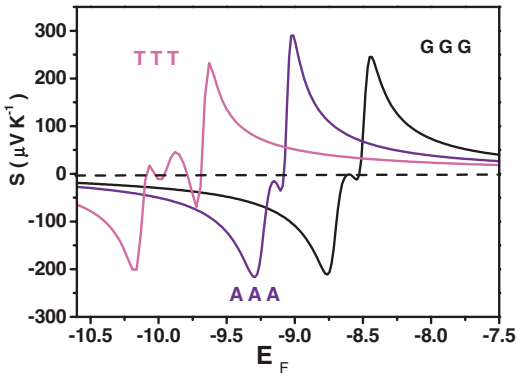

FIG. 2. (Color online) Dependence of the room-temperature thermopower as a function of the Fermi-level energy for TTT, AAA, and GGG codons with the same model parameters used in Fig. 1.

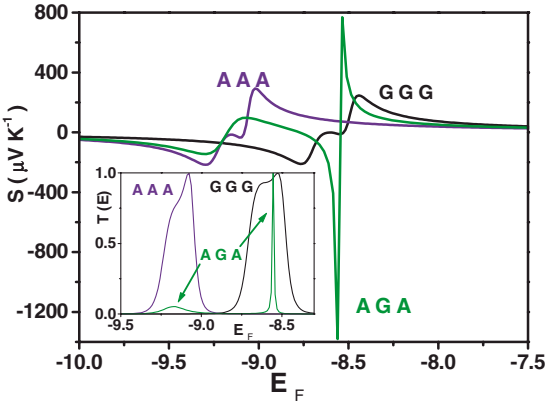

FIG. 3. (Color online) Dependence of the room-temperature thermopower (main frame) and transmission coefficient (inset) as a function of the Fermi energy for AGA, AAA, and GGG codons with τ=0.5 eV, t_{GG}=0.084 eV, t_{AA}=0.070 eV, and t_{GA}=t_{AG}=0.070 eV (Refs. 14 and 15). The arrows in the inset feature the narrow (E_{AGA}=−8.555 eV) and broad (E_{AGA}=−9.155 eV) peaks characteristic of XYX-type codons.

tion, its length and temperature the effective value of the hopping integrals $t_{\alpha\beta}$ can vary over a relatively broad range. We have considered the values derived from quantum chemistry/molecular-dynamics calculations, including dynamical effects on the orbital overlapping when available.[14,15] Modeling the geometry and bonding character of the contacts is a very delicate issue since detailed information on both the metal geometry and DNA chemical bonding at the contacts is poorly known to date. Consequently, in our work the parameter τ=0.5–1.0 eV deals with the tunneling probability between the frontier orbitals. Finally, we have considered a contact electronic configuration able to mimic reported experimental results indicating a E_F-HOMO separation ~1 eV,[7,8] with the tight-binding parameters t_M=2.0 eV and ε_M=−7.58 eV, determining the allowed spectral window [−11.58,−3.58] eV.

In the first place, we shall consider the transport properties corresponding to homomer GGG, AAA, CCC, and TTT codons, respectively, codifying for glycine, lysine, proline, and phenylalanine in the standard genetic code.[16] In Figs. 1 and 2, we show their transmission and thermopower curves as derived from Eqs. (5) and (7), respectively. The $S(E_F)$ curves corresponding to the GGG, AAA, and CCC (not shown since it peaks at about $\pm$2000 μV K^{-1}) trimers are characterized by two peaks and a crossing point, defining two different regimes exhibiting p-type or n-type thermopower, respectively. The Seebeck coefficient value attained at the peaks is significantly large and it compares well with values reported for benchmark thermoelectric materials. By inspecting Fig. 2, two main conclusions can be drawn. First, each homomer thermopower curve peaks at characteristic energy values, which are different enough to allow for a reliable energy resolution among different XXX moieties. Second, the thermopower curves corresponding to certain codons display complementary responses among them. Thus, the AAA thermopower curve shows a p-type behavior within the energy range $-9.1 \lesssim E \lesssim -8.6$ whereas the GGG curve takes negative values instead. The same occurs for CCC and AAA (not shown), or TTT and AAA in the energy windows $-9.4 \lesssim E \lesssim -9.1$, and $-9.7 \lesssim E \lesssim -9.1$, respectively. In addition, the $S_{TTT}(E)$ curve exhibits two subsidiary (positive) peaks located in an energy interval where the other three trimers take on negative values. These features provide an additional criterion highlighting the presence of TTT trimers as they stem from the stronger coupling between neighboring T-T bases, giving rise to a characteristic triplet structure in the transmission coefficient shown in Fig. 1.

Now we consider the thermopower curve corresponding to trimers of the form XYX as compared to that previously obtained for XXX codons. As a representative example, in Fig. 3 we compare the thermopower curves for the AGA (codifying for arginine), AAA, and GGG codons. The presence of a very strong thermopower signal at E=−8.555 eV is related to the very narrow feature in the transmission coefficient shown in the inset. This feature clearly distinguishes the AGA codon from its related homomers. On the other hand, the thermopower curve corresponding to the GAG codon (codifying for glutamic acid), which is obtained from AGA by exchanging A$\leftrightarrow$G bases, also exhibits a strong thermopower signal but it is now located at E=−9.150 (that is, where the AGA codon exhibits a small, broad transmission peak). Accordingly, as a result of the base exchange the positions of narrow and broad resonances are mutually interchanged in both S and $\mathcal{T}$ curves, which allows for a clear distinguishability among codons of similar chemical composition codifying for different amino acids. This property is a general feature of the thermopower curve corresponding to all codons of the form XYX, and is directly related to the algebraic structure of the $R_{\alpha\beta\gamma}(E)$ polynomial under base permutations. This is a remarkable result suggesting an underlying structure of the genetic code which expresses itself in a physical transport property of codon units. To further analyze this interesting issue, we will compare the thermopower spectral portraits corresponding to the codons TAG and TGA (codifying protein synthesis stops) versus the ATG codon (codifying for the initiation site). As we can see these codons are mutually related through a simple permutation operation. By inspecting Fig. 4, we realize that both stop codons exhibit strong Seebeck signals of similar intensity which are close in energy whereas the start codon ther-

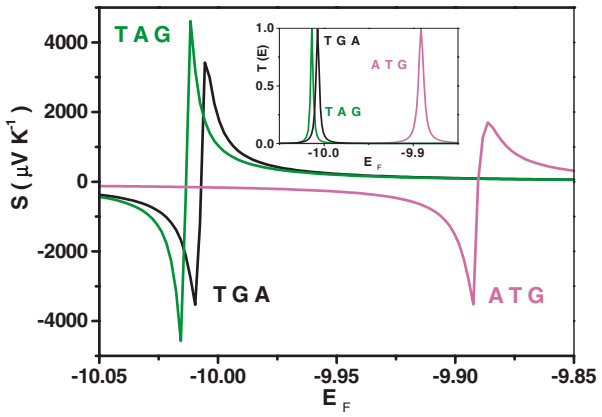

FIG. 4. (Color online) Dependence of the room-temperature thermopower (main frame) and transmission coefficient (inset) as a function of the Fermi-level energy for TAG, TGA, and ATG codons with τ=0.5 eV, $t_{TG}=t_{GT}$=0.137 eV, $t_{AT}=t_{TA}$=0.105 eV, and $t_{GA}=t_{AG}$=0.070 eV (Refs. 14 and 15). Full transmission energies (in eV) are given by E_{TAG}=−10.013, E_{TGA}=−10.007, and E_{ATG}=−9.891.

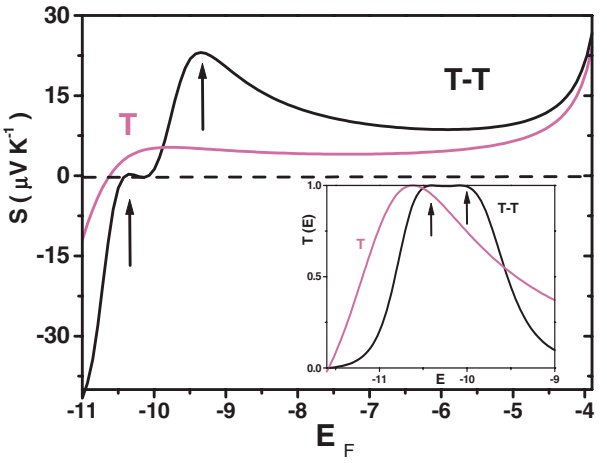

FIG. 5. (Color online) Room-temperature dependence of the Seebeck coefficient as a function of the Fermi energy for TT and T nucleobases with τ=1.0 eV, t_m=2.0 eV, t_{TT}=0.4 eV, and ε_m=−7.58 eV. (Inset) Transmission coefficient as a function of the Fermi energy for T and TT bases. The vertical arrows indicate the location of the resonance energy values (inset) and the related peaks in the thermopower curve (main frame).

mopower peak is significantly shifted toward higher energies and it exhibits smaller amplitude. Accordingly, codons related to complementary tasks in the protein synthesis can be properly assigned quite different thermoelectric responses.

IV. CONCLUSIONS

As it was mentioned in Sec. I, recent advances in single-molecule sequencing have considered the detection of electrical conductivity changes using nanopores.[4,5] This requires electrophoretical threading of DNA molecules through a nanoscale pore while a current is applied in the transversal direction (perpendicular to the DNA stack). The major challenges of this method are twofold. On the one hand, one has the stochastic motion of the DNA molecule as it passes through the pore which increases the background noise thereby reducing the potential for single-base resolution.[17] On the other hand, in phyisiological conditions the DNA motion might even be too fast to detect the single bases at all.[18] In this work, we consider an alternative approach based on the thermoelectric properties of DNA codon units forming a molecular junction. Since the molecule is fixed at the ends, the aforementioned kinematic drawbacks are properly circumvented in these case. In fact, rather than a one-by-one nucleotide reading, typical of DNA motion-based techniques, one *directly senses* triplet nucleobases associations as a whole by measuring their thermopower energy dependence. To illustrate this important issue in Fig. 5, we show the thermopower energy dependence for a single T nuclotide and a TT dimer, respectively, attached to metallic contacts at different temperatures at both ends. As we can see, the $S(E)$ curve corresponding to the single nucleotide is almost featureless whereas that corresponding to the TT dimer exhibits two features, which are, respectively, related to the transmission peaks shown in the inset. In turn, by comparing the $S(E)$ and $T(E)$ curves shown in Fig. 5 with those corresponding to the TTT triplet in Figs. 1 and 2, we realize that rather than detecting the presence of three individual thymine bases along the DNA chain (each one characterized by a featureless $S(E)$ curve as that shown in Fig. 5) this approach allows one to readily identify the presence of each kind of codon as a molecular entity by itself. This ability relies on resonance orbital overlapping among the different bases along the DNA stacking direction, ultimately determining the electronic structure of different triplets. Therefore, by properly comparing the spectral portraits corresponding to different codons, one can confidently disclose nucleotide associations in terms of their characteristic thermoelectric signature. Certainly, this approach could be readily extended to consider DNA oligonucleotides with $N>3$. In this work, we have focused on the case $N=3$ in order to show the very possibility of elaborating a *thermopower* signature *codon chart*. The main goal of this chart is to assign to each codon triplet a characteristic $S(E)$ curve, as it has been illustrated through Figs. 2–4 for the GGG (glycine), AAA (lysine), TTT (phenylalanine), AGA (arginine), TAG (stop), TGA (stop), and ATG (initiation).

Since the coding properties of DNA introns are closely related to codon triplet associations these preliminary results may enclose some biological relevance well deserving some experimental test. From a practical viewpoint, since the Fermi level is mainly determined by the contact leads and is not easily tunable, a third contact able to electrotrostatically gating the codon molecular orbitals should be properly included in a realistic experimental setup. In this way, it would be possible to complement electrical conductance transport measurements with thermopower measurements to study the electronic properties of short DNA sequences. Finally, the

obtained results also show that the presence of resonance effects involving molecular orbitals in codon's nucleobases and the contact's Fermi level can lead to a significan enhancement of the thermoelectric signal at certain characteristic energies, hence suggesting the potential of nucleic acids for thermoelectric energy conversion on molecular scale.

ACKNOWLEDGMENTS

I warmly thank Pramod Reddy for sharing useful information and M. Victoria Hernández for a critical reading of the manuscript.

[1] M. Xu, R. G. Endres, and Y. Arakawa, in *Charge Migration in DNA: Perspectives from Physics, Chemistry and Biology*, Nanoscience and Technology, edited by T. Chakraborty (Springer-Verlag, Berlin, 2007), p. 205.
[2] F. Sanger, S. Nicklen, and A. R. Coulson, Proc. Natl. Acad. Sci. U.S.A. **74**, 5463 (1977).
[3] M. Zwolak and M. Di Ventra, Rev. Mod. Phys. **80**, 141 (2008).
[4] J. Lagerqvist, M. Zwolak, and M. Di Ventra, Nano Lett. **6**, 779 (2006); M. Zwolak and M. Di Ventra, *ibid.* **5**, 421 (2005).
[5] H. W. C. Postma, Nano Lett. **10**, 420 (2010).
[6] M. Xu, R. G. Endres, and Y. Arakawa, Small **3**, 1539 (2007).
[7] P. Reddy, S. Y. Jang, R. A. Segalman, and A. Majumdar, Science **315**, 1568 (2007); K. Baheti, J. A. Malen, P. Doak, P. Reddy, S. Y. Jang, T. D. Tilley, A. Majumdar, and R. A. Segalman, Nano Lett. **8**, 715 (2008).
[8] A. Tan, S. Sadat, and P. Reddy, Appl. Phys. Lett. **96**, 013110 (2010).
[9] S. Roche and E. Maciá, Mod. Phys. Lett. B **18**, 847 (2004).
[10] E. Maciá, Nanotechnology **16**, S254 (2005); Rev. Adv. Mater. Sci. **10**, 166 (2005).
[11] E. Maciá, Phys. Rev. B **75**, 035130 (2007).
[12] M. Paulsson and S. Datta, Phys. Rev. B **67**, 241403(R) (2003).
[13] J. Ladik, A. Bende, and F. Bogár, J. Chem. Phys. **128**, 105101 (2008).
[14] A. A. Voityuk, J. Jortner, M. Bixon, and N. Rösch, J. Chem. Phys. **114**, 5614 (2001); J. Phys. Chem. B **104**, 9740 (2000).
[15] A. A. Voityuk, J. Chem. Phys. **128**, 115101 (2008).
[16] A. J. F. Griffiths, J. H. Miller, T. A. Suzuki, R. C. Lewontin, and W. M. Gelbart, *An Introduction to Genetic Analysis*, 7th ed. (Freeman, San Francisco, 1999).
[17] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, and X. Huang, Nat. Biotechnol. **26**, 1146 (2008).
[18] E. Treffer and V. Deckert, Curr. Opin. Biotechnol. **21**, 4 (2010).