

## Functionalized Nanopore-Embedded Electrodes for Rapid DNA Sequencing

Haiying He, Ralph H. Scheicher,<sup>\*,‡</sup> and Ravindra Pandey<sup>\*</sup>

*Department of Physics and Multi-Scale Technologies Institute, Michigan Technological University, Houghton, Michigan 49931*

Alexandre Reily Rocha<sup>†</sup> and Stefano Sanvito

*School of Physics and CRANN, Trinity College, Dublin 2, Ireland*

Anton Grigoriev and Rajeev Ahuja

*Condensed Matter Theory Group, Department of Physics, Box 530, Uppsala University, Sweden*

Shashi P. Karna

*U.S. Army Research Laboratory, Weapons and Materials Research Directorate, ATTN: AMSRD-ARL-WM, Aberdeen Proving Ground, Maryland 21005-5069*

*Received: December 6, 2007*

With the aim of improving nanopore-based DNA sequencing, we explored the effects of functionalizing the embedded gold electrodes with purine and pyrimidine molecules. Hydrogen bonds formed between the molecular probe and target bases stabilize the scanned DNA unit against thermal fluctuations and thus greatly reduce noise in the current signal. The results of our first-principles study indicate that this proposed scheme could allow DNA sequencing with a robust and reliable yield, producing current signals that differ by at least 1 order of magnitude for the different bases.

The determination of a patient's DNA sequence can, in principle, reveal an increased risk to fall ill with particular diseases<sup>1,2</sup> and help to design "personalized medicine".<sup>3</sup> Moreover, statistical studies and comparison of genomes<sup>4</sup> of a large number of individuals are crucial for the analysis of mutations<sup>5</sup> and hereditary diseases, paving the way to preventive medicine.<sup>6</sup> DNA sequencing is, however, currently still a vastly time-consuming and very expensive task,<sup>4</sup> consisting of preprocessing steps, the actual sequencing using the Sanger method, and postprocessing in the form of data analysis.<sup>7</sup>

The new possibility of passing a single-stranded DNA (ssDNA) through a so-called "nanopore" with a diameter of only a few nanometers has been explored over the past decade<sup>8–12</sup> for the purpose of DNA sequencing. As the nucleotides of DNA are migrated across the membrane, they will partially block the pore in different ways depending on their size.<sup>13</sup> Thus, monitoring the ionic blockade currents could lead to the determination of the DNA sequence.<sup>11,14</sup> To resolve the remaining ambiguity between the pyrimidine bases cytosine (C) and thymine (T) and the purine bases adenine (A) and guanine (G), it was recently suggested<sup>15,16</sup> that one could embed electrodes in the walls of a solid-state nanopore.<sup>17</sup> By applying a bias voltage across the electrodes a small electric current perpendicular to the DNA

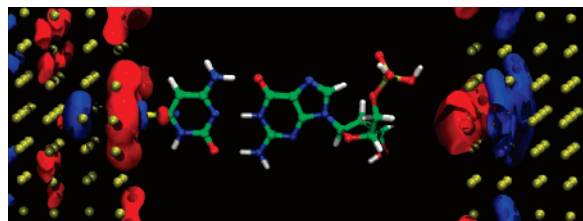
strand can be measured. Thus as the ssDNA translocates through the pore, time-dependent current–voltage signals from the electrodes would supposedly be specific enough to allow the unequivocal identification of the nucleotide sequence. It is, however, still unclear whether or not the resulting current–voltage signal from pore-embedded electrodes is sensitive enough to reliably distinguish the four nucleic acid bases of DNA. The probability distributions of current signals for different nucleotides are found to overlap significantly<sup>18,19</sup> largely due to the structural fluctuations of nucleotides between the embedded electrodes, which makes a reliable distinction of the different bases rather difficult. Although it may be possible to resolve the nucleic acid bases by applying a much smaller driving field and using statistical analysis, this approach requires a relatively large number of independent electrical current samplings of a given base. Even so, the resulting probability distributions of the signal currents span only 1 order of magnitude,<sup>20</sup> which may not be good enough for an accurate sequencing run in a realistic setup.

In this letter, we propose a new approach that relies on functionalized nanopore-embedded electrodes to achieve an unambiguous distinction of the four nucleic acid bases in the DNA sequencing process. The transport properties of the setup investigated by us, employing state-of-the-art density functional theory (DFT) together with the nonequilibrium Green's function (NEGF) method, lead to current responses that differ by at least 1 order of magnitude for different bases. These results are

\* Corresponding authors. E-mail: (R.H.S.) rhs@mtu.edu; (R.P.) pandey@mtu.edu.

<sup>†</sup> Present address: University of São Paulo, São Paulo, Brazil.

<sup>‡</sup> Present address: CMT Group, Uppsala University, Sweden.



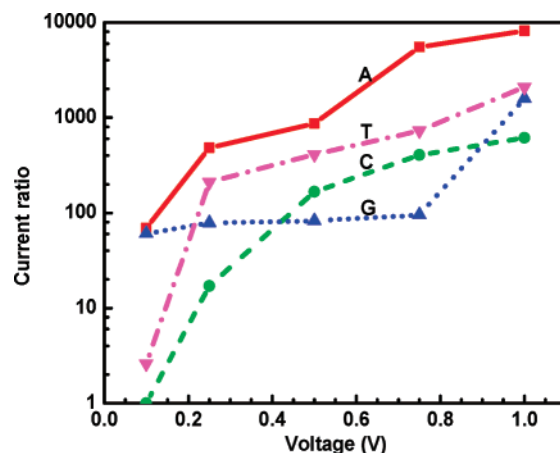
**Figure 1.** An illustration of the proposed device. ssDNA is passing through the nanopore with functionalized gold electrodes embedded. A probe molecule (e.g., Cytosine) is immobilized on the inner surface of the left electrode by a sulfur atom. As part of the ssDNA molecule, a guanosine monophosphate unit is shown as the target to be identified. The wave function of the first occupied contact state corresponding to the transmission peak labeled by \* in Figure 4 is also shown with the colors blue and red indicating the phase.

intriguing in that this approach would potentially achieve base contrasts large enough to discriminate the four DNA bases.

A schematic illustration of the key elements of the proposed device is shown in Figure 1. It consists of two gold electrodes embedded oppositely in a nanopore, a chosen DNA probe molecule anchored to the inner surface of one of the electrodes via a thiol group, and a target that is a segment of a ssDNA including base, sugar, and phosphate group. Because the ions and water can effectively screen the charge on the phosphate group in a solution,<sup>20</sup> it is kept uncharged. As an ssDNA is pulled through the nanopore via a driving electric field, the probing molecule will simultaneously fulfill two functions: (i) stabilization of the target base in the DNA sequence by forming weak hydrogen bonds, and (ii) detection of the target base by coupling electronically to it. Because the bases occurring in natural DNA possess an inherent ability to selectively bind to their respective complementary base partners, all four base molecules (i.e., A, C, G, and T) are considered as the probe molecules in calculations.

We have started our investigation by determining the most stable pairing geometries between the probe and the target molecule. For each pair, five initial positions were considered corresponding to the five matching position at the H-bonding edge (see Supporting Information for more details). We employed DFT within the generalized gradient approximation (GGA) of the exchange and correlation functional<sup>21</sup> incorporated in SIESTA.<sup>22</sup> Norm-conserving pseudopotentials and double- $\zeta$  basis sets with polarization orbitals were used on all the atoms. The tolerance in maximum density matrix difference is  $10^{-5}$  and the tolerance in maximum atomic force is  $0.04$  eV/Å. The bias-dependent tunneling current is then calculated from the NEGF method based on the Keldysh formalism, as implemented in SMEAGOL.<sup>23–25</sup> A central scattering region is defined to include the DNA probe-target molecules and four  $4 \times 4$  R Au layers on either side. In considering the computational cost, single- $\zeta$  basis sets with polarization orbitals were used for C, N, O, S, and H, while a single- $\zeta$  basis set was used for Au with only 6s as the valence electron. The charge density is obtained by integrating the Green's function over 200 imaginary and 1000 real energy points according to the scheme described in ref 25.

The binding energies ( $E_b$ ) of the “probe-target” base pairs formed temporarily in the nanopore are strongly correlated to the number of H-bonds formed in between with about  $\sim 0.4$  eV per H-bond. The calculated values for  $E_b$  show that the H-bonding can significantly stabilize the DNA molecule as it passes through the pore, thereby preventing drastic variations in the current due to thermal fluctuations of the structure and the uncertainty in the orientation of the base between the two electrodes. On the other hand, the H-bonding between probe-



**Figure 2.** The current–voltage curves for the device functionalized with a C-probe for all four possible target bases (A, red square; C, green circle; G, blue triangle; T, pink upside-down triangle) with current signals plotted in logarithmic scale. The current signals are taken as ratios with respect to the value of having C as the target at a bias of 0.1 V.

**TABLE 1: Summary of the Information Deducible for Different Probes from Current Measurements (at 100 mV and 250 mV)<sup>a</sup>**

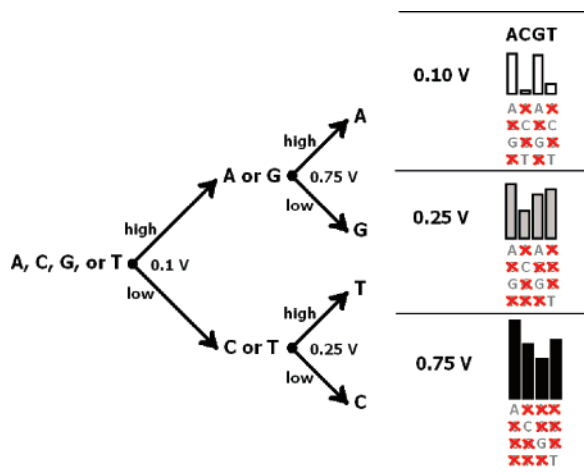
probe/bias	$V = 100$ mV	$V = 250$ mV
A	(A, C, G) from T	A from (C, G) from T
C	(A, G) from (C, T)	A from G, C from T
G	(A, C, G) from T	(A, C, G) from T
T	(A, C, G) from T	C from (A, G) from T

<sup>a</sup> Target bases that cannot be distinguished are combined in parentheses.

and target-base is weak enough to allow for an easy breakup when the ssDNA molecule is pulled through the nanopore by the driving field. The covalent bond of the probe-base's thiol group is, however, strong enough to keep the probe firmly placed on the surface of the gold electrode without the risk that it would be swept along with the passing DNA molecule.

The current–voltage curves calculated for the device involving the base C as a probe are shown in Figure 2. The vacuum gap between target base and gold electrode (Figure 1) leads to relatively small current values at low biases. Recalling that the figure of merit for distinguishing two different base molecules is that their associated currents differ by at least 1 order of magnitude, in Table 1 we present our recognition map. For instance, at a bias of 100 mV, using either A, G, or T as a probe, we can distinguish the set A, C, and G from T (while A, C, and G cannot be distinguished from each other). In contrast, when C is used as a probe, we can differentiate A and G from C and T. At an increased bias, the recognition properties of the proposed device change. For instance at 250 mV, the A probe provides different currents for A, T or C and G (C and G however remain indistinguishable from each other). Very importantly, our scheme adds the ability of identifying A and G, which used to be the major obstacle of nanopore DNA sequencing.

Functionalization of the electrodes with C as a probe appears to yield the best way for a reliable identification of all four base molecules in DNA (Figure 2). This can be achieved by carrying out three sequencing runs at three different bias voltages, namely 100 mV, 250 mV, and 750 mV. The flow diagram shown in Figure 3 gives the illustration of this proposed protocol of DNA sequencing. The first set of measurements at 100 mV would result in a series of current signals that fall into

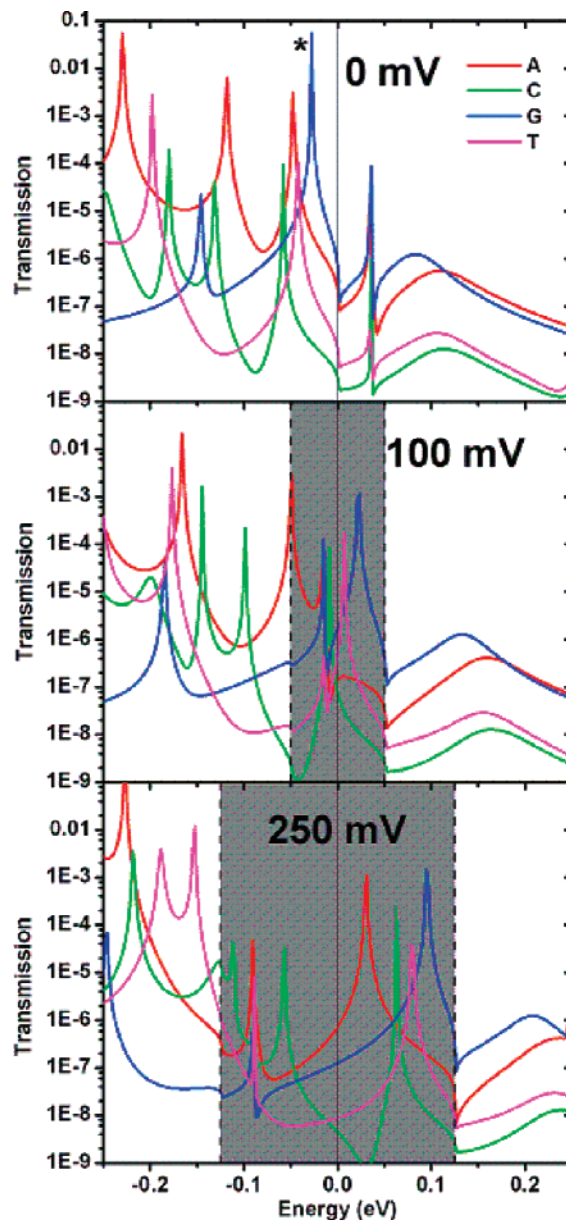


**Figure 3.** Flow diagram illustrating the decision-making process of a device involving C-probe leading to the identification of a target base in the sequence. Here, high and low refer to higher or lower current values at a given bias voltage. The height of the bars below the letters A, C, G, and T on the right side of the figure corresponds to the respective current signal (on a logarithmic scale). The crossed-out letters below the bars refer to possible target bases that have been ruled out.

two easily distinguishable categories: “high” current values if A or G is the target base, and “low” current values if C or T is the target base. The difference between the two categories is nearly 2 orders of magnitude, which should make the distinction extraordinarily robust. We now require additional information to resolve the remaining ambiguity of A/G and C/T. In a second measurement at 250 mV, it will be possible to distinguish between C and T, as their respective current values differ by 1 order of magnitude at that bias voltage. Thus, any high current value would lead to the identification of a T in the sequence, while any low current value means that a C is at this position in the sequence. Finally, a third measurement at 750 mV causes the current values for the bases A and G to differ by 2 orders of magnitude, leading to an easy distinction between the two, where high current values correspond to A, while low current values correspond to G.

Having demonstrated the principal capability of our hypothetical device for DNA sequencing, we now turn our attention to understand the electrical response for different target bases. In the following discussion, we concentrate on the most promising setup involving the C probe. The tunneling current (Figure 2) is obtained by integrating the transmission function shown in Figure 4 for each target base predominantly within the voltage window.<sup>25,26</sup> For the given nanopore (2.18 nm wide in this case), the similarity in the zero-bias transmission-function shape for all target nucleotides in the near- $E_F$  region can be attributed to the fact that the peaks near  $E_F$  and within the molecular HOMO–LUMO gap, away from specific molecular orbitals, are mostly associated with the contact states localized on the anchoring group and the gold electrodes (Figure 1). The size of the target base, thereby the gap between target and the (right) gold electrode, determines the transmission magnitude via the overlap between the electronic states localized on the molecule and the states on the right electrode. Nevertheless, we would like to emphasize here that it is the different effect of the target base on the bias-dependent shift of these peaks that allows one to distinguish between the four bases.

When the bias is increased from 0 to 250 mV, the transmission peaks associated with the contact states enter the voltage window and contribute to the increase of the current across the device (e.g., the potential on the left electrode increases, thus



**Figure 4.** Device with C-Probe. Transmission functions for four target bases at a bias of  $V = 0$  mV, 100 mV, and 250 mV with transmission values plotted in logarithmic scale. The Fermi level is aligned to zero. The voltage window of  $\pm V/2$  is indicated by the shaded area. The transmission peak for target “G” at zero bias corresponding to the molecular orbital shown in Figure 1 is labeled by \*.

causing a rise in the energy of the peaks). The position of the peak related to the contact state closely follows the shift of the electrode potential and is only slightly affected by charging/discharging of the state when it is driven between the occupied state on the left electrode and the unoccupied on the right.

In summary, we have shown that nanopore-embedded gold electrodes functionalized with molecular probes can lead to a dramatic improvement in the sensitivity of base molecules sensing in a pore-translocating DNA sequence. In STM experiments, researchers have indeed modified the probing tips with nucleic acid bases and succeeded in distinguishing their complementary base from others.<sup>27</sup> If the described setup in this study could be successfully implemented, then our findings might lead us closer to achieving the goal of rapid DNA sequencing at a low cost. We should, however, be aware that the solvent effect is not taken into account in the current simulation. A recent study from Di Ventra’s group<sup>20</sup> has shown

that water in the environment has little effect on the electrical response of the system. So we would expect that the potential distinguishability of these bases would not significantly be affected. It is also of great interest in the future to study the DNA translocation dynamics through the nanopore.

**Acknowledgment.** Helpful discussions with S. Gowtham and K. C. Lau are acknowledged. The work at Michigan Technological University was supported by DARPA (contract number ARL-DAAD17-03-C-0115). A.G., R.H.S., and R.A. gratefully acknowledge financial support from Carl Tryggers Stiftelse för Vetenskaplig Forskning, Göran Gustafsson Stiftelse, and Wenner-Gren Stiftelserna. A.R.R. and S.S. thank Science Foundation of Ireland (Grants SFI02/IN1/I175 and SFI05/RFP/PHY0062) for financial support.

**Supporting Information Available:** Details on the computational method, optimized geometries of base-pairs, and transmission functions for a larger bias range. This material is available free of charge via the Internet at <http://pub.acs.org>.

## References and Notes

- (1) Futreal, P. A.; et al. *Nature* **2001**, *409*, 850.
- (2) Jimenez-Sanchez, G.; Childs, B.; Valle, D. *Nature* **2001**, *409*, 853.
- (3) Chakravarti, A. *Nature* **2001**, *409*, 822.
- (4) Church, G. M. *Sci. Am.* **2006**, *294*, 47.
- (5) Stoneking, M. *Nature* **2001**, *409*, 821.
- (6) Hood, L.; et al. *Science* **2004**, *306*, 640.
- (7) International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860.
- (8) Kasianowicz, J. J.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770.
- (9) Akeson, M.; et al. *Biophys. J.* **1999**, *77*, 3227.
- (10) Meller, A.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1079.
- (11) Deamer, D. W.; Akeson, M. *Trends Biotechnol.* **2000**, *18*, 147 and references therein.
- (12) Khanna, V. K. *Biotechnol. Adv.* **2007**, *25*, 85.
- (13) Fologea, D.; et al. *Nano Lett.* **2005**, *5*, 1905.
- (14) Vercoutere, W. A.; et al. *Nucleic Acids Res.* **2003**, *31*, 1311.
- (15) Li, J.; Gershow, M., et al. *Nat. Mater.* **2003**, *2*, 611.
- (16) Rhee, M.; Burns, M. A. *Trends Biotechnol.* **2007**, *25*, 174.
- (17) Zwolak, M.; Di Ventra, M. *Rev. Mod. Phys.* **2008**, *80*, 141.
- (18) Zwolak, M.; Di Ventra, M. *Nano Lett.* **2005**, *5*, 421.
- (19) Lagerqvist, J.; Zwolak, M.; Di Ventra, M. *Nano Lett.* **2006**, *6*, 779.
- (20) Lagerqvist, J.; Zwolak, M.; Di Ventra, M. *Biophys. J.* **2007**, *93*, 2384.
- (21) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (22) Soler, J. M.; et al. *J. Phys. Condens. Matter* **2002**, *14*, 2745.
- (23) Rocha, A. R.; et al. Computer code SMEAGOL (Spin and Molecular Electronics in an Atomically Generated Orbital Landscape), [www.smeagol.tcd.ie](http://www.smeagol.tcd.ie) (accessed March 2007).
- (24) Rocha, A. R.; et al. *Nat. Mater.* **2005**, *4*, 335.
- (25) Rocha, A. R.; et al. *Phys. Rev. B* **2006**, *73*, 085414.
- (26) Grigoriev, A.; Ahuja, R. *Molecular Electronics Devices in Nano and Molecular Electronics Handbook*; Lyshevski, S. E., Ed.; CRC Press: Boca Raton, FL, 2007.
- (27) Ohshiro, T.; Umezawa, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10.