Efficient DNA Sensing with Fabricated Silicon Nanopores: Diagnosis Methodology and Algorithms

S. Bhattacharya, V. Natarajan, A. Chatterjee
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta GA 30332
{soumendu, vishwa, chat} @ece.gatech.edu

S. Nair
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology, Atlanta GA 30332
sankar.nair@chbe.gatech.edu

Abstract

Novel advances in fluidic MEMS sensor/sensor electronics design utilizing silicon nanopores have opened the possibility of accurate and "portable" field diagnosis of DNA molecules. Currently, a DC stimulus is used to "push (translocate)" each molecule through a silicon nanopore and the translocation time of the molecule through the nanopore is used to estimate its length. In this paper, it is shown that polymer transport models, combined with the knowledge of the electrical forces exerted on each molecule by application of a (calibrated) stimulus to the sensor electrodes, allows much more fine-grained diagnosis of DNA molecules than it is currently possible. Hence, by proper stimulus design, the efficiency of the DNA sensing using silicon sensors can be improved significantly without incurring any additional hardware overhead.

1. Introduction

Size-separation and sequencing of chain-like biomolecules (single stranded DNA, RNA and proteins) is a process of vital importance in biotechnology and medicine [1]. Sequencing speed is already a bottleneck in genomics and allied disciplines [2]. In particular, current DNA sequencing methods involve electrophoretic separation of DNA strands of varying sizes, generated by the well-established polymerase chain reaction (PCR) process [3]. The PCR process generates DNA strands of varying lengths from the original sample, such that the length of a generated strand reflects the identity (A, C, G or T) of the base at the fluorescently labeled termination position. The sequencing problem is thus reduced to size-separating (or sizing) DNA strands. This step is carried out via electrophoresis in a gel or capillary bundle, whereby the molecules are separated into bands by virtue of difference in their transport rates through the medium as a function of their size, under an applied electric field. Efforts to substantially raise the throughput rates of these devices are impeded by the 'short read length' problem, i.e. inefficient operation at long sequence lengths due to very slow transport of long strands through the medium [1]-[4]. In addition, the use of fluorescence-based optical techniques to detect the DNA bands increases the size and cost of DNA sequencing devices. However, overcoming the limitations of electrophoresis will result in a large technological impact, in terms of new applications such as 'personalized medicine' (routine, patient-specific genome sequencing to diagnose genetic health risks), and fast genotyping of new pathogens or biological warfare agents to allow a rapid response [5].

In recent years, research at the biology-nanotechnology interface has shown potential for creating revolutionary advances in speed, efficiency, reliability and portability of biomolecule sensors. An underlying advantage of a truly nano-scale biomolecule sensing technique is the ability to detect single molecules at a nanometer length scale and at very short time scales, using only small amounts of sample. Operation at short length and time scales would remove the transport limitations associated with electrophoresis technology. Of several proposed strategies for sizing DNA, the use of nano-scale ion channels is particularly attractive [6]-[10]. The sensing element is a nanometer-scale pore (~ 2-5 nm in diameter and a few nm long) embedded in a substrate. Sensing occurs by measuring the partial blockage of ionic current through a single nanopore as individual analyte molecules are driven through by an electric field (Figure 1), using measurement techniques originally developed to study ion channels in living cell membranes [11]. For chain molecules like DNA, RNA and proteins, the duration of the current blockages correlate with the chain length.

The advantages offered by this technique include very high sensitivity (single molecule levels), extremely rapid and reversible response due to the short detection length and small time scales (c.a. 1 nm and 1 ms respectively), good signal-to-noise ratio even at low analyte concentrations, as single molecules are detected irrespective of concentration, and concurrent multiple-analyte sensing using arrays of nanopores. It is estimated that these devices can increase sizing speeds from 10^4-10^5 bases per day on a single electrophoresis instrument, to levels of 10^7-10^8 bases per day; i.e. 3-4 orders of magnitude higher [12].

Different types of nanopores [6][13][14] have been proposed or demonstrated for use in the above devices, and are currently under further development in several research groups. Concurrently, there is a requirement for developing efficient diagnostic methodologies and algorithms for operating the nanopore sensor and analyzing its output. Preliminary experiments on the nanopores have invariably used a DC voltage [6]-[10] to demonstrate the translocation of biomolecules through the pore. However, more sophisticated sensing protocols must be developed to enable the processing of real samples in an efficient and reliable manner, and to optimize the characteristics of the sensor such as the sensitivity and signal-to-noise ratio. These diagnostic tools should be based on the essential physics of translocation of chain-like biomolecules through a nano-scale channel driven by a voltage waveform, and must take into account the stochastic nature of this single-molecule process.

2. Overview of DNA Diagnosis System

The physical basis of the present sensing technology is the measurement and interpretation of ionic current (pA to nA levels) through individually addressable nanoporous ion channels (of ~ 2-5 nm diameter) in a substrate (e.g. silicon nitride membrane of ~5 nm thickness). Figure 1 illustrates the operation of the sensor on the nano-scale. When a biomolecule is driven through the pore, the ionic current is partially blocked [6]. The duration of the current blockage correlates with the translocation time for the molecule through the pore, which is in turn is directly correlated to its length [6]-[10].

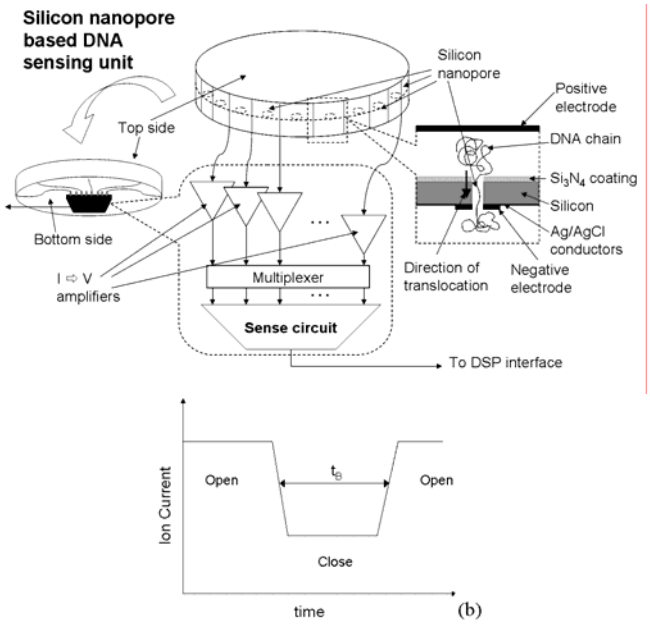

Figure 1. Internal diagram and the operating principle of the sensor: In 'open' mode, the pore permits a high current. During translocation, the current is reduced to a low level due to partial pore blockage by the molecule. The blockage duration (t_B) correlates to the molecule length.

The electrodes that are used to supply the driving stimulus can be positioned close to the nanopore to detect currents. In previous experiments, the nanopores and the electrodes were submerged in a polymeric (e.g., teflon, acrylic) sample cell of fluid volume approximately 0.1-1 mL [10]. Based on current research efforts, it is anticipated that this apparatus will be replaced in the near future with a nano/microfabricated system on a silicon chip, containing an array of individually addressable nanopores enclosed in fabricated micro-chambers with microelectrodes at appropriate locations. This system would be combined with a circuit chip that analyzes the signals from the nanopore array, as well as a micro-fluidic system for handling input and output of analyte samples.

3. Objectives

As DNA molecules are charged in aqueous solution, the applied potential forces them to move towards the nanopore and eventually pass through it. The motion of the molecule under the influence of an applied electrical potential is governed by three main forces: electrical, viscous drag and a rapidly fluctuating thermal force. The latter random force obeys a normal distribution about a zero mean. The random force and the drag force are highly correlated to each other, since the molecular origin of both forces is the same, viz. bombardment of the biomolecule by solvent molecules. The change in the channel resistance due to the passage of a DNA molecule through the nanopore causes a change (*ca.* 50-75 pA) in the observed ionic current. The DNA length can be calculated based on the duration for which the channel is blocked, known as *translocation time (t_B)*.

The translocation time for molecules of the same length (under an applied DC potential) shows a stochastic variation and follows a Gaussian distribution due to the random force effects operating during each translocation event. This reduces the resolution of the sensor for distinguishing between two different chain lengths that are close in value. Even for two chains that can be clearly distinguished, the process may become time consuming and inefficient due to the measurement time required for improving the statistical quality of the data.

We propose to use an AC stimulus on top of the applied DC potential. By doing so, the *difference between the translocation times of DNA strands of comparable lengths can be increased significantly compared to using DC stimulus alone*. As presented in the paper, this results in a more efficient methodology for sizing a mix of DNA molecules in an unknown solution.

4. Physics of DNA Transport in a Nanopore

The DNA molecule is considered as a rigid cylindrical rod in this paper. This is a reasonable assumption, as the sensor operates at an appropriate pH and salt concentration of the DNA solution [18], in which conditions the DNA chains remain uncoiled and in an extended form by

application of the electric field. For simplicity, it is also assumed that the motions along x and y directions are restricted and hence the only degree of freedom of movement for the molecule is along z-axis, i.e. along the nanopore axis. The DNA strand has N bases (monomers), each of length $l \sim 0.7$ nm [18], giving a total length of $L = Nl$. The number of bases can be up to 10,000. The DNA strand has a diameter $d \sim 1.5$ nm. It is also assumed that the length of the DNA chain is much greater than the length of the pore (~ 5 nm), so that the pore can be considered as a hole of essentially zero thickness separating two chambers.

4.1. Forces acting on the molecule

In this section, we discuss the three types of forces acting on the DNA molecule.

Electric force (F_E): Within the solution, each base holds approximately 4 electrons of charge. Thus, the total charge on the DNA strand is $Q = -4qN$, where $q = 1.6\times10^{-19}$ Coulomb. We assume a potential difference of $V(\omega,t)$ applied to the electrodes resulting in an electrical force given by:

$$\vec{F}_E = Q.\frac{V(\omega,t)}{s} \tag{1}$$

where s is the separation between the electrodes. Thus, the magnitude of electric force on a molecule is directly proportional to the molecule length.

Viscous Drag force (F_D): The magnitude of drag force acting on a molecule is given by [19]:

$$\vec{F}_D = -M\gamma\vec{v} \tag{2}$$

where M is the mass of the DNA strand ($M = Nm$, with m being the mass of a single monomer of DNA, *ca.* 330 a.m.u.), $\vec{v}$ is the velocity of the DNA strand, and γ is the drag coefficient of the solution given by [19]-[20]:

$$\gamma = \frac{3\pi\eta L}{M[\ln(L/d)+1]} \tag{3}$$

where, η is the viscosity of water (0.0089 Poise). The drag force is proportional to $N/\ln(N)$, and increases with increasing chain length.

Random Force (F_R): The random force F_R [20] acting on a molecule follows a gaussian distribution with zero mean and a variance given by $2M\gamma kT/\Delta t$. Here, $k = 1.38\times10^{-23}$ J/K is the Boltzmann constant and Δt is the time step used for numerical integration (typically ~1 ns). The random force is strongly related to the viscous force, as the drag coefficient γ appears in both force definitions.

4.2. Equation of Motion

The translocation of the DNA molecule through the nanopore in the presence of an electric field can be modeled using Newton's first law of motion:

$$M\frac{d\vec{v}}{dt} = \vec{F}_E + \vec{F}_D + \vec{F}_R \tag{4}$$

The initial velocity for the molecule was obtained by equating its one-dimensional kinetic energy to the one-dimensional thermal energy ½kT:

$$\frac{1}{2}kT = \frac{1}{2}Mv_0^2 \;\rightarrow\; v_0 = \sqrt{kT/M} \tag{5}$$

Using (1)-(3) and substituting in (4), we get a differential equation in $v(t)$, with the initial condition given by (5). Algebraic manipulation of (4) gives us the translocation time under the influence of the forces discussed. Due to the random force, the translocation times for molecules having the same length will be different for repeated simulation instances.

4.3. Effect of DC Stimulus on Translocation

The model was first simulated with a 100 mV DC driving voltage. The differential equation (4) was solved to obtain the translocation time for a molecules with different values of N and the variation of translocation time with change in base length (N) is shown in Figure 2.

The sensitivity of the translocation time of the molecule to the lengths of the translocating molecules is given by:

$$S = \frac{\Delta t_B / t_B}{\Delta N / N} \rightarrow \frac{N}{t_B}.\frac{dt_B}{dN} \rightarrow \frac{d(\ln t_B)}{d(\ln N)} \tag{6}$$

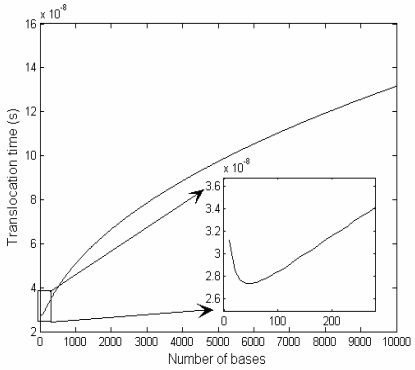

Figure 2. Variation in mean translocation times of the molecules with base length (Inset: short molecules).

The DC sensitivity curve is shown in Figure 3. As evident from the figure, the sensitivity of the sensor is practically constant for almost the entire range of molecule sizes investigated, and also becomes lower for short molecules.

Next, we used a DC stimulus to diagnose the composition of an unknown solution. The number of histograms and area under each histogram are calculated. This gives an estimate of the different DNA bases constituting the solution. However, due to the effect of random forces, if two molecules are almost of the same length and the proportion of one molecule is much larger compared to the other, it might completely mask the histogram of the other molecule. These factors result in a reduced accuracy of prediction, and show the need for alternate stimuli that can increase the sensitivity for molecules with comparable lengths.

4.4. Effect of AC Stimulus on Translocation

If a sinusoidal signal with carefully adjusted frequency, amplitude, and phase is added to the DC signal, the sensitivity for certain pairs of bases can be enhanced significantly. This is evident from Figure 4, where the translocation time vs. base length is plotted for an applied AC signal with the DC. For some regions of N, the slope of the translocation time curve (and hence the sensitivity) increases sharply when an AC stimulus is applied. However, for other regions, the sensitivity is lower than DC when an AC stimulus is applied.

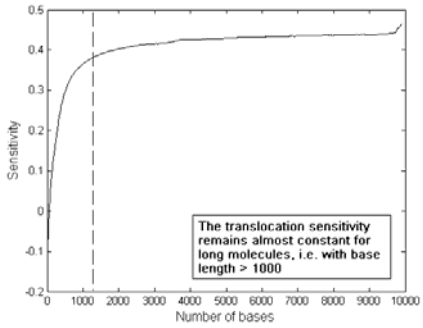

Figure 3. The sensitivity variation for a DC stimulus for different base lengths.

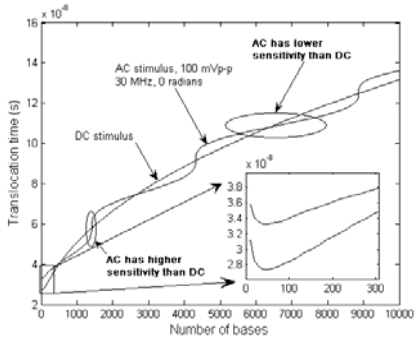

Figure 4. Mean translocation time versus N for AC stimulus.

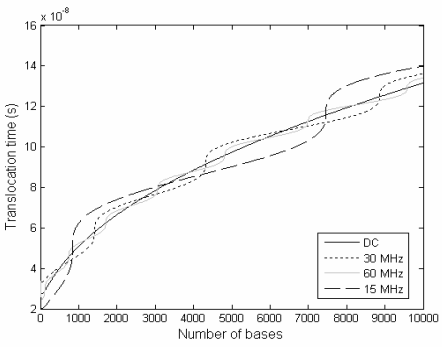

Figure 5. Mean translocation time vs. *N* for different AC amplitudes with zero phase.

We also found that the sensitivity obtained varies with the frequency, amplitude, phase and the DC offset of the input AC and DC signals (Figure 5). The data shows that a particular combination of frequency, phase, amplitude, and DC offset values provides better performance than DC for one set of base lengths, and at the same time is worse than DC for another set of base lengths. The histograms of translocation times for two different values of *N* are shown in Figure 6. The introduction of an AC stimulus separates the two histograms better than the use of a DC stimulus, providing better resolution and diagnosis.

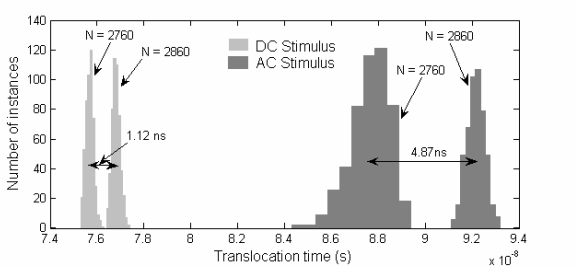

Figure 6. Histogram of stochastic translocation times for repeated translocation events of DNA molecules of two different base lengths (2760 and 2860) under DC (left) and AC stimulus (right).

5. Optimal Diagnosis Algorithm

The diagnosis algorithm optimizes the frequency, amplitude and phase of the AC to increase the sensitivity for a pair of molecules compared to DC. The algorithm consists of two main parts. The first part uses a DC stimulus to identify different *base length regions* that may be present in the sample. In the next part, the AC stimulus is optimized for these different regions. Each region is investigated by optimizing the values of input parameters, i.e. frequency, amplitude and phase of the input stimuli, to increase the sensitivity (i.e., the distance between the mean translocation times of the different DNA strands of comparable lengths). Figure 7 shows the pseudo-code for the proposed algorithm:

The function `decision` uses a gradient search (steepest descent) algorithm to determine the optimal set of parameters, i.e. frequency, amplitude and phase values for the AC stimulus. It also determines whether to increase or decrease the optimization parameters and the step size for the same. In addition, the algorithm maintains a history of the optimization steps. At some point, if the optimization parameters move away from a target value, it reverts to the previous point where best sensitivity was obtained, and changes the direction of search.

```
function [f,A,θ] = decision(solution)

Input ← solution with unknown composition of
DNA bases
Perform a DC simulation and generate histograms
Identify different regions
Compute mean base lengths for each region
(perform stimulus generation for each region
separately)

for Region R,
   Set initial frequency, amplitude and phase
   (f, A and θ respectively)
```

```
[N1,N2] = Calculate Base Pair(f,A,θ)

if more than one base found in region,
   S ← DC Sensitivity(N1,N2)

   while frequency not optimized
      i = current iteration
      S1  ← AC Sensitivity(f+Δf,A,θ)
      S2  ← AC Sensitivity(f-Δf,A,θ)
      Di  ← Calculate Direction(S1,S2)
      Sfi ← max(S1,S2);
      fstep ← Decision(Sfi,Sfi-1,Di,Di-1,f)
      Change f by fstep
      if Sfi>K.S or i > MaxIterations
         Frequency optimized
      end if
      i = i+1
   end while
   Repeat for A and θ ...
   if not all optimized
      continue
   else
      Output → optimized f, A, θ
   end if

  end if
end for
```

Figure 7. Psuedo-code for the proposed algorithm

The algorithm targets the attainment of a sensitivity (using AC stimulus) equal to K times that for a DC stimulus, where K is a user-specified performance enhancement factor, set to a value of 3 in this paper. The algorithm stops when each of the optimization parameters achieve this condition.

6. Results and Discussion

The optimization algorithm described in Section 5 was used to simulate and diagnose a solution of unknown bases. We assumed a solution with base lengths 1390, 1760, 2680 and 2760. The sensor output from this solution was simulated with a DC stimulus of 100 mV and the histograms shown in Figure 8 were obtained. While the two smaller molecules were easily distinguished, the molecules having lengths 2680 and 2760 bases were not distinguishable from each other. Therefore, three regions could be identified for separate optimization analysis to find the optimal AC stimulus for each region. In addition, the mean of each region and the corresponding base lengths were determined.

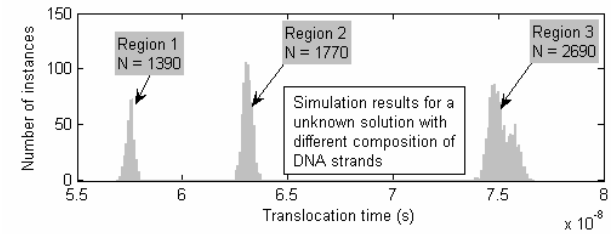

Figure 8. Histogram of translocation times for initial analysis (4 bases were present, but two of the bases were not distinguishable from each other).

Next, each region was simulated with an AC stimulus. The initial choice of frequency for the AC stimulus was calculated from the inverse of the mean translocation time of the region. The initial amplitude of the AC stimulus was chosen to be 50 mV. The Regions 1 and 2 (Figure 8) contained only one type of strand each, and as expected, did not show any more molecule lengths when simulated with an AC stimulus. Region 3 showed significant splitting (Figure 9) when the above-mentioned AC stimulus was applied.

The proposed optimization algorithm was then applied to refine the sensitivity in Region 3. We observed that the phase and the frequency opposed each other during optimization, so phase was not optimized further and was set at zero for all applied AC stimuli. The DC stimulus showed a spread of the translocation times by 2.9 ns (Figure 8), while the initial AC stimulus (f = 13 MHz, A = 50 mV) doubled the spread to 6.1 ns. Figure 10 shows the improvement in sensitivity as the optimization proceeds till the desired enhancement factor is obtained.

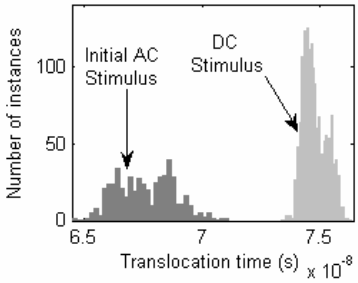

Figure 9. The translocation time histograms in Region 3 for the DC and the initial AC stimulus applied.

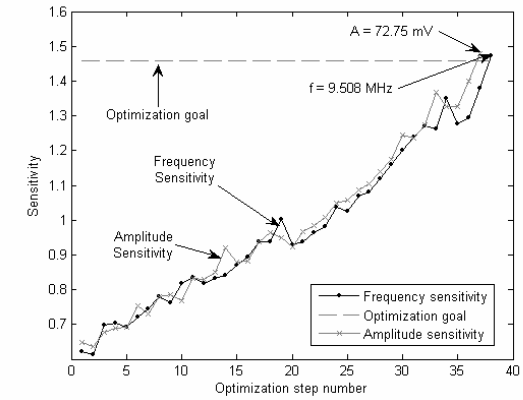

Figure 10. Change in sensitivity values with progress in optimization (downslopes indicate algorithm backtracking).

The final stimulus (f = 9.5 MHz, A = 72.75 mV) obtained from the optimization algorithm shows a spread of 27.6 ns, about ten times the spread for a DC stimulus. Using this stimulus, the two molecules can be easily distinguished in the histogram shown (Figure 11). Thus, the AC stimulus not only helps diagnosing the hidden

bases within each region, it also increases the resolution of the diagnosis process significantly. This first study of performance diagnostics in nanoscale stochastic single-molecule sensors shows clearly the potential for model- and algorithm-based diagnostic approaches towards optimizing the performance of this emerging technology. The short time scales make it feasible to implement the optimization process *'on the fly'* during operation, to analyze DNA mixtures of arbitrary complexity. Experimental translocation times of DNA strands are in the micro-to millisecond range, due to the additional degrees of freedom in three dimensions and deviation from rigid rod behavior. A replacement of the simple model with a more detailed simulation [15]-[17], with no change in the optimization methodology, will account for this effect. Our approach is therefore very general, and shows clearly the potential for exploiting the DNA transport physics *via* the proposed diagnostic method.

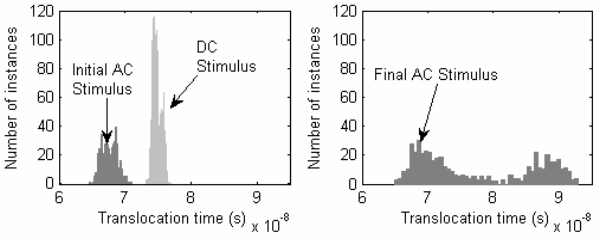

Figure 11. Performance for DC and the initial AC stimulus (left), and the optimized AC stimulus (right).

7. Conclusions

A diagnostic methodology for high-speed sizing of mixtures of chain biomolecules like DNA by fabricated nanopores has been developed, based on optimization of the externally applied electrical driving potential. An initial DC driving force produces a coarse size distribution assay of the sample, which can then be refined by optimized AC stimuli specific to each region of the size distribution. In particular, an AC stimulus that is *'tuned'* to the time scale of the transport process considerably increases the performance. Although a simplified transport model is used here, the approach is completely general and can be applied to atomistically detailed descriptions of the nanopore transport. The new diagnostic methodology, together with ongoing device fabrication research, strengthens the potential of nanoscale sensing to provide next-generation biomolecule analysis methods.

8. Acknowledgements

This work was funded in part by NSF ITR award CCR-0325555 and GSRC/MARCO 2003-DT-660.

9. References

[1] Marziali, A. and M. Akeson. New DNA sequencing methods. *Annual Review of Biomedical Engineering.* 3: 195-223, 2001.

[2] Franca, L.T.C., E. Carrilho, and T.B.L. Kist. A review of DNA sequencing techniques. *Quarterly Reviews of Biophysics.* 35 (2): 69-200, 2002.

[3] Sanger, F., et al. Nucleotide-Sequence of Bacteriophage PhiChi174 DNA. *Nature*. 265 (5596): 687-695, 1977.

[4] Vercoutere, W. and M. Akeson. Biosensors for DNA sequence detection. *Current Opinion in Chemical Biology.* 6(6): 816-822, 2002.

[5] Kasianowicz, J.J., et al. Applications for DNA transport in a single nanopore. *Biophysical Journal*. 80 (1): 1418-1423, 2001.

[6] Bayley, H. and P.S. Cremer. Stochastic sensors inspired by biology. *Nature.* 413 (6852): 226-230, 2001.

[7] Eisenberg, B. Ion channels as devices. *Journal of Computational Electronics*. 2: 245-249, 2003.

[8] Miller, S. A. and C. R. Martin. Redox modulation of electro-osmotic flow in a carbon nanotube membrane. *Journal of the American Chemical Society.* 126: 6226-6227, 2004.

[9] Khitrov, G. Single DNA molecules detected by nanopore technology. *MRS Bulletin*. 28 (2): 96, 2003.

[10] Meller, A., et al. Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences of the United States of America*. 97 (3): 1079-1084, 2000.

[11] Sakmann, B. and E. Neher, *Single-channel recording*. 2nd Edition. Plenum Press, New York, 1995.

[12] Chen, C.M. and E.H. Peng. Nanopore sequencing of polynucleotides assisted by a rotating electric field. *Applied Physics Letters.* 82 (8):1308-1310, 2003.

[13] Mukherjee, S., V. M. Bartlow and S. Nair. Phenomenology of the growth of single-walled aluminosilicate and aluminogermanate nanotubes of precise dimensions. *Chemistry of Materials* (in press).

[14] Li, J., D. Stein, C. McMullan, D. Branton, M.J. Aziz and J.A. Golovchenko. Ion-beam sculpting at nanometre length scales. *Nature.* 412 (6843): 166-169, 2001.

[15] Muthukumar, M. Translocation of a confined polymer through a hole. *Physical Review Letters.* 86 (14): 3188-3191, 2001.

[16] Tian, P. and G.D. Smith. Translocation of a polymer chain across a nanopore: A Brownian dynamics simulation study *Journal of Chemical Physics.* 119 (21): 11475-11483, 2003.

[17] Slonkina, E. and A.B. Kolomeisky. Polymer translocation through a long nanopore. *Journal of Chemical Physics.* 118 (15): 7112-7118, 2003.

[18] Lodish, H., et al. *Molecular Cell Biology*. 4th Edition. W.H. Freeman, New York, 2000.

[19] Kuyucak, S., O. S. Andersen and S.-H. Chung. Models ofpermeation in ion channels. *Reports of Progress in Physics.* 64: 1427-1472, 2001.

[20] Pecora, R. Spectral distribution of light scattered by monodisperse rigid rods. *Journal of Chemical Physics.* 48 (9): 4126-4128, 1968.