A Tandem Cell for Nanopore-based DNA Sequencing with an Exonuclease Enzyme

G. Sampath
sampath_2068@yahoo.com

Abstract. A tandem electrolytic cell is proposed for DNA sequencing in which an exonuclease enzyme cleaves bases (mononucleotides) from the DNA strand for identification inside a nanopore. It has two nanopores and three compartments with the structure [*cis1*, upstream nanopore (UNP), *trans1=cis2*, downstream nanopore (DNP), *trans2*]. The exonuclease is attached to the exit side of UNP in *trans1*/*cis2*. Separate potential differences are applied to the five sections to enable drift-driven translocation of DNA, cleaved bases, and detected bases. A cleaved base cannot regress into *cis1* because of the remaining DNA strand in UNP. A profiled electric field across DNP with positive and negative components is suggested for slowing down base translocation through DNP. A Fokker-Planck equation is used to model the system and a piecewise solution presented. Results from the model suggest that with probability approaching 1 bases enter DNP in their natural order, are detected without any loss, and do not regress into DNP after progressing into *trans2*. If this holds in practice then the only determinant of sequencing efficiency would be the level of discrimination among the four (or five, if methylation is included) base types in DNP. A hybrid biological-synthetic implementation is considered. Thus a biological pore (AHL or MspA) may be used for UNP because of the need to covalently attach or genetically fuse the exonuclease to the exit side of UNP. For DNP a multi-layered solid-state pore with interposed graphene sheets is suggested. Other implementation issues and potential extensions are discussed.

Keywords: nanopore, DNA sequencing, exonuclease, Fokker-Planck equation, tandem cell

I INTRODUCTION

A recent publication describes a nanopore-based DNA sequencing method in which bases (mononucleotides) are cleaved from a strand of DNA by an exonuclease enzyme situated in the *cis* compartment of an electrolytic cell, fall into a nanopore bridging the cell's *cis* and *trans* compartments, and are detected while passing through the pore into *trans* under the influence of an electric field [1]. The method, which has been modeled mathematically and analyzed in detail [2], appears to have several problems, including the potential for losing bases to diffusion, out-of-order sequencing, and high speeds of translocation through the pore. Here it is shown that an electrolytic cell structure with two nanopores connected in tandem may be able to resolve most of these problems.

The following is a summary of this report. Section II gives a brief account of DNA sequencing using a nanopore with or without exonuclease. Section III describes the tandem cell and presents a rationale for its use in exonuclease-based sequencing. In Section IV a Fokker-Planck equation is used to model the cell mathematically and a piecewise-analysis presented with numerical results. A case is made for the tandem cell being able to sequence bases efficiently at a rate that is compatible with reasonable detector bandwidths in the presence of noise. To achieve the latter an electrical method of slowing down a cleaved base when it passes through the second nanopore is suggested. Section V discusses implementation issues. Section VI concludes with notes on

potential extensions to the tandem cell and its possible modification for use in polypeptide sequencing.

To keep the report in focus references to the literature are kept to a minimum. For a recent review of nanopore-based DNA sequencing with an extensive list of current references see [3].

II NANOPORE-BASED SEQUENCING OF DNA USING EXONUCLEASE

Figure 1a shows the basic structure of a conventional nanopore-based electrolytic cell containing an electrolyte that is a salt (typically Kcl) solution. Negative (positive) ions flow through the pore from *cis* (*trans*) to *trans* (*cis*) under the influence of a positive (negative) potential difference. When a strand of DNA is introduced into *cis*, the negatively charged biomolecule is drawn into the pore (which is embedded in a membrane separating the two compartments) and modulates the ionic current through the pore by an amount that depends on the type of base (A, T, C, G) passing through the pore. If the changes are sufficiently discriminating of the base type, the current level can be used to identify successive bases (mononucleotides) and thereby sequence the strand [4]. This approach is commonly referred to as 'strand sequencing'.

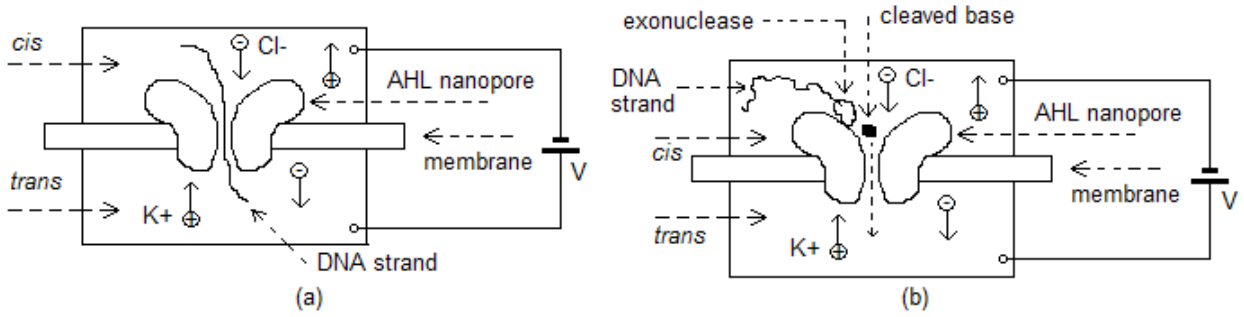

Figure 1. Schematic of nanopore DNA sequencing
a. *strand sequencing* **b.** *exonuclease sequencing*

In a second approach known as 'exonuclease sequencing' [1], exonuclease enzyme in the *cis* compartment of an electrolytic cell next to the vestibule of an *alpha-Hemolysin* (*AHL*) nanopore cleaves bases in a single strand of DNA and drops them into the nanopore where they cause a current blockade that is unique for each of the base types A, T, C, and G (Figure 1b). However some of the cleaved bases may diffuse back into the *cis* compartment either to be 'lost' to diffusion over time and/or be captured later and get called out of order at detection time. This problem is lessened considerably if the enzyme is covalently bonded close to the nanopore entrance. Also two or more cleaved bases could arrive in quick succession and occupy the pore at the same time. Another problem is the speed with which the base translocates through the pore. Inadequate circuit bandwidth may then make it impossible to detect every base passing through the pore. A mathematical model and detailed analysis of the method can be found in [2], which also describes numerical simulations conducted to study the efficiencies of the method and goes on to suggest several improvements.

In the following section a tandem electrolytic cell structure is proposed that has the potential to resolve some of the above-mentioned problems in exonuclease-based sequencing. Results from a mathematical model (see Section IV) support most of the expectations. With probability approaching 1 the proposed structure appears capable of identifying bases in the correct order without losing them to diffusion. A method to decrease the speed of base translocation through the pore is also suggested.

III A TANDEM CELL FOR DNA SEQUENCING: STRUCTURE AND RATIONALE

A tandem cell (Figure 2a) consists of two nanopores connected in tandem and exonuclease enzyme attached to the *trans* side of the first pore. The enzyme cleaves bases from a single-stranded DNA (ssDNA) molecule that has translocated through the first (upstream) pore or UNP, and the cleaved bases are detected during their passage through the second (downstream) pore or DNP. Formally the structure can be written as [*cis*1, UNP, *trans*1=*cis*2, DNP, *trans*2].

Potential differences are introduced in *cis*1, *trans*1/*cis*2, and *trans*2 that are distinct from the voltages applied across UNP and DNP. Their roles are described below. Operationally the tandem cell can be described as '*thread-cleave-deliver-detect-recover*'. Thus, *thread* ssDNA through the first pore, *cleave* the next (leading) base from ssDNA on the output side of the first pore, *deliver* the cleaved base to the second pore, *detect* the cleaved base as it passes through the second pore, and *recover* the detected base on the output side of the second pore. The tandem cell is modeled with five stages or sections in which the five potential differences play a major role.

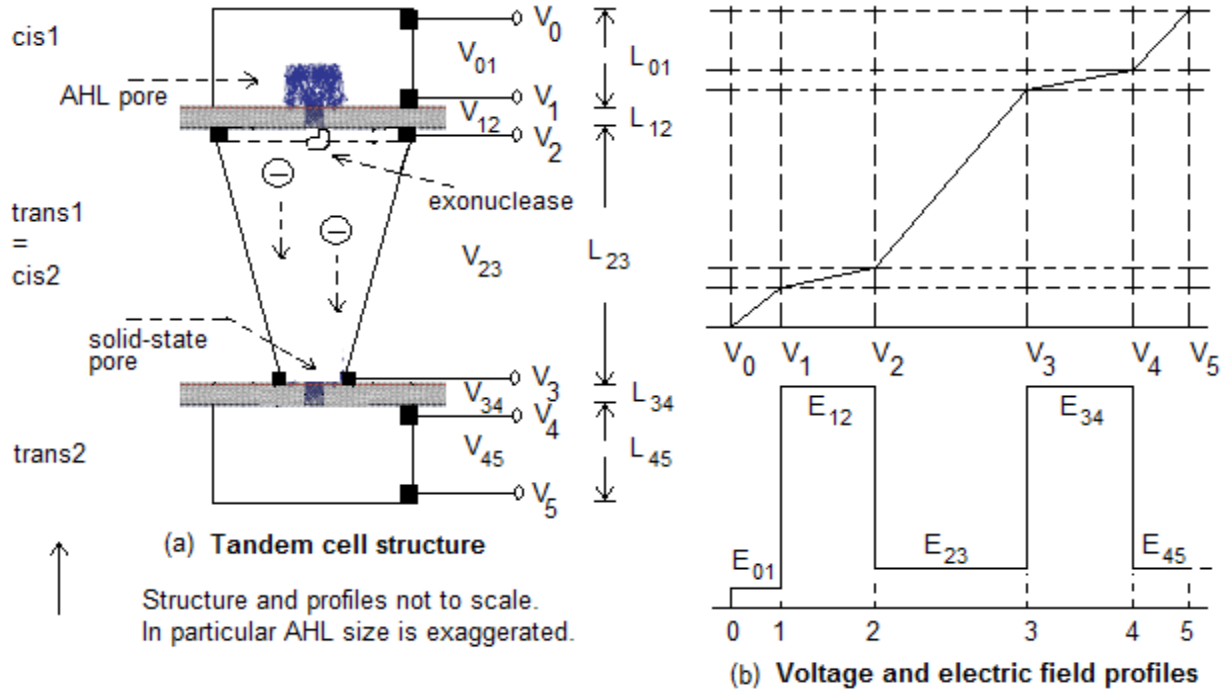

Figure 2 Schematic of tandem cell

Stage 1 (*Thread*). Single strands of ssDNA are drawn to the entrance of UNP under the influence of the electric field E_{01} due to V_0V_1. Since studies have shown that ssDNA is more easily threaded through the pore of a conventional cell with higher membrane voltages (see, for example, Figure 7 in [3]), the explicit field in *cis1* further increases the likelihood of capture. Here it is assumed that the ssDNA strand always threads through UNP and encounters the exonuclease on the exit side of UNP.
Stage 2 (*Cleave*). The exonuclease engine seizes the ssDNA as it emerges on the *trans*1/*cis*2 side of UNP and the resulting ratcheting action cleaves bases and (in concert with the potential difference V_1V_2) moves the strand at a rate of one base every 1 to 10 msec [1, 2]. The cleaved base drops into *trans1*/*cis2* and passes into Stage 3.
Stage 3 (*Deliver*). V_2V_3 considerably increases the probability of a cleaved base overcoming diffusion by drifting to and being drawn into DNP.
Stage 4 (*Detect*). V_3V_4 is the usual voltage applied to a nanopore to facilitate passage of an analyte through the pore for detection [4]. Once the cleaved base passes into DNP its successful detection depends on the blockading characteristics of the pore. When the base reaches the bottom of the pore it is considered to have been detected and then exit into *trans*2 (Stage 5, see next). To slow down translocation of the base through DNP, a profiled voltage with positive and negative components is applied over DNP.
Stage 5 (*Recover*). A sufficiently high V_3V_4 over DNP in combination with the potential V_4V_5 across *trans2* enables the detected base to exit into *trans*2 where it is recovered [5].

Two implementations may be considered. The first has AHL [6] or MspA [7] for both UNP and DNP, while the second is a hybrid structure with AHL/MspA for UNP, a solid-state pore for DNP, and a silicon microchannel for *trans*1/*cis*2. The downstream AHL pore in the first is assumed to have a cyclodextrin adapter [6] covalently bonded to it. The second structure appears to be better suited for a tandem cell, in part because integration of *trans*1/*cis*2 with DNP is easier and because transverse electrodes can be accommodated in the thicker pore. In an all-AHL/MspA structure an AHL/MspA pore replaces the solid-state pore.

III.1 Rationale

For sequencing to be accurate it is necessary (at least in a statistical sense) for the following conditions to be satisfied:
a) cleaved bases arrive at and are captured by DNP in their natural order;
b) DNP correctly identifies each and every base as it passes through;
c) the detected base exits from DNP without regressing.

The first condition can be satisfied if the drift potential over *trans*1/*cis*2 is large enough for the base to move at a rate that is greater than the exonuclease turnover rate. The second situation (assuming perfect discrimination among the four base types) will occur as long as two successive cleaved bases do not occupy DNP at the same time (this partly depends on the first condition being satisfied) and the time between successive bases passing through DNP is greater than 1/bandwidth of the detector circuit (plus any other processing time such as, for example, noise filtering). The third condition is usually satisfied if the voltage across DNP is sufficiently large, but it tends to conflict with the second condition.

It appears possible, at least in principle, that by using multiple explicit potential differences (see Figure 2 above) all three conditions can be satisfied. The following four-part rationale is presented in support of such a proposal. (More formal arguments based on mathematical models are given in Section IV.)

1) In most reported work on nanopore sequencing of DNA, potential differences are applied across a membrane with one electrode in *trans* and one in *cis* without specifying where the electrodes are located or how the potential is distributed in the compartment. At most 1% of the voltage between the two electrodes drops across the *cis* compartment of a conventional cell [2]. Using the optimum desired value [1] of 180 mV across a nanopore that is 10 nm thick, the maximum possible electric field with a *cis* height of 1 mm and 1.8 mV across *cis* is 1.8 V/m. The time taken for a mononucleotide to diffuse-drift through *cis* is then about 20000 sec. The nucleotide is virtually 'lost' to diffusion and/or successive bases entering *cis* can get hopelessly out of order. On the other hand, the time to diffuse-drift through the pore is about 20 nsec, which makes it nearly impossible for the electronics to detect bases without a substantial number of misses.

In contrast, a tandem cell has a potential difference applied between the top and bottom of a compartment using a pair of electrodes rather than just one electrode per compartment. This leads to 5 potential differences (Figure 2) explicitly defined over the cell, all of which can in principle be set independently. A base just after it has been cleaved by the exonuclease is subject to diffusion biased by the drift field in *trans*1/*cis*2. With a sufficiently large bias a cleaved base can arrive at the entrance of DNP and be captured. For a compartment height of 1 mm (Figure 2) and V_2V_3 = 1 V an electric field of 10^3 V/m would result, a substantial increase over the case with a single electrode in *trans*1/*cis*2. The time for a cleaved base to diffuse-drift through *trans*1/*cis*2 to the entrance of DNP is now only ~40 secs. While considerably lower in comparison to 20000 secs even this looks inordinately high, but results from the mathematical model in Section IV suggest that it may not be a serious problem (in part because of the pipeline structure). Assuming for now that cleaved bases arrive at DNP in their natural order a cleaved base that enters DNP can be detected well ahead of its successor because, given an exonuclease turnover interval of 1 to 10 msec [2], if bases arrive at DNP separated by more than the time taken by a base to translocate through DNP then two successive bases cannot be in the pore at the same time. The probability of bases entering DNP in their natural order and no more than one occupying DNP at the same time would then be close to 1. (See Section IV for statistical calculations relating to the above arguments.)

2) A similar rationale holds for *cis1* and the transport of ssDNA to the entrance of UNP for transport through the pore to the exonuclease for cleaving (and also for *trans2* and exit-recovery of a cleaved base that has passed through DNP). Thus with large enough V_0V_1 the DNA molecule can overcome the entropy barrier [5] at the entrance of UNP and with probability 1 be threaded through the pore to the exonuclease at the top of *trans*1/*cis*2.

3) In a tandem cell a cleaved base cannot diffuse back into *cis*1 because the remaining ssDNA is still in the pore and will block its passage. Thus there is zero loss of bases due to regression into *cis1*, and cleaved bases drop into *trans1*/*cis2* in their natural order.

4) A large enough V_4V_5 (in conjunction with a sufficiently large V_3V_4) can present a drift field that biases the detected base away from DNP into and through *trans2* for recovery.

IV ANALYTICAL MODEL

The proposed scheme can be modeled analytically using methods similar to existing studies of strand [5, 8] and exonuclease [2] sequencing. In [5, 8] the capture of a DNA strand in the entrance of a nanopore (followed by its translocation through the pore and subsequent detection) is modeled, while in [2] a detailed model and extended discussion of the efficiencies inherent in the exonuclease method of DNA sequencing [1] are presented along with simulation results. In the present work the model for a tandem cell is considered in five stages with the focus on Stages 3, 4, and 5 (*deliver-detect-recover*). Stages 1 and 2 (*thread-cleave*) are not considered in detail because it is assumed that the DNA always translocates through *cis1* and is captured by UNP (such capture being ensured by a high enough V_1V_2 [3], which is further reinforced by the presence of an explicit V_0V_1), threads successfully through UNP and arrives at the exonuclease enzyme where the leading base is cleaved at a rate determined by a set of controlled parameters (temperature, salt concentration, etc.). Existing models of DNA capture [5, 8] may be adapted for them by incorporating the effect of the imposed electric field due to V_0V_1.

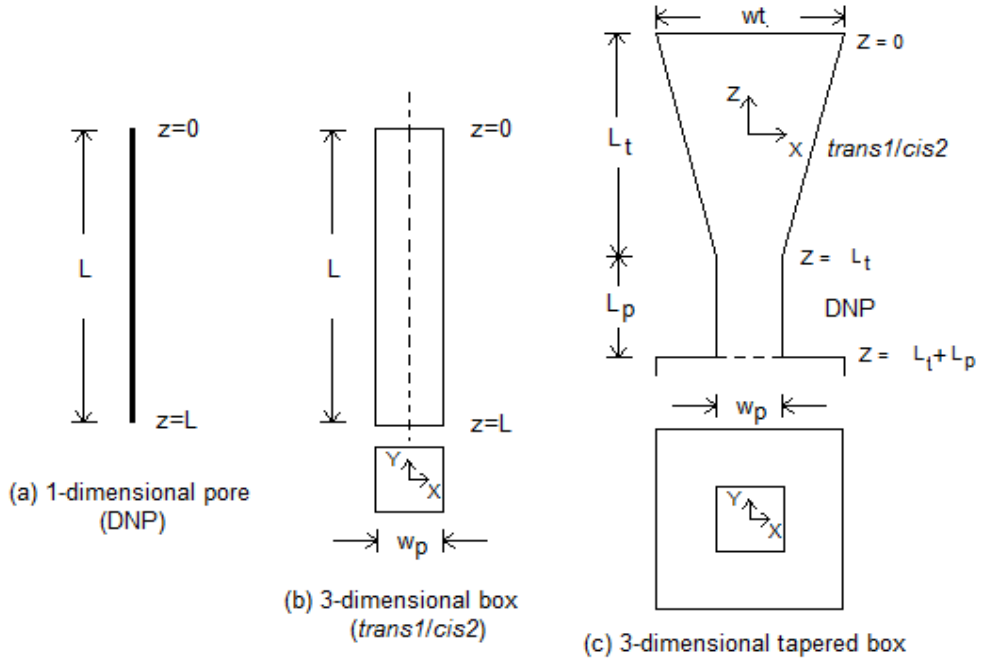

Figure 3 Coordinate systems for models

The behavior of a cleaved base is studied through the trajectory of a particle whose probability density or propagator function G (x,y,z,t) in Cartesian (or G(r,θ,z,t) in cylindrical) coordinates is given by a linear Fokker-Planck equation in three dimensions. The F-P equation is used for piecewise analysis of each stage based on modifications appropriate to each section. Each stage is modeled independently in its own coordinate system and the transition occurring at the interface between two stages studied separately. The coordinate systems used are shown in Figure 3. Standard methods from partial differential equations and Laplace transforms are used, only the essentials of the mathematics involved are given here. For convenience of discussion the stages are considered out of order.

IV.1 Model of Stage 4 (*Detection*)

Although the model of Stage 3 can be adapted for Stage 4, a one-dimensional approximation is presented first to simplify the discussion that follows (see Figure 3a). The trajectory of the cleaved based in DNP is described by the propagator function G(z, t) which satisfies

$$\partial G/\partial t + v_z\, \partial G/\partial z = D\, \partial^2 G/\partial z^2 \qquad z \in [0, L=L_{34}] \tag{1}$$

with initial and boundary value conditions

I.V. The particle is released at z = 0 at time t = 0:

$$G(0, t=0) = \delta(z) \tag{2}$$

B.C.1 The particle is captured at z = L:

$$G(L, t) = 0 \tag{3}$$

B.C.2 The particle is reflected at z = L:

$$D\, \partial G(z, t)/\partial z \,|_{z=0} = v_z\, G(z, t) \tag{4}$$

Here v_z, the drift velocity through DNP, is given by $v_z = \mu V_{34}/L$, and μ is the nucleotide mobility (assumed to be the same for all four types). Following standard procedures φ(t), the pdf of the first passage time (translocation time) for a particle to diffuse-drift from z=0 to z=L and be absorbed at z=L, can be obtained as

$$\varphi(t) = (2/(\sqrt{(\pi 4Dt^3)}\, [\, \sum_{k=0}^{\infty} ((2k+1)L + v_z t) \exp(-((2k+1)L + v_z t)^2/(4Dt)) + \sum_{k=0}^{\infty} ((2k+1)L - v_z t) \exp(-((2k+1)L + v_z t)^2/(4Dt)) \tag{5}$$

φ(t) can be computed numerically but the series oscillates and converges very slowly. Instead an alternative closed-form approach based on the earlier referenced model of exonuclease-based sequencing [2] and Laplace transforms is used after modifying it to accord with Equations 1-4.

In [2] a cell consists of single *cis* and *trans* compartments with the structure [*cis*-nanopore-*trans*]. The exonuclease is on the *cis* side and cleaves the leading base of ss-DNA as it approaches the pore entrance on the *cis* side and drops it into the pore, where the base modulates the pore current as it passes through to *trans* under the influence of a potential difference across the pore. The *cis* compartment is effectively a bulk medium coupled to the pore so that there exists the possibility of the cleaved base diffusing away without entering the pore. This is modeled by the boundary condition

$$D\, \partial G(z, t)/\partial z \,|_{z=0} = v_z\, G(z, t) - \kappa\, G(z, t) \tag{6}$$

where κ is a rate constant representing the rate of loss of a cleaved base to diffusion into the bulk (*cis*). Equation 6 is a slightly modified version of its equivalent in [2], where the z coordinate is the reverse of that in the present work, the cleaved base is released at z=L rather than at z=0 as in Equation 2, and detection is at z=0 instead of at z=L as is done here.

Now consider [*cis2* DNP] in the tandem cell. If κ is set to 0, Equation 6 reduces to Equation 4, after making allowance for the different coordinate system. Setting κ to 0 essentially decouples the bulk *trans1*/*cis2* compartment from DNP and effectively converts the entrance of DNP to a reflecting boundary. If it is assumed that a cleaved base always arrives at the entrance of DNP and DNP is considered in isolation then the cleaved base can be considered to originate at the entrance of DNP. If the latter is considered a reflecting boundary then the base cannot regress into *trans1*/*cis2* and Equations 1 through 4 apply.

The quantity of interest here is φ(t), the pdf of the first passage time T (or equivalently the time for a cleaved base to translocate through the pore and be detected at the end of its translocation), which is independent of the coordinate system. Its Laplace transform (with a slightly different notation than that in [2]) is

$$\varphi^*(s) = \exp(\alpha/2) \,/\, [\cosh(y) + (\beta + \alpha/2) \sinh(y)/y] \qquad (7)$$

where

$$\alpha = v_z L/D; \qquad y^2 = \sqrt{(\alpha^2/4 + 2\tau s)}; \qquad \tau = L^2/2D; \qquad \beta = \kappa LD \qquad (8)$$

(In [2] the parameter α is given as = $|v_z|$ L/D rather than as $\alpha = v_z L/D$, where the modulus around v_z ensures that the drift velocity is always directed in the -z direction (as defined in the coordinate system used in [2]). In the present analysis this restriction is removed: the drift velocity can be positive or negative.)

With κ set to 0 and using Equation 4 (rather than Equation 6), Equation 7 transforms to

$$\varphi^*(s) = \exp(\alpha/2) \,/\, [\cosh(y) + (\alpha/2) \sinh(y)/y] \qquad (9)$$

The mean E(T) can be obtained by differentiating $-\varphi^*(s)$ w.r.t. s and setting s to 0. After some algebra one gets

$$E(T) = -d\varphi^*(s)/ds\,|_{s=0} = (L^2/D\alpha)[1 - (1/\alpha)(1 - \exp(-\alpha))] \qquad (10)$$

Similarly, the second moment $E(T^2)$ can be obtained by differentiating $\varphi^*(s)$ twice:

$$E(T^2) = d^2\varphi^*(s)/ds^2\,|_{s=0} = (L^2/D\alpha)^2 (\alpha^2 - 2\alpha + 1 + \exp(-2\alpha) - 2\exp(-\alpha) + 2\alpha\exp(-\alpha)) \qquad (11)$$

From here the variance $\sigma^2(T) = E(T^2) - E^2(T)$ is obtained as

$$\sigma^2(T) = (L^2/D\alpha)^2 (1/\alpha^4)(2\alpha - 5 + 4\exp(-\alpha) + 4\alpha\exp(-\alpha) + \exp(-2\alpha)) \qquad (12)$$

where σ is the standard deviation.
For $v_z = 0$, these three statistics are given by

$$E_0(T) = L^2/2D \qquad E_0(T^2) = (5/12)\,(L^4/D^2) \qquad \sigma_0^2(T) = (1/6)\,(L^4/D^2) \qquad (13)$$

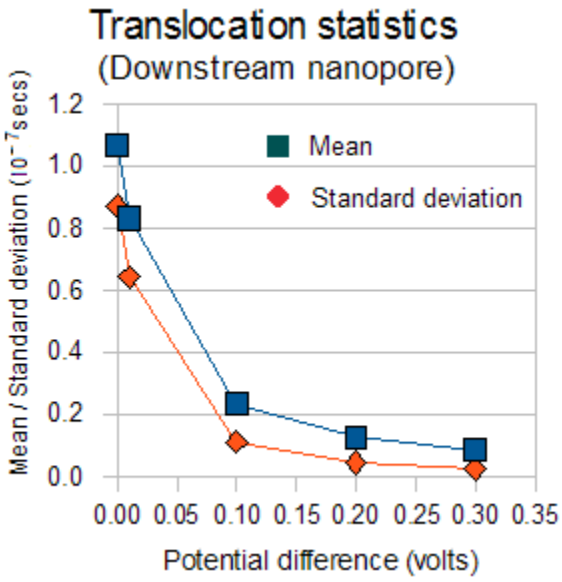

Figure 4 Translocation statistics for DNP

Figure 4 shows the mean and standard deviation of T for different values of $V_{34} \geq 0$ for a DNP with $L = 8 \times 10^{-9}$ m, $D = 3 \times 10^{-10}$ m²/sec, and $\mu = 2.4 \times 10^{-8}$ m²/volt-sec. In [1] the optimum potential difference across the nanopore for detecting a base dropped into the pore is noted as 0.180 V, for which the mean and standard deviation from the data for Figure 4 are respectively 1.3786×10^{-8} sec and 5.019×10^{-9} sec. At these levels the bandwidth required of the detection electronics to detect the corresponding blockade current is very high. One possible way to slow down the translocating base and thereby reduce the bandwidth requirement is to use negative V_{34}, see Section IV.6 for a discussion of this possibility.

IV.2 Model of Stage 3 (*Delivery*)

For simplicity the *trans1*/*cis2* compartment is assumed to be a rectangular box-shaped region (see Figure 3b; the tapered structure of Figure 2 and Figure 3c is considered later) in which a particle is released at the top and translocates to the bottom of the compartment where it is 'absorbed' . 'Absorption' here is taken to mean that the particle moves to the next section without regressing where it is subjected to the model pertaining to that section. The propagator function G(x, y, z, t) is given by a linear Fokker-Planck equation in three dimensions:

$$\partial G/\partial t + v_x\,\partial G/\partial x + v_y\,\partial G/\partial y + v_z\,\partial G/\partial z = D\,(\partial^2 G/\partial x^2 + \partial^2 G/\partial y^2 + \partial^2 G/\partial z^2) \qquad (14)$$

where v_x, v_y, and v_z are the drift velocities in the x, y, and z directions, and D is the diffusion coefficient. In *trans1*/*cis2* there is no drift potential in the x and y directions (Figure 3b) so

$$v_x = v_y = 0 \qquad (15)$$

in (14). The following initial value (I.V.) and boundary values (B.C.) apply:

1) The particle is released at position (0, 0, 0) at time t = 0. This is represented by a delta function δ(x,y,z):

$$\text{I.V.} \qquad G(0, 0, 0, t=0) = \delta(x,y,z) = \delta(x)\delta(y)\delta(z) \qquad (16)$$

2) It is absorbed at the bottom of *trans1*/*cis2* at t > 0:

$$\text{B.C. 1} \qquad G(x, y, L_{23}=L, t) = 0 \qquad (17)$$

3) It is reflected at the sides of *trans1*/*cis2* at t > 0:

$$\text{B.C. 2} \qquad D\,\partial G(x, y, z, t)/\partial x \,|_{x = \pm d/2} = 0 \qquad (18)$$

$$\text{B.C. 3} \qquad D\,\partial G(x, y, z, t)/\partial y \,|_{y = \pm d/2} = 0 \qquad (19)$$

3) It is reflected at the top of trans1/cis2:

$$\text{B.C. 4} \qquad D\,\partial G(x, y, z, t)/\partial z \,|_{z = 0} = v_z\, G(x, y, 0, t), \qquad t > 0 \qquad (20)$$

Since the initial value is a separable function of x, y, and z, the above boundary value problem in three dimensions can be considered mathematically as three boundary value problems [9], one in each dimension, and the propagator function viewed as the product of three independent propagator functions:

$$G(x,y,z,t) = G_x(x,t)\, G_y(y,t)\, G_z(z,t) \qquad (21)$$

where

$$G_x(x,t) = (2/d) \sum_{m=0}^{\infty} \cos \alpha_m x/\sqrt{D} \exp(-\alpha_m^2 t) \qquad (22)$$

$$G_y(y,t) = (2/d) \sum_{n=0}^{\infty} \cos \beta_n x/\sqrt{D} \exp(-\beta_n^2 t) \qquad (23)$$

and

$$G_z(z,t) = (2D/L) \exp(v_z z/2D + v_z^2/4Dt) \sum_{k=1}^{\infty} \sin \omega_k L \sin \omega_k (z-L) \exp(-D\omega_m^2 t)/N(\omega_k) \qquad (24)$$

with

$$\alpha_m = 2m\pi\sqrt{D}/d \qquad \beta_n = 2m\pi\sqrt{D}/d \qquad (25)$$

and

$$N(\omega_k) = (D/v_z)(\exp(v_zL/D)-1) - \{(v_z/D)(\exp(v_zL/D) - \cos 2\omega_kL) - 2\omega_k \sin 2\omega_kL)\}/(v_z/D)^2 + 4\omega_m^2) \quad (26)$$

If detection is defined to occur at z=L, the first passage time is the time the particle crosses z=L at any x and y, $0 \le x,y \le d/2$, so that its pdf $\varphi(t)$ can be written as

$$\varphi(t) = \int_{-d/2}^{d/2} \int_{-d/2}^{d/2} (-D\, dG(x,y,z,t)/dz\,|_{z=L})\, dx\, dy = \int_{-d/2}^{d/2} G_x(x,t)\, dx \int_{-d/2}^{d/2} G_y(y,t)\, dy\, \varphi_z(t) \quad (27)$$

where

$$\varphi_z(t) = 2D \exp(v_zz/2D-v_z^2/4Dt) \sum_{k=1}^{\infty} \omega_k \sin \omega_kL \exp(-D\omega_m^2t)/N(\omega_k) \quad (28)$$

and

$$N(\omega_k) = (D/v_z)(\exp(v_zL/D)-1) - \{(v_z/D)(\exp(v_zL/D) - \cos 2\omega_kL) - 2\omega_k \sin 2\omega_kL)\}/(v_z/D)^2 + 4\omega_m^2) \quad (29)$$

Just as separation of the three-dimensional boundary value problem defined by Equations 14-20 into three independent one-dimensional boundary value problems is mathematically justified [9], one can consider in physical terms a similar separation of diffusive effects in the three directions. With free diffusion given by Equations 14-15 and the initial condition in Equation 16 (thus the boundary conditions are dispensed with), the diffusion has a spatial mean of (0,0,0) and is independent in the three directions. Adding the reflective boundaries z=0, x= ±d/2, and y=±d/2 and a positive drift potential ($V_{23} > 0$) causes the mean of the first passage time to z=L (which is an absorbing boundary, where detection is considered to occur for any x and y; $0 \le x,y \le d/2$) to be less than the mean time with $V_{23} = 0$. Considering $\varphi_z(t)$ in isolation, this is in effect the one-dimensional first passage time distribution with mean $E(T_z)$ and standard deviation σ_z. Figure 6 shows the dependence of these quantities on V_{23} for L=0.1mm.

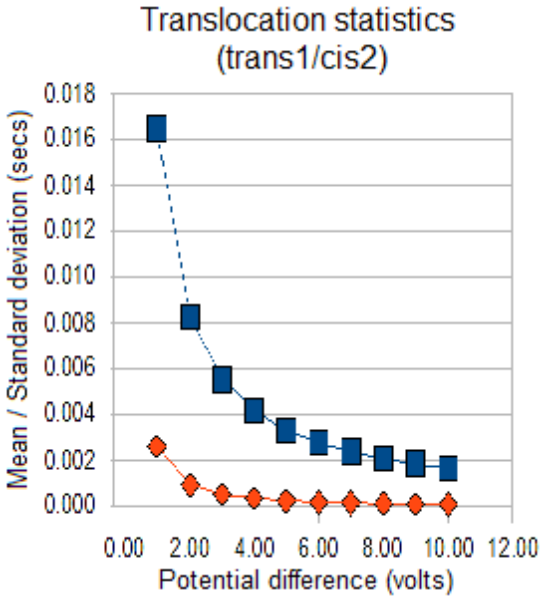

Figure 5 **Translocation statistics for** *trans1/cis2*

To see if diffusion in the x and y directions has any effect on G(x,y,z,t) consider the factor $\int_{-d/2}^{d/2} G_x(x,t)\,dx$ in Equation 27 (the behavior of $G_y(y,t)$ is identical owing to the symmetry in x and y). To compute its value the method of images [10] can be used. Thus start without any boundary conditions on x, which corresponds to free diffusion in the x dimension. $G_x(x,t)$ is then given by the heat kernel:

$$G_x(x,t) = (1/(\sqrt{(\pi 4Dt)})\exp(-x^2/(4Dt)), \qquad -\infty < z < \infty \qquad (30)$$

The reflecting boundaries at x=±d/2 result in a pair of images that add $(1/(\sqrt{(\pi 4Dt)}))\exp(-(x-d)^2/(4Dt))$ and $(1/(\sqrt{(\pi 4Dt)}))\exp(-(x+d)^2/(4Dt))$ to the right side of Equation (30):

$$G_x(x,t) \approx (1/(\sqrt{(\pi 4Dt)}))\,[\exp(-x^2/4Dt) + \exp(-(x-d)^2/4Dt) + \exp(-(x+d)^2/4Dt)], \qquad -d/2 \le x \le d/2 \qquad (31)$$

This is only approximate because the two images are themselves reflected at the opposite boundaries, leading to another pair of image functions centered at x = 2d and x = -2d, leading to

$$G_x(x,t) \approx (1/(\sqrt{(\pi 4Dt)}))\,[\exp(-x^2/4Dt) + \exp(-(x-d)^2/4Dt) + \exp(-(x+d)^2/4Dt) + \exp(-(x-2d)^2/4Dt) + \exp(-(x+2d)^2/4Dt)], \quad -d/2 \le x \le d/2 \quad (32)$$

These images occur ad infinitum (similar to opposing mirrors). The final result (which is exact) is

$$G_x(x,t) = (1/\sqrt{(\pi 4Dt)}[\exp(-x^2/4Dt) + \textstyle\sum_{k=1}^{\infty}\exp(-(x+kd)^2/4Dt) + \sum_{k=1}^{\infty}\exp(-(x-kd)^2/4Dt)], \quad -d/2 \le x \le d/2 \qquad (33)$$

However this is just the free diffusion heat kernel function folded inward an infinite number of times at the two reflecting boundaries. Because probability is conserved, the integral of $G_x(x,t)$ over $-d/2 \le x \le d/2$ is the area under the heat kernel function over $-\infty < z < \infty$, which is 1. A similar result holds for G(y, t) by symmetry. Hence Equation (27) reduces to

$$\varphi(t) = \varphi_z(t) \qquad (34)$$

This means that diffusion in the x and y dimensions does not affect the translocation time in the z dimension if it is assumed that arrival of the particle at any (x, y, z=L) is tantamount to detection.

IV.3 Model of Stage 5 (*Recovery*)

If it is assumed that the 'detected' particle moves into *trans2* without regressing into DNP, the analytical model for *trans1* can be used for Stage 5 as well.

IV.4 Model of Stages 1 and 2 (*Thread-Cleave*)

Studies have shown that ssDNA molecules are more easily threaded through the pore the higher the membrane voltage (see, for example, Figure 7 in [3]). Therefore it is assumed here that the ssDNA strand always threads through UNP and encounters the exonuclease on the exit side of

UNP. Existing DNA capture models [5, 8] may be used for these two stages after modification to include the effect of the electric field due to V_0V_1.

IV.5 Behavior at the interface between two sections

Since successive pairs of sections in the tandem cell are conjoined the boundary considered between a pair is artificial so that one needs to consider the behavior of the cleaved base at each interface (Stages 3-4 or Stages 4-5; see Figure 2). There are two possible approaches:

1) The first considers the probability currents at some fixed point (x,y,L±) on either side of an interface pair (*trans1*/*cis2*-DNP or DNP-*trans2*), where L is the length of the first section in the pair. Considering *trans1*/*cis2*-DNP, if there is an absorbing barrier at L- then the probability function on the *trans1*/*cis2* side would be

$$G_3(x,y,L_{23}\text{-},t) = 0 \tag{35}$$

On the DNP side if there is a reflecting barrier the probability current would be

$$J_4(x,y,L_{23}\text{+},t) = v_{z4}\, G_4(x,y,L_{23}\text{+},t) - D\, \partial G_4(x,y,L_{23}\text{+},t)/\partial z = 0 \tag{36}$$

$$v_{z3} = V_{23}/L_{23} \qquad v_{z4} = V_{34}/L_{34} \tag{37}$$

But there is really no barrier. The particle oscillates at the interface because of diffusion, before eventually passing into DNP, such passage being aided by the positively directed drift potentials in both compartments and indirectly by the reflecting boundaries in *trans1*/*cis2*. Thus

$$J_3(x,y,L_{23}\text{-},t) = v_{z3}\, G_3(x,y,L_{23}\text{-},t) - D\, \partial G_3(x,y,L_{23}\text{-},t)/\partial z \neq 0 \tag{38}$$

and

$$J_4(x,y,L_{23}\text{+},t) = v_{z4}\, G_4(x,y,L_{23}\text{+},t) - D\, \partial G_4(x,y,L_{23}\text{+},t)/\partial z \neq 0 \tag{39}$$

Continuity requires

$$J_3(x,y,L_{23}\text{-},t) = J_4(x,y,L_{23}\text{+},t) \tag{40}$$

In order for the particle to translocate successfully through DNP in the z direction so that it can be detected inside DNP, the net probability current at L_{23} must be in the positive z direction. This can be achieved with suitably large positive v_{z3} and v_{z4} by using positive voltages V_{23} and V_{34}. Thus

$$J_{34}(x,y,L_{23},t) = J_3(x,y,L_{23}\text{-},t) = J_4(x,y,L_{23}\text{+},t) > 0 \tag{41}$$

If $E_{23} > E_{34}$ or $E_{23} < E_{34}$ there is a jump in the potential field at the interface and therefore a jump in the drift velocity of the particle at the join $z=L_{23}$. The jump is negative if $E_{23} > E_{34}$ and positive if $E_{23} < E_{34}$. The electric fields however must not be so large as to exceed the breakdown potential of the electrolyte (roughly 70MV/m) in the respective compartment. Also if E_{34} is too

large it would cause the particle to translocate too rapidly through DNP making effective detection difficult.

While these conditions appear to be contradictory, simulations done with Gaussian particles (using the methodology in [2]) suggest that there is a band of values over which both can be satisfied. The optimum values for V_{23} and V_{34} to use in practice can be determined by experiment.

The behavior at the interface between DNP and *trans2* is similar.

Additionally the taper in *trans1*/*cis2* aids drift of the particle into DNP when positive drift fields are used in both compartments. The tapered geometry can be modeled with a Fokker-Planck equation just like in Equation 14 but with a trapezoidal (or conical, if cylindrical coordinates are used) frustum boundary. The resulting system of equations is not as easily solved as Equations 14 through 20 although it is amenable to numerical solution. Similar to the taper in *trans1*/*cis2* aiding capture of the base at the entrance of DNP the abrupt increase in diameter from DNP to *trans2* decreases the probability of a detected particle regressing into DNP from *trans2*. One can also think of these two behaviors in terms of entropy barriers [5]: the taper in *trans1*/*cis2* decreases the barrier for entry into DNP (below what it would be with a box that is not tapered), while the step change going from DNP to *trans2* effectively increases the barrier for regressing into DNP even while the positive drift potential across *trans2* provides the necessary inducement for the detected particle to pass through *trans2* and eventually be recovered.

2) A second possible approach may be based on a recent model of diffusion [11] in a tube of abruptly changing diameter in which the boundary between two neighboring sections is rendered quasi-homogeneous by varying the diffusion constant. As discussed next, one way to slow down translocation inside DNP is to use a negative electrical field. The negative change in voltage can also be modeled as an equivalent change in the diameter of DNP, which in turn can be modeled as a change in the diffusion constant. This approach is under investigation.

IV.6 Slowing down translocation

The speed with which a cleaved base translocates through DNP presents a problem to the detector electronics. Several methods to slow down the translocating base have been considered, examples include the use of 'molecular brakes' [12], magnetic or optical tweezers [13], and increased salt concentration [14]. Here an approach based on the use of an electric field is proposed.

Consider a negative potential over the length of DNP. Extending the computation of the mean and standard deviation of T to include $V_{34} < 0$ for a DNP with L=10 nm, $D = 3 \times 10^{-10}$ m^2/sec, and $\mu = 2.4 \times 10^{-8}$ volt/m^2 sec leads to Figure 6 (the data in this figure form a superset of the data in Figure 4). An increase in the mean translocation time occurs, indicating slowdown, but it is also accompanied by a significant increase in the variance. With V_{34} approaching -0.3 V, the mean has increased by 9 orders of magnitude over the mean for $V_{34} = 0.3$ V and the standard deviation is closely tracking the mean, indicating that diffusion has started to take over.

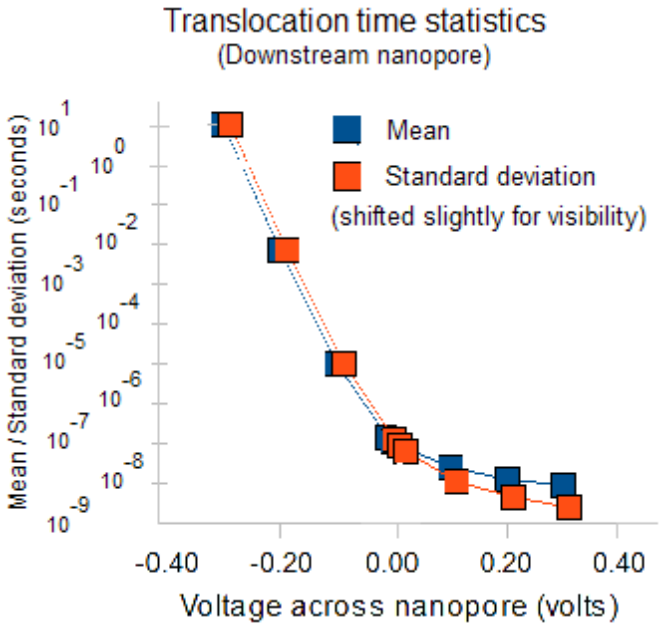

Figure 6

For this approach to work the following conditions have to be met:
1) A cleaved base entering DNP does not regress into *cis2*/*trans1*;
2) A detected base exiting into *trans2* does not regress into DNP;
3) The probability that there is more than one base in DNP approaches 0.

To satisfy condition 1 a base moving from *cis2*/*trans1* into DNP has to experience a positive drift field at the interface. This requires that E_{23} and E_{34} both be positive. To satisfy condition 2 a base moving from DNP into *trans2* has to experience a positive drift field at the interface. This requires that E_{34} and E_{45} both be positive. Slowing down the base inside DNP requires E_{34} to be negative. All three conditions may be satisfied if L_{34} is split into three parts $L_{34\text{-}1}$, $L_{34\text{-}2}$, and $L_{34\text{-}3}$ with respective electric fields $E_{34\text{-}1}$, $E_{34\text{-}2}$, and $E_{34\text{-}3}$ such that $E_{34\text{-}1} > 0$, $E_{34\text{-}2} < 0$, and $E_{34\text{-}3} > 0$. Such an electric field profile is shown in Figure 7.

The analysis in IV.3 may be extended to the behavior of a base that experiences this kind of profiled potential in DNP. There is a tradeoff among the need to reduce the translocation speed through DNP, the need to prevent regression from DNP into *trans1*/*cis2*, and the need to prevent regression into DNP from *trans2*. Let $L_{34\text{-}1}$=a L_{34}, $L_{34\text{-}2}$= bL_{34}, and $L_{34\text{-}3}$= cL_{34}, with a + b + c = 1. The first and second conditions require $E_{34\text{-}1}$ and $E_{34\text{-}3}$ both to be sufficiently positive. Since two electrodes are required to define the internal negative potential segment, each of a, b and c has a minimum value of $b_{min} = e_w + e_s$, where e_w = width of electrode and e_s = interelectrode spacing. This spacing along with the applied voltages $V_{34\text{-}1}$, $V_{34\text{-}2}$, and $V_{34\text{-}3}$ can be used to determine the span of the negative electric field over DNP (Figure 7).

It may be possible to achieve this kind of field profile if ultra thin graphene sheets (which have been studied for their potential use in strand sequencing [15]) are used as electrodes.

Implementation could take the form of two graphene sheets interposed between three layers of silicon pores. The resulting DNP would then have the structure [Si pore-graphene electrode-Si pore-graphene electrode-Si pore].

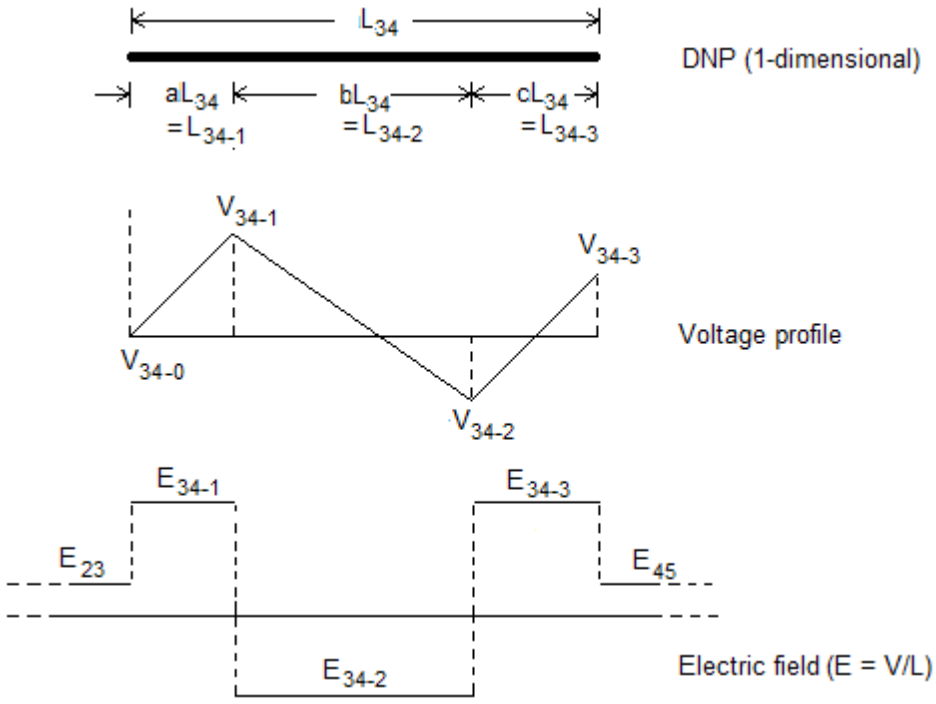

Figure 7 Voltage and electric field profiles over DNP

The objective is to minimize the translocation time of a particle through DNP where the drift velocity has the sign pattern positive-negative-positive. To find an optimum electric field profile over DNP, a mathematical optimization problem can be formulated with the following constraints:

1) $E_{23}, E_{45} > 0$

2) $E_{34-1} > 0, E_{34-2} < 0, E_{34-3} > 0$

3) $L_{34-1}=a\,L_{34}, L_{34-2}= bL_{34}, L_{34-3}= cL_{34}$; $a, b, c > e_w + e_s$; $a + b + c = 1$

This work is in progress.

Here an estimate of the translocation time is obtained by using Equations 10 and 12 from the one-dimensional problem of Section IV.2 and using the piecewise constancy of the electric field while ignoring the transitional behavior at the two ends (see Figure 8c). Let $V_{34-1} - V_{34-0} = V_a$, $V_{34-2} - V_{34-1} = V_b$, and $V_{34-3} - V_{34-2} = V_c$. With $V_a = V_c = 0.1$ volt and $V_b = -0.1$ volt the mean and standard deviation of the translocation time over each of the three segments of a nanopore of length $L_{34} = 10$ nm are shown in Table 1 for different values of a and b.

The translocation is considerably slowed down by the negative field over the segment [aL_{34}, $aL_{34} + bL_{34}$], which dominates the total translocation time over DNP. Figure 8 shows the mean and standard deviation of the translocation time due to the negative field for different values of b.

a = c (positive field segment)	Mean (10^{-8} sec)	Standard deviation (10^{-8} sec)	b (negative field segment)	Mean (10^{-3} sec)	Standard deviation (10^{-3} sec)
0.1	0.0365	0.0173	0.8	7.405078	7.405067
0.2	0.1458	0.0691	0.6	4.165356	4.165350
0.3	0.3281	0.1556	0.4	1.851269	1.851267
0.4	0.5834	0.2765	0.2	0.4628174	0.4628167

Table 1
Translocation times over positive and negative electric field segments of nanopore

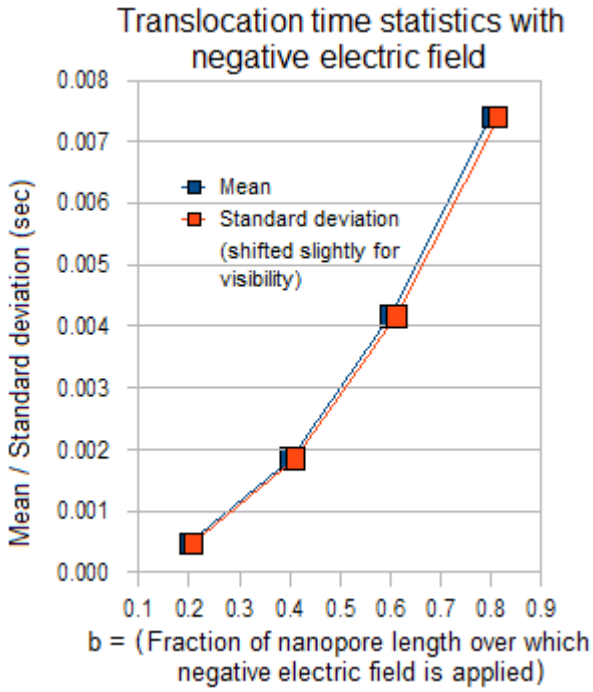

Figure 8 Effect of negative electric field over a segment of DNP of length 10 nm

The above analysis does not take into account any possible adverse effects the negative field might have on the ionic current.

IV.7 Probability of bases arriving at DNP out of order

Assume that bases are cleaved at a rate of one every T seconds. (The designer has some control over the value of T. Examples of control parameters include temperature and salt concentration.) Without loss of generality let base 1 be cleaved at time t=0 and base 2 at t=T. Let T_i = time for base i to diffuse-drift over *trans1*/cis2 and P = translocation time through the pore. T_i, and P are independent random variables with pdfs $f_{Ti}(.)$ and $f_P(.)$, mean $\mu_{Ti}=\mu_B$ and μ_P, and standard deviation $\sigma_{Ti}=\sigma_B$ and σ_P respectively, where the T_is are assumed to be i.i.d. If the two pdfs are

approximated to have finite support equal to six-σ, the supports would respectively be $[\mu_{Ti}-3\sigma_{Ti}, \mu_{Ti}+3\sigma_{Ti}]$ and $[\mu_P-3\sigma_P, \mu_P+3\sigma_P]$. Base 1 is cleaved at t=0, arrives at DNP in the time interval $[\mu_{T1}-3\sigma_{T1}, \mu_{T1}+3\sigma_{T1}]$ and exits DNP in the interval $[\mu_{T1}-3\sigma_{T1}+\mu_P-3\sigma_P, \mu_{T1}+3\sigma_{T1}+\mu_P+3\sigma_P]$. Base 2 is cleaved at time t=T, arrives at DNP in the interval $[\mu_{T2}-3\sigma_{T2}, \mu_{T2+}3\sigma_{T2}]$, and exits DNP in the interval $[\mu_{T2}-3\sigma_{T2}+\mu_P-3\sigma_P, \mu_{T2}+3\sigma_{T2}+\mu_P+3\sigma_P]$. The following sufficient condition holds for the two bases to arrive out of order:

$$T < (\mu_{T1}-3\sigma_{T1}) - (\mu_{T2}+3\sigma_{T2}) \quad (42)$$

or, conversely, for the bases to arrive in order:

$$T > (\mu_{T1}+3\sigma_{T1}) - (\mu_{T2}-3\sigma_{T2}) \quad (43)$$

In [16] the turnover rate for exonuclease under normal conditions is given as 10 to 80 msecs. Setting T=10 msecs and V_{23}=6V and interpolating over the data in Figure 5 gives $\mu_{T1}=\mu_{T2}=2.78$ msecs, $\sigma_{T1}=\sigma_{T2}= 0.18$ msec. Using this in the inequality in Equation 43 gives

$$10 > 2.78 + 3 \times 0.18 - (2.78 - 3 \times 0.18) = 1.08$$

Equation 43 is satisfied, which means that the bases arrive sequentially at DNP.

Using similar arguments, it can be shown that detected bases passing into and through *trans2* do so in their natural order.

IV.8 Probability of two bases being inside DNP at the same time

Let bases be cleaved every T seconds. Consider the random variables T_1 and T_2 corresponding to the arrival times at DNP of two successive bases following cleaving and the random variable P corresponding to the translocation time for a base through the pore. For bases 1 and 2 to be inside the pore at the same time base 2 must arrive in the pore before base 1 exits the pore. The set of events corresponding to this is

$$T1 < T + T2 < T1 + P \quad (44)$$

When the pdfs have finite support equal to six-σ this gives

$$[\mu_{T1}-3\sigma_{T1}, \mu_{T1}+3\sigma_{T1}] < T + [\mu_{T2}-3\sigma_{T2}, \mu_{T2}+3\sigma_{T2}] < [\mu_{T1}-3\sigma_{T1}, \mu_{T1}+3\sigma_{T1}] + [\mu_P-3\sigma_P, \mu_P+3\sigma_P] \quad (45)$$

For the bases to be in the pore at the same time

$$\mu_{T1}+3\sigma_{T1} < T + \mu_{T2} -3\sigma_{T2} \quad (46)$$

(which is the same as Equation 43)
and

$$T + \mu_{T2}+3\sigma_{T2} < \mu_{T1}-3\sigma_{T1} + \mu_P-3\sigma_P \quad (47)$$

If Equation 46 (same as 43) is assumed to be satisfied (corresponding to base 1 arriving at DNP after base 2) only Equation 47 needs to be examined. The condition for two bases not being in the pore at the same time is then obtained by logically inverting Equation 47:

$$T + \mu_{T2} - 3\sigma_{T2} > \mu_{T1} + 3\sigma_{T1} + \mu_P + 3\sigma_P \quad (48)$$

Setting T=10 msec and V_{34}=200 mV and interpolating the data in Figures 6 gives $\mu_{T1}=\mu_{T2}$=2.78 msecs, $\sigma_{T1}=\sigma_{T2}$= 0.18 msec, μ_P=0.000013 msec, and σ_P=0.000004 msec. Using this in the inequality in Equation 48 gives

$$10 + 2.78 - 3 \times 0.18 = 12.24 > 2.78 + 3 \times 0.18 + 0.000013 + 3 \times 0.000004$$

which satisfies Equation 48. The two bases cannot be in the pore at the same time. The bandwidth required would be on the order of 4 Mhz.

Conversely, from Equation 48 the minimum interval required between the release of two successive cleaved bases on the exit of UNP so that they do not occupy DNP at the same time is given by

$$T_{min} = 3\sigma_{T1} + 3\sigma_{T2} + \mu_P + 3\sigma_P \quad (48)$$

Using $\sigma_{T1}=\sigma_{T2}$= 0.18 msec, μ_P=0.000013 msec, and σ_P=0.000004 msec

$$T_{min} = 0.54 + 0.54 + 0.000013 + 0.000012 = 1.080025 \text{ msec}$$

In [1] T is reported to be in the range of 1 to 10 msec. Although T_{min} is greater than the minimum reported value, the range of T well exceeds the minimum so that with a suitable set of controls (temperature, salt concentration, etc.) the enzyme can be set to cleave bases at a rate well below one every 1.08 msec.

With a negative electric field over part of DNP, the standard deviation (Figure 8) is close to the mean, which means that the translocation time distribution is skewed toward 0. To obtain a rough bound on the probability of two bases being in DNP at the same time, consider for example L_{34} = 10 nm, b=0.4, and V_b = -0.1 volt. From the data in Figure 6 and Table 1, $\sigma_{T1}=\sigma_{T2}$= 0.18 msec, μ_P=1.851269 msec, and σ_P=1.851267 msec. Using Equation 48

$$T_{min} = 3\sigma_{T1} + 3\sigma_{T2} + \mu_P + 3\sigma_P = 3 \times 0.18 + 3 \times 0.18 + 1.851269 + 3 \times 1.851267 = 8.485 \text{ msec}$$

This is within the range of turnover rates achieved with exonuclease [1, 16]. With a mean translocation time of 1.851269 msec through a distance of $L_{34\text{-}2} = bL_{34}$ = 4 nm, the detector bandwidth required would be on the order of 1 Khz.

V IMPLEMENTATION ISSUES

V.1 Positioning the enzyme

The enzyme on the exit side of UNP need only be engineered so that it is in the path of the threading DNA sequence such that the first base of the remaining sequence is presented to it. If the leading base is not cleaved then it would mean that the ssDNA has either stopped moving (normally the voltage V_1V_2, which is designed to work with the ratcheting action of the enzyme, would prevent this from happening) or has slipped past the enzyme without being cleaved at all (since cleaving can only occur of the leading base). Failure to detect the characteristic pulses for the four (or five, with methylation) base types would indicate that no cleaving has occurred. Additionally, if occurrence of cleavage can be identified by the electronics then the turnover rate can be artificially controlled to an arbitrary degree by retracting and holding the remaining strand using V_1V_2.

V.2 Enzyme turnover rate

The critical factor determining the sequencing rate is the enzyme turnover rate [1, 16]. The delay between the time when the first base is cleaved and when it is detected is not important. As a result the height of *trans*1/*cis*2 can be large as long as *trans*1/*cis*2 tapers from its width at the top to the width of DNP. With the exonuclease covalently bonded to the exit side of UNP, turnover rates at the enzyme could be optimized using chemical design and appropriate environmental conditions such as temperature, salt concentrations, etc. A feedback circuit to control the voltage across UNP can be designed to work in conjunction with the detector circuit and the ratcheting action of the exonuclease to further ensure this.

V.3 Voltage drift

If an ion-selective pore (for example AHL favors anions over cations 3 to 2) is used for DNP, ion currents, which are typically on the order of a few 100 pA, can lead to an electroosmotic potential which with the passage of time can cause buildup of charge in the pore and lead to the pore voltage drifting over time. The resulting measurement errors can be more or less significant over time depending in part on the volume of the source compartment.

Methods commonly used in electronic measurements may be used to overcome the drift problem. One of these is based on the use of a stable reference voltage against which the drift is tracked and the difference subtracted from the recorded data (much like the moving average in statistical analysis of time series data). This may be done by signal processing hardware or by software in the form of a post-facto computation over the recorded signal. (Considering the times involved the latter can be done in real time with minimal delays.)

Two other alternatives are possible.

1) Increase the volume of the source compartment (*trans1*/*cis2* here) and/or the concentration. Thus if the source volume and concentration are large, the onset of drift may take several minutes or hours, a time span that is sufficient for read lengths (that is, number of bases detected in a single run) to be large. On the other hand a large volume means diffusion becoming dominant. With the tandem cell considered here, if the tapered box for *trans1*/*cis2* has a base of 1 mm × 1 mm tapering to 2 nm × 2 nm over a height of 1 mm, the compartment volume would be ~10^{-9} m^3, which, with 1M solvent, is sufficiently large for diffusion not to be a problem. However, it also increases the footprint of the cell, something that needs to be considered if several thousand cells are to be implemented in parallel on a single substrate.

2) Even if the time for a significant amount of drift to set in is only on the order of a few tens of seconds the number of bases that can be sequenced before voltage drift sets in is a few hundred

(based on the exonuclease turnover rate, which is the determining factor). This is comparable to the sequencing length capability of some commercial systems and of research systems under development and/or reported in the literature. In this case, one possible approach would be to drain the compartments (*trans1*/*cis2*, DNP, *trans2*) periodically and refill with electrolyte. To prevent the occurrence of deletion errors due to cleaved bases still in transit through *trans1*/*cis2* while draining is taking place, the draining step may be preceded by retraction of the strand in UNP (achieved by reversing the potential across UNP) and pausing until the cleaved bases in transit have passed into DNP and been detected. Routine design methods may be used for the required control electronics and mechanicals.

(The above discussion of voltage drift was developed after the issue was raised by a referee for a journal to which an earlier version of this manuscript was submitted.)

V.4 Solid-state pore for DNP

With a DNP based on AHL/MspA, the vestibular structure of these biological pores could lower the probability of capture of a nucleotide in DNP. For example, a nucleotide could find its way to the space between the wall of *trans*1/*cis*2 and the side of the vestibule away from the pore entrance from where it might not be able to diffuse-drift back to the pore, resulting in misses ('deletes') in the sequence. The problem would not be as serious with a solid state DNP because the latter is simply a hole in a substrate. The solid state pore, which also has the advantages of scaling and integration in fabrication, has been studied widely both experimentally and theoretically in the context of DNA sequencing. Whereas in strand sequencing single-nucleotide discrimination within a DNA strand cannot be achieved with solid-state membranes (currently available membranes are around 20 nm thick with an hourglass shape and actual pore thickness about 10 nm [17]), with exonuclease sequencing using a tandem cell this may not be a problem because of the near zero probability of two nucleotides being in the nanopore at the same time. A thicker pore reduces the detection bandwidth required and is also easier to fabricate. Also as noted in Section IV, if an electric field with a negative segment is used along the length of DNP to slow down translocation, it may be possible to fabricate multi-layered silicon structures in which graphene sheets can be interposed between layers to act as electrodes that are used to apply the negative field.

VI DISCUSSION

1. With a tandem cell, repeat bases (homopolymers) do not present a problem because of the time and space separation between successive cleaved bases.
2. Inclusion of an explicit bias in *trans*1/*cis*2 may be considered equivalent to increasing the access resistance ahead of DNP.
3. An arbitrary number of tandem cells could be implemented in parallel. With a sequencing rate of 100/second, an array of 10000 cells can potentially sequence a billion (10^9) bases in ~16 minutes.
4. A pipeline of tandem cells with a *cis trans-cis ... trans-cis trans* structure may be used for error checking and/or to obtain upstream-downstream correlations in real time. With an N-stage pipeline N× coverage is possible, requiring, at least in principle, little more time than the time needed for 1× sequencing. It may also be possible to resynthesize the original strand using the recovered bases in the last stage of the pipeline.

5. MspA with its narrower constriction may provide better discrimination than AHL as DNP.
6. A recent report describes the use of heavy tags attached to mononucleotides that are used by a processive enzyme to synthesize DNA threaded through the enzyme [18]. When a nucleotide is added to the growing DNA strand, the tag is cleaved and drops through a nanopore causing a blockade event that is unique for each of four different tag types. This method is in a sense the opposite of exonuclease-based sequencing. If the tags could be attached to bases in ssDNA, it could be adapted for use with the tandem-cell method proposed here. The heavier base-specific tags could lead to better discrimination among the four base types.
7. Recently nanopores in graphene nanoribbons have been considered for strand sequencing [15]. By measuring the changing transverse current through the sheet as the strand passes through the pore the sequence can be determined. The ultrathin sheet provides the kind of single-base resolution that makes this possible, at least in theory. However the translocation speed is too high for reliable detection. One possible solution to this problem could be embedding the graphene layer in the negative-field segment of a solid-state nanopore with an electric field profile similar to that in Figure 7. The resulting slowdown in the translocation could make it easier to measure changes in the transverse current with available detector bandwidths. The cell would then be a conventional single cell with the structure [*cis*-Si pore-graphene electrode-Si pore-graphene sheet-Si pore-graphene electrode-Si pore-*trans*], where the large thickness of the Si pore, by itself a disadvantage in sequencing, is no longer important.
8. A modified version of the tandem cell may be investigated for polypeptide sequencing. Such a structure would be more complicated than the one for DNA sequencing because a) peptides can be positively or negatively charged or charge neutral; and b) different peptidases have to be used for different types of peptides. This will require a pipeline of tandem cells one for each type of peptide as well as an algorithm to assemble the sequence from the signal data which would consist of multiple time series, one for each peptide type.
9. More generally the tandem cell, after suitable modifications, could be potentially used as a chemical analysis tool to study biomolecules in general.

REFERENCES

[1] J. Clarke, H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley. "Continuous base identification for single-molecule nanopore DNA sequencing." Nature Nanotech. **4**, 265-270 (2009).
[2] J. E. Reiner, A. Balijepalli, J. W. F. Robertson, D. L. Burden, B. S. Drown, and J. J. Kasianowicz. "The effects of diffusion on an exonuclease/nanopore-based DNA sequencing engine." J. Chem Phys. **137** (21), 214903 (2012).
[3] M. Wanunu. "Nanopores: a journey towards DNA sequencing." Phys Life Rev **9** (2), 125–158 (2012).
[4] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. "Characterization of individual polynucleotide molecules using a membrane channel." PNAS **93**, 13770-13773 (1996).
[5] M. Muthukumar. "Theory of capture rate in polymer translocation." J. Chem. Phys. **132**, 195101 (2010).
[6] Y Astier, O Braha, H Bayley. "Toward Single Molecule DNA Sequencing: Direct Identification of Ribonucleoside and Deoxyribonucleoside 5'-Monophosphates by Using an Engineered Protein Nanopore Equipped with a Molecular Adapter." J. ACS **128**, 1705-1710 (2006).

[7] I M Derrington, T Z Butler, M D Collins, E Manrao, M Pavlenok, M Niederweis, J H Gundlach. "Nanopore DNA sequencing with MspA." PNAS **107**, 16060-16065 (2010).
[8] D Y Ling and X S Ling. "On the distribution of DNA translocation times in solid-state nanopores: an analysis using Schrödinger's first-passage-time theory." J. Phys. Condens. Matter **25**, 375102 (2013).
[9] D Trim. *Applied Partial Differential Equations*. PWS, Boston, 1990.
[10] T Myint-U and L Debnath. *Linear Partial Differential Equations for Scientists and Engineers* (4th edition). Birkhauser, Boston, 2007.
[11] A M Berezhkovskii, A V Barzykin, and V Y Zitserman. "One-dimensional description of diffusion in a tube of abruptly changing diameter: boundary homogenization based approach." J Chem Phys. **131**, 224110 (2009).
[12] M Rincon-Restrepo, E Mikhailova, H Bayley, G Maglia. "Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes." Nano Lett. **11**, 746-750 (2011).
[13] E H Trepagnier, A Radenovic, D Sivak, P Geissler, J Liphardt. "Controlling DNA capture and propagation through artificial nanopores." Nano Lett. **7**, 2824-2830 (2007).
[14] R Samanthi, S de Zoysa, D A Jayawardhana, Q Zhao, D Wang, D W Armstrong, X Guan. "Slowing DNA Translocation through Nanopores Using a Solution Containing Organic Salts." J. Phys. Chem. B **113**, 13332-13336 (2009).
[15] K K Saha, M Drndic, B K Nikolic. "DNA base-specific modulation of microampere transverse edge currents through a metallic graphene nanoribbon with a nanopore." Nano Lett **12**, 50–5 (2012).
[16] J H Werner et al. "Exonuclease I Hydrolyzes DNA ..." Biophys. J. **88**, 1403–1412 (2005).
[17] S W Kowalczyk, A Y Grosberg, Y Rabin, C Dekker. "Modeling the conductance and DNA blockade of solid-state nanopores." Nanotechnology **22**, 315101 ((2011).
[18] S Kumar, C Tao, M Chien, B Hellner, A Balijepalli, J W F Robertson, Z Li, J J Russo, J E Reiner, J J Kasianowicz, and J Ju. "PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis." Scientific Reports **2**, Article no. 684, September 2012.