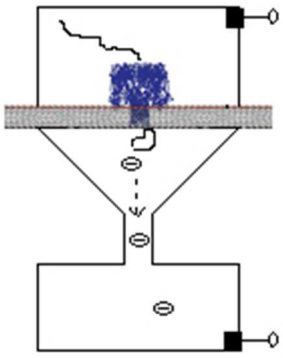
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances



37x46mm (96 x 96 DPI)

A tandem cell for nanopore-based DNA sequencing with exonuclease

G. Sampath[a]

Received (in XXX, XXX) Xth XXXXXXXXX 200X, Accepted Xth XXXXXXXXX 200X
First published on the web Xth XXXXXXXXX 200X
DOI: 10.1039/b00000000x

A tandem electrolytic cell with the structure [*cis*1, upstream nanopore (UNP), *trans*1=*cis*2, downstream nanopore (DNP), *trans*2], an exonuclease enzyme attached to the downstream side of UNP, and a chemical adapter in or a profiled voltage over DNP can be used with bandwidths of a few KHz to sequence bases in ssDNA in natural order with high accuracy and without loss in *trans*1/*cis*2 or regression into DNP from *trans*2.

An electrolytic cell with the structure [*cis,* membrane-with-nanopore, *trans*] and a potential difference between *cis* and *trans* can be used to sequence DNA using a minimal amount of preparation[1,2]. Pores may be biological (AHL[3], MspA[4]) or synthetic (silicon nitride[5], graphene[6-8], 'DNA transistors'[9], silicon-based gated MOSFET-like structures[10,11]).

In 'strand sequencing' negatively charged bases (A, T, C, G) in a single strand of DNA (ssDNA) passing from *cis* to *trans* are identified by their current blockade levels[12]. Base discrimination with this method is limited by the length of the pore and the speed of strand translocation. In 'exonuclease sequencing'[13] an exonuclease enzyme in *cis* adjacent to the nanopore cleaves single bases (more correctly dNMPs or 'deoxynucleoside monophospates'[13], where M = A, T, C, or G; the term 'base' is commonly used instead) in ssDNA and drops them into the pore for identification. This method can discriminate among the base types with 99.8% accuracy[13] but it has a major problem: a cleaved base may diffuse back into the *cis* bulk and be missed or called out of order[14].

In this communication an electrolytic cell structure with two nanopores in tandem is proposed for DNA sequencing. Formally it is written as the pipeline [*cis*1, upstream nanopore (UNP), *trans*1=*cis*2, downstream nanopore (DNP), *trans*2]. An exonuclease enzyme covalently bonded to the downstream side of UNP (assumed to be AHL) cleaves the leading base from ssDNA threading through UNP. The cleaved base is detected as it passes through DNP and is slowed down by a chemical adapter or a profiled voltage applied over a segment of DNP. A Fokker-Planck model of the tandem cell shows that it avoids the difficulties encountered in strand and exonuclease sequencing (as it is currently known[13]) and can, with probability approaching 1, sequence ssDNA without loss to diffusion or errors in sequence order. In particular, homopolymers (repeat bases) do not present a problem.

With this approach, sequencing efficiency depends only on the level of discrimination in DNP among the four (or more, if modifications like methylation[13] are considered) base types. Possible biological and biological-synthetic implementations are considered. With a biological DNP and a cyclodextrin adapter, 99.8% accuracy is possible[13]. Only the essential features of the proposed structure and its mathematical model are presented here, the details are given in a Supplement, which also contains brief notes on materials and a note on other two-pore sequencing methods that have been reported.

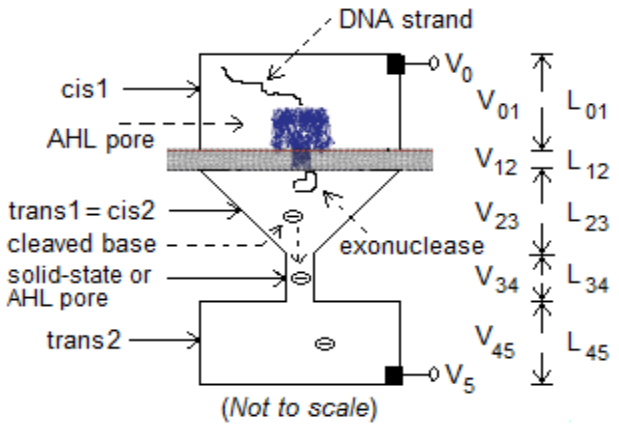


Figure 1. Tandem cell with five pipelined stages: *cis*1, UNP (depicted here as an AHL pore), *trans*1 (= *cis*2), DNP (depicted here as a solid-state pore), and *trans*2. Dimensions considered: 1) *cis*1: box of height 1 μm, side 1 μm; 2) UNP: AHL pore of length 8-10 nm, diameter 2 nm; 3) *trans*1/*cis*2: tapered box of length 1 μm tapering from 1 μm^2 cross-section to 4 nm^2; 4) DNP: solid-state pore of length 10-20 nm and diameter 2 nm, or AHL pore as in UNP; 5) *trans*2: box of height 1 μm, side 1 μm. Electrodes assumed inserted at top of *cis*1 and bottom of *trans*2.

Figure 1 shows a schematic of the tandem cell. Most of the potential difference V_{05} (~99%) drops across the two pores[14]. Here it is assumed that with a large enough V_{05} a strand of ssDNA is captured in the mouth of UNP, threads through UNP, and presents itself to the exonuclease for cleaving on the *trans*1/*cis*2 side of UNP. With V_{05} = 0.4 V and 49.5% of it dropping across each of the two pores, ~198 mV across UNP is sufficient to ensure capture-threading (see Figure 7, Reference 1), and a similar ~198 mV across DNP for a cleaved base to translocate through DNP and be detected there during its passage[13]. Sequencing will be accurate if: a) cleaved bases arrive at and are captured by DNP in their natural order; b) DNP identifies each and every base as it passes through; and c) the detected base exits DNP without regressing. All these conditions are satisfied in the tandem cell based on the following informal rationale (see below for more formal arguments based on a mathematical model): 1) If the leading base is cleaved by the enzyme at a rate of 1 every 10-80 ms[14,15], then with L_{23} = 1 μm, mobility $\mu = 2.4 \times 10^{-8}$ m/V-sec, and diffusion constant $D = 3 \times 10^{-10}$ m^2/sec, the mean time for a cleaved base to translocate through *trans*1/*cis*2 is < 1.667 ms (= $L_{23}^2/2D$, the mean with zero drift voltage; see Supplement). If the translocation time spread is not too high then successive

bases will not enter DNP out of order; 2) A cleaved base cannot regress into *cis*1 because the remaining DNA strand blocks its passage; 3) Similar to 1, the mean translocation time for a cleaved base through DNP is ~0.167 μs << 10 ms = minimum time between two successive bases arriving at DNP. Therefore two bases do not occupy DNP at the same time; 4) Likewise, a detected base exiting into *trans*2 under the influence of V_{05} cannot regress into DNP from *trans*2 for large enough V_{05}. The likelihood of regression can be further reduced with a reinforcing drift field over *trans*2 using an additional electrode V_4 at the top of *trans*2 and a voltage difference applied between V_4 and V_5.

More formally, the behavior of a cleaved base as it translocates through *trans*1/*cis*2 and DNP can be studied via the trajectory of a particle whose propagator function G (x,y,z,t) is given by a linear Fokker-Planck (F-P) equation[14,16] in one dimension (DNP) or three (*trans*1/*cis*2).

1) *Translocation of cleaved base through DNP*. A 1-d approximation is applied to DNP resulting in an initial-value boundary-value problem that can be solved analytically[16,17]. A cleaved base is treated as a particle that is released at z = 0, t = 0; reflected at z = 0, t > 0; and captured at z = L_{34}, t > 0.

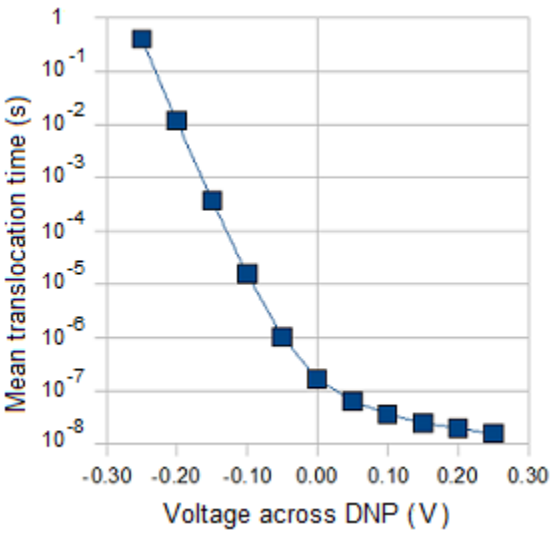


Figure 2. Mean of time for particle to translocate from time of entry into DNP (negligible cross-section and length L_{34} = 8-10 nm) to time of exit into *trans*2. (Standard deviation very close to mean, see Supplement.) Parameter values used: mononucleotide mobility μ = 2.4 × 10^{-8} m^2/V-sec, diffusion constant D = 3 × 10^{-10} m^2/sec. Calculations are for absolute potential differences in the range 0.0-0.3 V. Negative field across DNP results in markedly decreased translocation times, see text below.

The probability density function (pdf) of the first passage time (translocation time) for a particle to diffuse-drift from z = 0 to z = L_{34} and get absorbed at z = L_{34} is obtained using standard methods (see Supplement for details). Equivalently results from a recently published model[14] of exonuclease sequencing have been modified (see Supplement) to obtain analytical expressions for the translocation time (T) statistics (mean and standard deviation) of a cleaved base passing through DNP. Figure 2 shows the mean E(T) for different voltages (including negative voltages, see below for more on using these to slow down a translocating base) over DNP. (The standard deviation σ(T) is very close to the mean and is not shown; see Supplement.) With a biological DNP based on AHL and using the optimum potential difference of ~0.18 V across the pore of a conventional cell as determined in exonuclease sequencing studies[13], E(T) and σ(T) are ~10^{-8} sec, which is too fast for the detector electronics.

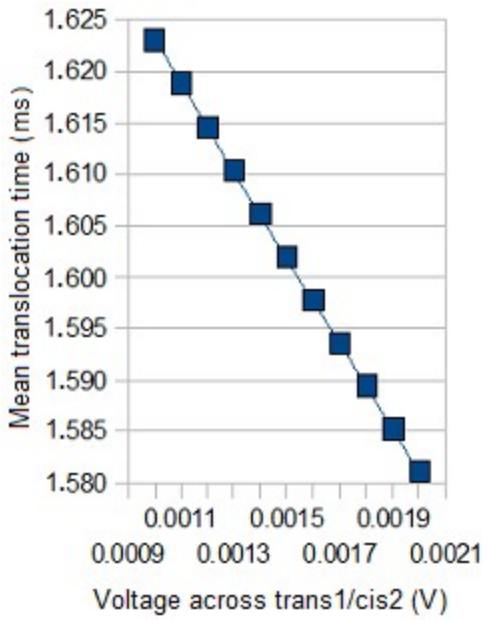


Figure 3. Mean of translocation time for particle (cleaved base) released by exonuclease at top of *trans*1/*cis*2 (3-dimensional box with height 1 μm and cross-section 1 $μm^2$) to move to entrance of DNP. (Standard deviation in same range as mean, not shown; see Supplement.) Parameter values used are same as in Figure 2. Calculations are for potential differences V_{05} in the range 0.0-0.3 V, with 1-2 mV dropping across *trans*1/*cis*2.

2) *Translocation of a cleaved base through trans*1/*cis*2. This is modeled in three dimensions. Assuming *trans*1/*cis*2 to be a rectangular box-shaped region (0 ≤ z ≤ L_{23}; -d/2 ≤ x, y ≤ d/2), a particle is released at the top and translocates to the bottom of the compartment where it is 'absorbed'. That is, the particle is detected when it reaches z = L_{23} independent of x and y and moves into DNP without regressing into *trans*1/*cis*2. Its behavior in DNP is described by a F-P equation in three dimensions with the following initial and boundary conditions: a) The particle is released at position (0,0,0) at t = 0; b) It is reflected at all x, y = ± d/2 and at all z = 0, t > 0; c) It is 'absorbed' at z = L_{23}, t > 0. Since the initial condition G(x,y,z,0) = δ(x,y,z) = δ(x) δ(y) δ(z) is separable in x, y and z, the propagator function G(x,y,z,t) can be written as the product of three independent propagator functions[16]. Based on the definition of 'absorption' as given above, it can be shown that diffusion in the x and y directions has no effect on G(x,y,z,t) so that the first passage time distribution in the three dimensional case reduces to that in the one-dimensional case. The derivation can be found in the Supplement. Figure 3 shows the dependence on pore voltage of the mean translocation time E(T) of a cleaved base through a *trans*1/*cis*2 compartment of length L_{23} = 1 μm. (Once again, the standard deviation is not shown; see Supplement.)

The F-P model summarized above is a piecewise model that

does not consider the behavior of the particle at the interface between two sections. Thus a particle oscillates at an interface because of diffusion. Formal probabilistic arguments can be used (see Supplement) to show in the case of *trans*1-DNP that with sufficiently large V_{05} the particle eventually passes into DNP, such passage being aided directly by the positively directed drift potentials in both sections (and indirectly by the reflecting boundaries in *trans*1/*cis*2). The behavior at the interface between DNP and *trans*2 is similar. The tapered geometry of *trans*1/*cis*2 shown in Figure 1 aids drift into DNP. As with the box geometry considered above it can be modeled with a F-P equation and boundary conditions appropriate to such a geometry. Similarly the abrupt increase in cross-section from DNP to *trans*2 decreases the probability of a detected particle regressing into DNP from *trans*2.

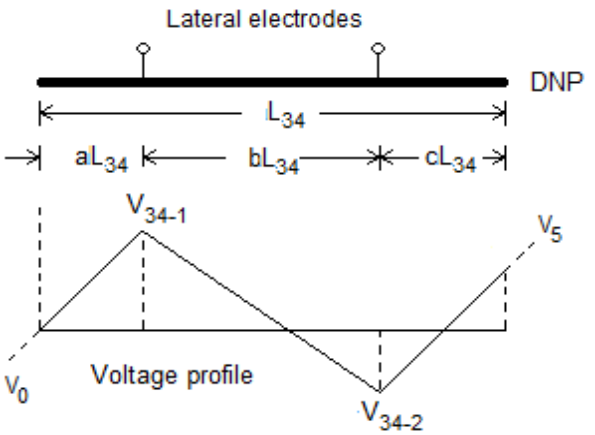


Figure 4. Example of a profiled voltage in which the pore length is divided into three segments L_{34-1}, L_{34-2}, and L_{34-3}, with lengths $L_{34-1} + L_{34-2} + L_{34-3} = L_{34}$. $V_{34-1} > V_0$ and $V_{34-2} < V_5$. Electric field is positive over L_{34-1} and L_{34-3}, negative over L_{34-2}.

The translocation of a base through DNP is too fast for the detector electronics. Methods to slow down translocation include the use of magnetic or optical tweezers[18], alternative electrolytes and/or increased salt concentration[19], and molecular brakes[20]. With a biological DNP (AHL) a covalently attached adapter[13] inside the pore can be used for slowdown; the bandwidth required is then ~20 KHz[13]. An alternative approach for use with a synthetic DNP based on a profiled electric field over DNP (Figure 4) is considered here. To obtain such a profile, voltages V_{34-1} and V_{34-2} are applied to lateral electrodes in DNP at aL_{34} and $aL_{34} + bL_{34}$ with $a + b + c = 1$, $V_0 < V_{34-1}$, $V_{34-1} > V_{34-2}$, and $V_{34-2} < V_5$. Calculations show that translocation over the segment $[aL_{34}, aL_{34} + bL_{34}]$ is considerably slowed down by the negative field, which also dominates the total translocation time over DNP. With L = 10 nm, a = c = 0.3, b = 0.4, $V_a = V_c = V_{34-1} - V_0 = 0.05$ V, and $V_b = V_{34-2} - V_{34-1} = -0.18$ V, the translocation time goes up from ~20 ns (for $V_{34} = 0.2$ V) to ~2.9 msec. See Supplement for table of calculated data.

The maximum voltage differences that can be applied over any stage in the tandem cell are set by the value of the breakdown field for distilled water (~70 MV/m), thus voltage differences of up to ~0.7 V are possible over a length of 10 nm. The optimum values for V_{05}, V_{34-1}, and V_{34-2} can be determined from experiment. With AHL for DNP the optimum over DNP [13] to prevent regression into DNP is ~0.18 V.

Results from the model show that with probability approaching 1 cleaved bases enter DNP in their natural order and that no more than one base occupies DNP at any time. The next three paragraphs apply to a biological or synthetic DNP.

Let bases be cleaved at a rate of one every T seconds, where T is a random variable with a distribution of values[15] in the range 10-80 ms. 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 *trans*1/*cis*2 and P = translocation time through the pore. Let T_i and P be independent and identically distributed (i.i.d.) random variables with respective means of μ_{Ti} and μ_P, standard deviation σ_{Ti} and σ_P, and finite support equal to 6σ. The following sufficient condition holds for the two bases to arrive in order:

$$T > (\mu_{T1}+3\sigma_{T1}) - \max(0, \mu_{T2}-3\sigma_{T2}) \qquad (1)$$

Using a minimum value of 10 ms for T and the data in Figure 3 (as well as standard deviation data from the Supplement), $V_{23} = 1.5$ mV, $\mu_{T1} = \mu_{T2} = 1.6$ ms, and $\sigma_{T1} = \sigma_{T2} = 1.3$ ms, Equation (1) is satisfied. Thus bases arrive sequentially at DNP. Using similar arguments, it can be shown that detected bases passing into and through *trans*2 do so in their natural order.

The condition for two bases not being in the pore at the same time can be obtained as

$$T + \max(0,\mu_{T2}-3\sigma_{T2}) > \mu_{T1} + 3\sigma_{T1} + \mu_P + 3\sigma_P \qquad (2)$$

Using T = 10 ms and the data in Figure 2 (plus standard deviation data from the Supplement), $V_{23} = 1.6$ mV, $V_{34} = 0.2$ V, $\mu_{T1} = \mu_{T2} = 1.6$ ms, $\sigma_{T1} = \sigma_{T2} = 1.3$ ms, $\mu_P \approx 2 \times 10^{-8}$ s, and $\sigma_P = 0.68 \times 10^{-8}$ s, Equation 2 is satisfied, so two bases cannot be in the pore at the same time. This also means that *repeat bases (homopolymers) can be identified without difficulty*.

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

$$T_{min} = \mu_{T1} + 3\sigma_{T1} + \mu_P + 3\sigma_P \qquad (3)$$

Using the data given earlier, $T_{min} \approx 5.5$ ms. With suitable controls (temperature, salt concentration, etc.) the enzyme can be set to cleave bases at time intervals > 10 ms[15].

With a negative electric field over part of a synthetic DNP a rough estimate can be obtained for the probability of two bases being in DNP at the same time. With $L_{34} = 10$ nm, $V_{23} = 1.6$ mV, b = 0.4, $V_b = -0.2$ V, $V_a = V_c = 0.05$V, $V_{23} = 1.6$ mV and using data from Figures 2 and 3 and standard deviation values from the Supplement, $\mu_{T1} \approx 1.6$ ms, $\sigma_{T1} \approx 1.3$ ms, and $\mu_P \approx \sigma_P \approx 0.46$ ms, from which $T_{min} \approx 7.35$ ms, which is within the range of turnover rates achieved with exonuclease[15]. The detector bandwidth required (including digital processing and noise filtering) would be on the order of 5 Khz, not far from the 1 msec translocation time criterion[2] for effective detection.

The tandem cell can be implemented in biological form

(AHL or MspA for both UNP and DNP, with a cyclodextrin adapter inside DNP for slowdown) or hybrid biological-synthetic form (AHL/MspA for UNP, synthetic pore for DNP). Several implementation issues are discussed next.

Accuracy. With AHL for DNP and a cyclodextrin adapter inside, base types can be distinguished with a 99.8% accuracy using bandwidths ~20KHz[13]. The accuracy with synthetic DNP may be obtained by experimentation.

Positioning the enzyme. The enzyme on the exit side of UNP needs to be in the path of the threading DNA sequence such that the first base of the remaining sequence is presented to it. Failure to cleave is indicated if current blockade pulses due to bases passing though DNP are totally absent or stop occurring.

Voltage drift. With an ion-selective DNP, ion current changes, which are typically < 100 pA, can lead to the pore voltage drifting over time. Methods commonly used in electronic measurements can be used to solve the voltage drift problem. Alternatively the *trans*1/*cis*2 and *trans*2 compartments and DNP can be drained periodically and refilled with electrolyte.

Solid state pore for DNP. A solid state pore has the advantages of scaling and integration in fabrication. It has been studied widely both experimentally and theoretically in the context of DNA sequencing. While such pores are not useful in strand sequencing for single-nucleotide discrimination because of their thickness (currently they have a minimum thickness of 20 nm and an hourglass shape[5] with actual pore thickness of 10 nm), 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.

Negative field over DNP. A negative field can be implemented over a synthetic DNP with a pair of electrodes in the form of graphene sheets with nanopores[8] alternating with three layers of silicon pores leading to the structure [Si pore, graphene electrode, Si pore, graphene electrode, Si pore]. Si++ or molybdenum sulphide[21] (MoS_2) layers with nanopores may also be used for the electrodes. Other possibilities are considered in the Supplement.

Sticky bases. A cleaved base may stick to a side wall while diffusing inside *trans*1/*cis*2 or DNP. The probability of a base sticking to *trans*1/*cis*2 can be calculated using the model described above. One way to prevent such sticking is to hold the side walls of *trans*1/*cis*2 at a slightly negative potential with respect to V_1 thereby creating a reflecting wall for the negatively charged base. Alternatively a compartment or pore wall may be chemically treated to prevent sticking, as shown in recent sequencing studies with solid-state pores[22] and graphene sheets[23]. Another option is to replace graphene with MoS_2, which is non-sticking and also provides better discrimination, as recent simulation studies have shown[21].

Recovering the original strand. The original strand may be re-synthesized with an extended tandem cell that has a third pore (TNP) and *trans*3 compartment. The cell would then have the structure [*cis*1, UNP, *trans*1=*cis*2, DNP, *trans*2=*cis*3, TNP, *trans*3]. In this extended cell a detected base would pass through TNP and be incorporated into a growing strand with a primer using a processive enzyme attached to the *trans*3 side of TNP. This effectively transforms exonuclease-based sequencing with a tandem cell into a non-destructive process. It is reminiscent of a recently reported sequencing-by-synthesis (SBS) method in which bases with heavy base-specific tags are incorporated into a strand in the *cis* side of a conventional cell using DNA polymerase[24]. The tags are cleaved during base incorporation and drop into the pore where the blockade levels are used to identify the base.

Parallelization. An arbitrary number of tandem cells could be implemented in parallel. With a sequencing rate of 30-100/sec (equal to the enzyme turnover rate), an array of 10000 cells can potentially sequence 10^9 bases in under an hour.

Additional information. Mathematical details and other implementation-related notes are given in a Supplement.

Acknowledgements. The author thanks the referees for their helpful comments.

References

1 M. Wanunu, *Phys Life Rev*, 2012, **9**, 125–158.
2 D. Branton, D. W. Deamer, et al., *Nature Biotech*., 2008, **26**, 1146-1153.
3 Y. Astier, O. Braha, and H. Bayley, *J. ACS*, 2006, **128**, 1705-1710.
4 I. M. Derrington, T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, and J. H. Gundlach, *PNAS*, 2010, **107**, 16060-16065.
5 S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Nanotechnology*, 2011, **22**, 315101.
6 H. W. Ch. Postma, *Nano Lett*. 2010, **10**, 420-425.
7 C. A. Merchant, K. Healy, M. Wanunu, V. Ray, N. Peterman, J. Bartel, M. D. Fischbein, K. Venta, Z. Luo, A. T. C. Johnson, and M. Drndic, *Nano Lett*., 2010, **10**, 2915-2921.
8 K. K. Saha, M. Drndic, and B. K. Nikolic, *Nano Lett*., 2012, **12**, 50–55.
9 S. Polonsky, S. Rossnagel, and G. Stolovitzky, *Appl. Phys. Lett*., 2007, **91**, 153103.
10 Y. He, M. Tsutsui, C. Fan, M. Taniguchi, and T. Kawai, *ACS Nano*, 2011, **5**, 5509–5518.
11 Y. Ai, J. Liu, B. Zhang, and S. Qian, *Anal. Chem*., 2010, **82**, 8217-8225.
12 J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, *PNAS*, 1996, **93**, 13770-13773.
13 J. Clarke, H-C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley, *Nature Nanotech*., 2009, **4**, 265-270.
14 J. E. Reiner, A. Balijepalli, J. F. Robertson, D. L. Burden, B. S. Drown, and J. J. Kasianowicz, *J. Chem Phys*. 2012, **137**, 214903.
15 J. H. Werner, H. Cai, R. A. Keller, and P. M. Goodwin, *Biophys. J*., 2005, **88**, 1403–1412.
16 D. Trim, *Applied Partial Differential Equations*. PWS, Boston, 1990.
17 D. Myint-U and L. Debnath, *Linear Partial Differential Equations for Scientists and Engineers*, Birkhauser, Boston, 4th edn., 2007.
18 E. H. Trepagnier, A. Radenovic, D. Sivak, P. Geissler, J. Liphardt, *Nano Lett*., 2007, **7**, 2824-2830.
19 R. Samanthi, S. de Zoysa, D. A. Jayawardhana, Q. Zhao, D. Wang, D. W. Armstrong, and X. Guan, *J. Phys. Chem. B*, 2009, **113**, 13332-13336.
20 M. Rincon-Restrepo, E. Mikhailova, H. Bayley, and G. Maglia, *Nano Lett*., 2011, **11**, 746-750.
21 A. B. Farimani, K. Min, and N. R. Aluru, *ACS Nano*, 2014, ASAP Article, July 9.
22 E. C. Yusko, J. M. Johnson, S. Majd, P. Prangkio, R. C. Rollings, J. Li, J. Yang, and M. Mayer, *Nature Nanotech*. 2011, **6**, 253–60.
23 G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, and C. Dekker, *Nano Lett*. 2010, **10**, 3163–3167.
24 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, *Scientific Reports*, 2012, **2**, Article no. 684.

[a] E-mail: sampath_2068@yahoo.com