Supporting Information High-speed detection of DNA Translocation in Nanopipettes

Raquel L. Fraccari¹, Pietro Ciccarella², Azadeh Bahrami¹, Marco Carminati², Giorgio Ferrari², Tim Albrecht^{1*}

¹Imperial College London, Department of Chemistry, Exhibition Road, London SW7 2AZ, UK

²Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, P.za Leonardo da Vinci 32, Milano, Italy

Contents

- 1. Gel electrophoresis of DNA samples
- 2. Low-noise amplifier characterization and comparison
- 3. Representative I V curves
- 4. Typical scatter plots of τ vs. ΔI for four L_{DNA} 's
- 5. Most probable τ values and scaling factors p for each V_{bias}
- 6. Comparison of analysis results: log-normal distribution vs. Schrödinger first-passage-based PDF
- 7. Statistical power of the scaling factor
- 8. Derivation of the extended model for the length dependence of the translocation time

1. Gel electrophoresis of DNA samples

The 4 kbp DNA sample was obtained by a long range polymerase chain reaction (PCR) amplification of the FOXA promoter using a Qiagen LongRnage PCR kit and the products purified using a Qiagen PCR purification kit.¹ Linearized 5.31 kbp DNA was produced by digestion of the pET-24a(+) plasmid (Novagen, Hertfordshire, UK) using the restriction enzyme BamH1 (Sigma-Aldrich, Dorset, UK) and purified using a Qiagen PCR purification kit (Hilden, Germany). The linear 48.5 kbp (λ) and 10 kbp DNA were purchased from Promega Limited (Southampton, UK) and New England Biolabs (NEB, Hitchin, UK) respectively. DNA concentrations were measured by UV Vis spectroscopy (NanoDrop 2000c spectrophotometer, Thermo Scientific,MA, U.S.A) and the purity of the DNA samples confirmed on a 0.5 % agarose gel for 48.5 kbp DNA alongside a 1kb extend DNA ladder (NEB) and on a 1 % agarose gel for 10, 5.31 and 4 kbp DNA alongside a 1kb DNA ladder (NEB).



Figure S1. 0.5 % agarose gel buffered with 1x TBE at 4.3 V/cm (left). 1% agarose gel buffered with 1x TAE at 3.8 V/cm (right). Lanes 1 and 3: 1kb extended DNA ladder; lanes 4 and 8: 1 kb DNA ladder; lanes 2, 5, 6 and 7: dsDNA, 48.5 kbp, 10 kbp, 5.31 kbp and 4 kbp, respectively.

2. Low-noise amplifier characterization and comparison

The measured frequency response of the amplifier with an input capacitance of 4pF is shown in Figure S2. The current-to-voltage conversion factor is correctly given by the feedback resistor of the TIA (R5 = $51k\Omega$) multiplied by the gain of the current amplifier (990). The -3dB cut-off frequency is approximately 3 MHz.



Figure S2. Measured frequency response of the fast path (AC output) confirming a bandwidth exceeding 1MHz.

Figure S3 shows the measured input referred current noise PSD for different external capacitances C_{ext} connected at the input node of the amplifier. The noise spectra are in agreement with the approximated theoretical formula:

$$S_i \approx S_{v,OP1} (2\pi f)^2 (C_{ext} + C_{amp})^2 + \frac{4kT}{R_1} + S_{i,ESD}$$
 (ES1)

where $S_{v,OP1}$ is the equivalent voltage noise of the input operational amplifier ($S_{v,OP1} = 7 \cdot 10^{-18} (1 + 50 k Hz/f) V^2/Hz$), C_{amp} is the amplifier input capacitance (including the connector) and the last two terms are the thermal noise of the feedback resistor for the DC current path (R₁=100M Ω) and the white noise added by the electrostatic discharge protection circuit of the CMOS current amplifier (S_{i,ESD} $\approx 10^{-28} \text{ A}^2/\text{Hz}$). The equivalent input noise added by the off-chip TIA is negligible in the useful frequency range of the amplifier.



Figure S3. Measurement of the input-referred current noise PSD for different values of the external capacitance.

The square root of the PSD integrated on a given bandwidth BW returns the rms current noise:

$$I_{rms} \approx \sqrt{S_{v,OP1}} \frac{2\pi}{\sqrt{3}} B W^{\frac{3}{2}} (C_{ext} + C_{amp}) + \sqrt{\left(\frac{4kT}{R_1} + S_{i,ESD}\right)} B W$$
(ES2)

The first term dominates in wide bandwidth measurements (fast pulses detection) and it is proportional to the total capacitance at the input node. Figure S4 and Table S1 show the rms current noise obtained from the measured PSD as a function of the low-pass filter bandwidth without external capacitance (open circuit condition) and with a capacitance of 6.8pF.



Figure S4. RMS current noise as a function of the low-pass filter bandwidth.

Table S1. Measured current resolution

External	Measured input-referred RMS current noise				
Capacitance Cext	10 kHz	100 kHz	200 kHz	500 kHz	1 MHz
0 pF	1.8 pA	6.6 pA	10.7 pA	23 pA	47 pA
6.8 pF	1.8 pA	7.3 pA	13.3 pA	35 pA	84 pA

The custom amplifier has ten times more bandwidth than the commercial instrument Axopatch 200B. Nevertheless, it maintains a lower noise when operated at similar bandwidth, as shown in Fig. S5 in the case of 60kHz.



Figure S5. Comparison between Axopatch 200B set-up (left panel) and AC and DC current from the custommade low noise amplifier (middle and right panel respectively) for a 4 kb ds DNA translocation in 1 M KCl, 10 mM Tris-hydrochloride (HCl), 1 mM Tris-ethylene diamine tetraacetic acid (EDTA) (pH 8) with $V_{bias} = -500$ mV, a filter frequency of 60 kHz and sampling rate of 4 μ s.

In comparison with the setup described in reference 2 (or 27 in the main text), we see the following advantages of this new design:

1) The noise level of our circuit is lower than their reported value by a factor \approx 2 (84 pArms versus 155 pArms) operating with the same bandwidth (1MHz) and capacitance of the sensor (6-7 pF).

2) Their circuit uses an active current source to manage the DC current. Differently from our circuital solution, the active current source adds a shot noise term proportional to the DC current. For a current of 10 nA, a common value for our nanopipettes, the shot noise adds white noise of 56 fA/VHz, 3.5 times higher than the white noise of our circuit.

3) Their platform has an electrode fabricated on the surface of the CMOS chip. Although this solution minimizes the stray capacitance of the wire connection, it requires a non-trivial post-processing of the CMOS chip to convert the standard aluminum pad into an Ag/AgCl electrode. During operation, the electrode is eventually exhausted, requiring the repetition of this fabrication step or even the change of the CMOS chip. Our platform is connected to the sensor using standard connectors without requiring post-processing of the CMOS chip and without imposing constraints to the chamber. The electrode is a simple Ag/AgCl wire.

4) Our circuit splits the processing of the input current in two path, one dedicated to the stationary current and one for the current variations given by the translocation events. Thus, the data acquisition of each path can be optimized in terms of gain, bandwidth and digital-to-analog conversion. Moreover, the digitized samples related to the input current variations have a mean value equal to zero irrespective of the

DC current. This facilitates the data analysis for the extraction of the main parameters related to a translocation event.

Taken together, this new design adds new flexibility and performance for low-current/high bandwidth electronics in nanopore sensing.

3. Representative I – V curves



Figure S 6. Two representatives I – V measurements for two different nanopipettes. Nanopipette pore conductance (G) was estimated from the slope of the I – V measurement (c(KCl) = 1 M). Blue: G = 60.6 nS, $d_{pore} \approx 21 \text{ nm}$. Green: G = 37.4 nS, $d_{pore} \approx 13 \text{ nm}$.

Using the conductance, the following equation, adapted from Steinbock et al. was used to estimate pore diameters:³

$$d_{pore} = \frac{4Gl + \frac{\pi}{2}GD_i}{D_i\pi g(c) - \frac{\pi}{2}G}$$
(ES3)

where d_{pore} is the estimated nanopore size, g(c) the KCl conductance, D_i the inner diameter at the base of the nanopipette and I the nanopipette taper length.

4. Typical scatter plots of τ vs. ΔI for four L_{DNA}'s - Separation of linear and folded DNA translocations

Clustering of DNA translocation events such that one cluster represents primarily linear DNA translocations and the other cluster primarily folded DNA translocations was achieved either by defining a cut-off in the peak amplitude histogram or by applying a Gaussian mixture model (GMM), as discussed below. In Figure S7, the peak amplitude histogram cut-off method was used for the 48.5 kbp sample while a GMM was applied for the remainder of the samples.



Figure S7. Scatter plots for L_{DNA} 's of 48.5, 10, 5.31 and 4 kbp with a V_{bias} of -800 mV and a filter frequency of 100 kHz for 48.5 kbp DNA and 200 kHz for 10, 5.31 and 4 kbp DNA. Events classified as linear DNA translocations are shown in red while those considered to be folded DNA translocations are shown in blue. Black data points are those excluded from the peak selection.

In cases where event clusters were not clearly separated in the event scatter plots, the peak amplitude histogram was used to separate events. DNA translocating in a linear conformation will have a smaller peak amplitude than DNA translocating in a folded conformation. Therefore a cut-off was defined just after the first peak (after accounting for short 'collision' events) in the peak amplitude histogram and all events below this cut-off defined as linear DNA translocations, while those above this cut-off as folded DNA translocations. In the example shown in Figure S8, the cut-off was taken as 300 pA, therefore events above 300 pA were attributed to mostly folded DNA translocations while those less than 300pA to mostly linear translocations as well as noise.



Figure S8. Peak-amplitude histogram for 48.5 kbp sample shown in Figure S7, where a cut-off of 300 pA was defined (top), based on the event current histogram. Bottom: Representative examples of linear (red) and folded (blue) DNA translocations.

5. Most probable τ values

V _{bias} (mV)	τ_{mp} (ms)					
	48.5 kbp	10 kbp	5.31 kbp	4 kbp		
-900	0.869 ± 0.024	0.112	0.051, 0.063	0.047, 0.034		
-800	0.901 ± 0.027	0.131	0.070, 0.060	0.043, 0.045		
-700	1.064 ± 0.046	0.153, 0.148	0.066, 0.082	0.045 ± 0.005		
-600	0.990 ± 0.017	0.176, 0.171	0.083 ± 0.007	0.050 ± 0.005		
-500	1.783, 1.390	0.251 ± 0.025	0.113 ± 0.010	0.061 ± 0.007		
-400	1.884 ± 0.108	0.286 ± 0.008	0.137 ± 0.009	0.080 ± 0.010		
-300	2.822	0.439 ± 0.041	0.104, 0.239	0.174, 0.109		
-200		0.524	0.304 ± 0.039	0.153, 0.186		

Table S2. τ_{mp} values from log-normal fits for the linear population of translocation events

N.B. Where there are three or more τ_{mp} values at a given bias, these have been averaged and the standard error is shown, where there are fewer than three τ_{mp} values at a given bias, the values are stated.

Table S3. Scaling factors p for different V_{bias}

V _{bias}	р	Δp (std. err.)
-900	1.23198	0.032
-800	1.20132	0.016
-700	1.24278	0.037
-600	1.16876	0.056
-500	1.26173	0.077
-400	1.2352	0.052
-300	1.22406	0.048
-200	1.16973	0.285

The average p is 1.22 ± 0.01 (standard error)

6. Comparison of analysis results: log-normal distribution vs. Schrödinger first-passage-based PDF

We used the following expressions for the translocation time distribution function, namely the log-normal distribution,

$$F_{LN}(t) = F_0 + \frac{A}{\sqrt{2\pi}wt} exp\left(-\frac{\left(ln\frac{t}{t_c}\right)^2}{2w^2}\right)$$
(ES4)

where F_0 is an offset, A the area under curve, w the standard deviation and t_c the mean. and the one taken from Ling and Ling, based on Schrödinger's first passage model:⁴

$$F_{FP}(t) = \frac{L}{\sqrt{4\pi Dt^3}} exp\left(-\frac{\left(L-vt\right)^2}{4Dt}\right)$$
(ES5)

As an example, we show the results for 4 kbp DNA, V_{bias} = -0.8 V:



Figure S8. Comparison of the fit functions, eqs. (ES4) and (ES5). The fit according to eq. ES(5) is broader at longer times and the maximum (most probable translocation time) is slightly shifted to the left (to shorter times). Fit parameters, for F_{LN} (std. errors in brackets): $F_0 = 0.00133 (\pm 0.00193)$; $t_c = 0.04259 (\pm 0.00121)$; $w = 0.4343 (\pm 0.02723)$; $A = 0.00463 (\pm 0.00024)$. Note that F0 is small, compared to the height of the peak. For F_{FP} : L = 1360 nm (fixed); D = 6.74 \cdot 10^6 nm^2/\mus; v = (-)25980 nm/ μ s.

7. Statistical power of the scaling factor

Matlab script used for statistical simulations:

```
% The idea of the simulation is to generate simulated data with a scaling
% factor of 1 (for example), based on experimental data, and then count the
\% number of observations that give a p-value in the range observed
\ensuremath{\$} experimentally in the present study. In other words, what is the
% likelihood of 'false positives', provided that the underlying scaling law
% is p = 1?
8
% * Determine three data points at a given experimental condition, from a
% Gaussian distribution with mean m and std. dev. s.
% * Determine the actual mean for these three data points and repeat for four
% different experimental conditions.
% * Plot the data in a log-log representation and fit a linear function;
% * Extract the slope and perform statistical analysis on the results (e.g.
% what fraction yields a slope of 1, rather than 1.3)
clear all; close all;
q=zeros(100,2);
A=zeros(3,4);
cc=0;
nmax=50000; % number of simulated log-log linear fits
ctot=zeros(nmax,1);
l=[4000 5310 10000 48502]; % DNA length in study, in bp
l log=log10(1);
m=[0.080 0.137 0.286 1.884]; % data set for 0.4 V bias (see Table S2 in SI)
s=[0.017 0.016 0.014 0.187]; % data set for 0.4 V bias
p=1.2352; % experimentally observed scaling factor for this data set
dp=0.052; % standard error of p
cut=p-dp; % cutoff value for scaling factor; here: observed value - std. err.
m=m.^{(1/p)}; % this scales the observed means to an exponent of 1
% Optional: scaling of std. errors (assume larger errors). s=s.^(1/p);
% The for loops make the code slow, but transparent. Replace by matri
% operations for speed.
for n0 = 1:nmax
    n0
    for n = 1:3 % Simulate 3 observations for each DNA length
        rng('shuffle');
        A(n,:)=random('norm',m,s,1,4);
        % One observation for each length, taken from a normal distribution
        % of mean m and std. dev. s (from experimental values)
    end
m ac=mean(A); % Calculate apparent mean
m ac log=log10(m ac);
q(n0,:)=polyfit(l log,m ac log,1); % fit linear eq. to double log plot
if q(n0,1) >= cut; \frac{1}{8} count number of simulated slopes that are beyond cutoff
    cc=cc+1;
end
ctot(n0)=cc/n0; % fraction of events beyond cutoff, during simulation run
end
res=['fraction beyond cutoff of ', num2str(cut),': ', num2str(ctot(nmax))];
disp(res)
plot(ctot)
```



Figure S9. Output of the script above: Number of simulated scaling factors > cutoff (1.1832), relative to the number of trials. Dataset for $V_{bias} = -0.4$ V (see above and main text for further details). The probability of observing a scaling factor such as in our experiments, despite the "true" scaling factor being p = 1, is negligibly small (no event in 50000 trials, 3 repeats). Inset: histogram of simulated scaling factors (mean = 1, FWHM ≈ 0.08)

8. Derivation of the extended model for the length dependence of the translocation time

Following Ghosal, we use the following expressions for the viscous and the electric force, respectively.⁵

$$F_{v} = 2\pi \cdot \varepsilon \cdot E_{0} \cdot a \cdot \phi'(a) + \frac{2\pi \cdot \mu}{ln\left(\frac{a}{R}\right)} \cdot v + \frac{2\pi \cdot \varepsilon \cdot E_{0}}{ln\left(\frac{a}{R}\right)} \cdot \left(\xi_{W} - \xi_{p}\right)$$

$$F_{e} = -2\pi \cdot \varepsilon \cdot E_{0} \cdot a \cdot \phi'(a)$$
(ES6)
(ES7)

where ε is the dielectric constant, ϕ' the Coulomb potential, a the cross-sectional radius of the polymer, R the radius of the (cylindrical) pore, μ the kinematic viscosity of the medium, v the translocation speed, E₀ the electric field inside the pore, and ξ_w and ξ_p the ξ -potentials of the wall and polymer, respectively. The friction force F_R is given by,⁶

$$F_R = N_{ads} \cdot \beta \cdot v \tag{FS8}$$

where N_{ads} is the number of monomers (base pairs) adsorbed to the membrane surface and β the friction coefficient.

 N_{ads} shall be related to the number of monomers on the *cis* side of the pore (i.e. before translocation), $N_c = N_p - N_t$, by $N_{ads} = N_c^{\alpha} = (N_p - N_t)^{\alpha}$. N_p is the total number of monomers in the polymer (DNA) strand, N_t the number of monomers that have already translocated (ranging from 1 to N_p) and a is a scaling coefficient ($0 \le \alpha \le 1$). If $\alpha = 1$, then every monomer on the *cis* side is in contact with the membrane surface, e.g. when the DNA is fully relaxed on the surface. If the adsorbed polymer is still globular when on the surface, e.g. if it is not (yet) equilibrated before the translocation process starts, then $\alpha < 1$, because only part of N_c is actually in contact with the surface and contributing to the friction term. We ignore hydrodynamic drag of the DNA globule itself here.⁷

The friction term F_R changes in magnitude with every monomer that translocates, so that it is maximal at the beginning of the translocation process, but eventually vanishes at the end, when $N_{ads} \rightarrow 0$. At each step, the sum of F_v , F_e and F_R shall be zero:

$$F_v + F_e + F_R = 0 \tag{ES9}$$

Using ES6-8 in ES9 and rearranging for v gives

$$v = -\frac{\frac{2\pi \cdot \epsilon \cdot E_0}{ln\left(\frac{a}{R}\right)} \cdot (\xi_W - \xi_p)}{\frac{2\pi \cdot \mu}{ln\left(\frac{a}{R}\right)} - (N_p - N_t)^{\alpha} \cdot \beta}$$
(ES10)

The translocation speed may also be written as

$$v = \rho_N \cdot \frac{dN_t}{dt} \approx \rho_N \cdot \frac{\Delta N_t}{\Delta t}$$
(ES11)

where ρ_N is the distance between two adjacent monomers (0.34 nm/bp for dsDNA). Δt is the time required to pass ΔN_t monomers through the pore. For $\Delta N_t = 1$, one obtains for $\Delta t(1)$:

$$\Delta t(N_t) = \rho_N \cdot \frac{1}{\nu(N_t)} \tag{ES12}$$

Thus, we calculate the time increment Δt for each monomer from N_t = 1 to N_p, and then sum over all N_p to give the translocation time τ :

10

$$\tau = \sum_{N_t=1}^{N_p} \rho_N \cdot \frac{1}{\nu(N_t)} = -\frac{\mu \cdot \rho_N}{\varepsilon \cdot E_0 \cdot (\xi_W - \xi_p)} \cdot N_p + \frac{\ln\left(\frac{\alpha}{R}\right)}{2\pi \cdot \varepsilon \cdot E_0 \cdot (\xi_W - \xi_p)} \cdot \beta \cdot \rho_N \cdot \sum_{1}^{N_p} (N_p - N_t)^{\alpha}$$
(ES13)

The summation on the right-hand side can be calculated approximately using Euler-Maclaurin summation,

$$\sum_{1}^{N_{p}} (N_{p} - N_{t})^{\alpha} = \int_{0}^{N_{p}} (N_{p} - N_{t})^{\alpha} dN_{t} - \frac{1}{2} \cdot N_{p} + \sum_{k=1}^{p} \frac{B_{2k}}{(2k)!} \cdot \left\{ f^{(2k-1)}(N_{p}) - f^{(2k-1)}(0) \right\} + R$$
(ES14)

where B_{2k} is a Bernoulli number, $f^{(2k-1)}$ the $(2k-1)^{th}$ derivative of $f = (N_p - N_t)^{\alpha}$, and R the residual or error term

$$R = \int_{0}^{N_{p}} f^{2p}(x) \cdot \frac{P_{2p}(x)}{(2p)!}$$
(ES15)

P_{2p} is a Bernoulli polynomial.

Inspection of eq. (ES14) however shows that the integral on the right-hand side is of largest order in N_p, compared to the remaining terms, namely α +1. Ignoring those terms, we are left with eq. (ES16), which is equivalent to replacing the sum in eq. (ES13) by an integral (a reasonable approximation for large N_p):

$$\sum_{1}^{N_{p}} (N_{p} - N_{t})^{\alpha} \approx \frac{N^{\alpha} + 1}{\alpha + 1}$$
(ES16)

Returning to eq. (ES13), for the translocation time τ , one obtains, in the limit of small N_p (or negligible β):

$$\tau \approx -\frac{\mu \cdot \rho_N}{\varepsilon \cdot E_0 \cdot (\xi_W - \xi_p)} \cdot N_p \text{ or } \log(\tau) \propto \log(N_p)$$
(ES17)

Note that $(\xi_{W}-\xi_{p}) < 0$ in the way the problem is set up, hence $\tau > 0$.

Or, for large N_p:

$$\tau \approx \frac{\ln\left(\frac{a}{R}\right)}{2\pi \cdot \varepsilon \cdot E_0 \cdot \left(\xi_W - \xi_p\right)} \cdot \beta \cdot \rho_N \cdot \frac{N^{\alpha+1}}{\alpha+1} \text{ or } \log(\tau) \propto (\alpha+1) \cdot \log(N_p)$$
(ES18)

Thus, in the former case the scaling factor is p = 1, in the latter case $p = \alpha + 1$. The transition between the two limits depends on b and, in principle, this result could be used to the adsorption of the polymer close to the pore region (within the constraints given by the model).

Figure 5 in the main text shows $log(\tau)$ vs. $log(N_p)$ as well as the scaling factor p vs. the friction coefficient β .

References:

- (1) Bahrami, A. Biophysical study of DNA at single molecule level using solid-state nanopores. Ph.D. Thesis, Imperial College London, January 2014.
- (2) Rosenstein, J. K.; Wanunu, M.; Merchant, C. A.; Drndic, M.; Shepard, K. L. *Nat. Methods* **2012**, *9*, 487–492.
- (3) Steinbock, L. J.; Lucas, A.; Otto, O.; Keyser, U. F. *Electrophoresis* **2012**, *33*, 3480–3487.
- (4) Ling, D. Y.; Ling, X. S. J. Phys. Condens. Matter **2013**, 25, 375102.
- (5) Ghosal, S. *Phys. Rev. Lett.* **2007**, *98*, 238104.
- (6) Kühner, F.; Erdmann, M.; Sonnenberg, L.; Serr, A.; Morfill, J.; Gaub, H. E. *Langmuir* **2006**, *22*, 11180–11186.
- (7) Lu, B.; Albertorio, F.; Hoogerheide, D. P.; Golovchenko, J. a. *Biophys. J.* **2011**, *101*, 70–79.