LEAST-SQUARES ESTIMATION OF NANOPORE CHANNEL CONDUCTANCE IN VOLTAGE-VARYING EXPERIMENTS

Christopher R. O'Donnell and William B. Dunbar

Baskin School of Engineering, University of California, Santa Cruz, U.S.A.

Abstract: Step-changing and sinusoidal voltage patterns have expanded the capabilities of the nanopore instrument for single molecule manipulation and measurement. A challenge with voltage-varying experiments is that capacitance in the system is excited and masks the contribution of the nanopore channel conductance in the measured current. The conductance is the parameter that can be used to infer the dynamics of the complex (e.g., DNA, or DNA-protein) in the pore. We present a least-squares parameter estimation (LSPE) algorithm for estimating the channel conductance under voltage-varying conditions, including step and sinusoidal voltages, with the objective of inferring the channel conductance parameter as continuously as possible. The algorithm is shown to recover the conductance faster than by waiting for capacitive transients to settle in step-voltage experiments, and provides accurate continuous conductance estimates in sinusoidal voltage experiments, with realistic noise levels superimposed on the measurements.

1 INTRODUCTION

Nanoscale pores are an established tool for measuring and manipulating individual DNA and DNA-protein complexes (Wilson et al., 2009), (Olasagasti et al., 2010). The nanopore device modeled in this work, shown in Figure 1, consists of a single α-hemolysin protein channel inserted in a lipid bilayer, which separates two chambers containing a buffered electrolytic solution. Voltage is applied across the bilayer creating an ionic current through the nanopore that is measured and passed through a 4-pole Bessel filter before being sampled and recorded. As DNA molecules are captured and driven through the nanopore, the conductance of the channel is reduced causing a drop in the measured ionic current. This change in current (or conductance) and its duration are used to characterize the state of the molecule captured in the nanopore.

Constant voltages have been used in the past to examine DNA and enzyme-bound DNA complexes (Benner et al., 2007). The use of time-varying voltages has expanded the capabilities of the nanopore. For example, active control with step-changing voltages has been used to measure nanopore-DNA interactions (Bates et al., 2003), and polymerase-DNA interactions on the nanopore (Wilson et al., 2009), (Olasagasti et al., 2010), at the single molecule level. Sinusoidal voltage patterns have also made it possi-

ble to monitor the presence of DNA in the pore at zero DC voltage (Ervin et al., 2008), with the assistance of custom hardware and filtering. A challenge with time-varying voltages is that the capacitive elements in the system contribute to the measured ionic current. In step-changing experiments, the true value of the conductance is obscured for the duration of the transient, restricting the time-resolution limits for detecting DNA or DNA-protein dynamics (Wilson et al., 2009). The LSPE algorithm presented in this paper uses the classical method of least-squares approximation. The derived LSPE is shown to provide efficient online estimation of the channel conductance during step-changing voltages, and continuous estimation during sinusoidal voltage inputs, with realistic noise superimposed on the measurements.

2 NANOPORE SYSTEM MODEL

The four-state model of the nanopore system has the transfer function $H(s)$ from the input voltage V_p to the output current I_p (i.e., $I_p(s)/V_p(s) = H(s)$ in the Laplace domain) given by

$$H(s) = \frac{C_\Sigma s + G_c}{a_1 s^4 + a_2 s^3 + a_3 s^2 + a_4 s + 1}$$
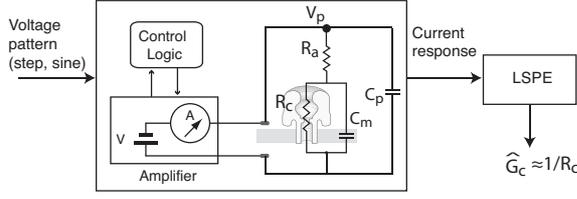
Figure 1: An amplifier applies voltage and measures the ionic current through the nanopore channel. Control logic is used to monitor the current and control the input voltage pattern. The known input signal and the measured current response are used by the LSPE algorithm to estimate $\widehat{G}_c \approx G_c = 1/R_c$, the conductance of the nanopore channel. In the circuit model of the system, R_c is the resistance of the channel, C_m and C_p are the membrane and parasitic capacitances, respectively, V_p is the voltage at the output of the amplifier, and R_a is the electrolytic access resistance.

where $C_\Sigma = C_p + C_m$ (pF) is the combined capacitance of the system (Fig. 1), G_c (nS) is the channel conductance of the nanopore and the coefficients a_1, a_2, a_3 and a_4 are characteristic of the Bessel filter. For consistency of units, time is in milliseconds and frequency is in kHz. We can ignore R_a in the model since it is negligible ($\sim 10^{-4}$ GΩ) compared to R_c (3 GΩ). In another work, we have used system identification tools to validate this model with experimental data (Garalde et al., 2011). The coefficients are defined in terms of the -3 dB cutoff frequency f_c as

$$(a_1, a_2, a_3, a_4) = \frac{(1, 10f, 45f^2, 105f^3)}{105f^4}, \quad (1)$$

with $f = (2\pi f_c)/2.113917675$ (numerator constant identified to match f_c with -3 dB frequency). The frequency domain representation of the system is converted to continuous time state-space (control canonical) form:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t); \ t \geq 0 \quad (2)$$

with column vector $x = [x_1; x_2; x_3; x_4]$ and matrices

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1/a_1 & -a_4/a_1 & -a_3/a_1 & -a_2/a_1 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} G_c/a_1 & C_\Sigma/a_1 & 0 & 0 \end{bmatrix}.$$

In the simulations in Section 4, white noise is added to u and y (with different variances). The system model (2) and LSPE algorithm can be extended to incorporate explicit models of noise (white or colored), with such noise models being experimentally identified. This extension is not done here for brevity.

2.1 Time Discretization of Equations

To perform estimation of parameter G_c by least-squares, the continuous equations of the model are first discretized. The solution to (2) is

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau$$

$$y(t) = Ce^{At}x(0) + \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau.$$

The sample period Δ defines sample times $t_k = k * \Delta$. The input signal is assumed to be piece-wise constant between the sample times: $u(t) = u(t_k)$ for all $t \in [t_k, t_{k+1})$. Using this, the continuous solution is converted to discrete time form as

$$x(t_{k+1}) = A_d x(t_k) + B_d u(t_k), \quad y(t_k) = C_d x(t_k), \quad (3)$$

with $A_d = e^{A\Delta}$, $B_d = \left(\int_0^\Delta e^{A(\Delta-\tau)}d\tau\right)B$, and $C_d = C$. The matrix A is invertible, so the matrix B_d can be rewritten as $B_d = A^{-1}\left(e^{A\Delta} - I\right)B$.

2.1.1 Delta Operator Form

Equation (3) is the traditional discrete time shift operator form, which models the absolute displacement of the state vector from sample to sample, whereas equation (2) models the infinitesimal increment of the state vector defined by the time derivative. This underlying characteristic of the continuous time state-space equations is more accurately modeled in discrete time using the delta operator form (Goodwin et al., 1992). Also known as the divided difference operator form, the delta operator form models the change in the absolute displacement of the state vector from sample to sample over a given sample period. Using the delta operator, the discrete time state-space model takes the form

$$\left.\begin{array}{ll} x_\delta(t_k) & = A_\delta x(t_k) + B_\delta u(t_k) \\ x(t_{k+1}) & = x(t_k) + x_\delta(t_k)\Delta \\ y(t_k) & = C_\delta x(t_k), \end{array}\right\} \quad (4)$$

with $A_\delta = (A_d - I)/\Delta$, $B_\delta = B_d/\Delta$, and $C_\delta = C_d = C$. Equation (4) is used in the remainder of the paper to construct the LSPE algorithm and simulate the response of the nanopore system.

3 LEAST-SQUARES PARAMETER ESTIMATION (LSPE) ALGORITHM

Algebraically, the sampled output can be written in terms of the system parameters, the state vector and

the initial condition by recursively evaluating equation (4). Beginning with t_1, the solution of the sampled output at t_n takes the form

$$y(t_n) = \frac{G_c}{a_1} \left[x_1(t_0) + \sum_{i=0}^{n-1} x_{\delta,1}(t_i)\Delta \right]$$
$$+ \frac{C_\Sigma}{a_1} \left[x_2(t_0) + \sum_{i=0}^{n-1} x_{\delta,2}(t_i)\Delta \right] \quad (5)$$

The matrix expression of interest that relates the output to the system parameters G_c and C_Σ can now be defined as

$$\begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_n) \end{bmatrix} = \begin{bmatrix} Q_1 & | & Q_2 \end{bmatrix} \begin{bmatrix} G_c/a_1 \\ \\ C_\Sigma/a_1 \end{bmatrix}$$

with

$$Q_1 = \begin{bmatrix} x_1(t_0) + x_{\delta,1}(t_0)\Delta \\ x_1(t_0) + x_{\delta,1}(t_0)\Delta + x_{\delta,1}(t_1)\Delta \\ \vdots \\ x_1(t_0) + \sum_{i=0}^{n-1} x_{\delta,1}(t_i)\Delta \end{bmatrix}$$

and

$$Q_2 = \begin{bmatrix} x_2(t_0) + x_{\delta,2}(t_0)\Delta \\ x_2(t_0) + x_{\delta,2}(t_0)\Delta + x_{\delta,2}(t_1)\Delta \\ \vdots \\ x_2(t_0) + \sum_{i=0}^{n-1} x_{\delta,2}(t_i)\Delta \end{bmatrix}$$

which is written in vector notation as

$$y^{1,n} = Qz$$

where the matrix $Q = [Q_1 \, Q_2] \in \mathbb{R}^{n \times 2}$ and the column vector $z = [G_c/a_1; \, C_\Sigma/a_1] \in \mathbb{R}^2$.

3.1 Least-squares Solution

The least-squares approximation problem is based upon finding the best estimate $\hat{z}$ of the vector z that minimizes

$$\|Qz - y^{1,n}\|^2$$

where $\| \cdot \|$ represents the Euclidean norm. Since the matrix Q has more rows than columns and has full column rank, the least-squares approximation problem has a unique solution (Boyd and Vandenberghe, 2004) in the form

$$\hat{z} = (Q^T Q)^{-1} Q^T y^{1,n}.$$

Once the least-squares solution $\hat{z}$ is computed, the estimates of the channel conductance and the system capacitance are $[\widehat{G_c}; \, \widehat{C_\Sigma}] = \hat{z} \times a_1$.

3.2 Sequential Implementation

The channel conductance of the nanopore changes when DNA is captured and translocates through the nanopore. These capture events occur on a micro-to-millisecond time scale (Benner et al., 2007). Thus, the LSPE algorithm must be able to estimate changes in G_c on these time scales. This is accomplished through sequential implementation of the algorithm on overlapping windows of length n that span the input and output data sets of length N, where $N \gg n$. We focus on an online implementation here that makes use of past windows of data to generate the estimate $\widehat{G_c}$. Offline implementation is acceptable when detecting protein-DNA dissociation events after an active control experiment is run, while online implementation allows superior active control during an experiment (Wilson et al., 2009), (Olasagasti et al., 2010).

Sequential implementation of the LSPE algorithm requires the initial condition $x(t_0)$ used in equation (5) to be reset after each iteration to reflect the starting point of the next window. This requires knowledge of the state vector x at every sample instance, which presents a problem since x cannot be directly measured or calculated from measurements. This problem is overcome by simulating x at every sample instance using the system model (4) with a known input signal.

4 RESULTS AND DISCUSSION

The performance of the LSPE algorithm was tested in simulations with step-changing and sinusoidal voltages. To emulate realistic experimental conditions, white noise was added to the input (0.2 mV RMS) and filtered output (1.5 pA RMS) with variances close to those observed experimentally (Wilson et al., 2009) (noise is white up to 10 kHz bandwidth). Also, the value of G_c was set to $1/3$ nS for positive voltages and $2/9$ nS for negative voltages, consistent with values for experiments performed in 0.3 M KCl buffered solution (Wilson et al., 2009). The performance of the LSPE algorithm is compared here to the performance of a simple "I/V method," defined as estimating the conductance by $I_p(t_k)/V_p(t_k)$ at each sample time t_k. When voltage is constant, the current is constant unless changes in G_c occur, for example, if DNA is captured in the nanopore, or polymerase bound to DNA dissociates from the DNA (Wilson et al., 2009), (Olasagasti et al., 2010). Thus, when V_p is constant for a sustained period, the I/V method produces an accurate estimate for G_c. To be of value in estimating G_c, the LSPE should perform comparably to the I/V method when V_p is constant, and outperform the I/V

method when V_p is time-varying.

4.1 Step-changing Input

For a step-changing input, the output current stays
constant except when the input transitions from one
level to another. The switching of the input voltage
produces a transient response in the output current,
the duration of which is dependent on the amplitude
of the input voltage, the amount of capacitance in the
system C_Σ and the Bessel filter cutoff frequency f_c. Of
these three effects, the post-step-change settling time
of the LSPE algorithm is most sensitive to the value of
f_c. An increase from 1 kHz to 10 kHz bandwidth re-
duces the time it takes the algorithm to settle to within
90% of its steady-state value from 0.996 ms to 0.212
ms (data not shown). However, in the presence of ad-
ditive noise, larger bandwidth also allows more noise
to contaminate the estimate $\widehat{G}_c$. To mitigate this trade-
off between minimizing settling time and minimizing
the standard deviation of the estimate, $f_c = 1$ kHz was
qualitatively chosen as the optimal bandwidth for the
LSPE algorithm. Settling time of the algorithm is
also effected by the length n of the sequential over-
lapping windows. Smaller n enables the algorithm
to more efficiently track the true value of G_c, but al-
lows more noise to contaminate the estimate. Again,
to mitigate the tradeoff between minimizing settling
time and standard deviation, $n = 250$ was chosen in
this paper. Future work will determine quantitative
metrics for establishing the optimal f_c and n choices.

Without noise and at 1 kHz bandwidth and 250
kHz sample rate, the step-response settling time of
the LSPE estimate $\widehat{G}_c$ is 0.996 ms, compared to 1.412
ms for the I/V method. That is, the LSPE estimate
converges faster (70%) than the output current does.
Practically, capacitance compensation on the record-
ing amplifier can speed the current settling time (and
thus the I/V method's estimate). However, the I/V
method with a compensated current will, in general,
not work in both step and sinusoidal conditions with-
out heuristic tuning of the compensation settings for
each set of conditions (voltage pattern, bandwidth),
while the LSPE algorithm works universally.

The performance of the LSPE algorithm for step
voltages is shown in Figures 2 and 3. In Figure 2, the
20 mV voltage step is always positive so $G_c = 1/3$
nS. The LSPE estimate $\widehat{G}_c$ has a much smaller stan-
dard deviation, and the I/V method produces a much
larger overshoot. In Figure 3, the 240 mV voltage step
changes polarity, causing a step change in G_c from
$1/3$ to $2/9$ nS. The LSPE algorithm has a larger over-
shoot than in the previous case, but the I/V method
estimate still has a larger overshoot and standard de-

viation at steady-state. In these voltage changes, we
ignore saturation of the measurement current, which
can occur if the recorded output gain is set to high.
Future work will examine and mitigate the effect of
output saturation for the LSPE algorithm.

The LSPE algorithm outperforms the I/V method
in that the estimate $\widehat{G}_c$ has a smaller standard devia-
tion. One might argue that the LSPE algorithm is sim-
ply acting as a filter, and the performance of the I/V
method could be improved if the current were first fil-
tered. In fact, the LSPE algorithm is not a filter but an
estimator, recursively computing the value of $\widehat{G}_c$ that
minimizes the error between the measured current and
current modeled by (4). Although additional low-pass
filtering of the current would reduce the standard de-
viation of the I/V estimate, the filter would further in-
crease the settling time of the estimate.
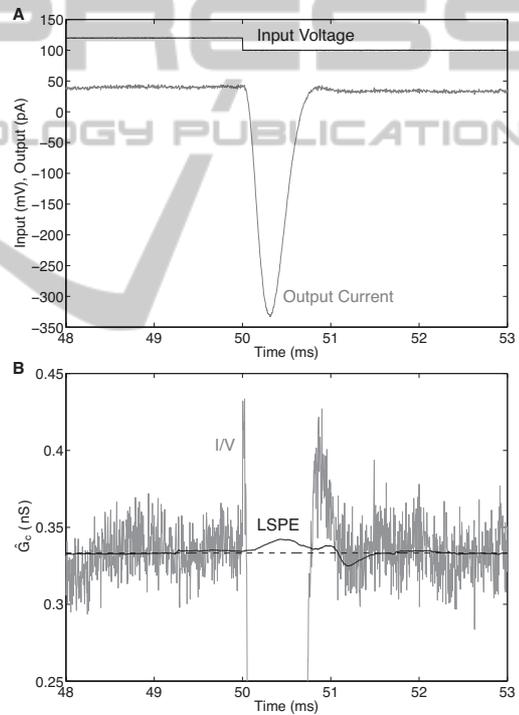
Figure 2: A) Voltage step response (120 to 100 mV) of the
nanopore system model. B) A comparison of the LPSE
and I/V methods for generating $\widehat{G}_c$. The I/V method has
a larger steady-state standard deviation (1.36×10^{-2} nS)
and a much larger overshoot (3.669 nS) in response to a
step change than the LSPE algorithm (7.927×10^{-4} nS and
9.708×10^{-3} nS).

4.2 Sinusoidal Input

For a sinusoidal voltage input, the output current is
constantly in a transient state, with the capacitive
elements in the system being persistently excited.
This has a positive effect on the LSPE algorithm in
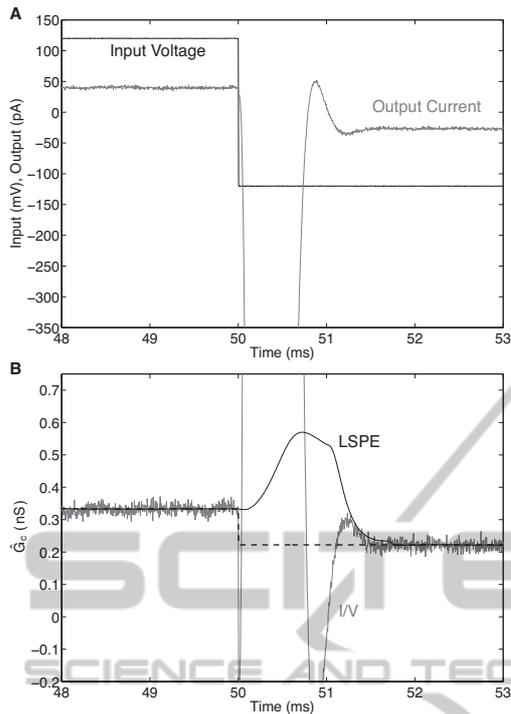
Figure 3: A) Voltage step response (120 to -120 mV) of the nanopore system model. B) A comparison of the LPSE and I/V methods for generating $\widehat{G}_c$. The voltage sign change at 50 ms causes a step change in G_c from $1/3$ to $2/9$ nS. The two methods have comparable settling times, with the LSPE algorithm having a smaller steady-state standard deviation (8.898×10^{-4} nS) and overshoot (0.349 nS) than the I/V method (1.34×10^{-2} nS and 36.57 nS).

that once $\widehat{G}_c$ converges, it does not diverge again, even though both input and output signals are non-constant. The settling time of $\widehat{G}_c$ is insensitive to changes in the sinusoidal frequency f_w. The standard deviation of the estimate increases modestly from 2.69^{-2} nS to 3.49×10^{-2} nS as f_w decreases from 10 Hz to 1 Hz. The sinusoidal frequency $f_w = 10$ Hz is used in the remainder of the paper.

The I/V method does not produce accurate values of $\widehat{G}_c$ for sinusoidal voltages, as expected, but we report the comparison here. Future work will compare LSPE to impedance spectroscopy methods (Katz and Willner, 2003). These methods are comparable to our estimator, and are designed specifically for sinusoidal voltage inputs. The performance of the LSPE algorithm for sinusoidal input voltages is shown in Figures 4 and 5. In Figure 4, $G_c = 1/3$ nS since the input stays positive. The I/V estimate has a large standard deviation and follows a 10 Hz sinusoidal pattern of the measurements, never converging to G_c. The I/V estimate crosses the true value of G_c only at the peaks of the sinusoidal input voltage. This also holds for a

sinusoidal input that changes polarity, shown in Figure 5. The change in polarity results in a step change in G_c, which the LSPE algorithm tracks well (Fig. 5). The LSPE estimate is noisier than when voltage remains positive (Fig. 4), but remains centered around the true values of G_c ($1/3$ nS and $2/9$ nS), whereas the I/V estimate ranges between 3.6×10^3 nS and -2.1×10^{-4} nS.
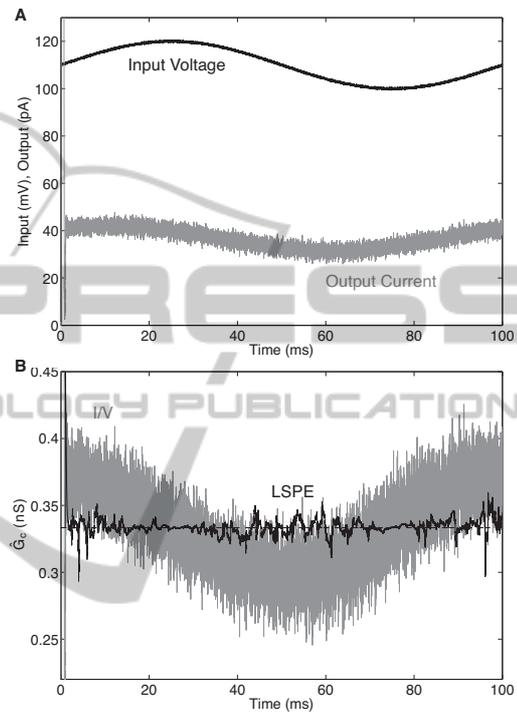
Figure 4: A) Sinusoidal voltage response (10 mV peak-to-peak, 10 Hz, 110 mV DC offset) of the nanopore system model. B) A comparison of the LPSE and I/V methods for generating $\widehat{G}_c$. The I/V method's estimate has a larger standard deviation (2.8×10^{-2} nS) than the LSPE algorithm (5.4×10^{-3} nS) and does not generate accurate estimates.

5 CONCLUSIONS

The LSPE algorithm presented in this paper provides an accurate means for estimating the channel conductance of a nanopore under voltage-varying conditions. The algorithm consistently achieves better performance (in terms of convergence time and standard deviation of the estimate) than the simple I/V method for both step-changing and sinusoidal input voltages. Since variance is improved, DNA or DNA-protein events that can be detected by the measured current (i.e., there is sufficient single-to-noise ratio) are easier to detect with our LSPE algorithm.

We focused on an online implementation here that
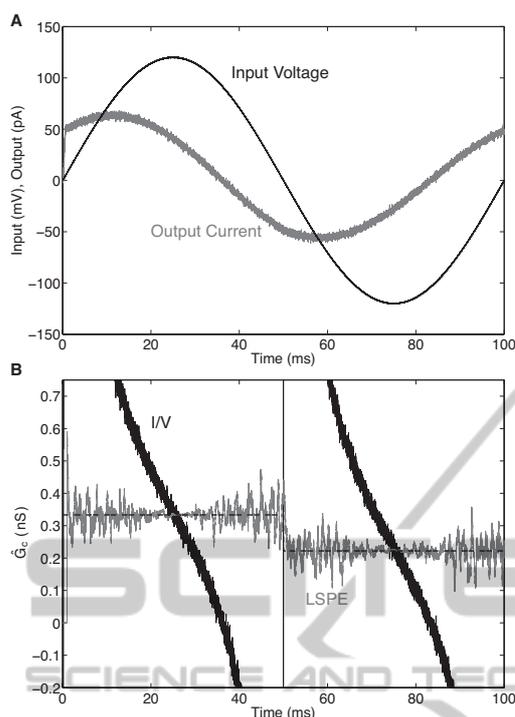
Figure 5: A) Sinusoidal voltage response (120 mV peak-to-peak, 10 Hz, 0 mV DC offset) of the nanopore system model. B) A comparison of the LPSE and I/V methods for generating $\widehat{G}_c$. The voltage sign change at 50 ms causes a step change in G_c from 1/3 to 2/9 nS. The I/V method does not generate accurate estimates, whereas the LSPE algorithm does track the change in G_c.

uses fixed-length windows of past data to generate the estimated conductance value. Future work will explore improving the algorithm's performance by varying the window length based on detected rates of change of the data (Jiang and Zhang, 2004), and by incorporating forgetting-factors in the sequential implementation (Ljung and Gunnarsson, 1990). Also, an offline implementation that makes use of future windows to compute the estimate will be developed to further improve the detection resolution of rapid DNA-protein dissociation events that follow voltage changes in active control experiments (Wilson et al., 2009), (Olasagasti et al., 2010).

The cited advantage of AC-signal detection (absent DC bias) is that nanopore/analyte interactions can be measured while reducing the effects of electroosmosis, electrophoresis, and protein deformation that accompany large DC biases (Ervin et al., 2008). In (Ervin et al., 2008), custom hardware (lock-in amplifier) and software permit high frequency (10–20 mV, 1–2 kHz f_w) sinusoidal voltage inputs. The LSPE derived here cannot track G_c at sinusoidal frequencies above 50 Hz (data not shown). Future work will ex-

plore if and how well the LSPE estimate may track the presence of DNA in the pore at sinusoidal voltages around 0 mV (no DC bias), at 5–50 Hz frequencies, as an alternative to the high frequency method in (Ervin et al., 2008).

REFERENCES

Bates, M., Burns, M., and Meller, A. (2003). Dynamics of DNA molecules in a membrane channel probed by active control techniques. *Biophysical Journal*, 84:2366–2372.

Benner, S., Chen, R. J. A., Wilson, N. A., Abu-Shumays, R., Hurt, N., Lieberman, K. R., Deamer, D. W., Dunbar, W. B., and Akeson, M. (2007). Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nature Nanotechnology*, 2:718–724.

Boyd, S. P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Ervin, E. N., Kawano, R., White, R., and White, H. (2008). Simultaneous alternating and direct current readout of protein ion channel blocking events using glass nanopore membranes. *Anal. Chem*, 80(6):2069–2076.

Garalde, D. R., Maitra, R. D., O'Donnell, C. R., Wang, G., and Dunbar, W. B. (2011). Modeling the biological nanopore instrument for biomolecular state estimation. *IEEE Trans. on Control Systems Technology*, in preparation.

Goodwin, G. C., Middleton, R. H., and Poor, H. V. (1992). High-speed digital signal processing and control. *Proceedings of the IEEE*, 80(2):240–259.

Jiang, J. and Zhang, Y. (2004). A novel variable-length sliding window blockwise least-squares algorithm for on-line estimation of time-varying parameters. *International Journal of Adaptive Control and Signal Processing*, 18(6):505–521.

Katz, E. and Willner, I. (2003). Probing biomolecular interactions at conductive and semiconductive surfaces by impedance spectroscopy: Routes to impedimetric immunosensors, DNA-sensors, and enzyme biosensors. *Electroanalysis*, 15(11):913–947.

Ljung, L. and Gunnarsson, S. (1990). Adaptation and tracking in system identification—a survey. *Automatica*, 26(1):7–21.

Olasagasti, F., Lieberman, K. R., Benner, S., Cherf, G. M., Dahl, J. M., Deamer, D. W., and Akeson, M. (2010). Replication of individual DNA molecules under electronic control using a protein nanopore. *Nature Nanotechnology*, 5(11):798–806.

Wilson, N. A., Abu-Shumays, R., Gyarfas, B., Wang, H., Lieberman, K. R., Akeson, M., and Dunbar, W. B. (2009). Electronic control of DNA polymerase binding and unbinding to single DNA molecules. *ACS Nano*, 3:995–1003.