Acquiring and Classifying Signals from Nanopores and Ion-Channels

Bharatan Konnanath, Prasanna Sattigeri, Trupthi Mathew, Andreas Spanias, Shalini Prasad, Michael Goryll, Trevor Thornton and Peter Knee

Department of Electrical Engineering,
Arizona State University,
Tempe, AZ 85287-5706,
USA.
{bkonnana, psattige, tmathew2, spanias, sprasad5, mgoryll, t.thornton, paknee }@asu.edu

Abstract. The use of engineered nanopores as sensing elements for chemical and biological agents is a rapidly developing area. The distinct signatures of nanopore-nanoparticle lend themselves to statistical analysis. As a result, processing of signals from these sensors is gaining importance, but this field is relatively less developed. In this paper we demonstrate a neural network approach to classify and interpret nanopore and ion-channel signals.

Keywords: Nanopore sensors, Principal component analysis, Sensing using nanopores and neural networks.

Introduction

Resistive pulse sensing or Coulter counting [1] is a wide research area centered on nanopores. Though originally developed for counting particles suspended in a fluid using micrometer sized pores, Coulter counting has recently been applied at the nanoscale level [2]. Through the reduction of device aperture size to the nanometer range, Coulter counting experiments of small particles such as DNA molecules [3, 4] and bovine serum albumin (BSA) [5] through solid state devices as well as virus particles [6] and DNA [7] through polymer materials have been demonstrated.

In the Coulter counting experiments, individual molecules are constrained to pass through a small constrained electric path in a suspending fluid as shown in Fig.1 (left panel). As the molecule passes through the orifice, it causes an increase in resistance which leads to a drop in the current as shown in Fig.1 (right panel). By observing the curvature of these spikes, the size, type and the concentration of the particles can be

This work was supported through the NSF EXP Award 0730810 program award.

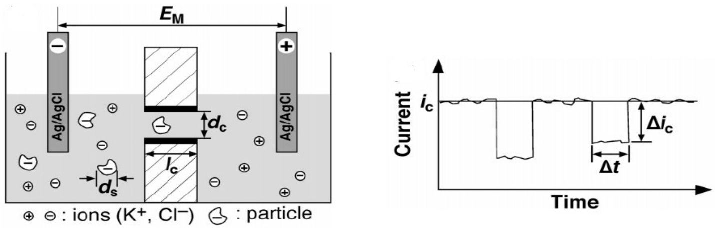

Fig. 1. Graphical Rendering of the Coulter counter (Reproduced from [8].)

determined [8, 9].

Ion channel proteins are naturally occurring nanopores that mediate the flow of ions and molecules across membranes. The utility of ion channels for stochastic sensing has been pioneered by Bayley and several of his collaborators [10]. Ion-channels can be engineered to act as biosensors that can detect metal ions and organic molecules such as proteins [10-12]. Potential applications of ion-channel sensing include detection of reactive molecules in pharmaceutical products, chemical weapons, pesticides and foodstuffs [13].

An example of an engineered pore is shown in Fig. 2. An applied potential to the pore creates a small current flow. A binding site for an analyte is engineered into each pore. An analyte binding event to the pore causes the current to be modulated as shown in the trace below the figure. The signature of the signal through the pore is generally different for distinct analytes. The frequency of occurrence of the binding events was shown to correlate with the concentration of the analyte while parameters of the current, such as the mean duration and amplitude correlate with the type of the analyte. Features of interest fall into two categories: switching and non-switching [12-14], both of which contain information that may be important in detecting an agent.

The conventional modeling procedures used to classify the ion-channel signals are dwell-time analysis [15-17] and Hidden Markov Models (HMMs) [18-21]. Feature extraction for ion-channel signals have been explored in [22, 23]. In this paper, we present a Wavelet based approach for denoising both nanopore and ion-channel signals. A 2 step feature extraction process using Walsh Transforms and Principal Component Analysis (PCA) is used for ion-channel signals. Robust analyte classification is carried out for both cases using Support Vector Machines (SVMs).

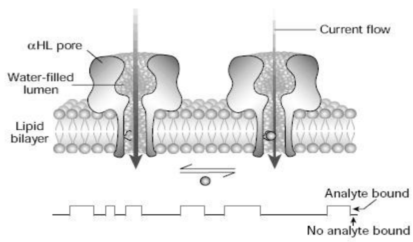

Fig. 2. Engineered Pore. (Reproduced from [8].)

Data Processing for Nanopore Signals

Data Generation

Nanopore data was generated using a Coulter counting element which was constructed using a Teflon chamber with two baths surrounding the nanopore. The two baths were filled with 0.1M KCl electrolyte solution. The nanopore used for the Coulter counting experiments was patterned to a diameter of 300nm but the measured diameter was 212 nm. The recordings were taken using the Axopatch 200B. Voltage traces were incremented from 0-200 mV in steps of 20 mV and each trace lasted 1s. The input signal was filtered with a 5 kHz low pass filter before the A/D conversion stage. The sampling rate was 50 kHz.

Wavelet Transform Based Signal Denoising

In the Coulter counting experiment considered in this paper, the baseline current is at the pA level. Due to the low signal-to-noise (SNR) ratio, the signal peaks, which indicate the transition events, can be easily corrupted. This creates difficulties in measurement of peak parameters such as peak height, width and shape. Hence signal denoising is essential to improve the sensitivity and accuracy of the Coulter counters.

Wavelet based denoising techniques for Coulter counting experiments have been discussed in [24]. Denoising using the Discrete Wavelet Transform (DWT) is a nonlinear operation and involves the following steps:

- Perform a suitable wavelet transform of the noisy data to produce the (noisy) wavelet coefficients.
- Select an appropriate threshold depending on the noise variance and perform a thresholding operation of the wavelet coefficients to remove the noise.
- Zero-pad the signal appropriately and perform the inverse DWT on the thresholded coefficients obtained from the previous step to obtain the signal estimate.

While traditional linear filtering techniques trade off noise suppression against a broadening of signal features, denoising using the DWT preserves the sharpness of features in the original signal. The type of wavelet function, the threshold limits and the level of decomposition is determined on a case by case basis. In our simulations, we determined using cross-validation that the biorthogonal wavelet gave the best performance. To capture most of the features in the signal, the level of decomposition was chosen to be 3. Fig. 3 shows the reproduced signal for a sample frame.

Feature Extraction: Baseline Current, Peak Height and Peak Width

The useful features to be extracted from the nanopore signals are the baseline current, peak height and peak width. The combination of baseline current and height of the peak indicates whether bead has passed through the nanopore completely or not. The peak amplitude is proportional to the baseline current-i.e. greater the baseline current

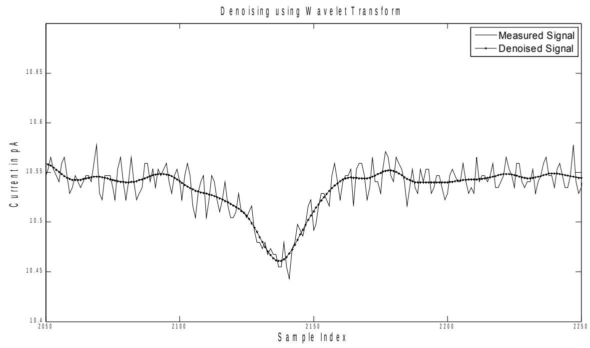

Fig. 3. Nanopore signal Denoising using the Discrete Wavelet Transform (DWT).

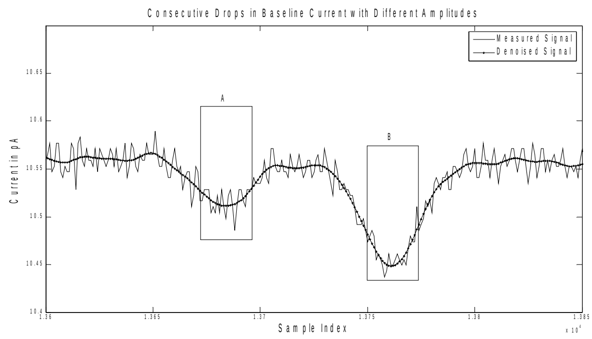

Fig. 4. Signal Denoising using the Discrete Wavelet Transform (DWT).

I, greater will be the drop in current ΔI for beads of the same diameter [9]. The width of the peak is proportional to the diameter of the bead [9]. Fig. 4 shows a sample event, where the bead collided with the pore but bounced back (A) and a few milliseconds later passed through the pore (B).

Event Classification Using Neural Networks

Support vector machines (SVMs) are widely used for solving binary classification problems [25]. SVMs are decision machines that rely on transforming lower-dimensional data into higher dimensional patterns, so that data from two categories can always be separated by a hyperplane, in accordance with Cover's Theorem [26].

The SVM uses the concept of the margin, which is defined to be the smallest distance between the decision boundary and any of the samples [27]. The support vectors are the training samples that are closest to the decision boundary and thus define the optimal separating hyperplane. In support vector machines the decision boundary is chosen to be the one for which the margin is maximized. It can be shown that the larger margin minimizes the total generalization error [26]. The choice of the nonlinear function that maps the input into a higher-dimensional space is usually dependent on the problem domain. Usually polynomial or radial basis functions are used to perform the mapping.

Experimental data for eight different bias voltages, ranging from 0-200mV, with 40,512 samples at each bias voltage are available. A rectangular window of size 1000 samples with no overlap was used to segment the data. In each segment, peaks were extracted using a *gradient* method. Each peak was labeled either as an *event* or a *non-event*. An event indicates that a bead passed through the nanopore completely whereas a non-event indicates whether:

(i) a bead bounced back instead of passing through the nanopore or

(ii) a spike due to noise.

In the given dataset, 3979 peaks were extracted from the signals, out of which 75 peaks indicated events. Peak width, mean baseline current and drop in current amplitude were chosen as features. The dataset was partitioned into a training set (containing 34 events and 1923 non-events) and a test set (containing 41 events and 2056 non-events).

Data Processing for Ion-Channel Signals

Data Generation

Multiple recordings of OmpF ion channels of E. coli in a lipid bilayer across a 50μm wide pore in silicon, sandwiched between reservoirs containing bathed in a 1M KCl solution are used for generating experimental data. Each recording is generated using a sampling rate of 10 kHz for 4 seconds and an applied voltage of 200 mV. The current amplifier employed was a HEKA EPC-8, operating at a gain of 1mV/pA, using a resistive feedback headstage. The input signal was filtered using an 8-pole Bessel filter with a corner frequency of 1kHz before the A/D conversion. The command voltage was generated by a National Instruments 6251 DAQ board.

As shown in Fig. 5, we demonstrate that the Discrete Wavelet Transform can be used for denoising ion-channel signals also using the same technique discussed in Sec 2.2. The Haar wavelet, with the level of decomposition set to 8, was found to give the best performance.

Since experimental data for two-analyte simulations are not yet available, we have used the QUB scientific package to generate synthetic data [28]. Fig. 6 (left panel) shows a sample 4-state Markov model used for generating data and a sample trace is shown in Fig. 6 (right panel). We constructed models to simulate responses of two highly similar analytes which closely resemble the authentic data. Utilizing multiple recordings, an input data matrix of dimension 400×10000 is formed by extracting four 10,000 point sequences of one second duration from each recording for a total of 50 input files for each analyte.

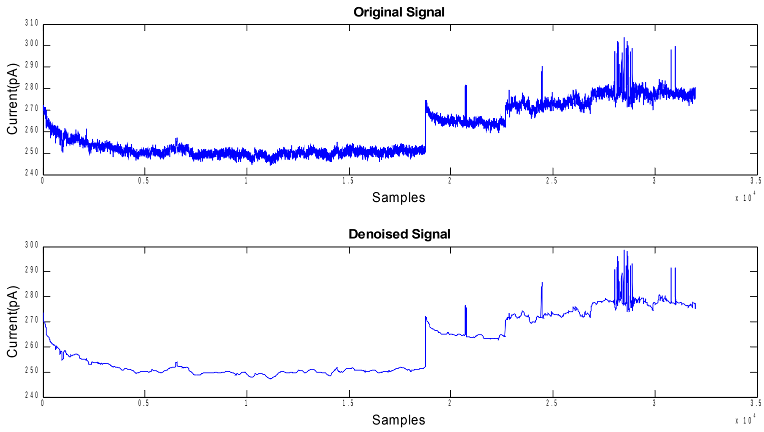

Fig. 5. Top panel: Current Recording for the OmpF ion channel in a membrane across a 50μm wide pore, bathed in 1M KCL solution**.** Bottom panel**:** Ion-channel Signal denoising using the Discrete Wavelet Transform (DWT).

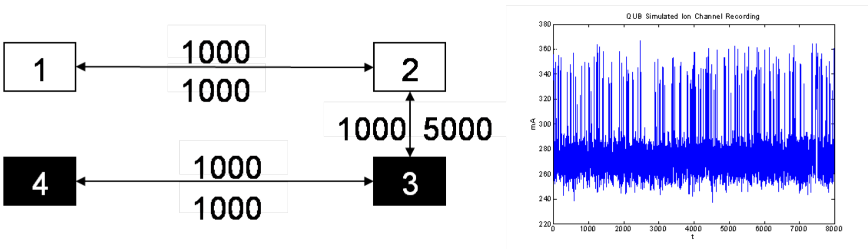

Fig. 6. Left Panel: Basic two class, four state model used by QUB to simulate two analyte ion channel. Right: 8000 *ms* simulation using QUB.

Feature Extraction using Principal Component Analysis

Signals are often processed in the transform domain as they offer attractive benefits like compactness, reduction in computational complexity and robustness to noise. Feature extraction from ion-channel sensor signals using the Walsh-Hadamard Transform (WHT) has been described in [22]. The WHT is able to represent signals with sharp discontinuities more accurately using fewer coefficients. For a given window size *N*, it was determined that 20% of the WHT coefficients represents 90% of the signal energy. Thus by discarding the coefficients that do not contribute significantly to the signal energy, the size of the dataset was reduced by 80%.

Even after WHT is performed, further dimensionality reduction is required on the dataset. For example, N=4096, the size of the transformed dataset is 400×819. It is likely that many of the selected coefficients are highly correlated and there is scope for further compaction.

Principal components analysis (PCA) is a commonly used linear technique for dimensionality reduction. It performs a linear mapping of multidimensional data to a lower dimensional space while retaining as much as possible of the data variability. It was determined that that the first 10 components account for 99% of the total variance. Thus we project the data on the bases represented by the 10 principal components. Now the dimension of the dataset is 400×10. This dataset is used as input to the pattern classification algorithms.

Analyte Classification

Since we are dealing with a binary classification problem, support vector machines (SVMs) were used in this case also. As mentioned earlier, the dataset consists of 400 vectors. The transformed dataset is randomly permuted and partitioned into a training set of 200 vectors and test set of 200 vectors. To compensate for the small size of the dataset, *m*-fold cross-validation was used for model selection [29].

Classification Results

All simulations were run on MATLAB version 7.5. The Spider toolbox [30] was used for classification using SVMs.

Nanopore Signals

The classification performance using SVMs are shown in Table 1. The best performance was obtained using RBF kernels using a kernel width of 5. All the events were captured correctly and only one non-event was wrongly labeled as an event. The best performance using a polynomial kernel function (of order 5) is also given below.

Table 1. Classification Performance on the Test Set

Kernel Used	***Classification Performance (%)***
RBF	***97.56***
Polynomial	***95.13***

Ion-Channel Signals

The goal of the SVM is to classify input data as quickly as possible and therefore a smaller window length would be preferable. However, there has to be sufficient transition data contained in the input window in order to be able to characterize the signal. For this reason, for each scenario, three different window lengths, N=4096, 2048 and 1024, were considered. The results of our simulations are shown in Table 2. Polynomial kernels (of order 8) and RBF kernels (of width 6) were found to yield the best results.

Our results indicate that as the window length decreases, the error rate increases for all classifiers. This is due to the fact that not only are less coefficient values being used to characterize the signal, but fewer binding events are occurring giving rise to the possibility that there is not enough signal data contained in the windowed segment.

Table 2. Classification Performance on the Test Set.

Algorithm Used	**Classification Performance (%)**		
	N=1024	***N=2048***	***N=4096***
RBF	***69.0***	***74.5***	***80.5***
Polynomial	***66.5***	***72.5***	***80.0***

Conclusions

Denoising signals using DWT was demonstrated for nanopore signals. Three features extracted from the peaks that occur in the signal- peak width, peak amplitude and mean baseline current were used to detect the passage of a bead through the nanopore. Classification was carried out using SVMs with 96% accuracy.

Denoising using DWT was demonstrated for experimental data. Feature extraction and pattern classification for discriminating between two highly similar analytes was carried out for ion-channel signals. Two-stage feature extraction using WHT and PCA provided feature vectors that could be used for classification using the four algorithms. Classification accuracy is at the 80th percentile for a frame length, N=4096. We plan to improve the accuracy of the classifiers using real data generated from experiments.

References

1. W. H. Coulter, "Means for Counting Particles Suspended in a Fluid," U.S. Patent Number 2656508, 1953.
2. R. R. Henriquez, T. Ito, L. Sun, and R. M. Crooks, "The resurgence of Coulter counting for analyzing nanoscale objects," *Analyst*, vol. 129, pp. 478-482, 2004.
3. J. B. Heng, C. Ho, T. Kim, R. Timp, A. Aksimentiev, Y. V. Grinkova, S. Sligar, K. Schulten, and G. Timp, "Sizing DNA using a nanometer-diameter pore," *Biophysical Journal,* vol. 87, pp. 2905-2911, Oct 2004.
4. J. B. Heng, A. Aksimentiev, C. Ho, P. Marks, Y. V. Grinkova, S. Sligar, K. Schulten, and G. Timp, "Stretching DNA using the electric field in a synthetic nanopore," *Nano Letters*, vol. 5, pp. 1883-1888, Oct 2005.
5. A. P. Han, G. Schurmann, G. Mondin, R. A. Bitterli, N. G. Hegelbach, N. F. de Rooij, and U. Staufer, "Sensing protein molecules using nanofabricated pores," *Applied Physics Letters,* vol. 88, 2006.

6. R. W. Deblois, C. P. Bean, and R. K. A. Wesley, "Electrokinetic Measurements with Submicron Particles and Pores by Resistive Pulse Technique," *Journal of Colloid and Interface Science,* vol. 61, pp. 323-335, 1977.
7. A. Mara, Z. Siwy, C. Trautmann, J. Wan, and F. Kamme, "An asymmetric polymer nanopore for single molecule detection," *Nano Letters,* vol. 4, pp. 497-501, Mar 2004.
8. T. Ito, L. Sun, R. R. Henriquez, and R. M. Crooks, "A Carbon Nanotube-Based Coulter Nanoparticle Counter", *Acc. Chem. Res*, vol. 37, no. 2, pp. 937-945, 2004.
9. L. Petrossian, "*Cylindrical Solid State Nanopores*", Ph. D Thesis, Arizona State University, Tempe, AZ -85287.
10. H. Bayley and C. R. Martin, "Resistive-pulse sensing-From microbes to molecules," *Chemical Rev.*, vol. 100, pp. 2575-2594, 2000.
11. O. Braha, L. Q. Gu, L. Zhou, X. F. Lu, S. Cheley, and H. Bayley, "Simultaneous stochastic sensing of divalent metal ions," *Nature Biotechnology*, vol. 18, pp. 1005-1007, 2000.
12. L. Q. Gu, O. Braha, S. Conlan, S. Cheley, and H. Bayley, "Stochastic sensing of organic analytes by a pore-forming protein Containing a molecular adapter," *Nature*, vol. 398, pp. 686-690, 1999.
13. T. Luchian, S. H. Shin, and H. Bayley, "Single-molecule covalent chemistry with spatially separated reactants," *Angewandte Chemie-International Edition*, vol. 42, pp. 3766-3771, 2003.
14. S. J. Wilk, M. Goryll, G. M. Laws, S. M. Goodnick, T. J. Thornton, M. Saraniti, J. Tang and R. S. Eisenberg., Teflon (TM)-coated silicon apertures for supported lipid bilayer membranes. *Appl. Phys Lett*, vol. 85, no.15, pp. 3307-3309, 2004.
15. O. B. McManus, A. L. Blatz, and K. L. Magleby, "Sampling, log binning, fitting, and plotting durations of open and shut intervals from single channels and the effects of noise," *Pflugers Arch.,* vol. 410, pp. 530-553, 1987.
16. F. J. Sigworth and S. M. Sine, "Data transformations for improved display and fitting of single-channel dwell time histograms," *Biophysical Journal,* vol. 52, pp. 1047-1054, 1987.
17. F. G. Ball, C. J. Kerry, R. L. Ramsey, M. S. P. Sansom, and P. N. R. Usherwood, "The use of dwell time cross-correlation functions to study single-ion channel gating kinetics," *Biophysical Journal*, vol. 54, pp. 309-320, Aug. 1988.
18. L. Venkataramanan and F. J. Sigworth, "Applying Hidden Markov Models to the analysis of single ion channel activity," *Biophysical Journal*, vol. 82, pp. 1930-1942, 2002.
19. L. Venkataramanan, J. Walsh, R. Kuc, and F. Sigworth, "Identification of HMM for ion channel currents-part I: colored noise," *IEEE Trans. on Sig. Proc*, vol. 46, no.7, pp. 1901-1915, Jul. 1998.
20. F. Qin, A. Auerbach, and F. Sachs, "Hidden Markov modeling for single channel kinetics with filtering and correlated noise," *Biophysical Journal*, vol. 79, pp. 1928-1944, Oct. 2000.
21. A. Spanias, S. Goodnick, T. Thornton, S. Phillips, S. Wilk and H. Kwon, "Signal processing for silicon ion-channel sensors," *Proc. IEEE SAFE '07*, pp.1-4, 2007.

22. H .Kwon, P. Knee, A. Spanias, S. Goodnick, T. Thornton and S. Phillips, "Transform-domain features for ion-channel sensors", *Proc. IASTED SPPRA 2008,* Paper 599-104, 2008.
23. B. Konnanath, P. Knee, A. Spanias, and G. Wichern, "Classification of Ion-Channel Signals using Neural Networks", *Proc. IASTED SPPRA 2009,* Paper 643-075, 2009.
24. A. V. Jagtiani, R. Sawant, J. Carletta and J.Zhe, "Wavelet transform-based methods for denoising of Coulter counter signals", *Meas. Sci. Technol*., vol. 19, no.6, pp. 1-15, 2008.
25. C. M. Bishop, *Pattern Recognition for Machine Learning* (Springer,2006).
26. S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Englewood, 1999.
27. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, Wiley, 2001.
28. "QuB: A software package for Markov analysis of single-molecule kinetics", Available online at http://www.qub.buffalo.edu/wiki/index.php/Main_Page.
29. B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
30. "The Spider version 1.71 for MATLAB", Available online at http://www.kyb.mpg.de/bs/people/spider/main.html