# CLASSIFICATION OF ION-CHANNEL SIGNALS USING NEURAL NETWORKS

B. Konnanath, P. Knee, A. Spanias, S. Mehta, M.Goryll, S. Goodnick, T. Thornton, S. Phillips
SenSIP Center, Department of Electrical Engineering
Arizona State University
Tempe, AZ 85287-5706, USA.

**ABSTRACT**
Ion-channel sensors can be used for detection of biochemical reagents. A silicon-based ion-channel platform has been developed for stochastic sensing for molecules. In this paper, we present techniques to extract appropriate features from sensor data using a combination Walsh-Hadamard Transform and Principal Component Analysis and use neural network techniques to discriminate between the analytes.

**KEY WORDS**
Ion-channel sensors, Walsh Transforms, PCA, Neural Networks, Feature Extraction and Pattern Classification.

## 1. Introduction

Ion-channels can be engineered to act as biosensors that can detect metal ions and organic molecules such as proteins [1,2,3]. Sensing is based on the modulation of single-channel current by molecules that bind reversibly to the pore of the channel [1]. The current fluctuations generated by the binding event are different for different analytes. Potential applications of ion-channel sensing include detection of reactive molecules in pharmaceutical products, chemical weapons, pesticides and foodstuffs [4].

A simple case of the engineered pore is shown in Fig. 1. The pore is situated in a planar lipid bilayer, with a permeable salt solution coating each side. An applied potential to the pore creates a 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 current signature through the pore is indicative of the individual analyte binding events. The frequency of occurrence of the binding events may reveal the concentration of the analyte while the modulated current signature, namely the mean duration and amplitude of the modulated signal, reveals the identity of the analyte. Features of interest fall into two categories: switching and non-switching components [3,5], both of which contain information that may be crucial in detecting an agent.

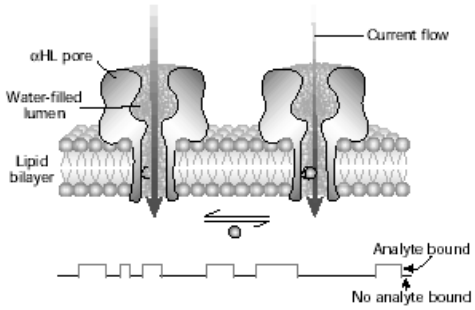


Fig.1. Single Engineered Pore. (Reproduced from [1].)

The engineered protein pores allow for single-molecule detection and give rise to the possibility of analyte classification based on the current signal [2].

The conventional modeling procedure used to classify the signal is referred to as dwell-time analysis. By analyzing the time intervals of open and closed states from ion-channel data, one is able to investigate the gating characteristics of the channel [6,7]. By setting a threshold value and determining the dwell times in the open and closed states, these values can be placed into a distribution and a Gaussian PDF can be fitted to the data using the Expectation-Maximization (EM) method [8]. Determination of the type of analyte can then be performed by examining the mean dwell-time in the closed state, which is taken to be the mean of the fitted Gaussian [9].

The difficulty in the conventional modeling approach arises when the signal-to-noise ratios (SNR) are too low. The mean-dwell time approach idealizes the data, basically separating it into two possible states, based on the threshold value, and analyzes this idealized data. However, when the SNR is too low, the inherent noise in the signal provides false information and has the potential to place the idealized data in the wrong state. Obviously, these errors in the idealized data will cause the mean dwell time to be wrong and the analyte to be classified incorrectly.

The second and most recent modeling approach uses Hidden Markov Models (HMM). The open-closed behavior of ion channels is described in terms of continuous-time Markov models in which model states

are taken to correspond to distinct states of binding [10]. Finding the best Markov model involves two steps: (a) Choosing the general topology of the model by specifying the number of states and the allowable transitions among states and (b) Optimizing the Markov model parameters. HMMs for ion-channel sensing have been further explored in [11, 12, 13].

In this paper, we present an alternative approach for classification which involves feature extraction using Walsh Transforms and Principal Component Analysis (PCA) and robust analyte classification using neural networks.

## 2. Feature Selection

Feature selection is the process of identifying relevant features from the original dataset to arrive at a reduced dimensionality representation. By using only the set of features instead of the raw data, the accuracy of the classification system can be improved and its complexity and training time is reduced. Here we present a two-stage feature extraction/dimensionality reduction approach-the Walsh-Hadamard Transform (WHT) in used in stage 1 and Principal Component Analysis (PCA) in stage 2. The proposed approach is shown in Fig. 2.

The use of Walsh-Hadamard Transform (WHT) for feature compaction of ion-channel sensor signals has been described in [14].The input signal is divided into frames of length $N$ and transformed using the WHT. A certain number of the Walsh-Hadamard transform coefficients are selected and are input to stage 2.

We use Principal Component Analysis (PCA) to carry out further dimensionality reduction in stage 2. PCA is an orthogonal projection of data from a higher dimensional space to a lower dimensional one such that the variance of the projected data is maximized [14].

PCA can be carried out using the following steps:

1) Determine the Covariance matrix of the dataset
2) Determine its eigenvectors and eigenvalues and sort them in descending order of eigenvalues.
3) Select $M$ eigenvectors corresponding to the first $M$ eigenvalues.

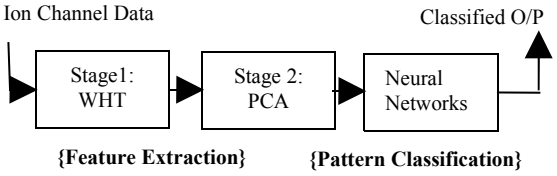


Fig.2. Feature Extraction and Pattern Classification Process

The selected eigenvectors form the M principal components. The value $M$ is chosen to be the number of eigenvectors adding up to $T$ % of the total variance of the signal where $T$ is a pre-determined threshold.

The data is projected onto the $M$ principal components and the projected data is used as inputs to the classification stage.

## 3. Analyte Classification

The following algorithms [14,15,16] are popular in special applications and can be used for signal classification from sensors[17]:

A. *Multi-Layer Perceptron (MLP)*

MLP is a supervised learning approach and is also the simplest type of Artificial Neural Network (ANN) for pattern classification. The output of the MLP $y$ is expressed as a weighed sum of the inputs $x_i$ passed through a nonlinearity $T$.

$$y = T(\sum w_i x_i) \quad (1)$$

The MLP learns the boundaries of non-linearly separable regions by minimizing a cost function (usually the mean-squared error between input and output). This is done using a gradient-descent algorithm. The output is modified at each iteration by adjusting the weights $w_i$ of the network.

B. *Linear/Quadratic Discriminant Analysis (LDA/QDA)*

The aim of LDA is to find a projection that minimizes distances within classes and maximizes the distances between classes, i.e., it seeks to maximizes the ratio of $S_b$ to $S_w$. where $S_b$ represents the *between-group* variances and $S_w$ represents the *within- group* variances.

$$S_b = \sum_{classes} P_i E[(\mu_i - \mu)(\mu_i - \mu)^T] \quad (2)$$

$$S_w = \sum_{classes} P_i E[(x_i - \mu_i)(x_i - \mu_i)^T] \quad (3)$$

C. *Radial Basis Function (RBF) Network*

The RBF Network is similar to an MLP except that it always has only 2 layers: a non-linear hidden layer and a linear output layer.

The hidden layer is made up of radial basis activation functions (usually Gaussian functions). The centers of the functions are selected randomly from the training data and the widths are proportional to the distance between the centers.

The weights of the output layer are arrived at using the equation

$$W = G^+ b \quad (4)$$

$G^+$ is the pseudo-inverse matrix of the RBF and $b$ is the desired output.

D. *Support Vector Machine (SVM)*

An SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. They are optimal in the sense that they maximize the separation between the classes hence they are known as maximum margin classifiers. The vectors that constrain the width of the margin between are the support vectors (they are encircled in the figure below).An SVM handles non-linearly separable signals by using kernel functions (E.g. Gaussian, polynomial, sigmoid) to map the data into higher dimensions where they can be separated by constructing an N-dimensional hyperplane that optimally separates the data into two categories.

## 4. Simulations and Results

Synthetic data was generated using the QUB scientific package [18]. QUB constructs Markov models to simulate single-channel kinetics. Fig. 3 shows a sample 4-state Markov model used for generating data. We construct models to simulate two analytes which are highly similar. For each model, 200 segments were created at a sampling rate of 10 kHz for a 1s duration. Thus, each segment consists of 10,000 data points for a particular model. The input data matrix is of dimension 400×10000.

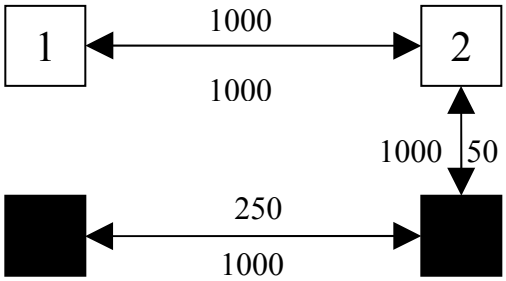


Fig. 3. Basic two class, four state model used by QUB to simulate single molecule ion channel.

The goal of the network is to classify input data as quickly as possible and thus 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 were considered, based on the number of input data samples N=4096, 2048 and 1024. Based on initial test data, feature extraction using the Walsh-Hadamard transform (WHT) required approximately 20% of the highest coefficient values in order to retain 90% of the signal energy. Thus, for each window length, the highest 20% of the transform coefficient values were retained. After applying WHT the transformed matrix is of size 400× (0.2×round (10000/N))

To further reduce the dimensionality PCA is performed on the constrained WHT dataset. It can be observed from Fig. 4 that that the first 3 components account for over 97% of the total variance. Thus, we project the data on the axes represented by the 10 principal components. Now the dataset dimension is 400×10. This dataset is used as input to the pattern classification algorithms described earlier.

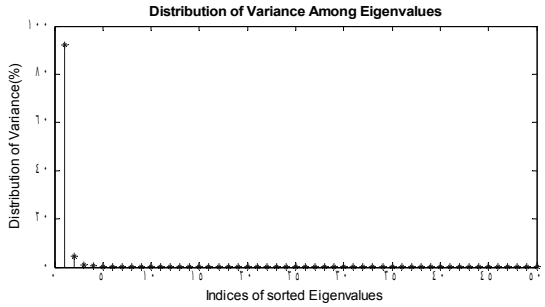


Fig.4. Distribution of Variance among the 819 eigenvalues (50 shown)

The transformed dataset is randomly permuted and partitioned into a training set of 200 vectors and test set of 200 vectors. We assign target classes to each of the vectors. The classification algorithms are trained using only the training set and performance is evaluated using the test set. Leave-*n* out procedure and *m*-fold cross-validation were used for model selection [16].

The results of our simulations are shown in Table 1.

TABLE I. CLASSIFICATION PERFORMANCE

| Algorithm Used | Classification Performance (%) | | |
|---|---|---|---|
| | *N=1024* | *N=2048* | *N=4096* |
| MLP | ***71.0*** | ***73.0*** | ***81.5*** |
| LDA/QDA | ***64.1*** | ***72*** | ***78.0*** |
| RBF | ***67.0*** | ***71.5*** | ***76.5*** |
| SVM | ***64.1*** | ***71.25*** | ***78.0*** |

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. MLPs gave the best performance for all frame lengths.

## 5. Conclusion

Feature extraction and pattern classification for discriminating between two highly similar analytes was carried out. 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 80$^{th}$ percentile for a frame length

N=4096. We plan to improve the accuracy of the classifiers using real data generated from experiments.

## Acknowledgements

The authors acknowledge NSF EXP Award 0730810, program award.

## References

[1] H. Bayley and C. R. Martin, "Resistive-pulse sensing-From microbes to molecules," *Chemical Rev.*, vol. 100, pp. 2575-2594, 2000.

[2] 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.

[3] 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.

[4] 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.

[5] 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*, 85(15): p. 3307-3309, 2004.

[6] 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.

[7] 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.

[8] T. K. Moon, "The Expectation-Maximization Algorithm", *Sig Proc.,* vol.13 (6), pp. 47-60,1996

[9] 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.

[10] 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.

[11] 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(7), pp. 1901-1915, Jul. 1998.

[12] 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.

[13] A. Spanias, S. Goodnick, T. Thornton, S. Phillips, S. Wilk and H. Kwon, "Signal Processing for Silicon Ion-Channel Sensors," *IEEE SAFE '07*, pp.1-4, 2007.

[14] H .Kwon, P. Knee, A. Spanias, S. Goodnick, T. Thornton and S. Phillips, "Transform-Domain Features for Ion-Channel Sensors", *IASTED SPPRA 2008,* Paper 599-104, 2008.

[14] C. M. Bishop, *Pattern Recognition for Machine Learning* (Springer,2006).

[15] R. O. Duda, P .E .Hart and D. G. Stork, *Pattern Classification* (Wiley, 2001).

[16] B. D. Ripley, *Pattern Recognition and Neural Networks* (Cambridge University Press, 1996).

[17] B. Konnanath, H. Kim, A. Spanias,B. Bakkaloglu, J. Wang, A. Mulchandani and N. Myung, "A Real-Time Monitoring System for Diesel and Gasoline Exhaust Exposure", submitted to *IEEE ISCAS-2009.*

[18] F. Qin. (Aug. 2001). QuB Question and Answer. http://www.qub.buffalo.edu/q_and_a_dtl.php?q=997483721&a= 997748872