Robust PSD Features for Ion-Channel Signals

P. Sattigeri, J. Thiagarajan, K. Ramamurthy, A. Spanias, M. Goryll and T. Thornton

SenSIP and CSSER Centers, School of ECEE, Arizona State University, Tempe, AZ-85287, USA.

{psattige,jjayaram,knatesan,spanias,mgoryll,t.thornton}@asu.edu

***Abstract*—Ion-channel sensors which mimic naturally occurring pore-forming proteins can be used to detect small metal ions and organic molecules. A chamber with a lipid bilayer hosting ion-channels produced by protein insertion constitutes such a sensor. Each analyte produces a characteristic signal pattern during its migration from one section of the chamber to another through the ion-channels. A four chamber ion-channel sensor array is built for accurate analyte detection. The power distribution information in the transform domain has been successfully used as discriminatory features for each chamber signal. However, these features are not robust to noise and hence result in a reduced classification performance. In this paper, we pose the stabilization of PSD features extracted from noisy segments as a matrix completion problem. Matrix completion with a low rank assumption provides the stabilized features. We demonstrate using a synthetic experiment that the proposed setup achieves improved classification performance in comparison to using the features directly. Furthermore, performing analyte detection in real ion-channel data, using the proposed robust features, provides reduction in false alarm rates.**

***Index Terms*—Ion-channel, Analyte detection, Matrix completion.**

I. Introduction

Pore forming proteins when inserted into a membrane made of lipid bilayer produce ion-channels. These pores or channels allow selective transport of analytes (certain ions and organic molecules) across the membrane. The ion-channels formed have a switching behavior [1]. The current across the membrane containing the ion-channels is measured using patch-clamp techniques. This current is called ion-channel signal. The closed state of the ion-channel leads to lower conductance and open state leads to higher conductance across the membrane. A baseline current is associated to the closed state and any increase in this current is associated with the opening of the channel. The kinetic rates of switching from open to close and vice versa are the characteristic of the protein that forms the pores. The interaction between the ion-channels and the analytes alter the switching pattern of the ion-channel signal [2]. Signal processing methods have been developed to extract such patterns in the signals and identify the analytes using neural networks [3] [4].

In our previous work [5], we analyzed the power distribution characteristics of ion-channel signals in the Fourier, Wavelet and Walsh-Hadamard domains. It was shown that the power distribution features, extracted from the frequency/sequency domains, can effectively discriminate different ion-channel signals. Here, we considered only signals generated using a single channel formed and operational in the membrane. The features were presented to a Support Vector Machine (SVM) classifier and the proposed setup achieved good classification rates under low noise conditions. Fourier domain features provided the best results among the three features proposed. For signals corrupted with noise, signal denoising was performed prior to feature extraction.

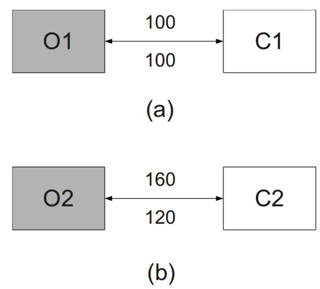

Fig. 1. Example state models used to generate ion-channel signals in QUB. Model (a) is used to generate a signal in the absence of an analyte and model (b) in the presence of an analyte with rate constants (100,100) and (160,120) respectively.

In ion-channel experiments, the number of channels formed is not controllable and hence it is typically unknown. In [6], we employed the PSD features to obtain an estimate of the number of channels operational using Support Vector Regression (SVR). It was shown that the number of active channels correspondingly alters the energy of the PSD features. Hence, a normalization of PSD features was performed by dividing the PSD features with the number of channels operational. The normalized PSD features of similar ion-channel signals (same state model) with different number of channels active were demonstrated to be close based on a weighted Euclidean distance.

In this paper, we propose to build robust power spectral density (PSD) features for ion-channel signals using matrix completion algorithms. Exact and noisy matrix completion algorithms under low rank conditions have been proposed in [7]. The feature vectors of the noisy signal segments are extracted and stacked into a matrix. Under noiseless conditions this matrix is typically low rank, since the feature vectors of consecutive segments are similar. However, the presence of noise in the acquired data does not guarantee the low rank behavior of the feature matrix. Hence, the entries of the matrix with high variances are removed to build an incomplete matrix. We now perform matrix completion under a low rank condition and the columns of the completed matrix contain the robust PSD feature vectors. Simulation results obtained

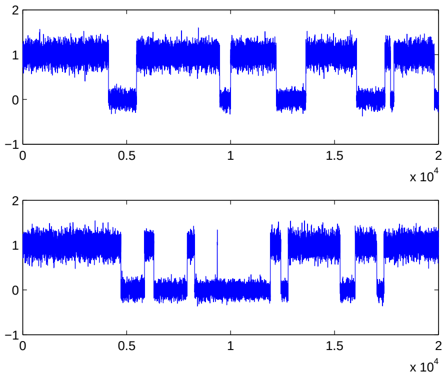

Fig. 2. Synthetic data generated from QUB for the two state models. The x-axis denotes the sample index and the y-axis the signal amplitude.

with synthetic single ion-channel data show that the stabilized features achieve improved classification performance in comparison to using the features extracted from the denoised signals. Furthermore, we demonstrate the effectiveness of the proposed robust features in reducing the false alarm rates when applied to analyte detection.

The rest of the paper is organized as follows. In Section II, we discuss the mathematical background of ion-channel signals and the PSD feature extraction procedure. The generation of synthetic data from QUB and the experimental data from the four chamber setup are also described in this section. The proposed framework that employs matrix completion to obtain robust PSD features is described in section III. Section IV presents the experimental results for classification and analyte detection. Section V concludes the paper.

II. Ion-channel signals

The ion-channel signals can be generated using a state model. The number of states and the switching rate parameters characterize the model. We use the QUB software package to generate synthetic ion-channel signals [8]. It allows the generation of signals with custom state models and multiple active channels. Two signals with a single channel and different state models are generated. Figures 1(a) and 1b show the models used to generate signals in the absence and the presence of an analyte respectively.

A. Power Spectral Density as Discriminatory Features

Any ion-channel signal $\mathbf{x}$ can be considered as a realization of a stationary Markov process with state transition matrix $\mathbf{A}$ under zero noise conditions [5]. The state transition probabilities are given by $p_{ij} = Pr(\tilde{x}_{t+1} = g(j)|\tilde{x}_t = g(i))$ where $i \in S$ and $j \in S$. S is the state space and $|S|$ gives the number of states. $p_{i,j}$ is the (i,j)th entry of $\mathbf{A}$ and $g(.)$ is the invertible map from the state space to random variable $\tilde{x}_t$. The stationary distribution of the process is given by $\pi_i = Pr(\tilde{x} = g(i))$. The PSD can be expressed as

$$F(z) = \mathbf{s}^T \mathbf{P}_\pi \mathbf{U} \hat{\mathbf{\Gamma}}(z) \mathbf{U}^{-1} \mathbf{s} \quad (1)$$

where $\mathbf{U\Gamma U}^{-1}$ is the eigen decomposition of $\mathbf{A}$ with eigen values $\{\gamma_k\}_{k=1}^{|S|}$. $\mathbf{s}$ is a vector with elements $g(i)$ and $\mathbf{P}_\pi$ is a diagonal matrix with π_i as the diagonal elements. $\hat{\mathbf{\Gamma}}(z)$ is a diagonal matrix with the (i,i)th entry as $(1-\gamma_i^2)/[(z-\gamma_i)(z^{-1}-\gamma_i)]$. Equation (1) shows that the average Power Spectral Density (PSD) of an ion-channel signal is only dependent on the eigen decomposition of the state transition matrix. Hence, the Fourier power spectrum contains sufficient discriminatory information to classify different ion-channel signals [5].

The PSD of an ion-channel signals exhibits low-pass characteristics and the corner frequency represents the average opening time of the channel [9]. The PSD is estimated using the Welch procedure given in [10]. We denote f_s as the sampling frequency and f_c as the frequency where the flat and sloping portions of the PSD intersect. $1/f_c$ represents the average opening time of the channel [11]. The DC value of the PSD is neglected and the PSD is divided into bins spaced in powers of two. The PSD values in each bin are summed and finally normalized by the total signal power, which results in the feature vector. Each bin represents the frequency range from $f_s/2^{l+1}$ to $f_s/2^l$ and the center frequency of the bin is given by $3f_s/2^{l+2}$, where $l = \{1, ..., L\}$ is the index of the bin.

B. Experimental Setup

Ion-channels produced from the outer membrane proteins of Escherichia coli can be used for analyte detection. These channels change their stochastic switching behavior in the presence of antibiotics such as Ampicillin [12].

The lipid bilayers were formed across apertures in silicon chips. These apertures were formed using photo-lithography and dry reactive ion etching. The channels have to be embedded in a lipid bilayer membrane as the membrane itself does not allow ions to penetrate. Details of the micro-fabrication process have been described in [13]. Samples were mounted in a custom-designed acrylic holder that allows vertical mounting of four silicon chips, providing free access to both sides of the individual chips. Lipid bilayers were formed using the bubble collapse (painting) method from a mixture of (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine and 1,2-dioleoyl-sn-glycero-3-phosphocholine) (DOPE:DOPC, 4:1) lipids, dissolved in n-decane (10 mg/ml). OmpF ion-channels were reconstituted into these membranes by adding 0.5 μl of OmpF stock solution to the cis compartment. We used four identical chips and compared the signals of two neighboring wells, one without Ampicillin added and the other with an Ampicillin concentration of 2.5 μM. The amplified signal was digitized using a National Instruments PCI-E 6021 DAQ card at a sample rate of 1 kHz. WinEDR [14] was used to acquire the signal as well as apply the stimulus voltage of 200 mV to the membrane containing the ion-channels. Figure 3 shows the plot of the data from two adjoining chambers one containing Ampicillin and the other acting as the base signal.

III. Generating Robust PSD features using

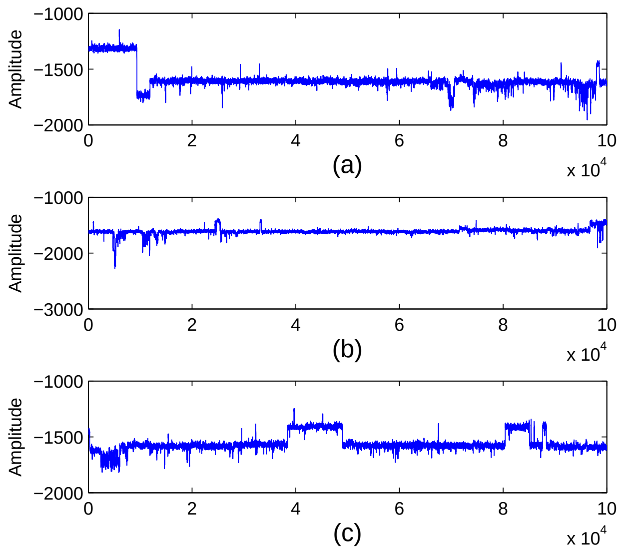

Fig. 3. Experimental data obtained from the sensor array with varying baseline current and noise levels. The x-axis denotes the sample index and the y-axis the signal amplitude.

Matrix Completion

In the problem of matrix completion, the missing entries of a matrix are inferred using a few observed entries, under some constraints. Assuming the matrix to be completed is of low rank and the observed entries are sampled from uniformly random locations in the matrix, exact recovery of the matrix is possible [7]. We pose the problem of stabilization of PSD features as a matrix completion problem. Stabilization here means that we eliminate the outliers in the PSD features make them robust.

Consider a matrix $\mathbf{M} \in \mathbb{R}^{n_1 x n_2}$ with missing entries. The indices of the observed entries $(i, j) \in \Omega$ where Ω is a subset of the cross-product set $\{1, \dots, n_1\} \times \{1, \dots, n_2\}$. The sampling operator P_Ω applied to a matrix $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$ is given by

$$[P_\Omega(\mathbf{Y})]_{i,j} = \left\{ \begin{array}{cc} Y_{i,j} & (i,j) \in \Omega \\ 0 & \text{otherwise} \end{array} \right\} \tag{2}$$

A unique low rank matrix $\mathbf{Y}$ consistent with the observed entries of $\mathbf{M}$ exists when the singular vectors of the latter matrix obeys certain conditions. Such a matrix $\mathbf{Y}$ can be obtained by solving the following optimization problem.

$$\begin{aligned} &\text{minimize } \operatorname{rank}(\mathbf{Y}) \\ &\text{subject to } P_\Omega(\mathbf{Y}) = P_\Omega(\mathbf{M}) \end{aligned} \tag{3}$$

The conditions on the singular vectors of M are expressed as

$$\|u_k\|_{l_\infty} \leq \sqrt{\mu_B/n_1}, \|v_k\|_{l_\infty} \leq \sqrt{\mu_B/n_2} \tag{4}$$

where $k \in [r]$, r is the rank of the matrix $\mathbf{M}$, u_k and v_k are singular vectors of matrix $\mathbf{M}$ obtained using singular value decomposition (SVD). When μ_B is small the singular values are well spread and are not spiky.

The rank minimization problem in (3) is non-convex and NP-hard. The rank can be replaced by the *nuclear norm* defined as the sum of the singular values of the matrix. It has been shown that this is the tightest convex relaxation to the rank minimization problem [7]. The relaxed problem is given by

$$\begin{aligned} &\text{minimize } \|\mathbf{X}\|_* \\ &\text{subject to } P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}) \end{aligned} \tag{5}$$

where $\|\mathbf{X}\|_* = \sum_k \sigma_k$ is the *nuclear norm* of the matrix $\mathbf{X}$.

Let assume the vectors $\mathbf{b}_1, \dots, \mathbf{b}_N$ are PSD features of N consecutive frames of an ion-channel signal. Note that each element in the feature vector corresponds to average PSD over a certain bin. These vectors are stacked column wise into a matrix $\mathbf{B}$. Ideally, this matrix should be low rank as consecutive frames are realization of the same Markov process and should have similar feature vectors. In order to identify the outlier feature samples and correct them, we assume every bin of the feature vector is a realization of an independent Gaussian random variable. In other words, each row of the matrix $\mathbf{B}$ contains realizations of a Gaussian random variable. Due to various types of noise, we may get some outliers in each bin. The outliers are identified by computing the variance of the entries in each column of $\mathbf{B}$ and identifying the samples whose values are more than an empirically decided threshold. We denote this incomplete matrix by $\mathbf{M}$. In cases where the entire column corresponding to frame has high variance, the column is removed altogether as this feature vector is not useful for classification. Furthermore, matrix completion algorithms cannot handle such scenarios.

Several algorithms have been proposed to solve the matrix completion problem efficiently. Few of the well known methods are Singular Value Thresholding (SVT) [15], Augmented Lagrange Multiplier (ALM) Method [16] and OptSpace [17]. We will now briefly describe the SVT algorithm which we use in this work. SVT algorithm iterates the following steps till stopping criterion is achieved.

$$\begin{aligned} \mathbf{Y}^k &= D_\tau(\mathbf{G}^{k-1}) \\ \mathbf{G}^k &= \mathbf{G}^{k-1} + \delta P_\Omega(\mathbf{M} - \mathbf{Y}^k) \end{aligned} \tag{6}$$

where $D_\tau(.)$ is the shrinkage operator which retains the singular values greater than τ of the argument matrix. Thus, the rank of $D_\tau(\mathbf{G})$ is considerably lower than that of matrix $\mathbf{G}$ if many of the singular values of $\mathbf{G}$ fall below τ. Further algorithmic and implementation details of SVT are given in [15]. The parameters τ and δ were experimentally set to 26 and 1.4 respectively. The stopping criterion was set to $1e-04$.

IV. Results

A. Classification

In order to evaluate the performance of the proposed robust features in ion-channel signal classification, we use the setup described in [5]. The signals generated using the models

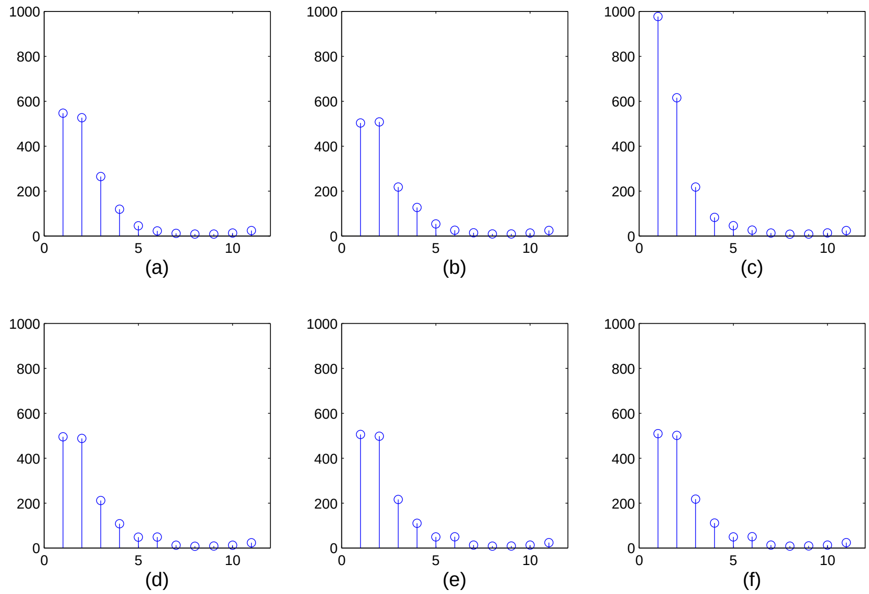

Fig. 4. Robust PSD features. (a)-(c) are the original PSD features for three segments of the data, while (d)-(f) are the corresponding stabilized PSD features. The x-axis denotes the frequency bins and y-axis shows the average power spectral density.

shown in Fig. 1, in the presence and the absence of the analyte, are divided into segments of $16,384$ samples. The dataset was randomly permuted to obtain a training set with 30 vectors and a test set with 30 vectors for each ion-channel. The Fourier domain PSD features and the proposed robust features are extracted and presented to a linear SVM for classification. Note that, the signals are denoised [5] prior to extracting the features. Table I shows the classification rates obtained with the original PSD features and the robust PSD features. Sensitivity and specificity measure the proportion of the correctly identified positives and negatives respectively. It can be clearly observed that the post processing of the PSD features leads to improved classification rates.

TABLE I
CLASSIFICATION PERFORMANCE USING THE ORIGINAL AND STABILIZED PSD FEATURES FOR QUB SIGNALS(LINEAR KERNEL).

Transform Domain	% Classification	% Sensitivity	% Specificity
Original	92.1	91.4	92.8
Stabilized	96.6	98.2	95.1

TABLE II
FALSE HITS PERCENTAGE IN DETECTION

Features	Percentage
Original	13.33
Stabilized	1.67

B. Analyte Detection

The classification setup described in the previous section cannot be directly used for analyte detection. The number of channels inserted in the lipid bilayer varies between experiments and training a classifier for all the possible cases is not possible. To overcome this problem, we proposed to use an array of ion-channel sensors [6] and detect the analyte (Ampicillin) by tracking the relative changes in PSD features among the sensors.

In the four chamber ion-channel sensor array described in Section II-B, each chamber holds an ion-channel sensor. Three of the chambers act as the base signals and the other chamber is used as the test signal in which the analyte is introduced. The change in the signal generated in the test chamber can be attributed to: (a) the change in the number of channels inserted in the lipid bilayer and (b) the change in the driving state model due to the presence of an analyte. Support Vector Regression (SVR) is used to estimate the number of channels inserted. Similar to the procedure in [6], the robust PSD features are normalized using the estimate of the number of channels. The PSD features are compared across all chambers using a weighted Euclidean distance (WED) measure. A larger distance measure indicates the presence of the analyte.

Similar to the classification setup, we extract both PSD features and the robust features from each of the signal segments. A detection hit is defined as the case when the WED goes above an empirically obtained threshold. A signal segment corrupted with noise can produce a high WED even when the analyte is absent. Such cases are referred to as false hits. Table II shows the percentage of false hits obtained using the original PSD features and the stabilized features.

V. Conclusions

We proposed a method to stabilize the PSD features for ion-channels using matrix completion. The performance of the robust PSD features were tested in a classification setup and a regression based analyte detection framework. The proposed features achieved better classification rates on the synthetic two class QUB data. These features were also used to analyze the signals obtained from the four chamber ion-channel sensor array device and detect Amplicillin. Lower false detection rates were observed in the case of the stabilized PSD features.

VI. Acknowledgment

The authors acknowledge the NSF EXP award 0730810 and the SenSIP center.

References

[1] B. Hille, *Ionic Channels of Excitable Membranes*, Sinauer Associates Inc., 1991.

[2] A. Spanias, S. Goodnick, T. Thornton, S. Philips, S. Wilk, and H. Kwon, "Signal processing for silicon ion-channel sensors," in *IEEE Workshop on Signal Processing Applications for Public Security and Forensics, Washington DC*, 2007.

[3] H. Kwon, P. Knee, A. Spanias, S. Goodnick, T. Thornton, and S. Phillips, "Transform-domain features for ion-channel sensors," in *Signal Processing, Pattern Recognition, and Applications, Innsbruck*, 2008.

[4] B. Konnanath, P. Knee, A. Spanias, and G. Wichern, "Classification of ion-channel signals using neural networks," in *Proceedings of Signal Processing, Pattern Recognition and Applications, Innsbruck*, 2009.

[5] K. N. Ramamurthy, J. J. Thiagarajan, P. Sattigeri, M. Goryll, A. Spanias, T. Thornton, and S. Phillips, "Transform domain features for ion-channel signal classification," *Biomedical Signal Processing and Control*, vol. In Press, Corrected Proof, Available online 29 October 2010.

[6] P. Sattigeri, K. N. Ramamurthy, J. J. Thiagarajan, M. Goryll, A. Spanias, and T. Thornton, "Analyte detection using an ion-channel sensor array, (accepted)," in *17th International Conference on Digital Signal Processing*, 2011.

[7] Emmanuel Cands and Benjamin Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, pp. 717–772, 2009, 10.1007/s10208-009-9045-5.

[8] "QuB software," Available online at: http://www.qub.buffalo.edu/wiki/index.php/Main_Page.

[9] Z. Y. Chen, X. P. Guo, X. L. Zhang, W. J. Cram, and Z. W. Li, "A novel method for analysis of single ion channel signal based on wavelet transform," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 559 – 562, 2007, Wavelet-based Algorithms for Medical Problems.

[10] P. Stoica and R. Moses, *Introduction to spectral analysis*, Prentice-Hall, 1997.

[11] Z. Chen, X. Guo, X. Zhang, W. Cram, and Z. Li, "A novel method for analysis of single ion channel signal based on wavelet transform," *Computers in Biology and Medicine*, vol. 37, pp. 559–562, 2007.

[12] M. Kreir, C. Farre, M. Beckler, M. George, and N. Fertig, "Rapid screening of membrane protein activity: electrophysiological analysis of ompf reconstituted in proteoliposomes," *Lab Chip*, vol. 8, pp. 587–595, 2008.

[13] S. J. Wilk, L. Petrossian, M. Goryll, T. J. Thornton, S. M. Goodnick, J. M. Tang, and R. S. Eisenberg, "Integrated electrodes on a silicon based ion channel measurement platform," *Biosensors and Bioelectronics*, vol. 23, no. 2, pp. 183 – 190, 2007.

[14] J. Dempster, "A new version of the strathclyde electrophysiology software package running within the windows environment," *Journal of Physiology-London*, vol. 504P, pp. 57, 1997.

[15] J.-F. Cai, E. J. Candes, and Z. Shen, "A Singular Value Thresholding Algorithm for Matrix Completion," *ArXiv e-prints*, Oct. 2008.

[16] Z. Lin, M. Chen, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," *ArXiv e-prints*, Sept. 2010.

[17] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh, "Matrix completion from noisy entries," *J. Mach. Learn. Res.*, vol. 99, pp. 2057–2078, August 2010.