DATA FUSION IN WIRELESS SENSOR ARRAY NETWORKS WITH SIGNAL AND NOISE CORRELATION MISMATCH

Karim G. Oweiss

Electrical & Computer Engineering Department
Michigan State University
koweiss@msu.edu

ABSTRACT

Data fusion from closely spaced, non-uniformly distributed sensor arrays with large sensor count requires a *smart* mechanism for minimizing information redundancy while maintaining high signal fidelity. In this work, we investigate the detection of multiple signal sources impinging on a non-uniformly spaced array of sensors when the additive noise has local spatial correlation characteristics. In particular, we show that if the signal spatial correlation length does not match the noise correlation length, significant performance loss occurs because of improper subarray selection. We propose to formulate the Log-Likelihood Function (LLF) in the multiresolution domain by spatially whitening the wavelet-transformed observations followed by local universal thresholding. The LLF obtained is independent of the noise covariance term, thereby yielding significant improvement over classical time domain LLF tests. Performance comparison with data averaging and decision fusion detectors is carried out to illustrate the potential advantages of the proposed method.

1. INTRODUCTION

Collaborative signal processing with multiple, spatially diverse sensor arrays has drawn much attention in recent years. Attributed mainly to advances in wireless communication, lower sensor cost, long-term sensor powering techniques, and the increased computational capability, much work has focused on distributed detection [1][2] and classification [3][4]. While classical array signal processing assumes that sensors communicate their acquired data to a *fusion center* to perform optimal detection and classification, distributed detection and classification is more powerful in cases where the information content of the phenomena being monitored is widespread [4]. Typically, there is a limited channel available to the sensor arrays for communication, making efficient bandwidth utilization critical. Conversely, dense sensor arrays offer superior spatial localization of transient events of interest, which leads to considerable system redundancy. The need to strip this redundancy requires certain level of *smartness* on the sensor side to minimize resource exhaustion, by requiring some data pre-processing to be carried out prior to communicating local decisions to the fusion center.

Among the numerous applications where collaborative, distributed signal processing maybe advantageous over centralized processing, we consider in this work a novel application, namely, neurophysiological recording of neural activity in the brain using an implantable microprobe array of electrodes [5]. In recent years, developments of microimplant fabrication technology for recording neural activity in the brain have been able to integrate large numbers of electrode sites on a single probe (typically in the hundreds [6]) to provide neuroscientists with implantable tools for simultaneous recording and stimulation of large populations of neural cells to better understand, model and restore neural function and connectivity [7]. A key issue in the development of these devices is the ability to equip them with sophisticated communication and signal processing capabilities to optimize the data transfer from the brain to the outside world. Due to limitations on the device size and resources to maintain a wireless communication link, and the nonstationarity in the neural signal environment [8], sensor array selectivity is strongly sought to maximize the information yield, increase reliability and minimize the bandwidth, without compromising the accuracy of detection and classification tasks. Motivated by the need to carry out computationally efficient algorithms for real-time functionality of wireless neural microimplants, and the unique characteristics of the neural signal environment, spatio-temporal characteristics of neural data were investigated in a novel multiresolution analysis framework [9][10].

Previous work on wireless sensor networks has focused on optimizing the decision rules for efficient distributed detection mechanisms with minimal use of resources, but were rather based on idealized assumptions, such as white noise, uncorrelated sources, etc…[2]-[3]. In this work, we focus on the distributed detection of multiple sources taking into account the unique characteristics of the neural medium. In particular, we consider the case where the underlying noise process has some local spatial correlation properties. This scenario typically occurs when the sensors are in close proximity to each other - a design requirement to provide adequate spatial sampling close to the Nyquist rate for optimal spatio-temporal source compression.

2. THEORY

2.1 Mathematical Preliminaries

It is assumed that P signal sources impinge on an array of M channels within the discrete time interval $t_1,\ldots,t_N$. Each source is characterized by a real valued zero-mean Gaussian distribution of the form $\boldsymbol{s}_p \sim \mathcal{N}(0,\Sigma_p)$, where $\Sigma_p \in \Re^{N\times N}$ denotes the temporal covariance of the p^{th} source. The sources are assumed to have equal temporal covariances Σ_p's, $p = 1,\ldots,P$, and banded spatial covariance $\boldsymbol{\Sigma}_S \in \Re^{P\times P}$ [11]. The additive noise model yields

$$\boldsymbol{y}[n] = \boldsymbol{x}[n] + \boldsymbol{z}[n] \qquad n = 1,\ldots,N \tag{1}$$

where $\boldsymbol{y} \in \Re^{Mx1}$ denotes the n^{th} snapshot of the array, $\boldsymbol{x} \in \Re^{Mx1}$ obeys the model $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}$, where $\boldsymbol{s} \in \Re^{P\times 1}$ denotes the signal vector, $\boldsymbol{A} \in \Re^{M \times P}$ denotes the mixing matrix whose columns express the array response to each source ($\boldsymbol{A}$ is assumed to be known), and $\boldsymbol{z} \in \Re^{Mx1}$ denotes a zero-mean additive Gaussian noise component with banded spatial covariance $\boldsymbol{\Sigma}_z$. The additive noise component is assumed temporally white, i.e., $z \sim \mathcal{N}(0, \boldsymbol{I})$.

Due to the limitations imposed by the computational complexity of the associated signal processing circuitry, only a group of sensors can be allocated to each signal processor to enable real-time functionality [13]. It is desired to split the M sensor array measurements into K subarrays, with m_k channels in each subarray such that $M = \sum_{k=1}^{K} m_k$. This model allows a variable number of sensors in each subarray, consistent with the non-uniform spatial sensor distribution assumption. Each subarray will be associated with a computational *node* that performs *local* processing tasks relevant to the signal activity 'sensed' by its m_k channels. Under this model, the goal is to select K and m_k such that the signal components in the m_k channels in a particular subarray are maximally correlated, but varies independently from subarray to another to minimize the system redundancy.

2.2 Optimal Subarray Selection

Each subarray does not "know" the statistics of the sources impinging on other subarrays, i.e. they do not "talk" to each other. This structural distribution yields a parallel topology with data fusion within nodes and decision fusion across nodes [1]. Within the measurement interval in subarray *k*, the following is assumed: P_k sources are present, sources can overlap between subarrays, i.e., $P \le \sum_{k=1}^{K} P_k$ (equality when sources do not overlap between subarrays and all sources are *heard* by at least one sensor), and the noise is correlated within each subarray, but is uncorrelated across subarrays. These assumptions imply that the signal and noise spatial covariances $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_z$ will have distinct banded structures and accordingly different correlation lengths.

Without loss of generality, let's consider the k^{th} subarray model under the composite hypothesis

$$H_{P_k}: \quad \boldsymbol{y}_k[n] = \boldsymbol{A}_k \boldsymbol{s}_k[n] + \boldsymbol{z}_k[n] \tag{2}$$

where $\boldsymbol{y}_k \in \Re^{m_k \times 1}$, $\boldsymbol{A}_k \in \Re^{m_k \times P_k}$ (a submatrix of $\boldsymbol{A}$), and $\boldsymbol{s}_k \in \Re^{P_k \times 1}$. A Maximum Likelihood (ML) detector/classifier is selected to map the observations to one of the *apriori* known mixture of P_k sources. The optimal detector will attempt to minimize information redundancy by simultaneously processing channels with maximally correlated signal activity while minimizing the inherent masking noise effect [3]. An optimal data fusion rule will be obtained if the subarrays are selected to maximize the joint Log-Likelihood Function (LLF)

$$C(x) = \underset{P_k \in \{1,\ldots,P\}}{argmax} \left(\mathcal{P}_{P_k}(x)\right) \tag{3}$$

where for P_k sources within the discrete interval $t_1,\ldots, t_N$, the joint distribution is given by:

$$\mathcal{P}_{P_k}(\boldsymbol{y}_k) = \frac{1}{\pi^{P_k N} \left|\boldsymbol{\Sigma}_{z_k}\right|^N} e^{-\sum_{n=1}^{N} \boldsymbol{y}_k^T[n] \boldsymbol{\Sigma}_{z_k}^{-1} \boldsymbol{y}_k[n]} \tag{4}$$

where $\Sigma_{z_k} = E[\boldsymbol{z}_k \boldsymbol{z}_k^T]$. The LLF under this model takes the form (the index *n* is omitted for simplicity)

$$L(\boldsymbol{y}_k) = -N\left(\log\left|\boldsymbol{\Sigma}_{z_k}\right| + \tfrac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}_k^T \boldsymbol{\Sigma}_{z_k}^{-1} \boldsymbol{y}_k\right) \tag{5}$$

which comprises a linear and a quadratic term.

2.3 Data Fusion

Since the noise is locally correlated within the subarray, the optimal detector attempts to reduce the subarray noise in the second term of (5) by *whitening* the observations within the subarray through the lower triangular transformation

$$\tilde{\boldsymbol{y}}_k = \boldsymbol{D}_{z_k}^{-1/2} \boldsymbol{U}_{z_k}^T \boldsymbol{y}_k \tag{6}$$

where the $m_k \times m_k$ diagonal matrix $\boldsymbol{D}_{z_k}$ and eigenvector matrix $\boldsymbol{U}_{z_k}$ can be obtained through the EigenValue Decomposition (EVD) of the ML estimate of the noise spatial covariance matrix in subarray *k*

$$\Sigma_{z_k} \cong \tfrac{1}{N} \sum_{n=1}^{N} \boldsymbol{z}_k[n] \boldsymbol{z}_k^T[n] = \boldsymbol{U}_{z_k} . \boldsymbol{D}_{z_k} . \boldsymbol{U}_{z_k}^T \tag{7}$$

Clearly, a key factor in (7) is the ability to obtain a *good* sample estimate of Σ_{z_k} in each subarray, which is usually computed using a subset of the observations lying outside the row space of the signal matrix $\boldsymbol{S}_k \in \Re^{P_k \times N}$. Many factors can contribute to a poor estimate of Σ_{z_k}, for example: improper selectivity of the channels in subarray *k*, nonstationarity of the noise process, distinct signal and noise spatial correlation lengths, etc…. In such case, the observation covariance Σ_{kk} can be used instead

$$\Sigma_{kk} = E[\boldsymbol{y}_k \boldsymbol{y}_k^T] \cong \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}_k[n] \boldsymbol{y}_k^T[n] = \boldsymbol{U}_k . \boldsymbol{D}_k . \boldsymbol{U}_k^T \tag{8}$$

If prior knowledge of P_k is available [12], then the transformation (6) can use a truncated version of (8), by using the right most $m_k - P_k$ columns of $\boldsymbol{U}_k$ associated with the smallest $m_k - P_k$ eigenvalues in $\boldsymbol{D}_k$, which span the noise subspace. Herein, we do not assume that knowledge of P_k is available. Therefore, the *local* LLF in this case takes the form

$$\begin{aligned} L(\boldsymbol{y}_k) &= -N\left(\log\left|\boldsymbol{\Sigma}_{kk}\right| + \tfrac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}_k^T \boldsymbol{\Sigma}_{kk}{}^{-1} \boldsymbol{y}_k\right) \\ &= -N\left(\log\left|\boldsymbol{\Sigma}_{kk}\right| + \tfrac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}_k^T \boldsymbol{U}_k^T . \boldsymbol{D}_k^{-1/2} . \boldsymbol{D}_k^{-1/2} . \boldsymbol{U}_k \boldsymbol{y}_k\right) \\ &= -N\left(\log\left|\boldsymbol{A}_k \Sigma_{s_k} \boldsymbol{A}_k^T + \Sigma_{z_k}\right| + tr(\hat{\Sigma}_k)\right) \end{aligned} \tag{9}$$

where $\hat{\Sigma}_k$ denotes the covariance of $\hat{\boldsymbol{y}}_k = \boldsymbol{D}_k^{-1/2} \boldsymbol{U}_k^T \boldsymbol{y}_k$. The fusion center attempts to optimally select the subarrays by maximizing the local LLF

$$L(\boldsymbol{y}) = \underset{k \in \{1,\ldots,K\}}{\arg\min} \left\{ \log\left|\boldsymbol{A}_k \Sigma_{s_k} \boldsymbol{A}_k^T + \Sigma_{z_k}\right| + tr(\hat{\Sigma}_k) \right\} \tag{10}$$

In low SNR conditions, the presence of Σ_{z_k} in the log term causes significant performance degradation due to the relatively large noise eigenvalues $\delta^k_{P_k+1} > \ldots > \delta^k_{m_k}$. In such case, "over-grouping" of m_k channels can occur. On the other hand, "under-grouping" can occur in the presence of strong sources that encompass the whole array if the contaminating noise is independent across channels. It is desirable for the optimal detector to work under the assumption that all the measurements are correlated within subarray k to subtract-out the correlated noise and to optimally segregate the P_k signals. The ideal situation would be if the signals and noise spatial correlation lengths exactly match, so that each subarray yields optimal performance. However, this is a rarely satisfied condition because of the variations in source signal power, nonstationarity of the medium and non-uniformity of sensor array spatial sampling of the surroundings.

2.4 Wavelet Subarray Detector

The ability of the wavelet transform to form an unconditional basis for a wide variety of signals makes it intuitively attractive for our analysis [14]. We've shown that many classical array processing problems can be overcome by exploring the representation of the array data in the wavelet domain [9]-[10]. In this work, we propose to use a Discrete Wavelet Transform (DWT) representation of the array data to remedy the above-mentioned problems. Let $\boldsymbol{W}$ denote an $N \times N$ orthonormal DWT matrix up to J levels[1], so that $\boldsymbol{s}^j_p = \boldsymbol{W}\boldsymbol{s}_p, \boldsymbol{z}^j = \boldsymbol{W}\boldsymbol{z}$. The set $\{\boldsymbol{s}^j_p\}, j = 1,\ldots J+1$ will comprise the DWT of $\boldsymbol{s}_p \sim \mathcal{N}(0, \Sigma_p)$ up to J levels. Note that an orthonormal wavelet transform will map $\boldsymbol{z}$ to $\boldsymbol{z}^j$ that is likewise zero-mean Gaussian with covariance $\sigma^2_j \boldsymbol{I}$, while compacting typical signals into a small number of large wavelet coefficients $\boldsymbol{s}^j_p$. The subarray model (2) yields

$$H^j_{P_k}: \quad \boldsymbol{y}^j_k = \boldsymbol{A}_k \boldsymbol{s}^j_k + \boldsymbol{z}^j_k \tag{11}$$

where $(.)^j_k$ denotes the j^{th} subband in subarray k. The whitening matrix $\boldsymbol{U}^j_k$ can be obtained through the spectral factorization of the covariance

$$\Sigma^j_{\boldsymbol{kk}} = E[\boldsymbol{y}^j_k \boldsymbol{y}^{jT}_k] \cong \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}^j_k[n]\boldsymbol{y}^{jT}_k[n] = \boldsymbol{U}^j_k.\boldsymbol{D}^j_k.\boldsymbol{U}^T_k \tag{12}$$

yielding the whitened array wavelet expansions

$$\hat{\boldsymbol{y}}^j_k = \boldsymbol{D}^{j^{-1/2}}_k \boldsymbol{U}^{jT}_k \boldsymbol{y}^j_k \tag{13}$$

The thresholding property of the DWT allows discarding the numerous small coefficients (presumably attributed to noise), while keeping the few large coefficients (presumably attributed to the signal) through the denoising operator $\boldsymbol{H}^j_k$

$$\bar{\boldsymbol{y}}^j_k = \boldsymbol{H}^j_k\{\hat{\boldsymbol{y}}^j_k\} \tag{14}$$

Various methods for estimating the threshold exist depending on the statistical model of the signal at hand [15]. Herein, the universal hard threshold $T_j = \sigma_{jk}\sqrt{2\log N}$ was used, with σ^2_j estimated from the finest wavelet scale. The outcome of (12)-(14) is the "noise free" wavelet coefficients that are used by the test statistic

$$L^j(\boldsymbol{y}) = \underset{k\in\{1,\ldots,K\}}{\arg\min}\left\{\log\left|\boldsymbol{A}_k \Sigma^j_{s_k} \boldsymbol{A}^T_k\right| + tr(\bar{\Sigma}^j_k)\right\} \tag{15}$$

where $\bar{\Sigma}^j_k$ denotes the covariance of $\bar{\boldsymbol{y}}^j_k$. Note that besides reducing the effect of noise in the first term, it can be argued that comparing (10) and (15), the term $tr(\bar{\Sigma}^j_k)$ is much smaller than $tr(\hat{\Sigma}_k)$ because of the compactness property of the DWT (the proof is omitted for the lack of space). Since the DWT yields orthogonal wavelet expansions, then the joint distribution can be obtained by averaging the LLFs across subbands

$$L(\boldsymbol{y}) = \underset{k\in\{1,\ldots,K\}}{\arg\min}\left\{\frac{1}{J}\sum_{j=1}^{J}\left(\log\left|\boldsymbol{A}_k \Sigma^j_{s_k} \boldsymbol{A}^T_k\right| + tr(\bar{\Sigma}^j_k)\right)\right\} \tag{16}$$

which yields an overall improvement in performance.

3. PERFORMANCE

The non-linearity property inherent in the thresholding operation in (14) makes the distribution of the test statistic $L(\boldsymbol{y})$ hard to obtain. The performance was accordingly evaluated through Monte-Carlo simulations. We compared the performance with two limiting cases. The first is a *Data Averaging* (DA) detector, which assumes that measurements are maximally correlated across the array. The second is a *Decision Fusion* (DF) detector that assumes all the measurements are independent across the array.

We present results for the proposed *Data Wavelet Fusion* (DWF) detector using simulated neural data from real experiments. The data was simulated using experimental neural noise superimposed on neural signal transients (*spikes*) extracted from the same experiment [9]. Typical characteristics of the neural signal environment are illustrated in Fig. 1, where the signal amplitude rolls-off rapidly indicating diminishing signal coherence between non-adjacent channels, while background noise spatial correlation remains relatively high.

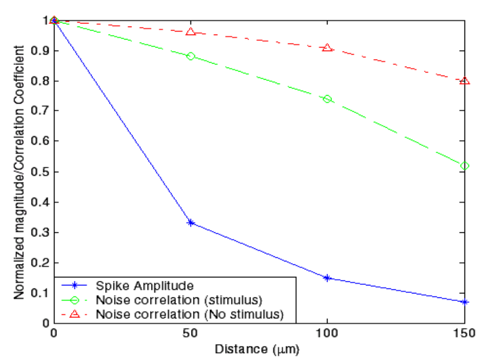

Fig.1: Neural signal amplitude roll-off and noise correlation along the array

For equal signal and noise correlation lengths, the performance of the three detectors was assessed for different SNRs (Fig.2). It is clear that under various SNR conditions, DWF always outperforms DA and DF because it exploits signal correlation without being affected by the inherent noise correlation. The performance was assessed for different signal and noise correlation

[1] For a J-level DWT binary tree, there are a total of $J+1$ terminal nodes, from which perfect signal reconstruction is feasible.

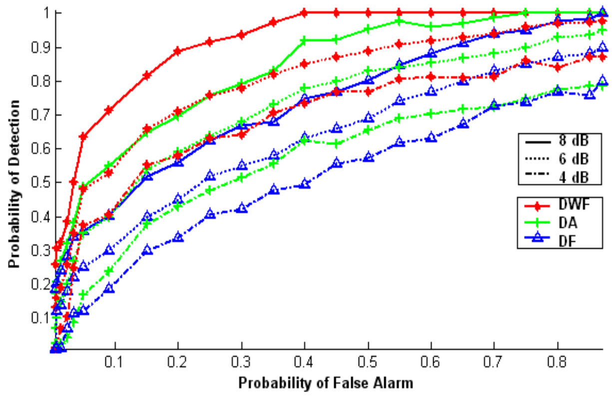

Fig.2: Performance of the DWF, DA, and DF detectors for different SNRs for equal signal and noise correlation lengths

lengths (Fig.3). A few remarks can be made: 1) DWF is almost independent of the signal and noise correlation length mismatch; 2) DA outperforms DF for higher noise correlation lengths. This last result is intuitive since data averaging within the subarray with highly correlated noise reduces noise variance thus achieving SNR gain compared to exploiting statistical independence across subarrays (DF) with shorter signal correlation length. The DF on the other hand incurs SNR loss depending on the fraction of the measurements that are correlated.

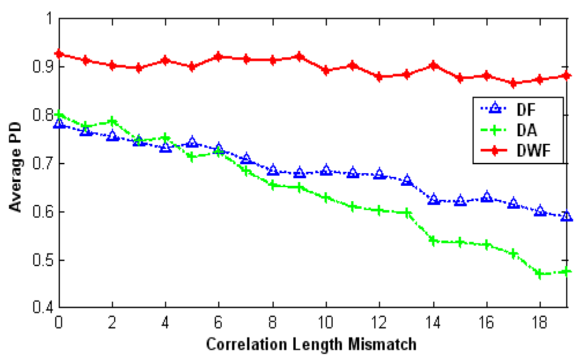

Fig.3: Performance of the three detectors versus correlation length mismatch

4. CONCLUSION

The analysis and examples presented in this work helped to gain some insight on the power of the proposed method to provide a robust and efficient scheme for selectively processing subarrays sharing a concoction of correlated and uncorrelated signal and noise processes. Signal coherence and noise spatial correlation in sensor array networks is clearly a limiting factor that has not gained considerable attention in the literature. Our emphasis mainly targeted the case where the spatial noise correlation length does not match that of the signal. To improve the performance, we've shown that the detector yields a sub-optimal solution when selecting subarrays with maximum signal correlation regardless of the noise statistics. Nevertheless, performance loss occurs if the noise is independent within subarrays especially in the presence of strong signal sources. Accordingly, spatial noise correlation cannot be ignored in non-uniformly distributed sensor networks because it can cause severe performance degradation. Our analysis was also carried out taking into account that signal sources may overlap across subarrays, which implies that a wide range of scenarios can be captured ranging from strong signal sources impinging on the whole array to weak sources that are local to specific subarrays. We proposed a novel wavelet-based array detector that was shown to yield improved performance because it is independent of the noise covariance term. Moreover, less variability of the test statistic is possible to achieve by averaging the test statistics across subbands. It is worthy to note that our framework differs significantly from spatial smoothing techniques by Evans *et al*. and Shan *et al.* in the sense that the later do not exploit noise correlation properties across sensors, but were rather focused on circumventing problems due to partial or full signal correlation (coherence) scenarios [16]. The outcome of the proposed technique will greatly help in optimizing neuroprosthetic microinterface modules.

5. REFERENCES

[1] Viswanathan R. and Varshney P. K., "Distributed detection with multiple sensors: Part I-fundamentals," *Proc. IEEE*, (85)-1:54–63, 1997.

[2] Blum R.S., Kassam S. A., and Poor H.V., "Distributed detection with multiple sensors: Part II-advanced topics," *Proc. IEEE*, (85)-1: 64–79, 1997.

[3] Ashwin D., Sayeed A., "Data vs. Decision Fusion in Wireless Sensor Networks," *Proc. ICASSP*, (IV): 832-835, 2003

[4] Kumar S., Zhao F. and Shepherd D., (*eds.*): "Special Issue on collaborative signal and information processing in microsensor networks," *IEEE SP mag.*, 17-29, March 2002

[5] Drake K.L. *et.al.* "Performance of planar multisite microprobes in recording extracellular activity," *IEEE Trans. BME*, (35): 719-732, 1988

[6] Wise K.D., "Wireless Implantable Microsystems: Coming Breakthroughs in Health Care," *Symp. on VLSI Circuits, Digest of Technical Papers*, 106–109, 2002

[7] Nicolelis M., "Actions from Thoughts, " *Nature*, (409): 403-407, Jan. 2001

[8] Fee M.S. *et. al.,* "Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability," *J. Neurosc. Meth.*, (69): 175-188, 1996

[9] Oweiss K.G. and Anderson D.J., " A Unified Framework for Advancing Array Signal Processing Technology of Multichannel Microprobe Neural Recording Devices, " *Proc. 2nd IEEE Microtech. in Med. & Bio.*, 245-250, May 2002

[10] Anderson D.J., Oweiss K.G. and Bierer S. M., "Sensor Arrays in the Micro-Environment of the Brain," *Proc. ICASSP*, 3433-3436, May 2001

[11] Chen W., Reilly J. and Wong K., "Detection of number of signals in noise with banded covariance matrices," *IEE Proc.-Radar, Sonar Navig.,* (143)-5: 289-295, Oct. 1996

[12] Oweiss K.G., "Source Detection in Correlated Multichannel Signal and Noise fields," *Proc. ICASSP*, (V): 257-260, 2003

[13] Oweiss K.G. *et al.* " Optimizing Signal Coding in Neural Interface *System-on-a-Chip* Modules," *Proc. 25th IEEE EMBS*, 2016-2019, Sept. 2003

[14] Vetterli M. and Kovacevic J., *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[15] Donoho D.L., "De-Noising by Soft Thresholding," *IEEE Trans. on IT*, (41): 613-627, 1995

[16] Shan T., Max W. and Kailath T., "On Spatial Smoothing for DOA Estimation of Coherent Signals," *IEEE Trans. On ASSP*, vol. 33, No.4, pp.806-811, Aug. 1985.