A Miniature, Low-Power, Intelligent Sensor Node for Persistent Acoustic Surveillance

Gert Cauwenberghs, Andreas Andreou, Jim West,
Milutin Stanacevic, Abdullah Celik, Pedro Julian, Thiago Teixeira
Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218

Chris Diehl
Johns Hopkins Applied Physics Laboratory, 11100 Johns Hopkins Road Laurel, MD 20723

Laurence Riddle
Signals Systems Corporation, 877 B&A Blvd Suite 210, Severna Park, MD 21146

ABSTRACT

The desire for persistent, long term surveillance and covertness places severe constraints on the power consumption of a sensor node. To achieve the desired endurance while minimizing the size of the node, it is imperative to use application-specific integrated circuits (ASICs) that deliver the required performance with maximal power efficiency while minimizing the amount of communication bandwidth needed. This paper reviews our ongoing effort to integrate several micropower devices for low-power wake-up detection, blind source separation and localization and pattern classification, and demonstrate the utility of the system in relevant surveillance applications. The capabilities of each module are presented in detail along with performance statistics measured during recent experiments.

Keywords: acoustic surveillance, sensor networks, micropower ASIC, acoustic target detection, localization and identification

1. INTRODUCTION

Intelligent, distributed sensing is critical in a variety of military reconnaissance and surveillance applications. The following three distinct, but closely related, questions are usually of interest:

Detection: Is a target present? Are there more than one? How many?

Localization: Where is it?

Identification: What is it?

While there has been great progress towards the design of sensor network systems that address these questions at least partially, major challenges remain in their realization:

Further author information: (Send correspondence to G. Cauwenberghs)
Gert Cauwenberghs: E-mail: gert@jhu.edu, Telephone: 1 410 516 5180
Andreas Andreou: E-mail: andreou@jhu.edu, Telephone: 1 410 516 8361
Jim West: E-mail: jimwest@jhu.edu, Telephone: 1 410 516 8546
Milutin Stanacevic: E-mail: miki@jhu.edu, Telephone: 1 410 516 7701
Abdullah Celik: E-mail: acelik@jhu.edu, Telephone: 1 410 516 7701
Pedro Julian: E-mail: pjulian@ieee.org, Telephone: 1 410 516 8361
Thiago Teixeira: E-mail: thiago@jhu.edu, Telephone: 1 410 516 8361
Chris Diehl: E-mail: Chris.Diehl@jhuapl.edu, Telephone: 1 410 516 5180
Laurence Riddle: E-mail: larry@signalsystemscorp.com, Telephone: 1 410 4317 148

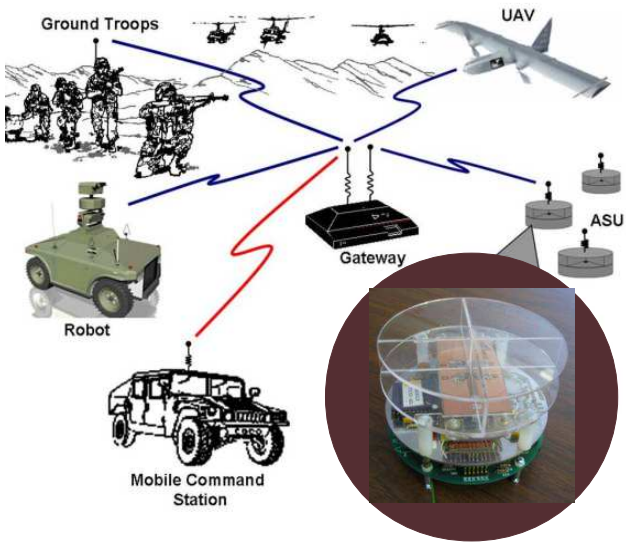


Figure 1. *Wireless Acoustic Surveillance Unit (wASU) in battlefield environment.*

1. The presence of multiple targets confounds the localization and identification of a single target of interest. Reverberation due to multipath propagation subject to environmental conditions further aggravates the mixing of target observations.

2. Intelligent ubiquitous and unobtrusive sensing of the environment requires a vast network of miniature sensing nodes, which pose stringent constraints on available power and bandwidth to process signals at the sensor level and communicate information between sensor nodes and to a base station.

These challenges call for an interdisciplinary, multi-pronged approach combining signal processing and pattern recognition algorithms, system-level power management, and sensor miniaturization and integration. The presented work addresses these challenges in the context of *acoustic target detection, localization and identification*. Acoustic sensing is a valuable, relatively low-power mechanism for cueing video sensing. Coarse localization of a target by an acoustic sensor may direct a video camera where to focus attention. While optics is prone to occlusion of targets, acoustics picks up anything present in the scene. The task of isolating targets from acoustic observations is therefore particularly challenging, requiring a significant amount of signal processing combining observations across multiple acoustic sensing nodes in the network, as illustrated in Figure 1.

2. WAKE-UP DETECTION

Sensor network nodes are subject to strict power budgets, as dictated by the need to prolong battery life. This requirement necessitates power-conscious design from high-level algorithms down to the circuit implementation.

One way to reduce power consumption is to employ a power management strategy. In the case of a surveillance application, the system may only be required to operate at full functionality in the presence of a novel object. For the remainder of the time, the system can persist in a "sleep" state, where the only required functionality is to detect the presence of the novel object. This power management scheme requires a "wake-up" front-end—a subsystem that detects the novel object and arouses the surveillance system to full functionality. It is crucial that the wake-up subsystem consumes very little power relative to the system as a whole if the power management strategy is to be effective. Finally but not the least important, the goal of the information processing in such systems is not the precise restitution of information (such as in modern multimedia communication), but rather the extraction of information in a timely manner.

Intelligent wake up detection systems are key to the success of low power acoustic identification and localization devices as they intricately relate to the task of *detection and identification*. Thus any sophisticated

architecture that will ultimate perform recognition needs a sub-system for wake-up not only for the purpose of saving power but also for reducing the amount of processing that subsequent stages have to perform.

2.1. Periodicity detection

We present a wake-up/classification system that is based on the degree of low-frequency periodicity of the sound signal, typical of sounds generated by vehicle engines. To this end, we have developed a periodicity estimation algorithm that maps particularly well to a low-power VLSI implementation. The time-domain algorithm is based on the "bumpiness" of the autocorrelation of a one-bit version of the acoustic signal. We describe a full-custom CMOS ASIC that implements the algorithm, and at its core consumes 835 nanowatt.[1]

The problem of periodicity estimation is closely related to that of pitch estimation. Pitch estimation asks the question, "What is the period of the signal", whereas periodicity estimation asks the question, "What is the degree of periodicity in the signal?" Much work has been done on pitch estimation, particularly in the context of speech recognition application.[2] A maximum likelihood approach has been developed by several authors.[3–5]

2.2. Autocorrelation alternative

While the MLE approach is the 'gold standard' and gives optimal estimation performance, it is computationally intensive, as it requires explicit computation of the period (pitch) of the signal. An alternative periodicity measure is based on the autocorrelation function (ACF), given by

$$R_{xx}[n] = \frac{1}{J}\sum_{j=0}^{J-1} x[j]x[j+n] \tag{1}$$

We will also refer to the normalized ACF:

$$\bar{R}_{xx}[n] = R_{xx}[n]/R_{xx}[0] \tag{2}$$

The ACF of a noise-like signal has low energy at higher lag values, whereas the ACF of a periodic signal has high energy at higher lag values. To quantify this, we define a range of lags of interest, $[N_{\min}, N_{\max}]$, and we sum the squares of the ACF over this range:

$$\mathcal{P}_{\text{alt,A}} = \sum_{n=N_{\min}}^{N_{\max}} (\bar{R}_{xx}[n])^2 \tag{3}$$

For practical reasons, our PM requires a modification. In reality, the signal is not stationary, enabling low-frequency fluctuations to introduce additional peaks into the ACF If the peaks corresponding to these fluctuations are far above the range of lags, they will appear as an offset in the ACF, causing the measure described above to indicate periodicity where none is present. We can eliminate this offset by first computing an approximate time derivative of the ACF, and then computing the sum of the squares. The expression for the PM becomes

$$\mathcal{P}_{\text{alt,A}} = \sum_{n=N_{\min}}^{N_{\max}-1} (\bar{R}_{xx}[n+1] - \bar{R}_{xx}[n])^2 \tag{4}$$

Because the derivative is a high-pass operation, this can be thought of as emphasizing higher-frequency components of the signal, and de-emphasizing lower-frequency components of the signal.

2.3. Simplified algorithm and architecture

The auto-correlation based alternative algorithm can be simplified a great deal and still give satisfactory results for the detection task. These simplifications have been made with a hardware implementation in mind, whether it be in an embedded processor or in a full-custom ASIC.

The signals used in the previous section were quantized with 16-bit precision. It turns out that one-bit precision is sufficient to give acceptable results for the detection task. This is because an infinitely clipped

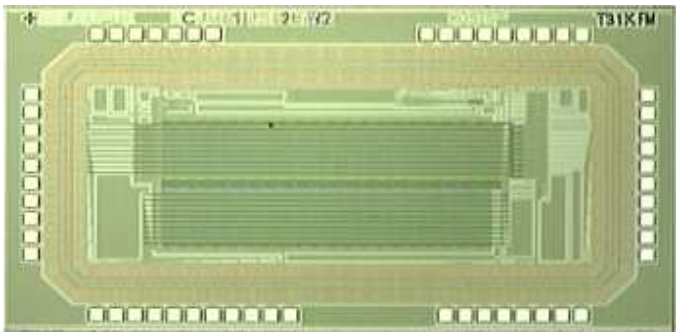


Figure 2. *Micrograph of the wake-up detector CMOS ASIC.*[1]

version of the signal retains the periodic structure of the original signal. In fact, the ACF of an infinitely clipped signal $\tilde{x}[n]$ is related to the ACF of the original signal $x[n]$ by[6]

$$R_{\tilde{x}\tilde{x}}[n] = \frac{2}{\pi}\sin^{-1}(R_{xx}[n]) \tag{5}$$

Provided that we encode the signal with zeros and ones (as opposed to -1s and $+1$s), the use of one-bit input signals greatly simplifies the hardware implementation. For example, the multiply operation in the correlation computation is reduced to an XNOR. The PM computation can be further simplified by computing the sum of the absolute value of the discrete differences rather than the sum of the squares.

The PM equations become

$$\tilde{R}_{\tilde{x}\tilde{x}}[n] = \sum_{j=0}^{J-1} \overline{x[j] \oplus x[j+n]} \tag{6}$$

$$\mathcal{P}_{\text{alt,B}} = \sum_{n=N_{\min}}^{N_{\max}} |\tilde{R}_{\tilde{x}\tilde{x}}[n+1] - \tilde{R}_{\tilde{x}\tilde{x}}[n]| \tag{7}$$

where $\oplus$ represents the XOR operation and the bar represents negation.

2.4. VLSI implementation

We designed a full-custom CMOS ASIC to implement the simplified wake-up detector algorithm described in the previous section. The chip was fabricated on a 3 mm $\times$ 1.5 mm die in a 0.5μ-process available from the MOSIS service. A micrograph of the chip is shown in Fig. 2.

The wake-up ASIC was connected to a microphone and signal conditioning circuitry (including a comparator for one-bit A/D conversion) and tested in a laboratory setting. The entire chip consumes 6.3 μW during operation. Because we use separate power supply pins for the I/O pads and the core, we can divide the total power consumption into its constituent parts. The I/O pads consume 5.5 μW and the core consumes 835 nW. The large power consumption of the pads is attributable to two factors: First, the clock input switches at 32 kHz, and is subsequently internally divided down to 1 kHz. This design decision was made in order to facilitate integration with a COTS oscillator. Secondly, approximately 40 non-essential pads were included for debugging purposes. Therefore, it is reasonable to assume that a next generation chip in the same process would consume 1 μW. This power consumption level is far smaller than that of the microphones and signal conditioning circuitry, which draw 300 μW. At these levels, the system will run for well over a year on 3 AA batteries.

3. GRADIENT FLOW SOURCE SEPARATION AND LOCALIZATION

A biologically-inspired gradient flow signal representation blindly separates and localizes targets using a miniature array of sub-wavelength aperture. Gradient flow constructs instantaneous, linearly independent signal

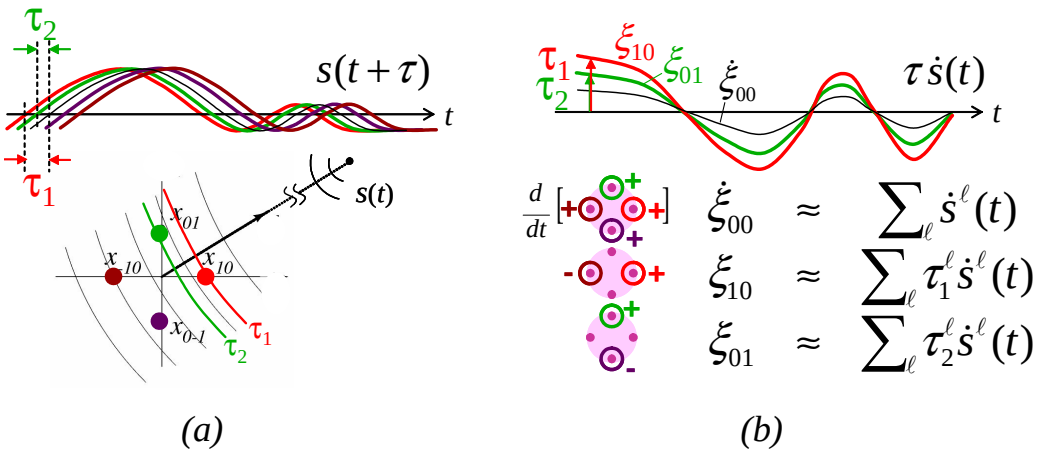


Figure 3. *Gradient flow principle.* (a) *Time-delayed observations of a wave signal* $s(t)$ *at four closely spaced sensor locations. The direction cosines of the source are determined by the time delays* τ_1 *and* τ_2*, shrinking with decreasing array aperture.* (b) *Gradient flow converts delays into relative amplitudes of the temporal derivative of the signal as observed in the spatial gradients of the wave signals over the array. As with correlation-based delay estimation, decreasing aperture causes shrinking gradient amplitudes. However, accurate localization can still be obtained for very small aperture by boosting differential gain in acquisition of the spatial gradients. In the presence of multiple sources* $s^\ell(t)$*, the gradient amplitudes combine linearly, where the coefficients directly represent the time delays* τ_1^ℓ *and* τ_2^ℓ*. Therefore, independent component analysis (ICA) can be applied in conjunction with gradient flow to blindly separate* and *localize multiple wave sources.*

observations of a mixture of traveling wave sources, from spatial and temporal differentials over a sensor array of dimensions much smaller than the coherence length of the source signals. The linear combinations of sources can then be effectively separated using tools of independent component analysis that express statistical independence based on entropy[7] or temporal decorrelation.[8] Identification of the mixture coefficients directly yields the direction coordinates of the sources relative to the array.[9]

3.1. Gradient flow independent component analysis

Gradient flow[9,10] is a signal conditioning technique for source separation and localization designed for arrays of very small aperture, *i.e.,*of dimensions significantly smaller than the shortest wavelength in the sources. Consider a traveling acoustic wave impinging on an array of four microphones, in the configuration of Figure 3 (a). The 3-D direction cosines of the traveling wave $\mathbf{u}$ are implied by propagation delays τ_1 and τ_2 in the source along orthogonal directions p and q in the sensor plane. Direct measurement of these delays is problematic as they require sampling in excess of the bandwidth of the signal, increasing noise floor and power requirements. However, indirect estimates of the delays are obtained, to first order, by relating spatial and temporal derivatives of the acoustic field:

$$\begin{aligned} \xi_{10}(t) &\approx \tau_1 \dot{\xi}_{00}(t) \\ \xi_{01}(t) &\approx \tau_2 \dot{\xi}_{00}(t) \end{aligned} \tag{8}$$

where ξ_{10} and ξ_{01} represent spatial gradients in p and q directions around the origin ($p = q = 0$), ξ_{00} the spatial common mode, and $\dot{\xi}_{00}$ its time derivative. Estimates of ξ_{00}, ξ_{10} and ξ_{01} for the sensor geometry of Figure 3 (a) can be obtained as:

$$\begin{aligned} \xi_{00} &\approx \tfrac{1}{4}\left(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}\right) \\ \xi_{10} &\approx \tfrac{1}{2}\left(x_{1,0} - x_{-1,0}\right) \\ \xi_{01} &\approx \tfrac{1}{2}\left(x_{0,1} - x_{0,-1}\right) \end{aligned} \tag{9}$$

A single source can be localized by estimating direction cosines τ_1 and τ_2 from (8), a principle known for years in monopulse radar, exploited by parasite insects,[11] and implemented in mixed-signal VLSI hardware.[12] However, little known is that multiple sources $s^\ell(t)$ can be jointly separated and localized using essentially the same principle:

$$\begin{aligned} \xi_{00}(t) &= \sum_\ell s^\ell(t) + \nu_{00}(t) \\ \xi_{10}(t) &= \sum_\ell \tau_1^\ell \dot{s}^\ell(t) + \nu_{10}(t) \\ \xi_{01}(t) &= \sum_\ell \tau_2^\ell \dot{s}^\ell(t) + \nu_{01}(t) \end{aligned} \tag{10}$$

where ν_{00}, ν_{10} and ν_{01} represent common mode and spatial derivative components of additive noise in the sensor observations.[9] Taking the time derivative of ξ_{00}, we thus obtain from the sensors a linear instantaneous mixture of the time-differentiated source signals,

$$\begin{bmatrix} \dot{\xi}_{00} \\ \xi_{10} \\ \xi_{01} \end{bmatrix} \approx \begin{bmatrix} 1 & \cdots & 1 \\ \tau_1^1 & \cdots & \tau_1^L \\ \tau_2^1 & \cdots & \tau_2^L \end{bmatrix} \begin{bmatrix} \dot{s}^1 \\ \vdots \\ \dot{s}^L \end{bmatrix} + \begin{bmatrix} \dot{\nu}_{00} \\ \nu_{10} \\ \nu_{01} \end{bmatrix}, \tag{11}$$

an equation in the standard form $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$, where $\mathbf{x}$ is given and the mixing matrix $\mathbf{A}$ and sources $\mathbf{s}$ are unknown. Ignoring the noise term $\mathbf{n}$, this problem setting is standard in independent component analysis (ICA) for blind source separation.

Various formulations of ICA exists. ICA algorithms typically specify some sort of statistical independence assumption on the sources $\mathbf{s}$ either in distribution over amplitude[7] or over time,[8] to obtain estimates of the unknown sources from the unknown mixtures. Most forms specify ICA to be *static*, in assuming that the observations contain static (instantaneous) linear mixtures of the sources. Note that this definition of *static* ICA includes methods for blind source separation that make use of temporal structure in the dynamics within the sources themselves,[8] as long as the observed mixture of the sources is static. In contrast, 'convolutive' ICA techniques explicitly assume convolutive or delayed mixtures in the source observations. Convolutive ICA techniques (*e.g.*,[13]) are usually much more involved and require a large number of parameters and long adaptation time horizons for proper convergence.

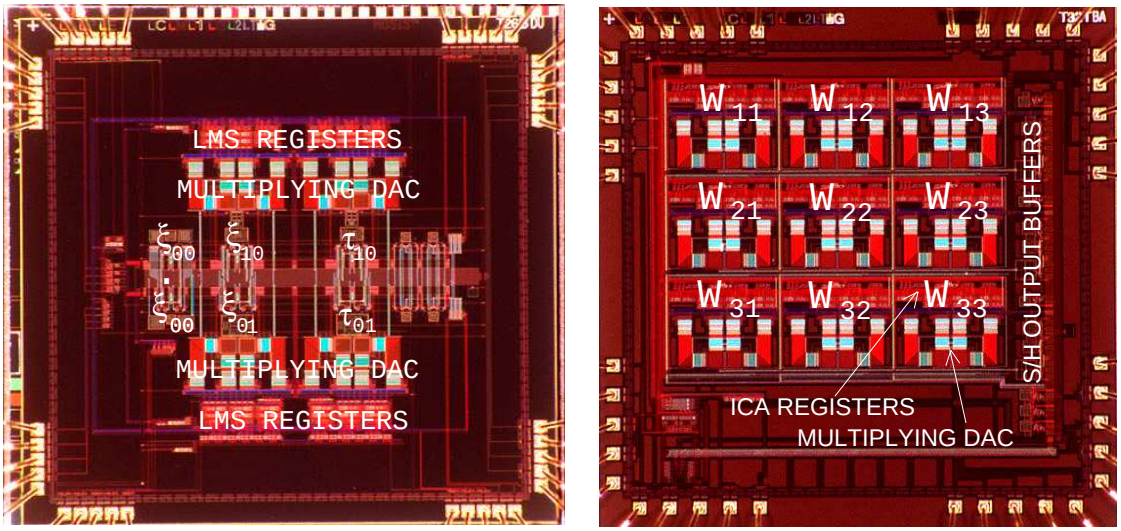
The instantaneous static formulation of gradient flow (11) is convenient,* and avoids the need for non-static (convolutive) ICA to separate mixtures of delayed signals in direct sensor array observations $x_{pq}(t) = \sum_\ell s^\ell(t + p\tau_1 + q\tau_2)$.

3.2. Micropower VLSI Realization

The gradient flow technique has demonstrated separation and localization of up to three acoustic sources using an array of four Knowles (IM series) hearing aid microphones within a 1 mm radius,[10] and lends itself to efficient mixed-signal VLSI implementation using correlated double-sampling (CDS) switched-capacitor (SC) circuits.[12] The mixed-signal VLSI chip shown in Figure 4(a) localizes a single source through LMS (least mean squares) adaptive regression of spatial and temporal gradient signals, according to Eq. (8) with power consumption of 32 μW at sampling frequency of 2 kHz.

The functionality of gradient flow can be extended to joint separation and localization of up to three acoustic sources through static ICA performed on spatial gradient signals. A prototype 3×3 mixed-signal ICA processor was integrated on a single $3mm \times 3mm$ chip fabricated in 0.5 μm 3M2P CMOS technology[14] and consumes 180 μW at sampling frequency of 16 kHz. A micrograph of the chip is shown in Figure 4(b). The architecture is digitally reconfigurable and implements a general class of ICA update rules in common outer-product form. Experiments on the prototype chip consisted of applying synthetic mixtures of speech signals as inputs. The estimated source signals converge towards the original (unseen) sources over time, with 30 dB separation at convergence.

*The time-derivative in the source signals (11) is immaterial, and can be removed by time-integrating the separated signals obtained by applying ICA directly to the gradient flow signals.



(a) *(b)*

Figure 4. (a) *Micrograph of* 3×3 mm^2 *mixed-signal VLSI gradient flow processor in 0.5 µm CMOS technology, applied to acoustic source localization.*[12] (b) *Micrograph of mixed-signal VLSI ICA processor chip in 0.5 µm CMOS technology.*[14]

4. SILICON SUPPORT VECTOR MACHINES FOR MICROPOWER ADAPTIVE PATTERN CLASSIFICATION AND SEQUENCE DECODING

A key component of our approach towards true sensor 'intelligence' and autonomy in wireless sensor networks is to embed pattern recognition directly at the sensor interface, implemented using micropower VLSI. The challenge in embedding pattern recognition intelligence onto sensing and communication interfaces is to balance requirements on precision, complexity and power consumption in the VLSI implementation.[15] Severe power constraints for miniaturized systems require power budget optimization at device, circuit, architecture and system levels.[16] Although most systems on chip designs are digital, analog design becomes attractive at very low power levels exploiting the analog sensory interface and computational primitives inherent in device physics.

Identification of an acoustic target requires an efficient way to represent and detect dynamic sequences of acoustic observations through tracking transitions between states in the classification. Viterbi-type decoding is a popular scheme implemented in analog VLSI for the purpose of sequence estimation[17] and speech recognition.[18] However the nemesis to analog VLSI implementation of Viterbi decoding is its algorithmic complexity, as it requires effectively a backward pass over a trellis. As the decoding depth increases, the dynamic range of path scores decreases and may exceed the noise margin of the system, making it harder to discriminate between classes.

A connectionist approach to *maximum a posteriori* (MAP) based decoding[19] offers an attractive alternative to Viterbi based decoding that relaxes the need for a backward pass over data. Forward-decoding Kernel Machines (FDKM) [20] are attractive for MAP sequence decoding because they provide an adaptive framework to elegantly trade off complexity and power-consumption of the system with limited dynamic range and system accuracy. At the core of FDKM is a support vector machine (SVM),[21] that makes the architecture noise robust by incorporating large margin principles.[22] The power-consumption of an FDKM is determined by number of templates (also called support vectors) stored in memory, which in turn is determined by the complexity of the discrimination boundary and the signal-to-noise ratio of the sensor interface.

We implemented FDKM in silicon, and demonstrated its operation for adaptive sequence detection and pattern recognition. The chip is fully configurable with parameters directly downloadable onto floating gate cell

array. By using PC-in-loop training, mismatches and non-linearities of analog implementation can be directly accounted for. The following briefly describes the system, implementation, and preliminary results.

4.1. FDKM sequence decoding

The problem of FDKM recognition is formulated in the framework of MAP (maximum a posteriori) estimation, combining Markovian dynamics with kernel machines.

The MAP forward decoder receives the sequence $\overline{\mathbf{X}}[n] = \{\mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[n]\}$ and produces an estimate of conditional probability measure of state variables $q[n]$ over all classes $i \in 1, .., S$, $\alpha_i[n] = P(q[n] = i \mid \overline{\mathbf{X}}[n])$. Unlike *hidden* Markov models, the states directly encode the symbols, and the observations $\mathbf{x}$ modulate transition probabilities between states.[19] Estimates of the posterior probability $\alpha_i[n]$ are obtained from estimates of local transition probabilities using the *forward-decoding* procedure[23]

$$\alpha_i[n] = \sum_{j=1}^{S} P_{ij}[n]\ \alpha_j[n-1] \tag{12}$$

where $P_{ij}[n] = P(q[n] = i \mid q[n-1] = j, \mathbf{x}[n])$ denotes the probability of making a transition from class j at time $n-1$ to class i at time n, given the current observation vector $\mathbf{x}[n]$. Forward decoding (12) expresses first order Markovian sequential dependence of state probabilities conditioned on the data.

The transition probabilities $P_{ij}[n]$ in (12) attached to each outgoing state j are obtained directly from SVM output (margin variable ($f_{ij}(\mathbf{x})$) and is given by :

$$P_{ij}[n] = [f_{ij}(\mathbf{x}[n]) - z_j[n]]_+ \tag{13}$$

where $[.]_+ = max(., 0)$ and the normalization factor $z_j[n]$ is subtractive rather than divisive and is obtained using a reverse- water-filling criterion with respect to a probability margin γ,[24]

$$\sum_i [f_{ij}(\mathbf{x}[n]) - z_j[n]]_+ = \gamma. \tag{14}$$

One of the advantages of normalization (14) besides improved robustness, is its amenability to current mode implementation as opposed to logistic normalization[25] which requires exponentiation of currents. The output $f_{ij}(\mathbf{x})$ is given by:

$$f_{ij}(\mathbf{x}) = \sum_{s}^{N} \lambda_{ij}^{s}\ K(\mathbf{x}, \mathbf{x}_s) + b_{ij} \tag{15}$$

where $K(\cdot, \cdot)$ denotes any symmetric positive-definite kernel[†] that satisfies the Mercer condition, such as a Gaussian radial basis function or a polynomial spline,[21] and $\mathbf{x}_s[m], m = 1, .., N$ are the support vectors stored in memory. The parameters λ_{ij}^{s} in (15) and the support vectors $\mathbf{x}_s[m]$ are determined by training on a labeled training set using a recursive FDKM procedure described in.[20]

4.2. Hardware implementation

A second order polynomial kernel $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}.\mathbf{y})^2$ was chosen for convenience of implementation. This inner-product based architecture directly maps onto an analog computational array, where storage and computation share common circuit elements. The FDKM system architecture is shown in Figure 5 (a). It consists of several SVM stages that generates state transition probabilities $P_{ij}[n]$, modulated by input data $\mathbf{x}[n]$, and a forward decoding block *maximum a posteriori* (MAP) estimation of the state sequence($\alpha_i[n]$).

A 14-input, 24-state, and 24×30-support vector FDKM was integrated on a 3mm×3mm FDKM chip, fabricated in a 0.5μm CMOS process, and fully tested. Figure 5 (b) shows the micrograph of the fabricated chip. Labeled training data pertaining to a certain task were used to train an SVM, and the training coefficients thus obtained were programmed onto the chip. Experiments with the processor trained for speaker verification and phoneme sequence estimation demonstrate real-time recognition accuracy at par with floating-point software, at sub-microwatt power.[26]

[†] $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}).\Phi(\mathbf{y})$. The map $\Phi(\cdot)$ need not be computed explicitly, as it only appears in inner-product form

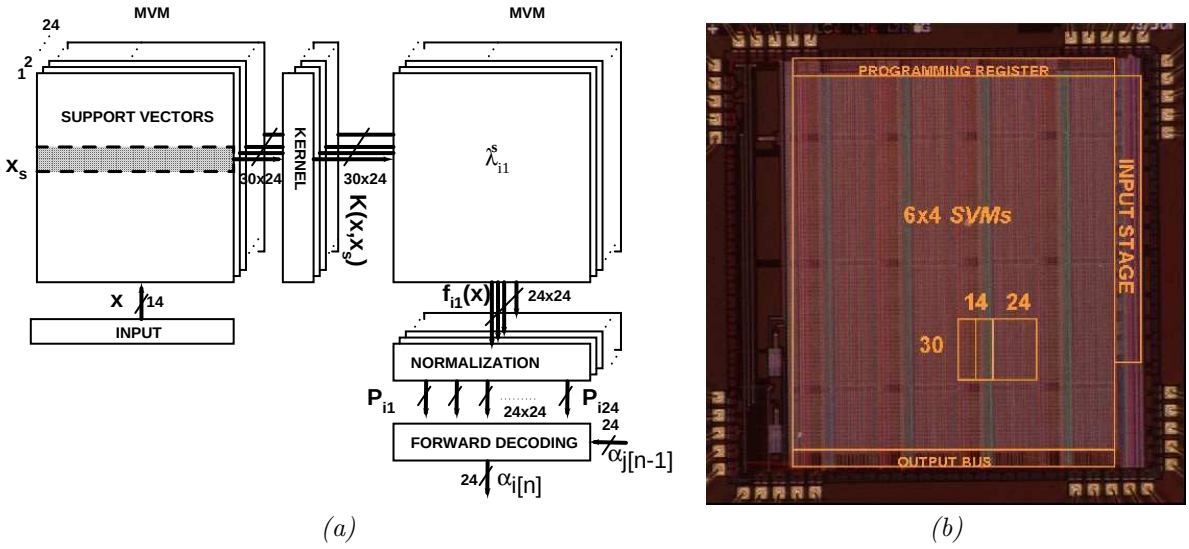


Figure 5. *(a) FDKM system architecture. (b) FDKM chip micrograph.*[26]

Figure 6. *The Acoustic Surveillance Unit (ASU),*[28] *a joint development project with Signal Systems Corporation. The unit houses the VLSI wake-up and localization processors and in field tests has demonstrated detection and localization performance on par with state-of-the art DSP systems at a mere fraction of their size and power dissipation.*

5. FIELD TEST RESULTS

The wake-up ASIC,[1] gradient flow bearing estimator ASIC[12] and a cross correlation bearing estimator ASIC[27] were integrated into acoustic surveillance unit (ASU)[28] enclosure. Measuring 11 cm in diameter, the ASU also contained four Knowles SiSonic MEMS microphones and signal conditioning circuitry. The ASU enclosure is depicted in Fig. 6. We conducted two field tests, one in a public park in Severna Park, Md. with synthesized sounds (field test 1), and another at Aberdeen Proving Ground, Aberdeen, Md. with a selection of ground-based military vehicles (field test 2).

Field test 1: Synthesized sounds In this test, synthesized sounds were played from a sub-woofer placed in an open field. In this setting, we had strict control of the frequency content and amplitude of the sounds. For target detection tests, in all trials, the ASU was placed 30 feet from the sub-woofer at an angle of 90°. We first performed a series of trials with a signal consisting of three simultaneous time-varying, harmonically related tones (125 Hz, 150 Hz, 175 Hz). This kind of signal is often used as a model for sound generated by a vehicle. The wake-up detector, with a threshold setting of 1024, was reliably triggered down to a narrowband SNR of 13 dB In trials with broadband white noise, the loudest possible volume (50 dB SPL) did not elicit a trigger.

For bearing estimation tests, two different sound sources, broadband bandlimited (20-300Hz) Gaussian signal and narrowband signal, were presented through a subwoofer under same conditions. The sampling system

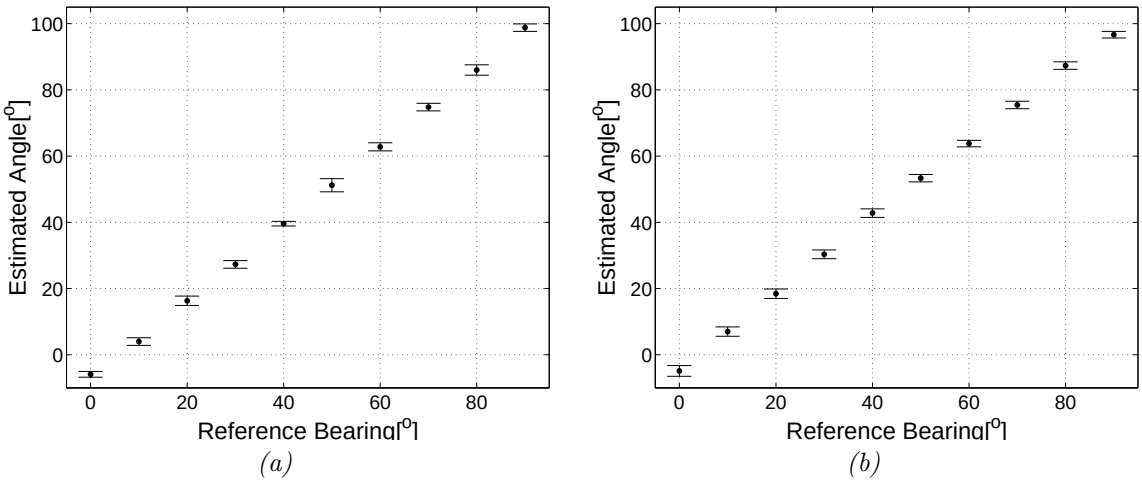


Figure 7. *Localization results for (a) broadband (b) narrowband source signal*

frequency was 2 kHz. Signal-to-noise (SNR) ratio was around 25-30 dB. Simple expressions can be obtained for the Cramer-Rao lower bound on the variance of bearing angle estimate, assuming Gaussian univariate distributions for the source and noise components.[29] For this experimental setup and the broadband acoustic source, calculated Cramer-Rao bound was around 1 degree. The assumption of uncorrelated noise is violated for sub-wavelength sensor geometries, and gradient flow technique exploits correlated noise and temporal dependencies to obtain superior bearing accuracies. For each bearing angle sound source was played for 10 seconds and at the end of each 1 second wide window the bearing estimate was reported by acoustic localizer ASIC. The bearing angle was changed from 0^o to 90^o in increments of 10^o. The mean and variance of estimated bearing angles are shown in Figure 7.

Field test 2: Ground-based vehicles In this test, vehicles were driven around a 662 m x 108 m oval-shaped track at different speeds, and tracked with three ASUs positioned at points S_1, S_2 and S_3 approximately 90 m apart as illustrated in Figure 8. The tests were performed with relatively loud ambient background noise. When the detector was triggered, the bearing estimation circuit on all of the ASUs localized and tracked the sound source. Table 1 summarizes the wake-up results on an assortment of vehicles in terms of the maximum distance that elicited a sustained detection. During the duration of the test, the false alarm rate was less than 2 per hour. Also, numerous unscripted targets were detected, such as helicopters, powerboats, and trucks. At one instance, an F/A-18 fighter jet flew overhead at approximately $10,000$ ft and no detection was elicited. As an example of localization performance, tracking of one vehicle moving clockwise around the oval is illustrated in Figure 8. The bearing ITD estimates were recorded at 1 s time intervals. Only the azimuth angles θ are shown for ground vehicles, since the ASU microphone arrays were oriented horizontally. Estimates of true bearing angles θ from GPS tracking of the vehicles, accounting for the approximate geometry of the track and ASUs and correcting for delays in acoustic wave propagation, are also depicted as 'ground truth' in Figure 8. In tests with multiple vehicles moving together or in opposite directions, the localizer chip tracked the loudest (received) target. For F/A-18 fighter jet flying overhead at approximately 3000 m was also successfully tracked by the localizer chip, accounting for elevation besides azimuth.

6. CONCLUSION

A compact, low-power integrated system for smart acoustic sensing was presented. The acoustic surveillance unit (ASU) comprises a sub-wavelength array of MEMS microphones, and micropower ASICs for acoustic source detection, separation, localization, and identification. A wireless ASU further provides for event-driven data transfer using the Crossbow Mica2 interface. Field tests with multiple ASU sensing nodes on detection and

Vehicle	Description	Distance
M60	Heavy Tracked	> 500 m*
HEMET	Heavy Wheeled	~ 250 m
M548	Light Tracked	~ 400 m
HMMWV	Light Wheeled	~ 55 m

Table 1. *Performance of the wake-up system on ground-based vehicles.* * *Test limited.*

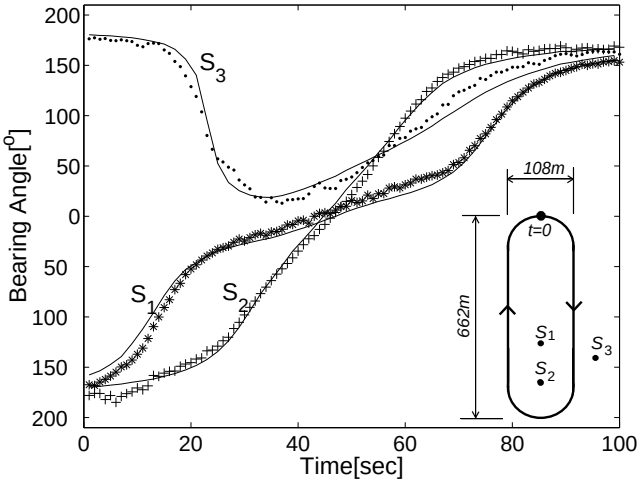


Figure 8. Experimental tracking of a ground vehicle using three acoustic localization sensor nodes positioned in the field as shown in the inset. Solid lines indicate ground truth estimates from GPS measurement.

localization of ground vehicles revealed a performance on par with real-time implementation on rack-mounted DSP systems, at a small fraction of the power. The long endurance, small form factor and autonomous operation of the ASU supports various applications for acoustic surveillance and reconnaissance in the digital battlefield.

ACKNOWLEDGMENTS

This work was supported by the Defense Intelligence Agency, National Science Foundation, Defense Advanced Research Projects Agency and Office of Naval Research.

REFERENCES

1. D. Goldberg, A. Andreou, P. Julián, P. Pouliquen, L. Riddle, and R. Rosasco, "A wake-up detector for an acoustic surveillance sensor network: Algorithm and vlsi implementation," *Information Processing in Sensor Networks (IPSN'2004)*, (Berkeley CA), May 2004.
2. L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-24(5)**, pp. 399–418, 1976.
3. A. M. Noll, "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate," *Proc. of the Symposium on Computer Processing in Communications* **19**, pp. 779–797, Wiley-Interscience, (Chichester, Sussex, UK), 1969.
4. J. D. Wise, J. R. Caprio, and T. W. Parks, "Maximum likelihood pitch estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-24(5)**, pp. 418–423, 1976.
5. D. H. Friedman, "Pseudo-maximum-likelihood speech pitch extraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-25(3)**, pp. 213–221, 1977.
6. J. H. Van Vleck and D. Middleton, "The spectrum of clipped noise," *Proceedings of the IEEE* **54(1)**, pp. 2–19, 1966.

7. A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comp* **7(6)**, pp. 1129–1159, Nov 1995.
8. L. Molgedey and G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters* **72**(23), pp. 3634–3637, 1994.
9. G. Cauwenberghs, M. Stanacevic, and G. Zweig, "Blind broadband source localization and separation in miniature sensor arrays," *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'2001)*, (Sydney, Australia), May 6-9, 2001.
10. M. Stanacevic, G. Cauwenberghs, and G. Zweig, "Gradient flow adaptive beamforming and signal separation in a miniature microphone array," *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP'2002)*, (Orlando FL), May 13-17, 2002.
11. D. Robert, R. Miles, and R. Hoy, "Tympanal hearing in the sarcophagid parasitoid fly emblemasoma sp.: the biomechanics of directional hearing," *J. Experimental Biology* **202**, pp. 1865–1876, 1999.
12. M. Stanacevic and G. Cauwenberghs, "Micropower mixed-signal acoustic localizer," *Proc. IEEE Eur. Solid State Circuits Conf. (ESSCIRC'2003)*, (Estoril Portugal), Sept. 16-18, 2003.
13. R. Lambert and A. Bell, "Blind separation of multiple speakers in a multipath environment," *Proc. ICASSP'97*, (Munich, Germany), 1997.
14. A. Celik, M. Stanacevic, and G. Cauwenberghs, "Mixed-signal real-time adaptive blind source separation," *Proc. IEEE Int. Symp. Circuits and Systems(ISCAS2004)*, (Vancouver Canada), May 23-26, 2004.
15. A. Wang and A. Chandrakasan, "Energy-efficient dsps for wireless sensor networks," *IEEE Signal Proc. Mag.* **19**(4), pp. 68–78, July 2002.
16. N. Shanbhag, "A mathematical basis for power reduction in digital vlsi systems," *IEEE Trans. Circuits and Systems II* **44**(11), Nov 1997.
17. M. Shakiba, D. Johns, and K. Martin, "Bicmos circuits for analog viterbi decoders," *IEEE Trans. Circuits and Systems II* **45**(12), Dec 1998.
18. J. Lazzaro, J. Wawrzynek, and R. Lippmann, "A micropower analog circuit implementation of hidden markov model state decoding," *IEEE J. Solid-State Circuits* **32**(8), Aug. 1997.
19. H. Bourlard and N. Morgan, *Connectionist Speech Recognition: A Hybrid Approach*, Kluwer Academic, 1994.
20. S. Chakrabartty and G. Cauwenberghs, "Forward decoding kernel machines: A hybrid hmm/svm approach to sequence recognition," *IEEE Int. Conf. of Pattern Recognition: SVM workshop. (ICPR'2002)*, (Niagara Falls), 2002.
21. V. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.
22. S. Chakrabartty and G. Cauwenberghs, "Power dissipation limits and large margin in wireless sensors," *Proc. IEEE Int. Symp. Circuits and Systems(ISCAS2003)* **4**, pp. 25–28, May 2003.
23. L. Bahl, C. J., J. F., and R. J., "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Inform. Theory* **IT-20**, pp. 284–287, 1974.
24. S. Chakrabartty and G. Cauwenberghs, "Margin propagation and forward decoding in analog vlsi," *Proc. IEEE Int. Symp. Circuits and Systems(ISCAS2004)*, (Vancouver Canada), May 23-26, 2004.
25. T. Jaakkola and D. Haussler, "Probabilistic kernel regression models," *Proc. Seventh Int. Workshop Artificial Intelligence and Statistics*, 1999.
26. S. Chakrabartty and G. Cauwenberghs, "Sub-microwatt analog vlsi support vector machine for pattern classification and sequence estimation," *Adv. Neural Information Processing Systems (NIPS'2004)*, MIT Press, (Cambridge, MA), 2005.
27. P. Julian, A. Andreou, L. Riddle, S. Shamma, and G. Cauwenberghs, "A comparison of algorithms for sound localization," *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'2003)*, (Bangkok Thailand), May 25-28, 2003.
28. L. Riddle, A. Andreou, G. Cauwenberghs, P. Julian, D. Goldberg, M. Stanacevic, S. Shamma, and T. Horiuchi, "Vlsi acoustic surveillance unit," *Proc. GOMAC'2004*, 2004.
29. B. Friedlander, "On the cramer-rao bound for time delay and doppler estimation," *IEEE Transactions on Information Theory* **30**(3), pp. 575–580, 1984.