

# Search and Discovery in an Uncertain Networked World

Erhan Baki Ermis Murat Alanyali Venkatesh Saligrama

Department of Electrical and Computer Engineering

Boston University, Boston MA 02215

Distributed sensing systems, which are networks of tiny sensing devices, hold the promise of providing an inexpensive, non-intrusive means to understand phenomena that exhibit spatial and temporal variations at multiple scales. Such systems are envisioned under two general architectures: Distributed systems that perform distributed sensing and in-network data processing to complete some desired tasks such as hypothesis testing or target tracking, and decentralized systems that perform distributed sensing but centralized data processing at a fusion center to achieve the desired task. The latter approach is taken in [4], and the former approach is taken in [5]. In this work we also focus on the architecture that performs in-network data processing.

While significant effort over the last decade has produced the necessary infrastructure for deployment, efficient monitoring requires development of new distributed signal processing and decision making approaches in wirelessly networked, power limited regimes. We consider the example of a sensor network for environmental monitoring as a motivating example to illustrate the set of problems and the issues involved. Here change is rarely predictable. Periods of relative stasis are punctuated by extreme events occurring on short time scales [23]. It is during these extreme events that some of the most interesting phenomena occur; yet almost all field monitoring programs rely on rigid time sampling protocols and sparse spatial sampling for recording environmental and ecological data [23]. Moreover, non-linear relationships between environmental and ecological processes are ubiquitous (e.g, photosynthesis is non-linear with light), and inappropriate averaging of environmental time-series data can lead to large errors in prediction of these processes. A network sensing system, in situ over extended areas and long times, that can monitor, evaluate, and adapt to its environment would greatly improve our ability to resolve extreme events in high temporal and spatial resolution while preventing an overwhelming and unneeded collection of data occurring during periods and at locations of relative stasis.

Consequently, environmental monitoring poses two fundamental challenges:

**(A) Lifetime:** For statistical accuracy, sensing systems need to be functional over long periods of time.

**(B) Burstiness:** Environmental phenomena are characterized by sudden events with periods of relative stasis punctuated by extreme events occurring over a short time.

Lifetime not only necessitates energy conservation but also requires combating sensor loss as well as degradation in sensing and communication processes over time. Burstiness entails adaptivity to sudden changes requiring the system to fuse spatial information over a wide region in a short time, which results in large energy expenditure. Together, these two requirements call for a natural structural hierarchy for distributed sensing. To conserve energy a sparse sensor network will perform the task of searching/sensor selection for localized isolated events (insignificant spatial correlation across sensors). Upon detection it will opportunistically trigger a sensor sub-network, which will then perform the task of discovering/estimation/tracking of interesting environmental processes by fusing spatially correlated information. The salient aspects of the article are:

**(A)** Demonstration of a distributed, bottom-up energy efficient systems approach for search/sensor selection and discovery/fusion. The two problems described above require diametrically opposite strategies. Indeed, fusion requires spatial integration of sensed information. In contrast, search entails strategies that can quickly rule out sensors that do not have the desired information.

**(B)** Asynchronous computational strategies to overcome effects of uncertain communication links, packet losses and channel errors.

**(C)** Development of distributed strategies that only utilize local information and models. This ensures robustness to possible model differences across sensors.

Based on the discussion above we focus on two specific classes of problems; *search* and *discovery*:

1. **Search Problem:** The objective is to select sensors containing relevant information, in scenarios where correlation between across sensors can be small (see Figure 1(a)). This problem arises in a number of other contexts such as labeling useful information in a data stream contaminated by clutter, matching a reference image to a target, sensor selection problems and target tracking in cluttered environments.
2. **Discovery Problem:** Here the main task is to fuse correlated information from multiple multi-modal networked sensors (see Figure 1(b)) over a multi-hop network to discover the underlying phenomena. This problem arises in a number of scenarios such as target tracking, change detection, and target

identification.

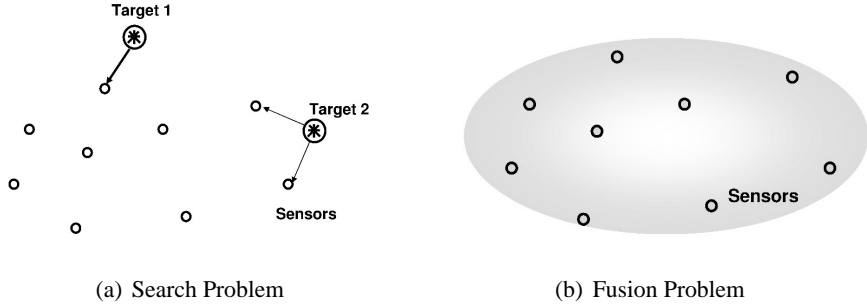


Figure 1: Schematic illustration of Search and Fusion Problems. In search each target is sensed by a small subset of sensors, i.e., the spatial correlations across the sensor network is small. In fusion problems the sensors can be viewing a single underlying phenomena and the spatial correlations across the sensor network is relatively large.

The mathematical and practical developments discussed here have appeared in a number of our prior publications. As a matter of reference we have dealt with distributed hypothesis testing using variants of belief propagation in [19, 2, 20], asynchronous distributed detection, estimation, and energy scaling issues in [14], distributed tracking and Kalman filtering over multi-hop networks in [1], and search and sampling in [22, 9, 10]. In parallel we have recently initiated practical implementation of these techniques at an environmental test site. A laboratory scale test-bed demonstrating these techniques was selected as a finalist in the *Crossbow Challenge Competition* [15].

**Collaboration:** While it is clear that fusion problem requires collaboration to integrate correlated information, it is not obvious why collaboration may be required for search problems. The following example points to some of the advantages of collaboration:

Suppose we have two sensors,  $S_1$  and  $S_2$ , with observations  $X_1$  and  $X_2$ . The objective is to map each observation to one of two hypotheses,  $H_0$  or  $H_1$ . The Maximum-Likelihood (ML) rule divides the observation space evenly for all possible sets of hypotheses, and for a given value of  $X_1$ , the mapping of that observation is done irrespective of the value of  $X_2$ . This is depicted in Figure 2 (a). Now consider a more complicated decision rule to reduce the probability of false alarm (which corresponds to declaring  $H_0$  when  $H_1$  is true), as in Figure 2 (b). As it can be seen, mapping of  $X_1$  depends on the value of  $X_2$ , and to be able to perform this joint mapping,  $S_1$  and  $S_2$  need to exchange some information about their observations.

This example clarifies why collaboration can be necessary or beneficial even in problems where the observations are independent and the decisions can be made completely in a decoupled fashion.

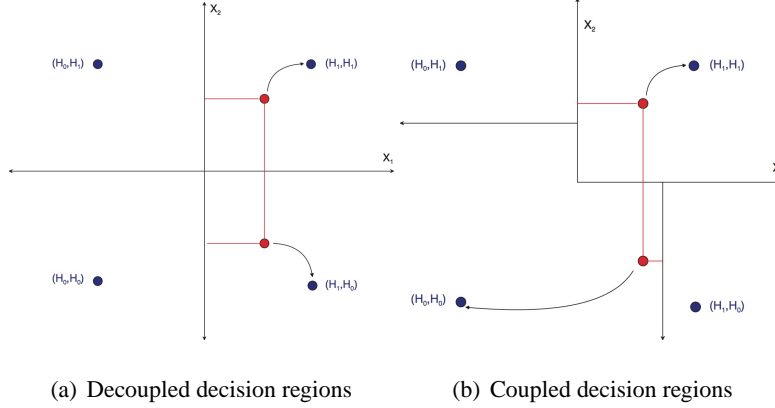


Figure 2: ML Rule vs. More elaborate rules: To be able to realize elaborate decision regions, a collaboration among sensors is necessary even if their observations are independent.  $X_1$  and  $X_2$  represent the observation spaces of sensors  $S_1$  and  $S_2$  respectively.

## 1 Adaptive Decision Strategies for the Search Problem

We develop methodologies that are inspired by recent developments in testing large number of hypothesis in the statistics literature.

We denote by  $H_1(v)$  (resp.  $H_0(v)$ ) the hypothesis that the observation of sensor  $v$  is generated in the presence (resp. absence) of a local target. Let  $V$  be the set of all sensors and let  $X(v)$  denote the true hypothesis for sensor  $v \in V$ . In the extreme situation, this model entails that variables  $X(v) : v \in V$  bear no statistical relation. But even in this case, as we have presented above, collaboration can be beneficial or even necessary.

We consider a problem where some local phenomena are independently distributed on a sensor field. The number of phenomena scattered on the sensor field is unknown, between zero and  $N$ . Also, we assume no prior knowledge of the probability of a phenomenon being at any given location on the sensor field. Therefore we assume a non-Bayesian structure on the problem. The overall objective is to devise a distributed decision rule such that it has large detection power, leads to small false alarms, and incurs a minimal communication cost to the system. We have the notional convention that a decision rule has larger detection power over another rule if it leads to smaller expected number of misses.

Given the above description of the problem, the general mathematical formulation is as follows:

$$\min \text{Expected}(\#misses) \text{ subject to: False Alarms \& Bit Budget Constraints} \quad (1)$$

In this formulation, the bit budget constraint can be chosen in a very straight forward manner. An upper

limit on the total number of transmissions can be set to the sensor network, and that limit can be enforced throughout the sensor network by keeping track of the number of transmitted bits.

Although there may be other possible strategies in terms of choosing a constraint to control the false alarms, in this work we explore two quantities: Probability of false alarm, and False Discovery Rate (FDR).

The probability of false alarm is a very common quantity in detection problems. It can be used when the cost of making a false alarm is very high. However, it turns out that in many detection problems, as the number of observations get large, using the probability of false alarm as a constraint can be very stringent in terms of detection power. In the problems where detection performance is a concern, FDR can be used to control the false alarms, where FDR is defined as the expected ratio of the number of false alarms to the number of observations that are declared to be significant. Although it is not so commonly used in detection problems, it is still a sensible quantity to consider. It is especially useful in problems where we are not concerned so much with making any false alarms, but rather a percentage of false alarms among the ones declared significant is allowed. We further discuss these two strategies below.

Among the scenarios that can be formulated within this framework is the boundary estimation problem. Here the problem is to determine the boundary separating two homogenous mediums, which is a generalization of localization of multiple targets. Similarly, the problem of determining the spatial region,  $R_\alpha$ , where the random field,  $\phi_s$  takes values larger than  $\alpha$ , i.e.,  $R_\alpha = \{s \mid \phi_s \geq \alpha\}$  is another scenario that can be addressed within this framework as well.

In solving the search problem we first focus on a setting where we are interested in optimizing probability of global detection subject to global false positive constraint. The optimal decision rule in this case is a standard likelihood ratio test, however when decisions are uncoordinated the tradeoff between detection power and false alarm rate scales unfavorably with increasing sensor population. More precisely, for a fixed local probability of false alarm at each sensor, as the number of sensors increase, the global probability of false alarm increases as well. Consequently, to control the global probability of false alarm at some level, the local probability of false alarms must be bounded at much smaller levels. An example will help clarify the point:

**Example:** Let us assume that there are 10000 sensors scattered on a field. We wish to have probability of false alarm less than or equal to 0.2. If we set the local probability of false alarm at level 0.2 at each sensor, by the law of large numbers, there would be about 2000 false alarms declared on the sensor network; furthermore, the probability of making any false alarm is near 1.

As the example illustrates, if the multiplicity of false alarms is not taken into account the global proba-

bility of false alarm cannot be controlled. A very common approach to resolving this multiplicity issue in the probability of false alarm is known as the Bonferroni procedure. From the union bound, to control the global probability of false alarm at level  $\gamma$ , it suffices to control the local probability of false alarm at level  $\gamma/N$ , where  $N$  is the number of sensors.

$$\text{Prob}\{\text{Local False Alarm}\} \leq \gamma/N \implies \text{Prob}\{\text{Global False Alarm}\} \leq \gamma$$

To understand how this local control of probability of false alarm affects the decision regions, we reconsider the two sensor example. Figure 3 depicts the ML decision regions and the modified decision regions corresponding to the Bonferroni procedure. We can see that although the region over which  $(H_0, H_0)$  is mapped has been enlarged, the decision regions remain decoupled. Hence the easy decentralization of this strategy follows.

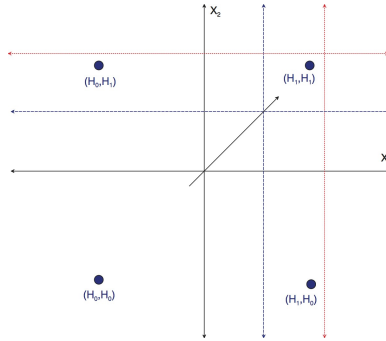


Figure 3: Modified ML regions: Global probability of false alarm decreases as the  $\{H_0, H_0\}$  decision region is expanded in the direction of diagonal arrow. These decision regions are still decoupled, thus they can be viewed as a variation of the original ML decision regions, partitioned by the solid lines.

It is important to notice that as the number of observations get large, the threshold for local probability of false alarm tends to zero, which in turn leads to arbitrarily low detection power.

Nevertheless, it can be shown that this is not just an issue of uncoordinated decisions but rather is an issue due to the stringent constraint of global probability of false alarm. In other words, the unfavorable tradeoff remains even if coordinated thresholds are used due to the fact that controlling the global false positive probability is not possible without forgoing detection power. For this reason a relaxed objective needs to be formalized. Inspired by recent developments in statistical theory (see [10] for details) we formalize the problem in terms of the False Discovery Rate (FDR).

To better explain this idea consider table below, where  $N$  is the number of samples (or sensor nodes) known in advance. The locations and the hypotheses are drawn with some probability distribution, which

may not be known.  $R$  is an observable random variable;  $U, V, S, T$  are unobservable random variables.

	Declared $H_0$	Declared $H_1$	Total
True $H_0$	U	V	$N_0$
True $H_1$	T	S	$N - N_0$
Total	$N - R$	R	N

The false discovery rate concept, instead of trying to control the probability of making any type one error, controls the expected ratio of the number of observations falsely declared as significant, ( $V$ ), to the total number of observations declared as significant, ( $R$ ), i.e.,

$$FDR = E\{V/V + S\} = E\{V/R\}.$$

This relaxation improves the detection power while maintaining FDR to within some level  $\gamma$ . It is easy to establish that the false alarm rate [6] is bounded from below by FDR, i.e.,  $Prob\{V \geq 1\} \geq E\{V/R\}$ . Although it is a weaker notion in terms of false alarm probability, the significant increase in the power of detection makes it a desirable approach in many problems. Mathematically, the problem is as follows:

$$\min E(T) \text{ subject to: } FDR = E\{V/R\} \leq \gamma, B \leq \alpha$$

where,  $B$  is the number of transmitted bits and the minimization is over all decision strategies.

Constraining FDR is very different from constraining false alarm probability. First, the two constraints are philosophically different in terms of their goals. False alarm probability controls the possibility of making any type one error, while FDR constraint controls the expected ratio of false alarms to the total number of observations that are declared to be significant. Therefore, the cost structure of these constraints are also very different, and this fact reflects in the decision regions. To better understand this point, we consider the two sensor example discussed earlier.

**Example** Suppose,  $\mathcal{A}$  denotes the class of one false alarm errors. This amounts to deciding  $\{H_0, H_1\}$  or  $\{H_1, H_0\}$  if  $\{H_0, H_0\}$  is true and in this case the ratio  $V/R$  is 1. Let  $\mathcal{B}$  now denote two false alarms, i.e., we decide  $\{H_1, H_1\}$  for  $\{H_0, H_0\}$ . In this scenario two observations are declared significant, while both are false alarms, but still  $V/R$  is 1. From an FDR perspective, the costs of class  $\mathcal{A}$  and  $\mathcal{B}$  errors are equal. However, if we consider detection power, class  $\mathcal{B}$  error is more desirable, because  $\{H_1, H_1\}$  decision will never miss, whatever the ground truth might be. Therefore, it makes sense to favor the  $\{H_1, H_1\}$  region over  $\{H_1, H_0\}$  and  $\{H_0, H_1\}$  regions. Figure 4 depicts the two classes of errors and a possible decision region

that is biased for  $\{H_1, H_1\}$  region. In general, the intuition leads us to the following fact: for the same ratio of  $V/R$ , decision regions that lead to larger values of  $R$  are desirable.

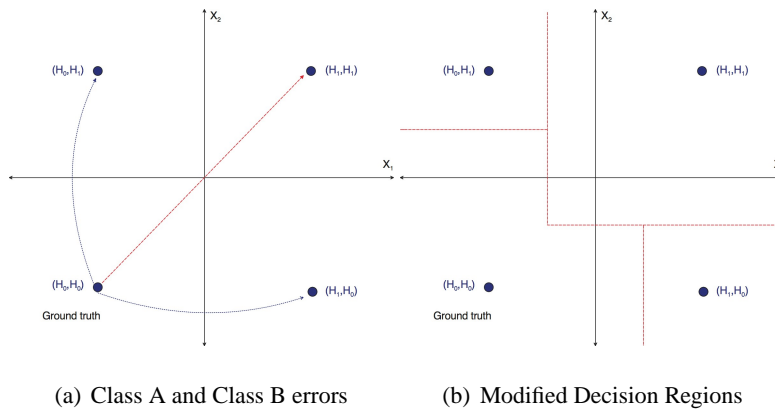


Figure 4: The nature of FDR constraint: Interestingly, two false alarms are more desirable than one false alarm, which leads to a biased decision regions as partitioned by dashed lines in (b).

In general, our objective reduces to optimizing the expected number of detections subject to FDR and communication-bit constraint. It turns out that the solution requires adaptively increasing the threshold and can be implemented by means of feedback. The main idea can be described as follows: All the sensors first fix a detection threshold, which corresponds to small false alarms. Based on the number of declarations corresponding to this threshold, a new threshold is locally computed by all the sensors. New declarations corresponding to this threshold are then made, which is then fed back to the sensor network and the thresholds are then updated again. Such an adaptive system constitutes a feedback structure, which is illustrated in Figure 5.

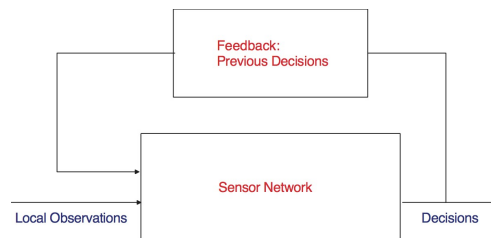


Figure 5: An adaptive decision making strategy for sensor networks. The previous decisions are fed back to the sensors, which then adjust their corresponding decision thresholds.

Only one-bit messages from sensors, which declare positive hypothesis, are required to be communicated. Counting the number of sensors that have significant observations throughout the sensor network to update the threshold is accomplished through the asynchronous message passing approach based on local



ad-hoc information coalescence, which is described in detail in [21]. The general algorithm is briefly described in Section 2.3.1 and it applies here by setting the initial values of *estimates* to 1 for sensors with a significant observation and to 0 for other sensors. This approach also requires each sensor to be aware of only the sensors that it can communicate directly, but it has significant reduction of energy consumption over other local ad-hoc messaging algorithms such as gossip and belief propagation.

The main insight behind the distributed FDR scheme is that the number of declarations at each stage provides a confidence level (in terms of false positives) for future declarations. To present the FDR procedure we need a transformation of the measured random variable. This transformation converts the distribution under  $H_0$  to a uniform distribution and is commonly referred to as the p-value:

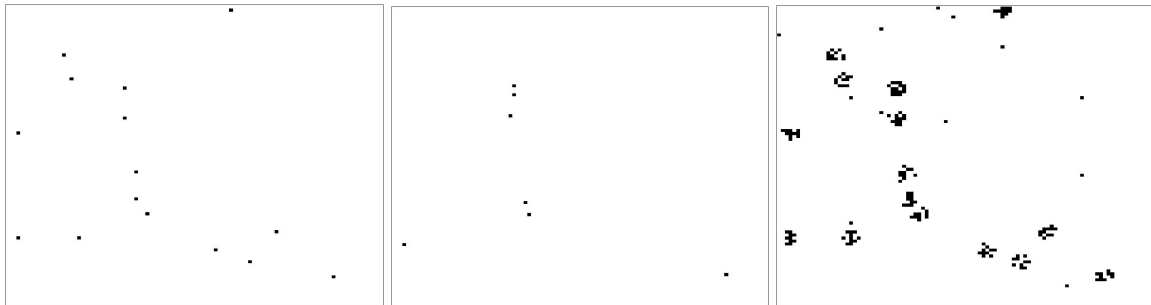
$$p_i(X) = \int_X^\infty f_{0i}(t)dt = 1 - F_{0i}(X) \quad (2)$$

where  $f_{0i}$  is the probability density function (pdf) of the observations under  $H_0$  at sensor  $i$ . The distributed FDR procedure in [10] is presented below:

1. Each sensor calculates the  $p$  value of its observation,  $p_i$ , and tests  $p_i$  with  $\frac{1}{m}\gamma$
2. The sensors with  $p_i \leq \frac{1}{m}\gamma$  declare their observations as significant, and communicate these decisions to other sensors by a suitable protocol, (assume  $l$  of them declare their observations significant)
3. Decisions of the  $l$  sensors are fed back to the network and sensors update their threshold to  $\frac{l+1}{m}\gamma$
4. Remaining sensors test their  $p$  values by the new threshold and declare their observations significant accordingly, (assume  $k$  more sensors declare their observations significant)
5. The new significant decisions are fed back to the system again, and threshold is updated to  $\frac{l+k+1}{m}\gamma$
6. Steps 4 and 5 are repeated until when there is no more sensors that declare their observations as significant under the most current threshold, which is when the process terminates.

The idea is that through the feedback system depicted in Figure 5 we retain the dynamic, linearly increasing structure of the centralized thresholding strategy. Furthermore, only the sensors that have significant observations declare themselves to the network with a one bit message. The procedure we present in [9, 10] requires a further local transformation of each p-value, which we refer to as *Domain Transformed FDR (DTFDR)* to achieve optimal detection power and energy scaling. The significance of this transformation is that it clusters all the significant observations to the same part of observation space, and it is applicable at each sensor node with knowledge of the local observation model. It is due to this clustering that DTFDR procedure can achieve minimal expected number of misses.

As it can easily be seen from Figure 6, the detection performance is much larger with an adaptive approach using FDR as the control constraint as opposed to the non-adaptive approach with probability of false alarm as the control constraint. As a trade off, the number of false alarms has increased by choosing FDR over probability of false alarm in the problem formulation.



(a) Objects of interest in sensor field (b) Detection via Bonferroni procedure (c) Detection via FDR procedure

Figure 6: Detection using non-adaptive Bonferroni procedure and adaptive FDR procedure: For simplicity, the sensor field is a 100x100 grid with a sensor at each pixel. In (a) the black pixels are the objects of interest whereas in (b) and (c) the black pixels are the sensors that detect an object of interest.

Figure 7 illustrates the tradeoff between communication bits and number of detections generated using 100 observations, 30 of which were significant. The experiments were repeated for 1000 times to get the average behavior of the systems. Now consider the following problem for illustration: The FDR constraint is set to  $\gamma = .15$ , and the bit budget to  $\alpha = 5$ . In this setting, the system using DTFDR procedure can detect 5 targets on average, while the system using FDR procedure cannot detect any. The advantage of using DTFDR procedure is the linear relationship induced between the number of bits used and the number of detections, which is absent if FDR procedure is used.

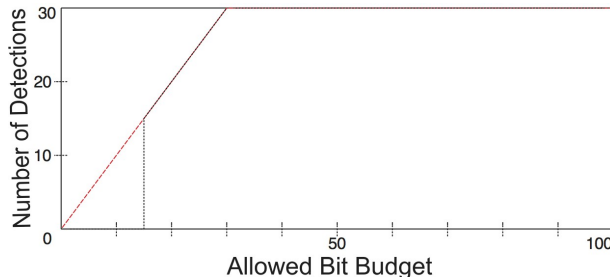


Figure 7: Communication Costs for DTFDR Procedure (dashed) and FDR procedure (dotted): DTFDR procedure leads to a linear relationship between bit budget and the number of correct detections, where this relationship lacks for FDR procedure.

In a separate simulation study, we generated a sensor field of size 100x100, where each pixel corresponds to a sensor. Then, fixing the FDR threshold, we varied the communication bit budget available to the network,  $\alpha$  to investigate the detection performance of FDR procedure and the DTFDR procedure in a communication constrained distributed detection setting. The results are presented for illustrative cases in Figure 8 and 9.



Figure 8: Detection performance of a sensor network using FDR and DTFDR procedures with a 600-bit communication budget: The sensor field is a 100x100 grid with a sensor at each pixel. Here, due to stringent communication bit budget, by using FDR procedure the sensor network detects no objects of interest whereas by using DTFDR procedure many sensors can detect an object of interest within their vicinity. In (a) the black pixels are the objects of interest and in (c) the black pixels are the sensors that detect an object of interest.

For  $\alpha \leq 800$ , implementation of the FDR procedure was unable to detect the observations from positive hypothesis, whereas the DTFDR procedure was able to do so. As the communication constraint was loosened, the performance of DTFDR procedure increased accordingly, almost in a linear fashion.

The predicted behavior of FDR procedure analogous to the one presented in Figure 7 in terms of having a necessary minimum amount of bit budget to detect any significant observations manifests itself in these figures as well. Until the bit budget is increased to 800, there was no detection.

It is also noteworthy that even when FDR procedure starts to declare observations as significant with sufficient amount of bit budget as in Figure 9, the number of false alarms are substantially more than that of DTFDR procedure. This is a very clear demonstration of DTFDR procedure's capability of minimizing  $E(\#misses)$ , since for the same number of significant declarations the number of false alarms are minimized.

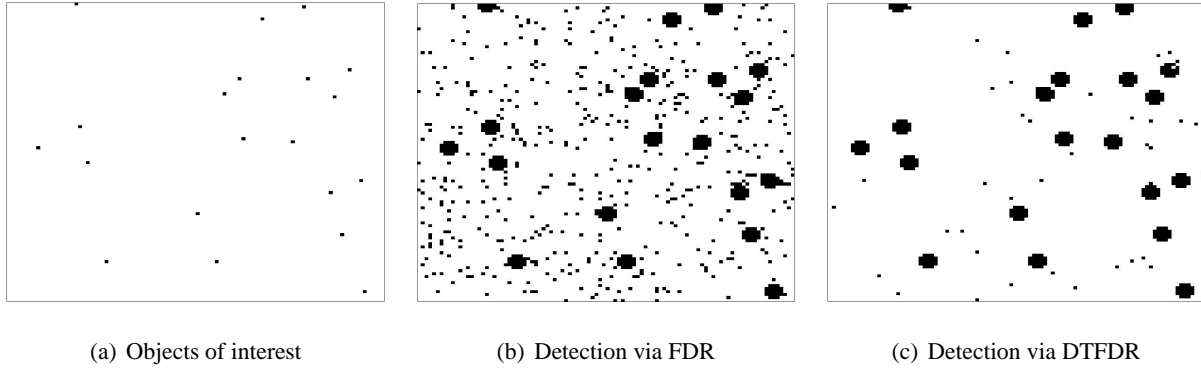


Figure 9: Detection performance of a sensor network using FDR and DTFDR procedures with a 1200-bit communication budget: The sensor field is a 100x100 grid with a sensor at each pixel. In (a) the black pixels are the objects of interest, in (b) and (c) the black pixels are the sensors that detect an object of interest. Using FDR procedure the sensor network detects the objects of interest, however there are many false alarms as well. DTFDR procedure minimizes the number of false alarms, and minimizes the number of bits wasted on communicating false alarms.

## 2 Discovery Problem: Distributed Detection, Estimation & Tracking

An appealing modeling framework for capturing important correlations among features generated by distributed sensors with different modalities is the use of graphical statistical models.

A graphical model is a collection of random variables whose joint distribution is specified by a set of potentials. Potentials are real-valued functions of realizations of these random variables and they determine the nature and the strength of correlations. Markov random fields (MRFs) are special random fields that possess a graphical structure. Namely, potentials of a MRF satisfy certain properties that lead to conditional independence relations with respect to cut-sets of the associated graph, as illustrated in Figure 10. Here we adopt such models with the interpretation that each node of the graph denotes a random quantity that pertains to a sensor measurement, and the graph structure connecting the nodes reflects first-order dependencies between the measurements of various sensors. In general graphical model structures are well-suited to account for the underlying state of the sensed medium. In particular, the dominance of local connectivity in such a network maps naturally to a sparsely connected graph structure. The idea of statistical inference on graphical models has been applied to situations that exhibit spatial variation using centralized processing, most notably in coding theory [13] and image processing [24] and more recently for fusion of audio-visual data for tracking moving objects [3]. A detailed discussion of graphical models and some applications to

distributed fusion in sensor networks is presented in [5].

Figure 10 illustrates a graphical model associated with the audio-visual fusion. Such a system may use video data, captured by a camera, to track the spatial location of the target based on its continually shifting image. If the target emits sound, such a system may use audio data, captured by a microphone pair (or array), to track the object location using the time delay of arrival of the audio signals at the different microphones. In principle, however, a tracker that exploits both modalities may achieve better performance than one that exploits either one or the other. The reason is that each modality may compensate for weaknesses of the other one. Thus, whereas a tracker using only video data may mistake the background for the target or lose the target altogether due to occlusion, a tracker also using audio data could continue focusing on the target by following its sound pattern.

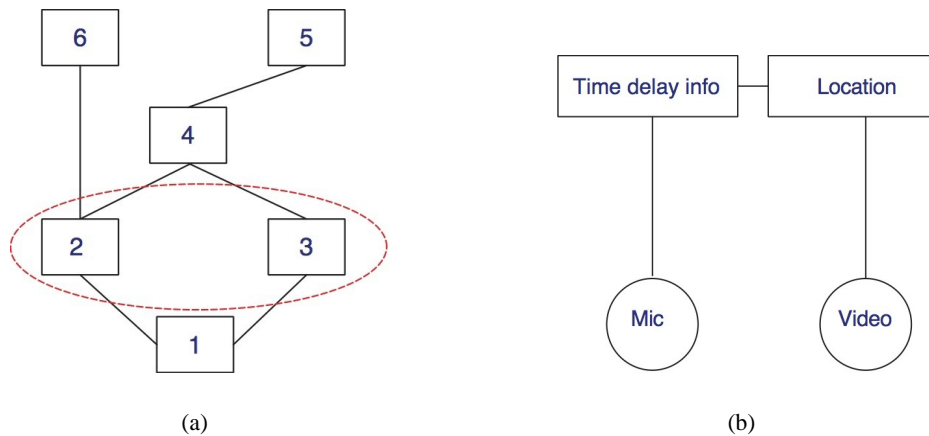


Figure 10: Illustration of independence in Graphical Models; (a): Feature 1 is independent of features 4,5,6 given the values of features 2 and 3. (b) illustrates a typical graphical model used in audio-visual fusion.

## 2.1 Fusion Algorithms

We present a distributed and dynamic approach that relies on local message passing and local filtering by individual sensors for fusion of multi-modal sensor information. Our technique uses belief-propagation to address problems of target localization, target identification, and target tracking.

The general question of dealing with distributed data in the context of detection has been an active topic of research (see [17, 18, 8] and references therein). Previously proposed techniques can be broadly categorized into two groups: The fusion-centric approach assumes that each sensor has a communication link to a data fusion center. Quantization of sensor data in this model was addressed by [17] and effects of power constraints on noisy communication channels were considered in [8]. The ad-hoc approach, on the other hand,

involves no designated fusion center but focuses on establishing consensus within the network via message exchanges. This approach is arguably more suitable to address energy issues in large-scale networks and also appears to have robustness advantages. Early work [7] establishes that consensus is achieved if messages are conditional expectations adopted to local measurements and messages, however the agreement itself is in general sensitive to the relative timing of messages, and computing the conditional expectations is not practically appealing. Message specifications and rigid messaging schedules that lead to consensus on optimal decisions were given in [16] for the special case of a completely connected communication topology.

We present an ad-hoc model and develop a data-oriented communication strategy in order to overcome the alluded issues. One of the main advantage of our scheme is that there is no single point of failure and that there are sufficiently large number of independent information channels that will not be degraded simultaneously. A natural idea for collaboration is to exchange a vector of individual sensor beliefs (that is, a sensor's estimate of the correct posterior probabilities) for different hypothesis between linked sensors at any instant of time. This idea is formalized by the belief propagation algorithm [12] in the context of artificial intelligence. Kalman smoothing is a special case of this idea applied to linear graphs with Gaussian potentials. The main problem is that belief propagation is known to work only for non-loopy network topologies, a situation that is quite restrictive and difficult to impose in a network sensing system. Nevertheless, our preliminary investigation points to the fact that for scenarios involving target tracking and detection, this issue does not arise. We explain this idea by means of an example involving distributed detection involving a collection of distributed noisy sensors observing a single phenomena, a situation that can be conveniently represented by Markov random fields. Details can be found in some of our recent papers [2, 20, 19, 1, 14]. We focus on a Bayesian hypothesis testing problem, which involves a set  $\{H_1, H_2, \dots, H_M\}$  of hypotheses, along with a known prior distribution  $\pi_o$  (the formulation extends easily to distributed estimation). We are interested in estimating the true hypothesis based on a collection of observations  $(Y_1, Y_2, \dots, Y_V)$  where each entry represents a measurement taken by a distinct sensor. Consider the case when the measurements are subject to uncorrelated observation noise, a situation captured by conditional independence of the observations given the true hypothesis. That is, the conditional distribution of the observation vector  $(Y_1, Y_2, \dots, Y_V)$  factors and it turns out that the posterior probability  $\pi(H_m)$  of hypothesis  $H_m$  based on these measurements satisfies

$$\pi(H_m) \propto \pi_o(H_m) \prod_{v=1}^V f_v(Y_v|H_m),$$

Our objective is to compute a MAP estimate, which is the hypothesis with the largest posterior probability.

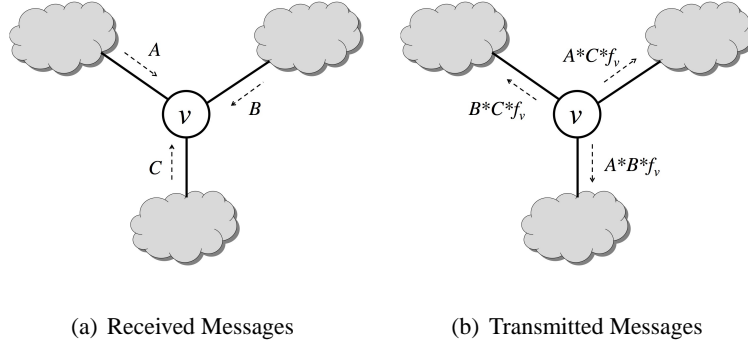


Figure 11: Illustration of Belief Propagation

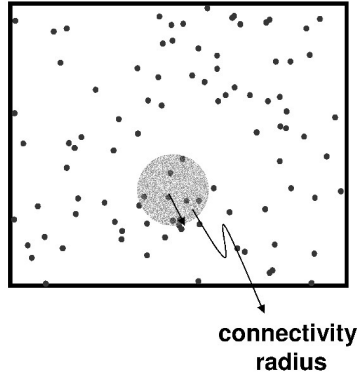
The collaboration between sensors is limited by a communication network structure represented by a weighted connected graph  $G = (V, E, C)$ . The vertices  $V$  of this graph correspond to sensors, the edges  $E$  correspond data links, and the weights  $C$  specify link capacities. Our focus in this section is to apply belief propagation in the context of our classification problem. In the mechanics of the algorithm, at each time step  $k$ , each sensor node,  $v'$ , forwards a message  $m_k^{v' \rightarrow v}(H_m)$  to the neighboring (linked) sensor node  $v$  about each hypothesis  $H_m$ . Namely, sensor node,  $v'$ , computes the product of all the messages pertaining to each hypothesis  $H_{m'}$  (excluding the message from  $v$ ), and averages this product across all hypothesis with weighting to reflect correlations between the hypothesis  $H_m$  and  $H_{m'}$ . In the instantiation of belief propagation pertaining to the present classification problem, these messages satisfy

$$m_{k+1}^{v \rightarrow \hat{v}}(H_m) = f_v(Y_v | H_m) \prod_{\text{neighbors } v' \text{ of } v \text{ except } \hat{v}} m_k^{v' \rightarrow v}(H_m) \quad (3)$$

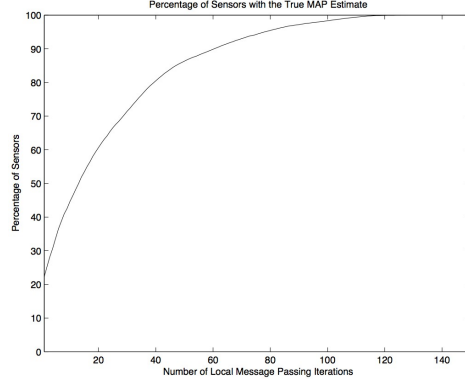
The nature of the messages are illustrated in Figure 11. Our focus is whether the scheme converges for the special structures endowed by the detection problem for general graphs. In contrast to other applications, belief propagation admits an exact analysis in the present instantiation because messages pertaining to distinct hypotheses evolve in an uncoupled fashion. Namely, it is evident from definition (3) that the logarithm of a message composed at the  $k$ th stage can be expressed as a linear function of the logarithms of the messages from the previous stage. This recursion is conveniently represented with a discrete-time linear system

$$x_{k+1} = Ax_k + b \quad (4)$$

where  $x_k = [x_k(e)]$  is a vector such that  $x_k(e)$  denotes  $\log(m_k^{v' \rightarrow v}(H_m))$  ( $e$  being the directed edge from  $v'$  to  $v$ ),  $b$  is a vector containing log-likelihoods  $\log(f_v(Y_v | H_m))$ , and  $A = [a_{e,e'}]_{E \times E}$  is binary matrix such that  $a_{e,e'} = 1$  if edge  $e$  leads to the origin of edge  $e'$ . Note that  $A$  is determined solely by the communication graph  $G$ , and it determines asymptotic properties of the recursion (4).



(a) Randomly distributed sensors with constant radius of communication connectivity



(b) Rate of convergence to consensus; Y-axis denotes percentage of sensors that have achieved consensus; X-axis denotes time index

Figure 12: Simulation of Distributed Detection for a 400-node randomly deployed sensing system for 10 different target hypotheses.

Also, since individual sensors could choose to either fuse their information or simply behave as relay, it follows that the dynamics of the fusion process is partially controllable. Based on these ideas we have established the following general result:

*Every communication graph topology admits a decentralized message passing algorithm, which converges to centralized MAP estimate under sufficiently small message quantization.*

A key feature of the present architecture is that individual sensor operation is based on local information. Namely, the message forwarding does not require global knowledge of sensor models. This stands in sharp contrast with previously proposed decentralized estimation schemes [8, 17, 7, 18]. Furthermore, the method is scalable to event-driven operation since it can be implemented by programming the sensor to send out its  $k$ th message only after receiving the  $(k - 1)$ th messages from all of its neighbors. Figure 12(b) illustrates how fast a consensus about the underlying phenomena/target arises in a connected network of randomly placed sensors.

**Distributed target tracking:** We close this section by pointing out that the principle of collaborative processing can be applied to dynamic problems wherein sensors take periodic measurements pertaining to an evolving phenomena. The canonical objective here is to track targets whose evolution is described by a linear dynamical system through sensors that observe noisy locations and exchange messages by means of a multi-hop protocol over a communication network. Several in-network processing techniques were studied



for this scenario in [1] with emphasis on the relative time-scales of measurements and communication. These techniques differ from existing techniques in two important aspects: a) there is no designated leader/fusion node and each sensor attempts to optimally track the system trajectory based on its local observations and time-dependent information available from other sensors in the network; b) the message computation at each sensor is structurally identical, where the computed message from each sensor is the innovation in the state conditioned on all the information available upto that time at each sensor. Consequently, the sensor network can be queried at any time and at any node to obtain optimal estimates for the state of the dynamical system.

## 2.2 Channel Errors and Packet Losses

Typical applications of sensor networks involve inexpensive and unreliable sensors that communicate through low capacity wireless links. Viable sensing systems should therefore be robust against attendant effects of these restrictions. We illustrate graceful degradation of distributed detection with such channel imperfections. First, it is relatively easy to establish robustness to finite bit rate constraints by resorting to quantization of messages. We have analyzed effects of such quantization and established design guidelines to achieve prescribed insensitivity to quantization in [14]. It turns out that the maximum admissible quantization level is proportional to the minimum Kullback divergence between any two conditional probability density functions of the observation  $Y$  under each hypothesis  $H_i$ .

Another aspect of wireless transmission is that only a few messages can be reliably transmitted along each communication link at each round of the algorithm. Our aim here is to account for the following two effects: First, messages may be corrupted and lost due to imperfections in point-to-point communication. Although link layer protocols would provide some relief against this issue, robustness of network operation against message losses needs to be addressed, especially if the physical communication medium is wireless. Secondly, one can imagine situations where some sensors operate on a slower time-scale than others, thereby slowing down the network under the lock-step message-passing algorithm outlined in Section 2.1. This limitation may be overcome if each sensor contributes to the collaborative effort at its own time-scale. In both cases described above the network operation is asynchronous in the sense that not all links are necessarily active at each round of the algorithm. In addition, the network could also be time-varying. The main idea used in overcoming packet loss effects (see [14] for more details) is to normalize the messages based on the number of outgoing functional links (i.e., the number of receivers of the message) in the same round. We point out that such an algorithm is consistent with the currently employed wireless protocols. This leads

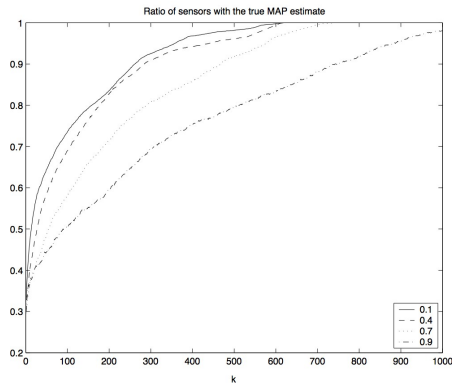


Figure 13: Convergence rate of consensus for different packet loss probabilities; Y-axis denotes percentage of sensors that have achieved consensus; X-axis denotes time index

to a situation, wherein each message is not only weighted equally but also produces information channels that will not be degraded simultaneously. Our results on graceful impact of packet losses is illustrated in Figure 13.

### 2.3 Energy Scaling

A fundamental question that arises in a distributed sensing system is energy required to form fusion estimates. Energy not only impacts the communication bit rate but also the lifetime of a sensing system. In this thrust we explore energy scaling issues for large-scale deployment of distributed sensing systems. We have conducted preliminary investigation for the fusion schemes presented in Section 2.1, which we describe below.

The fusion scheme is based on refinement of information at each time step and at each sensor location. In effect, we differentiate between data and information, and successively blend refinement and transportation of data in the course of the algorithm. The time to reach consensus via this operation is on the order of the diameter of the communication network [20, 19], a quantity that lower bounds the time to collect all observations at any designated node for centralized processing. Hence the distributed approach does not entail significant performance costs, whereas it evidently provides advantages in conforming to the scant transport capacity of wireless networks.

Figure 12(b) illustrates a network of 400 sensors in additive Gaussian noise. The task is to classify among 10 different hypotheses on a mean field. Figure 12(b) illustrates the dramatic convergence to consensus in a typical realization of noisy sensor measurements. Comparisons between conventional decentralized detection scheme and the consensus based scheme appears in the Table 1. An energy requirement of  $d^4 E_b$

Scheme	Energy (Joules/Node)
Decentralized Broadcast	$\mathcal{O}(N^{\alpha/2} d_0^\alpha E_b)$
Decentralized Multihop	$\mathcal{O}(\sqrt{N} d_0^\alpha E_b)$
Belief Propagation (grid)	$\mathcal{O}(N \log N d_0^\alpha E_b)$
Coalescent Random Walks	$\mathcal{O}(\log(N) d_0^\alpha E_b)$

Table 1: Energy scaling for different schemes;  $d_0$  is the internode distance;  $E_b$  is the energy required to transmit 1-bit over a unit distance;  $\alpha$  is the attenuation coefficient for transmit energy in free space

is assumed for transmitting a single bit over a distance  $d$ . Decentralized detection operates under a fusion network and local decisions from individual nodes are transmitted. The consensus approach shows dramatically improved performance. Furthermore, a with exponentially better performance in terms of energy scaling can be obtained as seen in the last row of Table 1 through a token based approach that we discuss in the following section. These facts establish that *utilizing high local data rates combined with information refinement algorithms such as Kalman Filtering can lead to significant gains in energy efficiency.*

### 2.3.1 Token-based Algorithms for Data Aggregation

The discovery problem in Section 2.1 dealt with a special type of data aggregation, namely, computing the sum of each sensor’s log-likelihoods. In general data aggregation in sensor networks typically involve distributed computation of functions such as maximum, sum, product, and weighted averages of individual sensor values. Such functions admit flexible decompositions in terms of pairwise operations, and this property can be exploited to obtain distributed computation algorithms that display substantial advantages in terms of energy consumption. We give a description of such algorithms by specializing to case of computing the sum of sensor values, for notational convenience.

Each sensor maintains a variable *estimate* that is updated as the sensor exchanges messages with its neighbors. The initial value of *estimate* is equal to a locally generated value whose network-wide sum is of interest. Messages carry the most recent value of *estimate* at the transmitter; the receiving sensor adds this value to its own *estimate* whereas the transmitting sensor sets its *estimate* to 0. Note that the sum of *estimates* of all sensors remains fixed at the number of sensors with significant observations, and eventually a single sensor will have a non-zero *estimate*. The algorithm terminates at this point. To ensure that this computation entails minimal energy consumption, message transmissions in the network are regulated by *tokens*. Namely, messaging decisions are taken independently and asynchronously at individual sensors, but

a sensor can transmit a message only if it is holding a token. The token then moves with the transmitted message. If the receiver also has a token at that time then the two tokens coalesce into one. The algorithm is guaranteed to terminate in a connected network, and it provides dramatic gains in message complexity relative to other distributed algorithms in planar topologies [21].

From an analytical perspective, the termination time of the algorithm is on the order of the cover time of a random walk if there is a single token in the network, and of the hitting times of a random walk if there are multiple tokens. Termination times of belief propagation and gossip algorithms, on the other hand, are on the same order as the mixing time of random walks. The net effect is that token based algorithms typically take longer to terminate but they entail far fewer message transmissions as only a few sensors are active at any time. Application of this algorithm for the setting of discovery problem leads to exponential savings of energy as illustrated in Table 1.

### **3 Conclusions**

In this article we identify two classes of sensor network problems that arise naturally in many applications. Fusion problem involves correlated data over the sensor network, while search/sensor selection involves selecting sensors that have useful information. The inherent assumption in fusion problem is that the sensor locations are known as well as the statistical relationships between the data and the features of observable phenomenon. The use of graphical models along with asynchronous message passing algorithms prove to be a very natural approach to solving target tracking, change detection, and target identification problems, all of which can be formulated in the fusion problem framework. Search problems in its extreme case involves problems of making decisions when sensor data could be uncorrelated across sensors. Nonetheless, we demonstrate that uncorrelated data across sensors still necessitates collaboration. In fact, depending on the cost structure, elaborate decision regions need to be devised, which require joint mapping of observations to decisions. This joint mapping in turn requires collaboration of sensors. We further describe an adaptive decision making system, which feeds back the previous decisions of the sensor network to make more informed decisions in the succeeding steps of the algorithm. The ideas that are developed for fusion problems prove to be useful in successfully executing the described algorithm. Problems such as target tracking in clutter and sensor selection can be fitted into the framework of this class.

A very important aspect of our work is the assumption that the sensors have only local information. For example, no sensor has access to the measurement model of another sensor. Therefore, all the solutions that

are outlined here are robust to variation of observation models across sensors. This fact has a very strong implication: to solve either the search problem or the fusion problem no communication of information is necessary other than that relevant to the observable phenomena.

## 4 Acknowledgements

This work was supported by PECASE grant no. N00014-02-100362, NSF grants CCF 0430983, CCF 0530572, ANI 0238397, and ECS 0449194.

**Erhan B. Ermis** received his B.S. degree from Purdue University, Indiana in 2003 and his M.S. degree from Boston University, Massachusetts in 2005 both in Electrical Engineering. He is currently pursuing a Ph.D. degree in electrical engineering at Boston University. His research interest is focused on distributed signal processing and statistical methods for decision making in sensor networks. He is a student member of IEEE.

**Murat Alanyali** received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1996. He was with Bell Laboratories, Holmdel, NJ, during 1996-1997, and with the Department of Electrical and Electronics Engineering at Bilkent University, Ankara, Turkey, from 1998 to 2002. Since 2002 he has been an Assistant Professor in the Department of Electrical and Computer Engineering in Boston University. His research interests are primarily in communication networks, with emphasis on their control and analysis. Dr. Alanyali is a member of IEEE and a recipient of NSF CAREER award.

**Venkatesh Saligrama** is an Associate Professor in the Department of Electrical and Computer Engineering at Boston University. Prior to joining Boston University in 2001, he was a Research Engineer at United Technologies Research Center. He received his Ph.D. from MIT and the B.Tech degree from IIT Madras. He is the recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE), the Office of Naval Research Young Investigator Award, the NSF CAREER Award and the Outstanding Achievement Award from United Technologies. He has been an Associate Editor for IEEE Transactions on Signal Processing since 2004. His research interests are in information & control, and distributed signal processing.

## References

- [1] M. Alanyali and V. Saligrama, Distributed Tracking with Multi-hop Networks. In *IEEE Statistical Signal Processing Workshop*, Bordeaux, France, July 2005.
- [2] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron, Distributed Bayesian Hypothesis Testing in Sensor Networks. In *American Control Conference*, Boston, MA, July 2004.
- [3] M. J. Beal, N. Jovic and H. Attias “A Graphical Model for Audiovisual Object Tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, July 2003
- [4] B. Chen, L. Tong, P. Varshney, ”Channel Aware Distributed Detection in Wireless Sensor Networks,” *IEEE SPM Special Issue on Distributed Signal Processing in Sensor Networks*, 2006
- [5] M. Cetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, A. S. Willsky, ”Distributed Fusion in Sensor Networks: A Graphical Models Perspective,” *IEEE SPM Special Issue on Distributed Signal Processing in Sensor Networks*, 2006
- [6] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society, Series B*, vol. 57, pp. 289-300, 1995.
- [7] V. Borkar, P. Varaiya, “Asymptotic agreement in distributed estimation,” *IEEE Transactions on Automatic Control*, 27, 1982
- [8] J. F. Chamberland and V. V. Veeravalli, “Decentralized detection in sensor networks,” *IEEE Transactions on Signal Processing*, 2003.
- [9] E. Ermis, and V. Saligrama, Decentralized Detection and Estimation of localized phenomena In *ICASSP*, Philadelphia, PA, March 2005.
- [10] E. Ermis, and V. Saligrama, Adaptive Sampling Strategies for Distributed Detection of Localized Phenomena. In *Information Processing for Sensor Networks*, Los Angeles, CA, April 2005.
- [11] E. Ermis, Adaptive Statistical Strategies for Distributed Detection of Localized Phenomena in Sensor Networks. *M.S. thesis*, Department of Electrical and Computer Engineering, Boston University, May 2005.
- [12] M.I. Jordan, Ed., “Learning in graphical models,” MIT Press, 1998.

- [13] R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's "belief propagation" algorithm," *Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 140–151, Feb. 1998.
- [14] V. Saligrama, M. Alanyali and O. Savas, Asynchronous Distributed Detection. *BU Technical Report 2005-03. To appear in IEEE Transactions on Signal Processing.*
- [15] D. Slovin, L. Pham, M. Alanyali and V. Saligrama, Location tracking through Belief Propagation with Applications to Ecology and Environmental Biology. Available at: [http://www.xbow.com/General\\_info/Info\\_pdf\\_files/Smart\\_Dust\\_finalists.pdf](http://www.xbow.com/General_info/Info_pdf_files/Smart_Dust_finalists.pdf), January 2005.
- [16] P. F. Swaszek and P. Willett. Parley as an approach to distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, pages 447–457, January 1995.
- [17] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas Eds, vol. 2.
- [18] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer, 1997.
- [19] S. Venkatesh and M. Alanyali, M-ary hypothesis testing in sensor networks. In *Conference on Information Sciences and Systems*, Princeton, NJ, February 2004.
- [20] S. Venkatesh, M. Alanyali, O. Savas, and S. Aeron, Classification in sensor networks. In *International Symposium on Information Theory*, Chicago, IL, June 2004.
- [21] O. Savas, M. Alanyali, V. Saligrama, Efficient In-Network Processing through Local Ad-hoc Information Coalescence *BU Technical Report 2006.*
- [22] S. Venkatesh, Y. Shi, and W. Karl, Performance Guarantees in Sensor Networks In *ICASSP*, Montreal, Canada, May 2004.
- [23] M. Williams, E. B. Rastetter, D. N. Fernandes, "Modelling the soil-plant-atmosphere continuum in a Quercus-Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties," *Plant, Cell and Environment*, 19, 911-927, 1996.
- [24] J. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Information. Theory*, Vol. IT-18, pg. 232-240, 1972