ASYMPTOTICALLY CONSISTENT ONE-BIT DETECTION IN LARGE SENSOR NETWORKS

Paolo Braca, Stefano Marano, and Vincenzo Matta

Department of Electronic and Computer Engineering (DIEII), University of Salerno
via Ponte don Melillo I-84084, Fisciano (SA), Italy
email: {pbraca, marano, vmatta}@unisa.it

ABSTRACT

Consider a wireless sensor network where n remote nodes deliver observed data to a common fusion center. A recent work by Blum and Sadler [1] proves that, by means of an ordered transmission policy, optimal detection performance is achieved when only a subset of the n samples has been received at the fusion center.

Here we show that, when the network size n is made large enough, arbitrarily small detection errors can be achieved with a fusion center receiving just the first sample out of n: One-bit detection can be asymptotically consistent. In our scheme the "winner takes all": the network global decision is the local decision of the first firing sensor, and key is the sample selection criterion — ordered transmissions — where more informative observations are delivered first.

1. INTRODUCTION

Consider a distributed detection system in which the remote nodes collect observations about a commonly monitored state of the nature for binary detection purposes [2–4]. These observations are sent to a fusion center to which the final decision is demanded, and a certain level of detection performance is achieved that clearly depends upon the network size, i.e., the total number of sensors or collected samples. In a recent work Blum and Sadler [1] showed that, surprisingly, nothing is lost in terms of detection capability when only a subset of the observations are delivered to the fusion center, provided that such subset contains the most "informative" samples. This can be achieved in practice by exploiting the idea of ordered transmissions: Each remote sensor delivers the observed sample towards the fusion center after a time interval that is inversely proportional to the informativeness of the sample. As soon as a certain number of sensors' messages are received, the fusion center can inhibit further transmissions, making its final decision with the same performance that would be obtained using the whole set of data. The number of sensors' messages (i.e., the number of samples used for the decision) is a random variable whose value is smaller than the network size: transmissions can be saved without degradation in the detection performance.

The focus of [1] is to maintain good performance given the number of available sensors. Many wireless sensor networks are composed by relatively simple, tiny and cheap remote devices such that a large number of such devices can be easily deployed. The constraint on the total number of sensors is not that tight and, therefore, one can improve performance by increasing the network size. In this case, it makes sense to relax the network size constraint and to take to one extreme the ordered transmission idea of [1]. In other words, reversing the perspective, it makes sense to make the final decision using *only the first, most informative, received sample*: we wonder if acceptable performance can be achieved in this way, provided that the network size is large enough. The main finding of this paper is that in this distributed detection system the decision based on one single sample is asymptotically consistent with increasing network sizes.

The provoking title of this work emphasizes that *just one sample* (loosely speaking "one bit") is sufficient for *asymptotically* optimal performance, provided that the ensemble from which the sample is extracted is sufficiently large (the term asymptotically is referred to the cardinality of the ensemble) and that the selection criterion is suitably chosen. Clearly, since one single observation is to be collected, there is the further benefit that the sensor can send its local decision — thus delivering just one bit — rather than the continuous-valued observed sample. It is worth noting that the communication task is one of the major, and often the largest, source of energy expense in sensor networks [5]. Our one-bit system is extremely efficient in this respect.

It is evident that the approach of this work bears similarities to the censoring detection schemes initially proposed by [6]. The censoring idea consists of inhibiting the transmission of poorly informative sensors, by using some convenient measure of informativeness at the remote nodes. Our detector can be thought as one where just one sample survives to the censoring, which is implemented by appropriately tuning the delivering time.

In this paper we address the problem from a rather theoretical viewpoint emphasizing methodological and analytical tools. Practical implementations of the proposed idea would require addressing several important issues, such the following. First, it should be noted that some feedback mechanism from the fusion center to the nodes must be implemented in order to inform all the sensors that one sample has been received and therefore the detection task is terminated, i.e., all the sensors can be switched off. This halting protocol can be conceived in many ways, for instance by sending a broadcast message or by a multi-hop message passing. Certainly this implies an energy burden, but at the expense of the fusion center that is usually much more powerful than the remote nodes. In addition, one should consider the specific network topology, as well as possible channel impairments. Special attention should be paid also to the system synchronism: a common time origin must exist, and local clocks must be aligned, thus ensuring that the fusion center actually receives the *first* fired message (even when multi-hop protocols are assumed). Finally, although our analysis is asymptotic, we assume independence among the measurements, for mathematical tractability. Some of these issues are dealt with in [7].

2. PROBLEM FORMALIZATION

Consider a distributed system made of n remote units that sense the environment to solve a binary hypothesis test, i.e., to decide which of two possible states of the nature $\mathcal{H}_0$ or $\mathcal{H}_1$, is in force. Each node, say node i, collects a scalar random variable X_i taking values in the reals with associated probability density function (pdf) $f_X(x)$ with unbounded support, i.e., $\sup_x\{f_X(x) > 0\} = \infty$ and $\inf_x\{f_X(x) > 0\} = -\infty$. More specifically, X_i is drawn from $f_X(x;\mathcal{H}_0)$ if the nature manifests $\mathcal{H}_0$, while it is drawn from $f_X(x;\mathcal{H}_1)$ under $\mathcal{H}_1$. The statistical test can be cast in the usual form: for $i = 1, 2, \dots, n$,

$$\begin{array}{ll} \mathcal{H}_0: & X_i \sim f_X(x;\mathcal{H}_0), \\ \mathcal{H}_1: & X_i \sim f_X(x;\mathcal{H}_1). \end{array}$$

To this detection problem two errors are associated

$$\begin{aligned} \alpha_n &= \mathbb{P}\left(\text{decide } \mathcal{H}_1, \text{ when } \mathcal{H}_0 \text{ is actually true}\right), \\ \beta_n &= \mathbb{P}\left(\text{decide } \mathcal{H}_0, \text{ when } \mathcal{H}_1 \text{ is actually true}\right), \end{aligned}$$

that are usually referred to as the false alarm and the miss detection probabilities, and it is well known that the log-likelihood test, i.e., one that compares

$$\sum_{i=1}^{n} L(x_i) = \sum_{i=1}^{n} \log \frac{f_X(x_i;\mathcal{H}_1)}{f_X(x_i;\mathcal{H}_0)}.$$

with some threshold, optimally solves the problem under a variety of optimality criteria based on α_n and β_n. On the other hand, when computing the log-likelihood is difficult or impossible (as in nonparametric tests) or in some specific context such as locally optimum detection, one approach is to seek for a suitable data transformation, say $T(x)$, as a surrogate of the optimal data-transforming nonlinearity $L(x)$. One obvious choice for $T(x)$ is the identity and in fact in the examples developed in Sect. 5 we consider the two possibilities

$$\begin{array}{ll} T(x) = x & \text{identity}, \\ T(x) = L(x) & \text{log-likelihood}. \end{array} \tag{1}$$

Note, however, that the general results of this work hold for an arbitrary nonlinearity $T(x)$.

Recall that in the designed system the fusion center does not receive the entire vector $(X_1, X_2, \dots, X_n)$ because of the time-based censoring strategy, so that neither the optimal $\sum_i L(x_i)$ nor the surrogate $\sum_i T(x_i)$ is available. Rather, sensor i after computing $T(X_i)$ makes a local decision according to $u\left(T(X_i) > \gamma_n\right)$, where $u(\cdot)$ is the unit step function and γ_n is a properly set threshold.

Furthermore, sensor i attempts to contact the fusion center at a time instant proportional to $1/|T(X_i)|$, measured with respect to a common time origin. We ignore the constant of proportionality since it is immaterial in our mathematical analysis, even though in practice it must be carefully chosen as function of the network size, to comply with real-time and time-resolution constraints of the system.

When, at time $1/|T(x)|$, sensor i contacts the fusion center, it delivers not the observed sample but the local decision, i.e., the binary digit $u\left(T(X_i) > \gamma_n\right)$ based on the single sample X_i. But this one-bit delivering actually takes place only if a halting broadcast message from the fusion center is not heard. Since this halting message is broadcast as soon as the fusion center receives the first message from the sensors, there is only one node that actually delivers its decision, and that *local* decision is taken as the *global* decision of the system. The point here is that the informativeness of the sample is just measured in terms of $|T(X_i)|$, so that the first firing sensor is expected to convey the maximum of information.

Since the delivering time is related to the modulus of the transformation, we refer to our transmission policy as the MO (modulus ordered) strategy; when in addition $T(x) = L(x)$ the network system is referred to as the ℓ-MO (log-likelihood modulus ordered) system.

We now introduce some definitions. First, let $Z_i{:=}T(X_i)$ be the transformed variable and let $F_Z\left(x;\mathcal{H}_j\right)$ and $f_Z\left(x;\mathcal{H}_j\right)$ be its cumulative distribution function (cdf) and pdf , respectively, when hypothesis $\mathcal{H}_j = 0, 1$, is in force. The modulus ordering can be formalized in terms of the index permutation $\pi(\cdot)$ defined by the property that

$$\left|Z_{\pi(1)}\right| \le \left|Z_{\pi(2)}\right| \le \cdots \le \left|Z_{\pi(n)}\right|. \tag{2}$$

We define $\mathcal{M}_n := Z_{\pi(n)}$ the sample with maximum modulus, which plays a special role in the analysis. Indeed, since the decision statistic of the whole system is that of the first (and unique) firing sensor, the global decision amounts to

$$\mathcal{M}_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_n, \tag{3}$$

where γ_n is a suitable decision threshold. Letting

$$\mathcal{M}_n^+ {:=} \max_{k \le n} Z_k, \qquad \mathcal{M}_n^- {:=} -\min_{k \le n} Z_k = \max_{k \le n}(-Z_k),$$

the decision statistic in (3) can be written as

$$\mathcal{M}_n = \left\{ \begin{array}{rll} \mathcal{M}_n^+ & \text{if} & \mathcal{M}_n^+ \ge \mathcal{M}_n^- \\ -\mathcal{M}_n^- & \text{if} & \mathcal{M}_n^+ < \mathcal{M}_n^- \end{array} \right. . \tag{4}$$

Note that choosing the delivering time on the basis of the modulus of $T(x)$ has the precise rationale that the firing sensor corresponds either to the largest sample $\mathcal{M}_n$ or to the smallest $-\mathcal{M}_n^-$, a fact that is key for the subsequent analysis.

Below we make use of the following results [8]:

$$f_{\mathcal{M}_n^+}\left(x;\mathcal{H}_j\right) = n\, F_Z^{n-1}\left(x;\mathcal{H}_j\right) f_Z\left(x;\mathcal{H}_j\right), \tag{5}$$

$$f_{\mathcal{M}_n^-}\left(x;\mathcal{H}_j\right) = n\left[1 - F_Z\left(-x;\mathcal{H}_j\right)\right]^{n-1} f_Z\left(-x;\mathcal{H}_j\right), \tag{6}$$

where capital letters denote the cdf and the corresponding lowercase the pdf of the random variable specified as subscript. Also we shall use [9]

$$f_{\mathcal{M}_n}(x;\mathcal{H}_j) = n\, h_j^{n-1}(x)\, f_Z(x;\mathcal{H}_j), \tag{7}$$

where

$$h_j(x) = F_Z(|x|;\mathcal{H}_j) - F_Z(-|x|;\mathcal{H}_j). \tag{8}$$

3. RELEVANT EVT BACKGROUND

Some relevant facts from the classical literature of Extreme Value Theory (EVT) are now recalled. Let $Y_1, Y_2, \dots, Y_n$ be a collection of iid random variables with unbounded support, and let $M_n = \max_{i\le n} Y_i$. Then, under mild regularity conditions, there exist sequences of normalizing constants a_n, b_n such that [10]

$$\lim_{n\to\infty} F_{M_n}(a_n\, x + b_n) = G(x), \tag{9}$$

where $G(x)$ is either the Gumbel or the Fréchet distribution[1]:

$$\text{Gumbel} \quad \begin{cases} G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty, \\ b_n = F_Y^{-1}\left(1-\frac{1}{n}\right), \qquad a_n = \frac{1}{n\, f_Y(b_n)}, \\ \lim_{n\to\infty}(b_n/a_n) = \infty, \end{cases}$$

$$\text{Fréchet} \quad \begin{cases} G(x) = \begin{cases} 0 & x \le 0 \\ \exp(-x^{-\xi}) & x > 0, \end{cases} \qquad \xi > 0, \\ b_n = 0, \quad a_n = F_Y^{-1}\left(1-\frac{1}{n}\right), \\ \lim_{n\to\infty} a_n = \infty. \end{cases}$$

When eq. (9) holds, we say that M_n (or equivalently its distribution) is attracted to the limiting distribution $G(x)$ with normalizing constants a_n and b_n. Finally, given a random variable Y, we have the following definitions. If

$$\lim_{x\to\infty} \frac{1-F_Y(x)}{F_Y(-x)} = \begin{cases} \infty & \\ & \text{then} \\ 0 & \end{cases} \begin{matrix} Y \text{ is right-tail dominant,} \\ \\ Y \text{ is left-tail dominant.} \end{matrix}$$

4. ASYMPTOTIC ANALYSIS

The main result is now stated. Consider a network of size n with independent and identically distributed $Z_i = T(X_i)$, $i = 1, 2, \dots, n$, that have unbounded support. Suppose that $\mathcal{M}_n^+$ is attracted under $\mathcal{H}_1$ and let a_n^+, b_n^+, and $G^+(x)$ be the normalizing constants and the limiting distribution, respectively. Similarly, suppose that $\mathcal{M}_n^-$ is attracted under $\mathcal{H}_0$ and let a_n^-, b_n^-, and $G^-(x)$ be the normalizing constants and the limiting distribution, respectively. Note that attraction is guaranteed under very broad technical conditions so that these assumptions are by no means restrictive for practical applications [10].

Theorem 1 *If the independent and identically distributed random variables Z_i's are right-tail dominant under $\mathcal{H}_1$ and left-tail dominant under $\mathcal{H}_0$, then the following convergences in probability hold*

$$\frac{\mathcal{M}_n}{\mathcal{M}_n^+} \to 1, \quad \mathcal{M}_n - \mathcal{M}_n^+ \to 0, \qquad \text{under } \mathcal{H}_1, \tag{10}$$

$$\frac{\mathcal{M}_n}{\mathcal{M}_n^-} \to -1, \quad \mathcal{M}_n + \mathcal{M}_n^- \to 0, \qquad \text{under } \mathcal{H}_0, \tag{11}$$

and the following attractions hold

$$\lim_{n\to\infty} F_{\mathcal{M}_n}\left(a_n^+\, x + b_n^+;\mathcal{H}_1\right) = G^+(x), \tag{12}$$

$$\lim_{n\to\infty} F_{\mathcal{M}_n}\left(a_n^-\, x - b_n^-;\mathcal{H}_0\right) = 1 - G^-(-x). \tag{13}$$

$\diamond$

Proof idea. The proof is given in [7] and cannot be offered here for space reasons. The main idea behind the proof, however, is simple. The local decision of the firing sensor (3) is based either on the largest $\mathcal{M}_n^+$ or on the smallest $-\mathcal{M}_n^-$ of the n (transformed) samples collected by the system. However, for n sufficiently large, if the right tail of the distribution $f_Z(x;\mathcal{H}_1)$ dominates over the left tail, then, with high probability, the decision statistic is $\mathcal{M}_n^+$. With left-tail dominance, conversely, the decision statistics is highly likely to be $-\mathcal{M}_n^-$. Under the assumptions of the theorem, we conclude that, roughly speaking, the hypothesis test compares a very large positive sample against a very small negative value, yielding good test performance.

This reasoning is based on the assumption of "large" n and, therefore, one expects that the detection error can be controlled as desired by choosing a sufficiently large n. To turn these informal arguments into a rigorous proof, in [7] we elaborate on sequences of events like $\{\mathcal{M}_n^+ \ge \mathcal{M}_n^-\}$ and investigate their probability as n grows. By judicious use of the convergences in probability and in distribution, we are finally able to give a formal proof of the theorem. $\bullet$

From the above arguments we argue that a reasonable threshold setting would be $\gamma_n = 0$, for any n, because for sufficiently large n the two decision statistics tend to be of different sign, under the two hypotheses. This is very convenient for approaching certain nonparametric problem where the distributions of the data are unknown or only partially known. (In these problems, we recall, $T(X)$ cannot be chosen as the log-likelihood.) In the following we refer to $\gamma_n = 0$ as the nonparametric threshold setting.

When knowledge of the statistical model is available, the most natural approach to set a threshold is perhaps to enforce a desired level of false alarm (in the spirit of the Neyman-Pearson criterion), i.e., $\alpha_n = \alpha$ for any finite n. Unfortunately, in many cases, this is not possible because the exact knowledge of the cdf of the detection statistic for any finite n would be required, that is usually unavailable. On the other hand, it is possible to use the asymptotic "similarity" (under $\mathcal{H}_0$) between $\mathcal{M}_n$ and $-\mathcal{M}_n^-$ to set the threshold as $(1 - F_Z(\gamma_n;\mathcal{H}_0))^n = \alpha$. Alternatively, exploiting Theorem 1 we can enforce an asymptotic false-alarm level α of the detection system using only the asymptotic distribution under $\mathcal{H}_0$. This yields the threshold of the test

[1] There exists a third limit distribution, namely the Weibull, which we do not consider here because its class of attraction is ruled out by the assumption that the support of the random variables is unbounded.

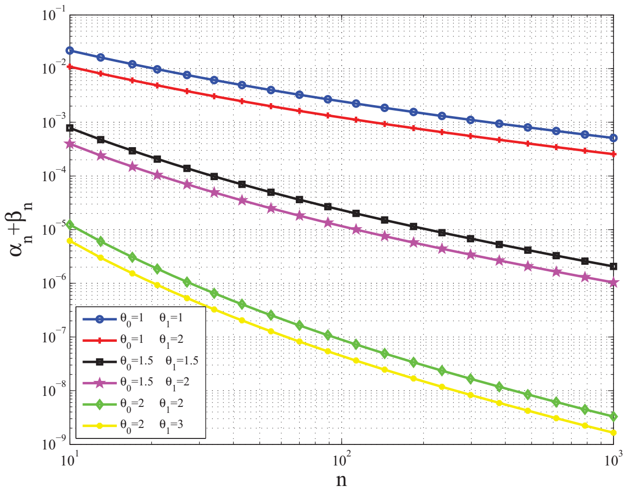

Figure 1: MO transmission policy. Error $\alpha_n + \beta_n$ for the Gaussian nonparametric example, with $\sigma = 1$ and different combinations of θ_0 and θ_1.

in the form $\gamma_n = a_n^- \gamma - b_n^-$, with

$$\alpha = G^-(-\gamma) \quad \gamma = \begin{cases} \log\log(1/\alpha) & \text{Gumbel,} \\ -(\log(1/\alpha))^{-\frac{1}{\xi}} & \text{Fréchet.} \end{cases} \tag{14}$$

We are now ready to present the result that shows the asymptotic consistency of the one-bit detection scheme. The proof, founded on the results of Theorem 1, is omitted here and the reader is remanded to [7].

Theorem 2 *Under the assumptions of Theorem 1 we have the following.*
i) Setting $\gamma_n = 0$ (nonparametric threshold), we get

$$\alpha_n + \beta_n \to 0. \tag{15}$$

ii) Let the detection threshold be either

$$\gamma_n = a_n^- \gamma - b_n^- \tag{16}$$

where γ is as in (14), or

$$\gamma_n = F_Z^{-1}\left(1 - \alpha^{\frac{1}{n}}; \mathcal{H}_0\right). \tag{17}$$

Then

$$\alpha_n \to \alpha, \quad \text{and} \quad \beta_n \to 0. \tag{18}$$

◇

The above theorems are valid for a general MO transmission policy, but for specific detection problems and/or local transformations, further results can be derived. For instance, consider the case of an ℓ-MO strategy (namely, when Z_i is the log-likelihood of the observations) applied to the shift-in-mean problems of the kind

$$f_X(x;\mathcal{H}_0) = \phi(x+\theta_0), \quad f_X(x;\mathcal{H}_1) = \phi(x-\theta_1) \tag{19}$$

where $\phi(x)$ is an even function, $\phi(x) > 0\ \forall x$, $\theta_0 \geq 0$ and $\theta_1 > 0$. For this case, along with the previous results, we can more or less easily bound β_n as follows:

$$\beta_n \leq e^{\gamma_n}. \tag{20}$$

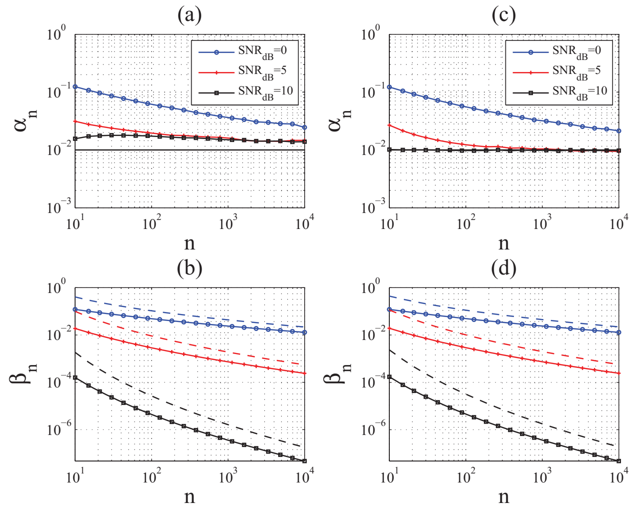

Figure 2: ℓ-MO transmission policy. Panels $(a)-(b)$ and $(c)-(d)$ refer respectively to the threshold setting given in (16) and (17). In the lower plots are shown, as dashed lines, the miss detection bounds (20).

5. EXAMPLES

We now illustrate by computer experiments the theoretical results obtained in the previous section. Let $\mathcal{N}(a,b)$ be a shortcut to denote the Gaussian distribution with mean a and standard deviation b, and consider the observation model

$$\begin{aligned} \mathcal{H}_0 &: X_i \sim \mathcal{N}(-\theta_0, \sigma), \\ \mathcal{H}_1 &: X_i \sim \mathcal{N}(\theta_1, \sigma), \end{aligned} \tag{21}$$

where θ_0, θ_1 and σ are positive parameters. Let us focus on the MO policy with $T(x) = x$ first. The main claim of Theorem 2 is that the one-bit distributed detector is consistent, in the sense that both the error probabilities go to zero when the network size n grows, and in fact this is true even in the case where all the model parameters θ_0, θ_1, σ and n, are unknown. Setting $\gamma_n = 0$ as prescribed in part $i)$ of Theorem 2 and exploiting expression (7) for numerical integration we get the results summarized in Fig. 1.

When the statistical model (21) of the observations is perfectly known, the ℓ-MO transmission policy can be implemented, with the results shown in Fig. 2. By defining the signal to noise ratio SNR=$(\theta_1+\theta_0)/\sigma$ we see that the larger is SNR, the faster the error probabilities go to zero as indicative of the fact, not unexpected, that tail dominance is emphasized at large SNRs. Panels (a) and (b) refer to the threshold setting given in (16), while (c) and (d) refer to the threshold in (17). We see that α_n converges to the desired asymptotic value; however, in (c) the convergence is somehow faster than that in (a), suggesting that the threshold setting (17) provides, in this example, some advantage. Note that the curves in (b) and (d) are very similar, which reveals that the threshold selection is not critical with respect to the miss error probability. Since the observation model belongs to the shift in mean setting of (19), we can invoke bound (20), which in this case can be approximated (for any of the two threshold selections) by the simple expression

$$\beta_n \leq \exp(-\text{SNR}\sqrt{2\log n})$$

and that is also shown in the figure.

6. SUMMARY

In this paper we illustrate a distributed detection system where n remote nodes collect observations to be delivered to a fusion center for the final decision. In a recent paper by Blum and Sadler [1] it has been shown that optimal detection performance can be guaranteed even though not all the n observations are sent to the fusion center, provided that the more informative are. Here we relax the constraint of maintaining the optimal performance for fixed network size n, and stress the energy saving implied by the reduction of the number of sensors' transmissions. Taking the approach to one extreme, we propose that just one single sensor delivers its observation (or, equivalently, local decision) to the fusion center. In this way, what is actually a local decision becomes the global decision of the system and $(n-1)$ out of n sensors remain silent. The key idea is that of ordered transmissions with more informative sensors that attempt to deliver their message first, combined with a halting protocol that sleeps down the whole system as soon as the quickest sensor send its decision.

The main question of this paper concerns the performance of such a one-bit distributed detector: can we achieve any desired level of performance provided that the network size n is large enough? In technical jargon, we are asking if such an extreme system design still ensures asymptotic consistency, and that the answer is affirmative should be considered our main result.

REFERENCES

[1] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3229–3235, Jul. 2008.

[2] J. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, pp. 102–114, Aug. 2002.

[3] Z.-Q. Luo, M. Gastpar, J. Liu, and A. Swami, "Distributed signal processing in sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 14–15, Jul. 2006.

[4] C. Chong and S. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proc. IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.

[5] B. M. Sadler, "Fundamentals of energy-constrained sensor network systems," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, Aug. 2005.

[6] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: a low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.

[7] P. Braca, S. Marano, and V. Matta, "Asymptotic properties of one-bit distributed detection with ordered transmissions," arXiv:1103.5142v1, Mar. 2011

[8] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*. SIAM, 2008.

[9] I. G. Bairamov and O. L. Gebizlioglu, "On the ordering of random vectors in a norm sense," *Journal of Applied Statistical Science*, vol. 6, no. 1, pp. 77–86, 1997.

[10] B. V. Gnedenko, "Sur la distribution limite du terme maxiumum d'une série aléatoire," *Ann. Math.*, vol. 44, pp. 423–453, 1943.