EXACT ASYMPTOTICS OF DISTRIBUTED DETECTION OVER ADAPTIVE NETWORKS

Vincenzo Matta(1) *Paolo Braca*(2) *Stefano Marano*(1) *Ali H. Sayed*(3)*

(1) Department of Information & Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano (SA), Italy
(2) Research Department, NATO-STO Centre for Maritime Research and Experimentation, La Spezia, Italy
(3) Electrical Engineering Department, University of California, Los Angeles, CA, USA

ABSTRACT

In [1], an important step toward the characterization of distributed detection over adaptive networks has been made by establishing the fundamental scaling law of the error probabilities. However, empirical evidence reported in [1] revealed that a refined asymptotic analysis is necessary in order to capture the exact impact of network *connectivity* on the detection performance of each individual agent. Here we address this open issue by exploiting the framework of exact asymptotics.

Index Terms— Distributed detection, adaptive network, diffusion strategy, large deviations, exact asymptotics.

1. MOTIVATION AND RELATED WORK

We consider distributed detection problems over adaptive networks, where dispersed agents learn continually from streaming data by means of local interactions. The requirement of *adaptation* allows the network of detectors to track drifts in the statistical conditions of the data and models. The requirement of *cooperation* allows each agent to deliver performance that is superior to what would be obtained if it were acting individually. These simultaneous requirements can be achieved by employing diffusion algorithms with constant step-size μ.

While the general topic of distributed detection is rich (see, e.g., [3–9] as entry points on the subject), the aforementioned setting is less explored, where *continuous* learning and adaptation needs to be embedded into the operation of the detectors. With reference to decentralized networks, solutions based on consensus strategies with *decaying* step-size have been proposed in [10–15]. However, to enable adaptation, it has been shown that *diffusion strategies* with *constant* step-size offer advantages in terms of enhanced stability and mean-square-error performance [16–20]. This is due to an inherent asymmetry in the update equations for consensus implementations, which can cause an unstable growth in the state of the network even when each individual agent is stable.

The problem of using diffusion strategies for detection purposes has been considered in [21], with reference to a Gaussian problem. More recently, the general problem of distributed detection over adaptive networks under more relaxed conditions has been pursued in [1, 2]. Resorting to the theory of large deviations [22, 23], it has been shown in [1, 2] that, as μ goes to zero, the steady-state error probabilities of each agent vanish exponentially as functions of $1/\mu$, and that all agents share the same error exponents. However, numerical evidence reported in [1] shows that, depending on their connectivity, different agents will exhibit different error probabilities (even if they exhibit the same scaling law to the leading exponential order), and that large-deviations theory fails to capture the impact of network connectivity on performance. This deficiency can be ascribed to a known limitation of large-deviations analysis, namely, to the fact that it neglects *sub-exponential* terms. A simple example can be useful to illustrate the practical implications of this aspect. Assume network agents 1 and 2 exhibit asymptotic error probabilities P_1 and P_2 of the form:

$$P_1 = e^{-\frac{1}{\mu}}, \qquad P_2 = 2\, e^{-\frac{1}{\mu}} = e^{-\frac{1}{\mu}[1+o(1)]}, \qquad (1)$$

where $o(1)$ stands for any correction such that $o(1) \to 0$ as $\mu \to 0$. These two probabilities have the same error exponent multiplying $-1/\mu$, but the error probability at agent 2 is always twice that of agent 1, a feature that is lost if the sub-exponential corrections are neglected.

To overcome these issues, in this work we exploit the framework of *exact asymptotics* [24], and extend the results of [1, 2] in several directions. We provide accurate analytical formulas for the error probabilities, leading to a powerful understanding of the *universal behavior of distributed detection over adaptive networks*: as functions of $1/\mu$, the error (log-)probability curves corresponding to different agents i) stay nearly-parallel to each other, and ii) are ordered according to the connectivity of each agent. In a nutshell, the more connected an agent is, the lower its error probability curve will be. We also enlarge the setting from the case of doubly-stochastic combination policies considered in [1, 2], to the general setting of right-stochastic combination policies. Our results allow a refined analysis of the interplay between network structure and inference performance, showing interesting and somehow unexpected behavior, and the lesson learned is that *connectivity matters*.

*The work of A. H. Sayed was supported in part by NSF grants CCF-1011918 and ECCS-1407712.

Notation. We use boldface letters to denote random variables, and normal font letters for their realizations. Capital letters refer to matrices, small letters to both vectors and scalars. Exceptions to these rules will be obvious from the context. The r-th derivative of a function $f(t)$ will be denoted by $f^{(r)}(t)$, with the convention $f^{(0)}(t) = f(t)$. When convenient, the first three derivatives will be alternatively denoted by $f'(t)$, $f''(t)$, and $f'''(t)$. The notation $f_\mu = \mathcal{O}(\mu)$ means that the ratio f_μ/μ stays bounded as $\mu \to 0$, while $o(1)$ denotes a term such that $o(1) \to 0$ as $\mu \to 0$.

2. PROBLEM FORMULATION

A network of agents collects observations about a physical phenomenon of interest. Data are assumed to be spatially and temporally independent and identically distributed (i.i.d.). From the observation measured at time n, the k-th agent computes its local statistic (the observation itself, or a suitable function thereof), which is denoted by $\boldsymbol{x}_k(n)$, $k = 1, 2, \ldots, S$. Following the distributed detection framework developed in [1, 2], we focus on the class of diffusion strategies for adaptation over networks, and in particular on the ATC (Adapt-Then-Combine) implementation [18]. In the ATC algorithm, each node k updates its state from $\boldsymbol{y}_k(n-1)$ to $\boldsymbol{y}_k(n)$ through local cooperation as follows:

$$\boldsymbol{v}_k(n) = \boldsymbol{y}_k(n-1) + \mu[\boldsymbol{x}_k(n) - \boldsymbol{y}_k(n-1)], \quad (2)$$

$$\boldsymbol{y}_k(n) = \sum_{\ell=1}^{S} a_{k,\ell} \boldsymbol{v}_\ell(n), \quad (3)$$

where $0 < \mu \ll 1$ is a small step-size parameter. It is seen that node k first uses its *local* statistic, $\boldsymbol{x}_k(n)$, to update its state from $\boldsymbol{y}_k(n-1)$ to an intermediate value $\boldsymbol{v}_k(n)$. The other network agents simultaneously perform similar updates using their local statistics. Subsequently, node k aggregates the intermediate states of its neighbors using nonnegative convex combination weights $\{a_{k,\ell}\}$ that add up to one. Again, all other network agents perform a similar calculation. Collecting the combination weights into a square matrix $A = [a_{k,\ell}]$, then A is a right-stochastic matrix, namely, the entries on each row add up to one. Formally:

$$a_{k,\ell} \geq 0, \quad A\mathbb{1} = \mathbb{1}, \quad (4)$$

with $\mathbb{1}$ being a column-vector with all entries equal to 1. We shall assume that A has second largest eigenvalue magnitude strictly less than one, which yields [19, 25]:

$$B_n = [b_{k,\ell}(n)] \triangleq A^n \stackrel{n\to\infty}{\longrightarrow} \mathbb{1}p, \quad (5)$$

where the row vector $p = [p_1, p_2, \ldots, p_S]$, usually referred to as the Perron eigenvector of A (see, e.g., [19, 20]), satisfies:

$$pA = p, \quad p_\ell > 0, \quad \sum_{\ell=1}^{S} p_\ell = 1. \quad (6)$$

The required condition on A is automatically satisfied by network topologies that are strongly-connected [20], i.e., when there is always a path with nonzero combination coefficients between any pair of nodes, and at least one node has a self-loop ($a_{k,k} > 0$ for some agent k).

In order to characterize an inference system based upon the diffusion output $\boldsymbol{y}_k(n)$, knowledge of the distribution of $\boldsymbol{y}_k(n)$ is crucial. This knowledge is seldom available, except for special cases (e.g., Gaussian observations). A common and well-established approach in the adaptation literature [18, 26] to address this difficulty is to focus on i) the steady-state properties (as $n \to \infty$), and ii) the small step-size regime ($\mu \to 0$).

We start by considering the steady-state behavior of $\boldsymbol{y}_k(n)$ for a given step-size μ. The existence of a steady-state random variable characterizing the diffusion output has been established in [1, 2] for doubly-stochastic matrices. The result can be extended to right-stochastic matrices, and is (the symbol $\rightsquigarrow$ means convergence in distribution):

$$\boxed{\boldsymbol{y}_k(n) \stackrel{n\to\infty}{\rightsquigarrow} \boldsymbol{y}^\star_{k,\mu} \triangleq \sum_{i=1}^{\infty} \sum_{\ell=1}^{S} \mu\,(1-\mu)^{i-1} b_{k,\ell}(i) \boldsymbol{x}_\ell(i)} \quad (7)$$

In our distributed detection formulation, the statistical properties of $\boldsymbol{x}_k(n)$ will depend upon an *unknown* binary state of nature, say, $\mathcal{H}_0$ or $\mathcal{H}_1$. The statistics $\boldsymbol{x}_k(n)$ are spatially and temporally i.i.d., *conditioned* on the hypothesis that gives rise to them. When needed, a subscript 0 or 1 will be appended to the statistical operators to denote the particular hypothesis in force. The decision rule of agent k at time n is of the form:

$$\boxed{\boldsymbol{y}_k(n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lesseqgtr}} \gamma} \quad (8)$$

so that the (steady-state) Type-I and Type-II error probabilities are defined as, respectively:

$$\alpha_{k,\mu} \triangleq \mathbb{P}_0[\boldsymbol{y}^\star_{k,\mu} > \gamma], \qquad \beta_{k,\mu} \triangleq \mathbb{P}_1[\boldsymbol{y}^\star_{k,\mu} \leq \gamma]. \quad (9)$$

3. MAIN RESULT

Our main contributions are now collected in two theorems, stated without proofs for space limitations. A primary role in the forthcoming analysis will be played by the logarithmic moment generating functions of the local statistic $\boldsymbol{x}_k(n)$ and of the steady-state variable $\boldsymbol{y}^\star_{k,\mu}$. These quantities are defined as, respectively:

$$\psi(t) \triangleq \ln \mathbb{E}[e^{t\boldsymbol{x}_k(n)}], \qquad \phi_{k,\mu}(t) \triangleq \ln \mathbb{E}[e^{t\boldsymbol{y}^\star_{k,\mu}}]. \quad (10)$$

The first theorem furnishes a characterization of $\phi_{k,\mu}(t)$. It provides analytical formulas for $\phi_{k,\mu}(t)$ and all its derivatives, in terms of $\psi(t)$ and of the combination weights, and it establishes the asymptotic behavior of $\phi_{k,\mu}(t)$ and all its derivatives as $\mu \to 0$. As we shall see later (Theorem 2), this detailed characterization of $\phi_{k,\mu}(t)$ is crucial to obtain the exact asymptotics of the error probabilities.

THEOREM 1 (Fundamental properties of $\phi^{(r)}_{k,\mu}(t)$). *Assume that $\psi(t) < \infty$ for all $t \in \mathbb{R}$, and let*

$$\phi(t) \triangleq \sum_{\ell=1}^{S} \int_0^{p_\ell t} \frac{\psi(\tau)}{\tau} d\tau. \tag{11}$$

Then, for $k = 1, 2, \ldots, S$, and for $r = 0, 1, \ldots$, it holds that

$$\boxed{\begin{aligned}\phi^{(r)}_{k,\mu}(t) = \sum_{i=1}^{\infty}\sum_{\ell=1}^{S} [\mu(1-\mu)^{i-1} b_{k,\ell}(i)]^r \times \\ \psi^{(r)}\big([\mu(1-\mu)^{i-1} b_{k,\ell}(i)]t\big)\end{aligned}} \tag{12}$$

and that (as $\mu \to 0$):

$$\boxed{\frac{\phi^{(r)}_{k,\mu}(t/\mu)}{\mu^{r-1}} = \phi^{(r)}(t) + \mathcal{O}(\mu)} \tag{13}$$

□

Before stating our second theorem, it is useful to introduce the Fenchel-Legendre transform of $\phi(t)$ (denoted by the corresponding capital letter), along with its essential domain, namely [22, 23]:

$$\Phi(\gamma) \triangleq \sup_{t\in\mathbb{R}}[\gamma t - \phi(t)], \quad \mathcal{D}_\Phi \triangleq \{\gamma \in \mathbb{R} : \ \Phi(\gamma) < \infty\}. \tag{14}$$

The notation $\mathcal{D}^o_\Phi$ will represent the interior of the set $\mathcal{D}_\Phi$.

We focus, without loss of generality[1], on the probability $\mathbb{P}[\boldsymbol{y}^\star_{k,\mu} > \gamma]$ for $\gamma > \mathbb{E}[\boldsymbol{x}]$. For doubly-stochastic combination matrices, it was established in [1, 2] that

$$\mathbb{P}[\boldsymbol{y}^\star_{k,\mu} > \gamma] = e^{-\frac{1}{\mu}[\Phi(\gamma)+o(1)]}. \tag{15}$$

This form highlights the fact that sub-exponential terms are neglected by large-deviations analysis. A refined study can be pursued by seeking an asymptotic approximation, $\mathscr{P}_{k,\mu}(\gamma)$, that ensures the stronger conclusion:

$$\mathbb{P}[\boldsymbol{y}^\star_{k,\mu} > \gamma] = \mathscr{P}_{k,\mu}(\gamma)[1 + o(1)], \tag{16}$$

a kind of asymptotic equivalence that will be denoted by

$$\mathbb{P}[\boldsymbol{y}^\star_{k,\mu} > \gamma] \sim \mathscr{P}_{k,\mu}(\gamma). \tag{17}$$

This framework is commonly referred to as *exact asymptotics*, and has been originally studied in [24] with reference to the simplest setting of normalized sums of i.i.d. random variables — see also [22, 23]. Unfortunately, the adaptive framework is fundamentally different from the latter setting, such that we cannot directly use the existing results, and the required analysis is more involved. The main result established in this article can now be formally stated as follows.

THEOREM 2 (Exact asymptotics of $\boldsymbol{y}^\star_{k,\mu}$ as $\mu \to 0$). *Assume that the distribution of $\boldsymbol{x}_k(n)$ is not of lattice type, and that $\psi(t) < +\infty$ for all $t \in \mathbb{R}$. Let $\gamma \in \mathcal{D}^o_\Phi$, with $\gamma > \mathbb{E}[\boldsymbol{x}]$, and let $\theta_\gamma > 0$ be the unique solution to the stationary equation $\phi'(\theta_\gamma) = \gamma$. Then, for $k = 1, 2, \ldots, S$, the asymptotic equivalence $\mathbb{P}[\boldsymbol{y}^\star_{k,\mu} > \gamma] \sim \mathscr{P}_{k,\mu}(\gamma)$ holds with*

$$\boxed{\mathscr{P}_{k,\mu}(\gamma) = \sqrt{\frac{\mu}{2\pi\theta_\gamma^2\,\phi''(\theta_\gamma)}}\, e^{-\frac{1}{\mu}[\Phi(\gamma)+\epsilon_{k,\mu}(\theta_\gamma)]}} \tag{18}$$

where

$$\boxed{\epsilon_{k,\mu}(t) \triangleq [\phi(t) - \mu\phi_{k,\mu}(t/\mu)] + \frac{[\phi'(t) - \phi'_{k,\mu}(t/\mu)]^2}{2\,\phi''(t)}} \tag{19}$$

□

The key ingredients to computing $\mathscr{P}_{k,\mu}(\gamma)$ in (18) are the logarithmic moment generating function $\psi(t)$ of the local statistics, and the combination matrix A. Indeed, the quantities $\Phi(\gamma)$ and θ_γ depend on the function $\phi(t)$, which in turn depends on $\psi(t)$ and on the *limiting* combination weights p_ℓ. The correction term $\epsilon_{k,\mu}(t)$ depends on $\psi(t)$ and on the *actual* combination weights $b_{k,\ell}(i)$. Despite its apparent complexity, Eq. (18) possesses a well defined structure. First, observe that, from (12) applied with $r = 0, 1$, both terms on the RHS of (19) vanish as $\mu \to 0$, so that we can write

$$\mathscr{P}_{k,\mu}(\gamma) = e^{-\frac{1}{\mu}[\Phi(\gamma)+o(1)]}. \tag{20}$$

Then, let us examine in more detail the terms of order $o(1)$ that collect all the sub-exponential corrections appearing in (18). They can be conveniently separated into two categories. The first term is $\sqrt{\frac{\mu}{2\pi\theta_\gamma^2\,\phi''(\theta_\gamma)}}$. This kind of sub-exponential refinement is typical in the framework of exact asymptotics, and is a consequence of a local Central Limit Theorem — see [22–24]. Observe that this correction is related to the network topology only through the Perron eigenvector, and is therefore *independent* of the particular agent index k. The second correction $\epsilon_{k,\mu}(\theta_\gamma)$, instead, depends on the agent index k, and takes into account the entire network topology and combination weights. In summary, through (18) we arrive at a detailed assessment of the behavior of distributed detection over adaptive networks: as functions of $1/\mu$, the error (log-)probability curves corresponding to different agents not only stay nearly-parallel to each other (as already discovered in [1, 2]), but they are also ordered following a criterion dictated by the term $\epsilon_{k,\mu}(\theta_\gamma)$.

4. NUMERICAL EXAMPLES

We consider a network made of $S = 10$ agents, with the topology shown in the inset of Fig. 1. In the following, $\mathcal{N}_k$

[1]Theorem 2 holds unchanged for the complementary case $\mathbb{P}[\boldsymbol{y}^\star_{k,\mu} \le \gamma]$ with $\gamma < \mathbb{E}[\boldsymbol{x}]$, but for the fact that $\theta_\gamma < 0$.

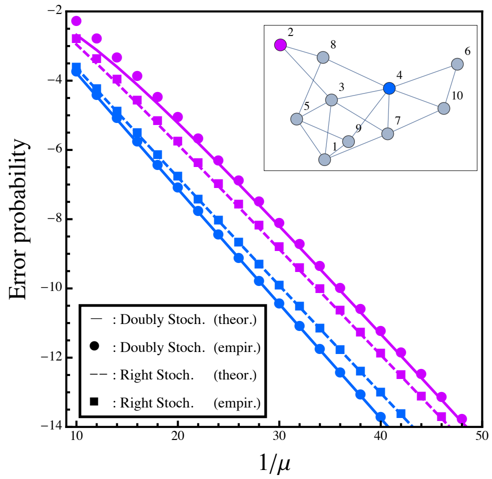

Fig. 1. Laplace example in (24), with $\rho = 0.6$, and combination weights following the Metropolis (Eq. (21), doubly stochastic) and the uniform averaging (Eq. (22), right stochastic) rules. The steady-state performance of agents 2 (magenta) and 4 (blue) is displayed.

is the neighborhood of the k-th agent (including k itself), and n_k the cardinality $|\mathcal{N}_k|$. Moreover, we assume that $a_{k,\ell} = 0$ if $\ell \notin \mathcal{N}_k$. Two different combination policies are tested. The first one is the Metropolis rule [20]:

$$a_{k,\ell} = \begin{cases} 1/\max\{n_k, n_\ell\}, & \ell \in \mathcal{N}_k \setminus \{k\}, \\ 1 - \sum_{m \in \mathcal{N}_k \setminus \{k\}} a_{k,m}, & \ell = k. \end{cases} \tag{21}$$

This choice provides a doubly-stochastic A. The second combination policy is the uniform averaging rule [20]:

$$a_{k,\ell} = 1/n_k, \qquad \ell \in \mathcal{N}_k, \tag{22}$$

yielding a right-stochastic A, whose Perron eigenvector is $p_\ell = \frac{n_\ell}{\sum_{m=1}^S n_m}$ — see [20].

We consider a canonical shift-in-mean detection problem with Laplace noise:

$$\mathcal{H}_0 \quad : \quad \boldsymbol{d}_k(n) \sim \mathscr{L}(d) \triangleq \frac{1}{2} e^{-|d|}, \tag{23}$$

$$\mathcal{H}_1 \quad : \quad \boldsymbol{d}_k(n) \sim \mathscr{L}(d - \rho), \tag{24}$$

where $\boldsymbol{d}_k(n)$ denotes the measurement collected by agent k at time n, $\mathscr{L}(d)$ is the unit-scale Laplace density, and $\rho > 0$ is the shift-in-mean parameter. The local statistics $\boldsymbol{x}_k(n)$ are chosen as the local log-likelihood ratios:

$$\boldsymbol{x}_k(n) = \ln\left(\frac{\mathscr{L}(\boldsymbol{d}_k(n) - \rho)}{\mathscr{L}(\boldsymbol{d}_k(n))}\right) = |\boldsymbol{d}_k(n)| - |\boldsymbol{d}_k(n) - \rho|, \tag{25}$$

whose logarithmic moment generating functions are [1, 14]:

$$\psi_0(t) = \frac{e^{-t\rho}}{2} + \frac{e^{(t-1)\rho}}{2} + \frac{e^{-\rho/2}}{2} \frac{\sinh[\rho(t - 1/2)]}{t - 1/2}, \tag{26}$$

and $\psi_1(t) = \psi_0(-t)$. Let us now examine the distributed network of detectors in operation. To work in the high-performance regime, we resort to Monte Carlo simulation with importance sampling techniques [27] (details omitted for space constraints). We refer to a sufficiently large time horizon, such that the steady-state assumption applies. Following [1], we set the detection threshold to $\gamma = 0$. By symmetry arguments, the error probabilities of first and second kind coincide, so that the terminologies “error probability” and “error exponent” refer to any of these errors.

In Fig. 1 we display the theoretical (Theorem 2) and empirical probability curves of agents 2 and 4, for the Metropolis matrix in (21) (doubly stochastic) and for the uniform averaging matrix (22) (right stochastic). We see that, for a given matrix, the curves of the two agents stay nearly-parallel to each other for sufficiently small values of the step-size μ, exhibiting an exponential decay with the same error exponent, which matches perfectly the results in [1,2]. However, the first-order characterization provided by large-deviations analysis is not powerful enough to capture another important feature: for a given matrix, the curves of the two agents are ordered, and the ordering reflects the degree of connectivity of each agent. This deficiency is addressed by the theoretical formulas given by Theorem 2, which are closely approached by the empirical probability points as $\mu \to 0$, and the agreement is excellent. *What in [1] was only a partial evidence arising from a particular experiment, emerges now as the universal behavior of distributed detection over adaptive networks.*

To get further insights, we compare the performance corresponding to the combination matrices (21) and (22). Here the comparison is made for two systems operating with the same value of the step-size μ. Clearly, the analysis should be complemented by examining the transient behavior, i.e., the adaptation properties, of the two combination policies. To begin with, let us evaluate the error exponents pertaining to the two systems, which yields $\Phi_{\text{doubly}}(0) \approx 0.75 > \Phi_{\text{right}}(0) \approx 0.7$. The doubly-stochastic combination policy asymptotically outperforms the right-stochastic one, which is perhaps not unexpected, in view of the asymptotic equipartition of the doubly-stochastic weights. However, examining Fig. 1, we see that the relative performance of the different combination policies depends on the connectivity of the individual agent. For instance, for the well-connected agent 4, the doubly-stochastic policy delivers superior performance, while exactly the converse is true for the sparsely connected agent 2. An explanation for this behavior can be as follows. The second largest magnitude eigenvalues of the doubly-stochastic and of the right-stochastic combination matrices are: $\lambda_{\text{doubly}} \approx 0.83 > \lambda_{\text{right}} \approx 0.7$, suggesting that convergence to the pertinent Perron eigenvector will be faster for the right-stochastic weights. Our results indicate that, for the sparsely connected agent 2, the benefits of the higher (doubly stochastic matrix) exponent are more than compensated by the faster (right stochastic matrix) convergence.

5. REFERENCES

[1] P. Braca, S. Marano, V. Matta, and A. H. Sayed, “Asymptotic performance of adaptive distributed detection over networks,” *submitted for publication*. Available as arXiv:1401.5742v2 [cs.IT], Jan. 2014.

[2] P. Braca, S. Marano, V. Matta, and A. H. Sayed, “Large deviations analysis of adaptive distributed detection,” in *Proc. IEEE ICASSP*, Florence, Italy, May 2014, pp. 6112–6116.

[3] M. Longo, T. D. Lookabaugh, and R. M. Gray, “Quantization for decentralized hypothesis testing under communication constraints,” *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 241–255, Mar. 1990.

[4] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, New York, 1997.

[5] R. Viswanathan and P. K. Varshney, “Distributed detection with multiple sensors: Part I–Fundamentals,” in *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[6] R. S. Blum, S. A. Kassam, and H. V. Poor, “Distributed detection with multiple sensors: Part II–Advanced topics,” in *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

[7] J. F. Chamberland and V. V. Veeravalli, “Decentralized detection in sensor networks,” *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.

[8] J. F. Chamberland and V. V. Veeravalli, “Wireless sensors in distributed detection applications,” *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 16–25, May 2007.

[9] B. Chen, L. Tong, and P. K. Varshney, “Channel-aware distributed detection in wireless sensor networks,” *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 16–26, Jul. 2006.

[10] P. Braca, S. Marano, and V. Matta, “Enforcing consensus while monitoring the environment in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jul. 2008.

[11] P. Braca, S. Marano, V. Matta, and P. Willett, “Asymptotic optimality of running consensus in testing statistical hypotheses,” *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.

[12] ——, “Consensus-based Page’s test in sensor networks,” *Signal Processing*, vol. 91, no. 4, pp. 919–930, Apr. 2011.

[13] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, “Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4396, Sep. 2011.

[14] D. Bajovic, D. Jakovetic, J. M. F. Moura, J. Xavier, and B. Sinopoli, “Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations,” *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5987–6002, Nov. 2012.

[15] D. Jakovetic, J. M. F. Moura, and J. Xavier, “Distributed detection over noisy networks: Large deviations analysis,” *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4306–4320, Aug. 2012.

[16] C. G. Lopes and A. H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[17] F. S. Cattivelli and A. H. Sayed, “Diffusion LMS strategies for distributed estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[18] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, “Diffusion strategies for adaptation and learning over networks,” *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.

[19] A. H. Sayed, “Diffusion adaptation over networks,” in *Academic Press Library in Signal Processing*, vol. 3, R. Chellapa and S. Theodoridis, *Eds.*, pp. 323–454, Academic Press, Elsevier, 2014.

[20] A. H. Sayed, “Adaptive networks,” *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[21] F. S. Cattivelli and A. H. Sayed, “Distributed detection over adaptive networks using diffusion adaptation,” *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.

[22] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.

[23] F. den Hollander, *Large Deviations*. American Mathematical Society, 2008.

[24] R. R. Bahadur, and R. Ranga Rao, “On deviations of the sample mean,” *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 1015–1027, Dec. 1960.

[25] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.

[26] A. H. Sayed, *Adaptive Filters*. Wiley, NJ, 2008.

[27] J. A. Bucklew, *Large Deviations Techniques in Decision, Simulation and Estimation*. Wiley, 1990.