Detection in Sensor Networks: The Saddlepoint Approximation

Saeed A. Aldosari, *Member, IEEE*, and José M. F. Moura, *Fellow, IEEE*

Abstract—This paper presents a computationally simple and accurate method to compute the error probabilities in decentralized detection in sensor networks. The cost of the direct computation of these probabilities—e.g., the probability of false alarm, the probability of a miss, or the average error probability—is combinatorial in the number of sensors and becomes infeasible even with small size networks. The method is based on the theory of large deviations, in particular, the saddlepoint approximation and applies to generic parallel fusion sensor networks, including networks with nonidentical sensors, nonidentical observations, and unreliable communication links. The paper demonstrates with parallel fusion sensor network problems the accuracy of the saddlepoint methodology: 1) computing the detection performance for a variety of small and large sensor network scenarios; and 2) designing the local detection thresholds. Elsewhere, we have used the saddlepoint approximation to study tradeoffs among parameters for networks of arbitrary size.

Index Terms—Decentralized detection, Lugannani-Rice approximation, parallel fusion, quantization, saddlepoint approximation, sensor fusion, sensor networks.

I. INTRODUCTION

THE potential for large-scale sensor networks is attracting great interest in many applications in recent years due to emerging technological advancements. Increasing levels of electronics and RF circuits integration lend themselves to the deployment of affordable, yet reliable sensing systems, which are envisioned as networks of autonomous densely distributed sensor nodes [1]. Individually, each sensor node may not accomplish much, but, working cooperatively, they have, for example, the potential to monitor large areas, detect the presence or absence of targets, or track moving objects.

The design and analysis of sensor networks for detection applications has received considerable attention in the past decade [2]–[4]. A major difficulty usually encountered in such applications is the high computational cost associated with evaluating the detection error probabilities of the network, which

is a combinatorial problem in the number N of sensors. Direct evaluation of these probabilities is possible only for rather small networks. In this paper, we develop a computationally fast and accurate methodology to evaluate the error, detection, and false alarm probabilities for networks of arbitrary size—small, medium, or large number of sensors. Our method is based on large deviation theory approximations to these probabilities, in particular, the saddlepoint approximation.

We illustrate the saddlepoint based methodology by considering a binary hypothesis detection problem in which the environment assumes one of two possible states (e.g., a target is present or absent). We focus on a parallel network architecture in which the sensors make *local* decisions based on their own measurements and then deliver these local decisions to a fusion center. The local measurements are quantized to b bits, so the local detectors can be thought of as b-bit local quantizers.

In this particular architecture, called parallel fusion (see Fig. 1), there is no communication among the local sensors and the fusion center does not sense the physical phenomenon. Fundamental results on distributed detection with a parallel architecture date back to the early work of Tenney and Sandell [5]. For an introduction and overview of the area of decentralized detection, interested readers are referred to [2] and [4].

Designing the network detector and evaluating the global performance probabilities is a complicated task requiring high computational costs that grow as N^{2^b-1}, where N is the number of sensors and b is the number of bits per local detector. This renders their direct evaluation infeasible, except when the number of sensors N or the number of bits b per sensor is small [6], [7]. The literature usually avoids the direct computation of the performance probabilities by evaluating their asymptotic exponential decay rate, e.g., given by the Chernoff and Kullback-Leibler (KL) distances, [8]–[11]. These are in certain cases simple to compute, but, we emphasize, such measures estimate the asymptotic exponential decay rate of the performance probabilities, not the probabilities themselves. Chernoff and KL distances do not help with evaluating the receiver operating characteristics (ROC), or designing the fusion rule, say under the Neyman-Pearson (NP) criterion, since both require the actual detection and false alarm probabilities and not their decay rates. To evaluate the detection performance probabilities, some authors use the normal approximation [12]–[14]. The normal approximation can handle many practical problems, but may not provide acceptable accuracy, especially when the points to be approximated are in the tail regions and far from the mean of the decision variable. Simulations show that the normal approximation performs better with smaller networks but its accuracy deteriorates rapidly as the network size increases.

We take in this paper a different approach that enables the analysis and design of networks of arbitrary size (small or large) by considering a large deviation theory based approximation to the error probabilities that is both simple to compute and accurate. We adopt the *saddlepoint* approximation [15], which has been used in many applications such as optical detection, bootstrapping, and queuing analysis. It can also be related to the method of stationary phase, which is used widely in physics. Although based on asymptotic expansions, the saddlepoint approximation is highly accurate even for networks of few sensors. In addition, remarkably, the computational complexity of the saddlepoint approximation is independent of the number of sensors. We provide numerical comparisons to illustrate the advantage of the saddlepoint approximation over other approximation methods under different conditions. We show that the saddlepoint formulas are an accurate approximation in practical scenarios involving identical or nonidentical observations, identical or nonidentical local detectors, and reliable or unreliable communication links between the sensors and the fusion center.

The organization of the paper is as follows. In Section II, we present the parallel fusion network and state the problem. At this stage, we take a simplifying approach, assuming independent, identically distributed (i.i.d.) measurements, identical local quantizers, and noiseless communications channel. In Section III, we address the difficulty of evaluating the global performance probabilities, and we explain two commonly used methods for approximating it—the normal approximation and the asymptotic exponential decay rate based analysis. In Section IV, we present the saddlepoint approximation for densities and distributions, and we discuss its complexity and theoretical accuracy. In Section V, several numerical studies illustrate the advantage of the saddlepoint approximation over other techniques by comparing 1) the error probabilities; 2) the fusion threshold; and 3) the local detectors' thresholds. In all cases, the quantities computed by the saddlepoint approximation compare much more favorably with the true values than the corresponding quantities computed with other methods. In Section VI, we extend the saddlepoint approximation to sensor networks with nonidentical measurements, nonidentical local detectors, and noisy channels. Finally, concluding comments are in Section VII.

II. MODEL AND PROBLEM STATEMENT

We consider the binary detection problem of deciding between H_0 or H_1 with a network of N parallel sensors. The network acquires N measurements, one per sensor, y_n, $n = 1, 2, \ldots, N$, makes a local b-bit decision u_n, i.e., quantizes each sensor measurement into b-bits, delivers all N local decisions u_n, $n = 1, 2, \ldots, N$, through a multiple access channel to a single fusion center. This fusion center makes the final determination $\widehat{H}$, see Fig. 1. In this particular model, the local sensors have no means of communicating their local decisions, except to a single fusion center, and the fusion center does not acquire independent measurements. Conditioned on H, the observations y_n are assumed to be independent. The quantization performed locally in each sensor is according to local decision rules γ_n, $n = 1, 2, \ldots, N$, which can be considered as mappings from the continuous observation space $\mathbb{R}$ to the discrete classification
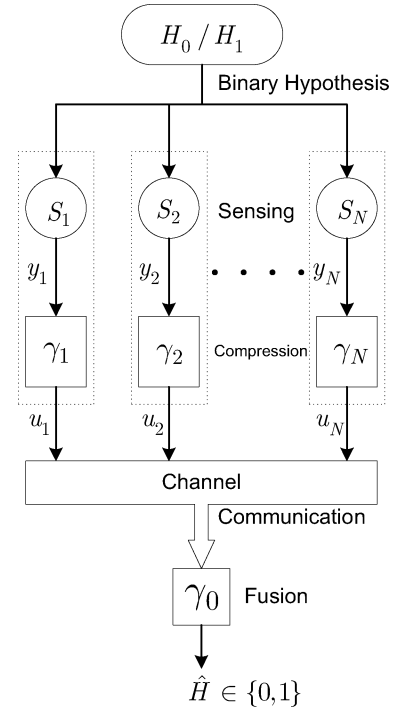

Fig. 1. Parallel fusion network.

space $\mathbb{U}$, i.e., $\gamma_n : \mathbb{R} \longrightarrow \mathbb{U}$, where $\mathbb{U} = \{1, 2, \ldots, M\}$, and M is the number of quantization levels. Upon receiving the N local decisions, the fusion center fuses them together according to a fusion rule $\gamma_0 : \mathbb{U}^N \longrightarrow \{0, 1\}$ to reach the final decision $\widehat{H}$.

We consider likelihood ratio (LR) based fusion rules, i.e., rules that rely on the LR of the local decisions

$$LR_u(\mathbf{u}) = \frac{\Pr(\mathbf{u}|H_1)}{\Pr(\mathbf{u}|H_0)}$$

where $\mathbf{u} = (u_1, \ldots, u_N)$ is the vector of all quantized local decisions received at the fusion center. Of particular interest are Bayesian and NP detectors for which optimal fusion relies on the LR statistic. In Bayesian detection problems, the fusion center minimizes the global average probability of error

$$P_e = \pi_0 P_0 + \pi_1 P_1$$

at the fusion center, where P_j is the probability of error $\Pr(\widehat{H} \neq H | H_j)$ under H_j, and $\pi_j = \Pr(H_j)$, $j = 0, 1$ are the prior probabilities. In NP detection, the fusion center minimizes the miss probability ($P_m = P_1$) subject to a constraint on the false alarm probability ($P_f = P_0$), i.e., $\min P_1$ subject to $P_0 \leq \alpha$, where α is called the size of the test.

We address the problem of evaluating, for example, the error probabilities

$$P_0 = \Pr(u_0 = 1|H_0) \text{ and } P_1 = \Pr(u_0 = 0|H_1)$$

at the fusion center, where u_0 is the global decision at the output of the fusion center, where $u_0 = j$, $j = 0, 1$ corresponds to a global decision of $\widehat{H} = H_j$. These "global" probabilities are the relevant quantities for detection in sensor networks. They

are important to various studies such as performance assessments (e.g., false alarm and detection probabilities, Bayesian cost, receiver operating characteristics) and sensor network optimization with respect to the global detection performance (e.g., finding simultaneously all the optimal local decision rules γ_n and the fusion rule γ_0, optimal in a Bayes or NP sense).

Initially, we assume that the measurements y_n have identical conditional distributions. We also restrict the local decision rules γ_n, $n = 1, 2, \ldots, N$, to be identical. It was shown in [16] that the performance loss due to this restriction of identical rules is negligible when the number of sensors is large. In Section VI, we extend our proposed techniques to cases involving nonidentical measurements and nonidentical local detection rules. The assumptions of i.i.d. measurements and identical detectors result in local decisions u_n that are also i.i.d. The conditional probabilities of the local decisions u_n are assumed to be known (or computable given the measurement statistics and the local decision rules) and they are denoted by

$$Q_{mi}^n = \Pr(u_n = m|H_i),$$
$$n = 1, 2, \ldots, N, \ \ m = 1, \ldots, M, \ \ i = 0, 1.$$

In the sequel, we suppress the n superscript and adopt the simplified notation

$$Q_{mi} = \Pr(u = m|H_i)$$

to denote the conditional probability of any of the n local decisions under hypothesis H_i.

In Section V, for simplicity, though not required by our methodology, we test the saddlepoint approximation with particular observation models and local quantization rules, not necessarily optimal. We pay special emphasis to local likelihood ratio detectors or quantizers (LRQ), for which the local decision of each sensor is done based on the measurement likelihood ratio

$$LR_y(y_n) = f_1(y_n)/f_0(y_n)$$

where $f_i(y) = f(y|H_i)$, $i = 0, 1$, are the conditional probability density functions (pdf) of the measurements. These rules may or may not be optimal, but will be used based on practical considerations; still, the method we develop applies in either case to compute the detection error probabilities. We recall that, for binary sensors ($b = 1$ bit per sensor), the optimality of likelihood ratio detectors with respect to Bayesian and the NP criteria has been established in [17] and [18]. For the general case ($b \geq 1$ bits per sensor), Tsitsiklis in [19] showed that the likelihood ratio detector is optimal with respect to Ali-Silvey distances (including Chernoff and KL distances), which implies the asymptotic optimality (as $N \to \infty$) of such tests under the NP and Bayesian criteria.

When the local measurements have monotonic likelihood ratios, quantizing the likelihood ratios $LR_y(y_n)$ is equivalent to direct quantization of the measurements y_n, in which case the local detectors reduce to

$$u_n = i \text{ if } \lambda_{i-1} < y_n \leq \lambda_i \tag{1}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{M-1})$ are the local detectors thresholds, and $\lambda_0 = -\infty$, $\lambda_M = +\infty$. In Section V, we will consider specific instantiations of these conditional probability densities.

III. PRELIMINARIES

A. Likelihood Ratio Fusion

Given the independence of the sensor observations, the likelihood ratio fusion rule is of the form

$$s = \sum_{n=1}^{N} \ell_n \underset{0}{\overset{1}{\gtrless}} v, \text{ where } \ell_n = \log \frac{\Pr(u_n|H_1)}{\Pr(u_n|H_0)} \tag{2}$$

is the LLR of the decision of the nth sensor, and v is the fusion threshold. In this paper, we are not necessarily concerned with the optimality of (2). However, we discuss briefly when to expect it to be optimal. Reference [4] shows that this fusion rule optimizes both the Bayesian and the NP criteria for independent observations. For both, the Bayesian and NP problems, we can consider nonrandomized tests of the LLR statistic s even though it is discrete. It was shown in [20] that the optimal fusion rule does not randomize as long as the unquantized measurements y_n are independent and their likelihood ratios $LR_y(y_n)$ contain no point-masses of probability, i.e., $\Pr[LR_y(y_n) = x] = 0$ for all $x \in \mathbb{R}$.

In Bayesian detection, the fusion threshold v can be computed directly given the priors π_0 and π_1 and the cost associated with the different decisions as follows:

$$v = \log \left(\frac{\pi_0(c_{10} - c_{00})}{\pi_1(c_{01} - c_{11})} \right) \tag{3}$$

where c_{ij}, $i, j \in \{0, 1\}$ is the cost associated with making a global decision $u_0 = i$ when H_j is present. For the minimum probability of error detector ($c_{10} = c_{01} = 1$ and $c_{00} = c_{11} = 0$), the threshold reduces to $v = \log(\pi_0/\pi_1)$. For NP detection, on the other hand, the fusion threshold v is determined by solving

$$P_f(v) \leq \alpha \tag{4}$$

for a given size α, where the false alarm probability is $P_f(v) = \Pr(u_0 = 1|H_0) = \Pr(s > v|H_0)$.

The local LLRs ℓ_n, $n = 1, \ldots, N$, are discrete random variables that take values in $\{L_1, \ldots, L_M\}$, where

$$L_m = \log \left(\frac{Q_{m1}}{Q_{m0}} \right), \ m = 1, \ldots, M \tag{5}$$

is the value of the LLR whenever a local decision in favor of m is made. By grouping the local decisions, the fusion rule in (2) can also be written as a weighted majority rule

$$s = \sum_{m=1}^{M} n_m L_m \underset{0}{\overset{1}{\gtrless}} v \tag{6}$$

where $n_m \in \{0, 1, \ldots, N\}$, $\sum_{m=1}^{M} n_m = N$, is the number of sensors making a $u_n = m$ decision and L_m is the LLR value of that local decision as defined in (5). Unlike the LLR variables ℓ_n, which are independent, the discrete variables n_m,

$m = 1, 2, \ldots, M$ are dependent with joint multinomial probability mass function (pmf) given by

$$\Pr(n_1, n_2, \ldots, n_M | H_i) = (n_1, \ldots, n_M)! \prod_{m=1}^{M} (Q_{mi})^{n_m},$$

when $\sum_{m=1}^{M} n_m = N$ and $\Pr(n_1, n_2, \ldots, n_M | H_i) = 0$ otherwise. The multinomial coefficient in the above equation is given by

$$(n_1, n_2, \ldots, n_M)! = \frac{N!}{n_1! n_2! \cdots n_M!}.$$

Although (6) involves a sum of only M random variables, analyzing the distribution of s through (2) is easier due to the independence of the LLR variables ℓ_n, $n = 1, \ldots, N$.

The local decision probabilities Q_{mi}, $m = 1, \ldots, M$, $i = 0, 1$, are assumed to be known or computable given the statistics of the observations and the local detection rules. For example, when the local decision at the sensors is done according to (1), the local decision probabilities can be computed given the conditional pdfs $f_i(y)$ of the unquantized measurements as follows:

$$Q_{mi} = \Pr(u = m | H_i) = \int_{\lambda_{m-1}}^{\lambda_m} f_i(y) dy \qquad (7)$$

under hypothesis H_i. This local decision rule is used later in Section V, but, as mentioned before, the saddlepoint formulas derived in Section IV approximating the global error probability are not limited to this particular quantization rule.

B. Exact Computation of the Error Probability

The global error probabilities P_0 and P_1 at the fusion center can be computed from the right and left tails of the global LLR s as follows:

$$P_0 = \Pr(s > v | H_0) \text{ and } P_1 = \Pr(s < v | H_1).$$

This can be accomplished by going through all M^N possible outcomes $(u_1, u_2, \ldots, u_N)$ of the N local detectors, considering only those combinations for which the global LLR s satisfies the particular tail condition ($s > v$ for P_0, and $s < v$ for P_1), then summing up their probabilities, where

$$\Pr(u_1 = k_1, \ldots, u_N = k_N | H_j) = \prod_{n=1}^{N} Q_{k_n j}.$$

Alternatively, one may use the i.i.d. and the identical detectors assumption to simplify this computation by noting that the fusion center only cares about the counts of distinct decisions as in (6). Hence, the probability of error under H_i is given by

$$P_i = \sum_{n_1, \ldots, n_M} \left[(n_1, n_2, \ldots, n_M)! \prod_{m=1}^{M} (Q_{mi})^{n_m} \right]$$

$$\text{s.t.} \sum_{m=1}^{M} n_m = N \text{ and } \sum_{m=1}^{M} n_m L_m \underset{\text{if } i=1}{\overset{\text{if } i=0}{\gtrless}} v. \quad (8)$$

The sum in the last equation considers all possible ways of selecting M integers such that their sum is equal to N. The number of such combinations is equal to the number of ways of arranging N identical balls into M distinct buckets, which is given by

$$\binom{N + M - 1}{N} \simeq \sqrt{\frac{M}{2\pi}} \left(\frac{e}{M} \right)^M N^{M-1},$$

where we used Stirling's approximation when N is large. We see that the number of terms in (8) is $O(N^{M-1}) = O(N^{2^b-1})$, which is too large for values of N and/or b of interest. Therefore, the direct evaluation of the error probability is not appropriate for practical use except when M and N are considerably low. For instance, computing the error probability in a network of $N = 100$ sensors where the local decisions are quantized into $b = 3$ bits requires evaluating about 2.6×10^{10} terms. The difficulties become even more prominent when embedding such exact expressions in an optimization algorithm that adapts the network parameters with the goal of minimizing the error probabilities. In Section IV, we introduce a method that avoids this large computational burden.

C. Normal Approximation

Instead of computing directly the probability of error, one can approximate it to avoid the associated high computational costs. Since s is a sum of N i.i.d. random variables, we may use the central limit theorem (CLT) to approximate the true distribution. Under H_i, $i = 0, 1$, the mean and variance of the local LLR ℓ are given by

$$\mu_{\ell,i} = E(\ell | H_i) = \sum_{m=1}^{M} Q_{mi} L_m$$

$$\sigma_{\ell,i}^2 = \text{Var}(\ell | H_i) = -\mu_{\ell,i}^2 + \sum_{m=1}^{M} Q_{mi} L_m^2$$

where $E(\cdot)$ and $\text{Var}(\cdot)$ denote the expectation and variance, respectively. The mean and variance of the global LLR s are given by

$$\mu_{s,i} = N \mu_{\ell,i} \text{ and } \sigma_{s,i}^2 = N \sigma_{\ell,i}^2$$

respectively. From the central limit theorem, the distribution of the global LLR s converges to that of a Gaussian random variable, i.e.,

$$\Pr(s_0 > v_0 | H_0) \overset{N \to \infty}{\longrightarrow} \Phi(v_0)$$

$$\Pr(s_1 < v_1 | H_1) \overset{N \to \infty}{\longrightarrow} \Phi_c(v_1)$$

where $\Phi(x) = \int_x^{\infty} \varphi(y) dy$ is the normal right-tail integral, $\Phi_c(x) = 1 - \Phi(x)$ is the normal left-tail integral, $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the normal pdf, and v_i, $i = 0, 1$, is given by

$$v_i \triangleq \frac{v/N - \mu_{\ell,i}}{\sigma_{\ell,i}} \sqrt{N} \qquad (9)$$

$$s_i \triangleq \frac{s/N - \mu_{\ell,i}}{\sigma_{\ell,i}} \sqrt{N}. \qquad (10)$$

When using the normal approximation, we have to be careful about its convergence speed. For instance, assuming v_1 is fixed, the absolute difference between the true distribution

$P_1 = \Pr(s < v|H_1) = \Pr(s_1 < v_1|H_1)$ and its normal approximation $\Phi_c(v_1)$ can be bounded using the Berry-Esséen theorem [21, sec. 2.5]

$$|P_1 - \Phi_c(v_1)| < \frac{c\rho_3}{\sigma_{\ell,1}^3} \frac{1}{\sqrt{N}} \tag{11}$$

where $\rho_3 = E(|\ell|^3|H_1)$ and c is a constant independent of the distribution of s. However, this only bounds the absolute difference; it does not quantify the relative difference $|P_1 - \Phi_c(v_1)|/P_1$. For example, if the true probability P_1 is 10^{-4} at $N = 100$, then, unless ρ_3 is very small, the $O(1/\sqrt{N})$ bound in (11) is not of much help since it is much larger than the true probability P_1, [15]. Notice that v_i, $i = 0,1$, in (9) grows as $\sqrt{N}$ and the true distribution (also its normal approximation) decays at an exponential rate with respect to N, which is faster than the $N^{-1/2}$ decay rate of the absolute difference. The normal approximation is inappropriate when the absolute difference $|P_1 - \Phi_c(v_1)|$ becomes larger (or comparable) to the approximated value P_1. We see that, when N is large, the normal approximation may not provide adequate accuracy.

D. Asymptotic Decay Rate

As the number of sensors N grows, the probability of error at the fusion center decays exponentially fast. In Bayesian problems, the decay rate of the average probability of error is defined by

$$C = -\lim_{N \to \infty} \frac{1}{N} \log P_e(N). \tag{12}$$

where $P_e(N) = \pi_0 P_0(N) + \pi_1 P_1(N)$ is the average probability of error when the number of sensors is N. In the asymptotic regime $(N \to \infty)$, the best rate of decay C is given by the *Chernoff distance* [11]

$$C = \max_{a \in [0,1]} -\log \sum_{m=1}^{M} [Q_{m0}]^a [Q_{m1}]^{1-a}. \tag{13}$$

The decay rate in NP detection can also be estimated. Let $\beta(N, \alpha)$ be the minimum miss probability such that the false alarm probability is less than α when the number of sensors is N, i.e,

$$\beta(N, \alpha) = \min_v P_1 \text{ subject to } P_0 < \alpha. $$

Then, as $N \to \infty$ and $\alpha \to 0$, the probability of miss decays exponentially fast with a rate D defined by

$$D = \lim_{\alpha \to 0} \lim_{N \to \infty} \frac{1}{N} \log \beta(N, \alpha).$$

From Stein's lemma, the exponential decay rate D is given by the KL distance [11]

$$D = \sum_{m=1}^{M} Q_{m0} \log \frac{Q_{m0}}{Q_{m1}}. \tag{14}$$

The Chernoff and KL distances provide the decay rates of the probability of error and miss probability, i.e., rough estimates of these probabilities in the form of $P_e \approx e^{-NC}$ and $P_1 \approx$ e^{-ND}, respectively. In contradistinction to these, in this paper, we are interested in more precise and direct estimates of the error probability and miss probability for particular finite values of N, v, and α.

IV. SADDLEPOINT APPROXIMATION

Saddlepoint techniques are powerful tools to derive accurate approximations to densities and tail probabilities of sums of independent random variables. In the context of the current problem, we use these techniques to approximate the tail probabilities of the global LLR $s = \sum_{n=1}^{N} \ell_n$. The evaluation point (in this case, it is the fusion threshold v) is usually bounded away from the mean of s and, hence, the normal approximation may not provide adequate accuracy. Saddlepoint techniques avoid this problem by relating the original density of s to a new so-called tilted density. The tilted density is chosen such that it is centered at the evaluation point v. An accurate approximation can then be obtained using the normal approximation. In many cases, the saddlepoint approximation of discrete random variables is similar to that of continuous random variables [22]. In what follows, we highlight results from the continuous saddlepoint theory that are relevant to the problem at hand. Accuracy differences resulting from applying continuous-based saddlepoint approximations to our discrete problem will be discussed afterward. For more details on saddlepoint approximations, refer to [15].

A. Saddlepoint Approximation of Densities

Denote the density of s under H_i by $f_{s,i}(x) = f_s(x|H_i)$, $i = 0,1$. The normal approximation may not be accurate when x is far from the mean $\mu_{s,i}$. In saddlepoint techniques, the original density $f_{s,i}(x)$ is embedded in a conjugate exponential family

$$f_{\hat{s},i}(x, \hat{\theta}_i) = e^{x\hat{\theta}_i - K_i(\hat{\theta}_i)} f_{s,i}(x) \tag{15}$$

where $\hat{s}$ is the transformed random variable, $\hat{\theta}_i$ is a fixed parameter, and $K_i(\theta)$ is the cumulant generating function (CGF) of s, which, for our detection problem, is given by

$$K_i(\theta) = N \log G_i(\theta) \tag{16}$$

where $G_i(\theta)$ is the moment generating function (MGF) of ℓ under H_i defined by

$$G_i(\theta) = E(e^{\theta\ell}|H_i) = \sum_{m=1}^{M} Q_{mi} e^{\theta L_m}. \tag{17}$$

The transformation performed in (15) is often called *exponential tilting*. The objective here is to shift the mean of the transformed variable $\hat{s}$ so that it becomes close to x. The normal approximation can then be applied safely to estimate the density of $\hat{s}$ at x. To do this, notice that the cumulant generating function of $\hat{s}$ under H_i is $K_{\hat{s},i}(\theta) = K_i(\theta + \hat{\theta}_i) + K_i(\hat{\theta}_i)$. Taking the first derivative with respect to θ and evaluating it at $\theta = 0$, we get the mean of the transformed variable, i.e.,

$\mu_{\hat{s},i} = E(\hat{s}|H_i) = K_i'(\hat{\theta}_i)$, where $K_i'(\theta)$ is the first derivative of the cumulant generating function of s given by

$$K_i'(\theta) = \partial K_i(\theta)/\partial\theta = NW_{1,i}(\theta)/G_i(\theta),$$
$$W_{k,i}(\theta) = \sum_{m=1}^{M} Q_{mi} e^{\theta L_m} (L_m)^k. \tag{18}$$

The mean $\mu_{\hat{s},i}$ of the transformed variable can be made precisely equal to x if we find $\hat{\theta}_i$ such that

$$\mu_{\hat{s},i} = K_i'(\hat{\theta}_i) = x. \tag{19}$$

The variance of $\hat{s}$ is obtained from the second cumulant, which is equal to the second derivative of $K_i(\theta)$ at $\theta = 0$, i.e.,

$$\sigma_{\hat{s},i}^2 = K_i''(\hat{\theta}_i)$$

where

$$K_i''(\theta) = \partial^2 K_i(\theta)/\partial\theta^2 \tag{20}$$
$$= N\left[G_i(\theta)W_{2,i}(\theta) - W_{1,i}(\theta)^2\right]/G_i(\theta)^2. \tag{21}$$

Since we set $x = \mu_{\hat{s},i}$, the density of $\hat{s}$ at x can be accurately approximated using the normal approximation as follows:

$$f_{\hat{s},i}(x,\hat{\theta}_i) = 1/\sqrt{2\pi K_i''(\hat{\theta}_i)}.$$

Dividing by $e^{x\hat{\theta}_i - K_i(\hat{\theta}_i)}$, we carry out the reverse transformation in (15) to get the density of the original variable s at x

$$f_{s,i}(x) \simeq \tilde{f}_{s,i}(x) = \frac{e^{K_i[\hat{\theta}_i(x)] - x\hat{\theta}_i(x)}}{\sqrt{2\pi K_i''[\hat{\theta}_i(x)]}}. \tag{22}$$

This is the saddlepoint approximation for densities, and, under H_i, $\hat{\theta}_i(x)$ is the *saddlepoint* at x, which can be found by solving (19). It can be shown that the *relative* error in (22) is $O(N^{-1})$ [23]. Using Edgeworth expansion, a correction term may be added to (22) to further reduce the relative error to $O(N^{-2})$, [24]. To ensure that $\tilde{f}_{s,i}(x)$ in (22) is a valid density, it should be normalized such that $\int_{-\infty}^{\infty} \tilde{f}_{s,i}(x)dx = 1$.

The saddlepoint approximation in (22) can also be written in terms of the Gaussian pdf as follows:

$$\tilde{f}_{s,i}(x) = \frac{\hat{\theta}_i(x)}{g_i[\hat{\theta}_i(x)]} \varphi\left\{r_i\left[\hat{\theta}_i(x)\right]\right\}$$

where

$$r_i(\theta) = \mathrm{Sgn}(\theta)\sqrt{2\left[\theta K_i'(\theta) - K_i(\theta)\right]}, \tag{23}$$
$$g_i(\theta) = \theta\sqrt{K_i''(\theta)}, \tag{24}$$

where $\mathrm{Sgn}(\cdot)$ denotes the sign operator. The saddlepoint $\hat{\theta}_i(x)$ at x is the solution of $K_i'(\hat{\theta}_i) = x$.

B. Saddlepoint Approximation of Tail Probabilities

The left and right tail probabilities can be approximated by direct integration of the saddlepoint density approximation. Here, we highlight the main steps involved in deriving the right tail

probability $P_0 = \mathrm{Pr}(s > v|H_0)$. For details on the relative error, as well as other derivation techniques, interested readers are referred to [15, sec. 3.3] and [25].

We find the probability of false alarm $P_0 = \mathrm{Pr}(u_0 = 1|H_0)$ by integrating the approximate density $\tilde{f}_{s,0}(x)$, i.e.,

$$P_0 = \mathrm{Pr}(s > v|H_0) \simeq \int_v^{\infty} \tilde{f}_{s,0}(x)dx$$
$$\simeq \int_v^{\infty} \frac{e^{K_0[\hat{\theta}_0(x)] - x\hat{\theta}_0(x)}}{\sqrt{2\pi K_0''[\hat{\theta}_0(x)]}}dx. \tag{25}$$

A change of variables is performed by using the saddlepoint equation $x = K_0'(\hat{\theta}_0)$. Since $dx = K_0''(\hat{\theta}_0)d\hat{\theta}_0$, P_0 can be written as

$$P_0 \simeq \int_{\hat{\theta}_0(v)}^{\infty} \frac{1}{\sqrt{2\pi}}\sqrt{K_0''(\theta)} e^{K_0(\theta) - \theta K_0'(\theta)}d\theta,$$

where $\hat{\theta}_0(v)$ is the saddlepoint at v obtained by solving $K_0'(\hat{\theta}_0) = v$. We add and subtract $1/r_0(\theta)$ followed by a change of variables using (23) and $r_0 dr_0 = \theta K_0''(\theta)d\theta$, from which we can write

$$P_0 \simeq \int_{\hat{\theta}_0(v)}^{\infty} \frac{1}{\sqrt{2\pi}}\theta K_0''(\theta)\left[\frac{1}{r_0(\theta)} - \frac{1}{r_0(\theta)} + \frac{1}{\theta\sqrt{K_0''(\theta)}}\right]$$
$$\times e^{-r_0(\theta)^2/2}d\theta \tag{26}$$
$$P_0 \simeq \int_{\hat{\theta}_0(v)}^{\infty} \frac{1}{\sqrt{2\pi}}\left[\frac{1}{g_0(\theta)} - \frac{1}{r_0(\theta)}\right]\theta K_0''(\theta)e^{-r_0(\theta)^2/2}d\theta$$
$$+ \int_{r_0[\hat{\theta}_0(v)]}^{\infty} \frac{e^{-r^2/2}}{\sqrt{2\pi}}dr \tag{27}$$

where $g_0(\theta)$ is defined in (24). The first integration is the normal right tail probability $\Phi\{r_0[\hat{\theta}_0(v)]\}$ while the second term in (27) can be integrated by parts by letting $dV = \theta K_0''(\theta)e^{-r_0(\theta)^2/2}d\theta$ and $U = 1/g_0(\theta) - 1/r_0(\theta)$. Since $V = -e^{-r_0(\theta)^2/2}$, and neglecting the small error term (see [25] for details on its magnitude) that results from the second integration, the simplified approximation formulas are

$$P_0(v) = \mathrm{Pr}(s > v|H_0)$$
$$\simeq \Phi\left\{r_0\left[\hat{\theta}_0(v)\right]\right\} + \varphi\left\{r_0\left[\hat{\theta}_0(v)\right]\right\}$$
$$\times \left[\frac{1}{g_0\left[\hat{\theta}_0(v)\right]} - \frac{1}{r_0\left[\hat{\theta}_0(v)\right]}\right], \tag{28}$$
$$P_1(v) = \mathrm{Pr}(s < v|H_1)$$
$$\simeq \Phi\left\{-r_1\left[\hat{\theta}_1(v)\right]\right\} - \varphi\left\{r_1\left[\hat{\theta}_1(v)\right]\right\}$$
$$\times \left[\frac{1}{g_1\left[\hat{\theta}_1(v)\right]} - \frac{1}{r_1\left[\hat{\theta}_1(v)\right]}\right] \tag{29}$$

where the second formula is obtained by following the same procedure as above to approximate the left tail integral. Equation (29) is often called the Lugannani-Rice formula, and it is one of the most popular and easy forms of the saddlepoint approximation of distributions. In summary, the approximation starts by solving the saddlepoint equation

$$K_i'(\hat{\theta}_i) = v$$

to find the saddlepoint $\hat{\theta}_i$. Then r_i and g_i are computed from (23) and (24) given the saddlepoint $\hat{\theta}_i$. Finally the left and right tail probabilities are approximated through (29).

C. Existence and Uniqueness of the Saddlepoint

The saddlepoint is obtained by solving the saddlepoint equation $K_i'(\hat{\theta}) = x$, under H_i, $i = 0,1$. The function $K_i'(\hat{\theta})$ is strictly increasing in $\hat{\theta}$ since its derivative $K_i''(\hat{\theta}) > 0$ is a variance. To verify this for the current problem, notice that the denominator in (20) is positive while the numerator can be written as

$$\left(\sum_{m=1}^{M} \chi_{mi}\right)\left(\sum_{m=1}^{M} \chi_{mi}L_m^2\right) - \left(\sum_{m=1}^{M} \chi_{mi}L_m\right)^2$$

where $\chi_{mi} = Q_{mi}e^{\hat{\theta}L_m} > 0$, $m = 1, \ldots M$. This can be simplified to

$$\sum_{j=1}^{M-1} \sum_{k=j}^{M} \chi_{ji}\chi_{ki}(L_j - L_k)^2$$

which is positive since $\chi_{ji}, \chi_{ki} > 0$ and, hence, $K_i''(\hat{\theta}) > 0$, and $K_i'(\hat{\theta})$ is strictly increasing in $\hat{\theta}$. Therefore, if a solution to the saddlepoint equation exists then it is unique.

Existence of the saddlepoint approximation depends on the interval on which the cumulant generating function $K_i(\theta)$ is defined, and on the form of the interval for the support of $f_{s,i}(x)$, [21]. For the decentralized detection problem at hand, the global LLR s takes values in $[NL_{\min}, NL_{\max}]$, where $L_{\min}$ and $L_{\max}$ are the minimum and maximum of L_m, $m = 1, \ldots M$. The CGF $K_i(\theta)$, on the other hand, is defined in $(-\infty, \infty)$. From (18), the limits of $K_i'(\theta)$ are given by

$$\lim_{\theta \to -\infty} K_i'(\theta) = NL_{\min} \text{ and } \lim_{\theta \to +\infty} K_i'(\theta) = NL_{\max}.$$

Therefore, a solution for the saddlepoint equation exists for any $x \in (NL_{\min}, NL_{\max})$. Further, since $K_i'(\theta)$ is strictly increasing, the solution is unique and can be found numerically using the Newton-Raphson method.

D. Accuracy of the Saddlepoint Approximation

The form presented in (29), (23), and (24) is often used to approximate the tail probabilities of sums of continuous random variables. For this reason, we refer to it as LR-Cont (i.e., the continuous form of the Lugannani-Rice approximation). However, the problem that we are considering involves the sum of discrete random variables ℓ_n. So, the question is: Is it still a good approximation when the random variables are discrete? Applying the same approximation above for discrete random variables has

been justified by Booth *et al.*, [22], by showing that the relative error of the approximation decays rapidly as the number of samples N grows. It is shown in [22] that, in almost all cases, the *relative* error is $o(N^{-1})$ or $O(N^{-3/2})$ when $M = 5,6$, or $M > 6$, respectively. We use the notation $A(N) = o(B(N))$ to mean that $A(N)/B(N) \to 0$ when $N \to \infty$, and the notation $A(N) = O(B(N))$ to mean that $A(N)/B(N)$ is bounded when $N \to \infty$). Recall that the sensor nodes produce M-ary local decisions, where $M = 2^b$ and b is the number of bits per sensor. The cases $M = 3,4$ (corresponding to ternary and quaternary sensor nodes, respectively) are not considered in [22], although numerical results show that the approximation performs well for these cases too. The case of $M = 2$ (binary sensor nodes) is a little different, since the distribution of ℓ becomes always lattice valued as we explain next.

The LLR random variable ℓ is lattice distributed when every possible value of it (L_m, $m = 1, \ldots, M$) is in the form $L_m = \delta_o + \delta_\Delta(m - 1)$, where $m = 1, \ldots, M$, and $\delta_o, \delta_\Delta \neq 0$ are, respectively, the offset and span of the lattice. When $M = 2$ (binary sensors), it is easy to see that ℓ is always lattice, regardless of the values of L_1 and L_2. When $M > 2$, the distribution of ℓ can also be lattice, but only under specific conditions on the noise distribution and the local detection rules.

We raise the issue of lattice versus nonlattice conditions for two reasons. First, it can be shown that, when ℓ is lattice, the fusion rule in (2) can be replaced with a simpler form of the majority rule, which makes its decisions based on the *integer* sum of the received local classifications $u_n \in \{1, 2, \ldots, M\}$ as follows:

$$\sum_{n=1}^{N} u_n \underset{0}{\overset{1}{\gtrless}} \frac{v - N\delta_o}{\delta_\Delta}. \tag{30}$$

This simplifies further the structure of the detector, which may be necessary for power and complexity constrained sensors nodes. Second, when the random variable ℓ is lattice-valued, there are other forms of the saddlepoint approximation that are specific for lattice-valued variables. One such approximation is obtained by using the same equations as before ((29) and (23)) but where g is replaced now with

$$g_i(\theta) = \frac{1}{\delta_\Delta}(1 - e^{-\delta_\Delta\theta})\sqrt{K_i''(\theta)} \tag{31}$$

where δ_Δ is the lattice span. This particular form has a relative error of $O(N^{-3/2})$ for any M provided that ℓ is lattice distributed, [15]. We refer to this approximation as LR-Latt (i.e., the lattice form of the Lugannani-Rice approximation). This approximation is valid only at the lattice edges. When the evaluation point falls in between, the approximation should be computed at the nearest right or left lattice edge depending on whether it is desired to compute the right or the left tail probability, respectively.

E. Complexity of the Saddlepoint Approximation

In addition to its high accuracy, the saddlepoint approximation is much more economical to compute than the direct approach. Evaluating the probability of error for a given local quantization rule requires computing few simple expressions in addition to finding the saddlepoint $\hat{\theta}$. The saddlepoint can be

obtained using numerical univariate techniques (e.g., Newton-Raphson), a much simpler task than the exact evaluation in (8), which has a computational complexity of order N^{2^b-1}. To solve for the saddlepoint, the computation of $K'_i(\theta)$ and $K''_i(\theta)$ at each step of the Newton-Raphson algorithm requires at most $M+1$ additions, $M+3$ multiplications, three divisions, and M exponentiations. For example, in a network of 100 sensors with $b = 3$ bits/sensor local detectors $(M = 8)$, the LLR computation requires a combined total of roughly 500 operations to reach the saddlepoint. Compare this with the exact computation, which, for the same example, requires roughly 2.6×10^{10} additions, 1.8×10^{11} multiplications, 1.8×10^{11} exponentiations, and 2.6×10^{10} binomial coefficients.

V. Experimental Results

In this section, we demonstrate experimentally the accuracy and usefulness of the saddlepoint formulas derived in Section IV by calculating in Section V-B the detection performance, i.e., the receiver operating characteristic, of a given network detector, not necessarily optimal, by computing in Section V-C the average error probability, the metric of interest in Bayes detection, by evaluating in Section V-D the probability of false alarm and the probability of a miss, quantities of interest in NP detection, and in Section V-E the thresholds for the local detectors. We start in Section V-A by clarifying the difference between the lattice and the nonlattice saddlepoint conditions.

Shift in Mean Model: The measurements are modeled by an additive noise, shift-in-mean observation model where

$$\text{under } H_i: \ y = \mu_i + \xi \tag{32}$$

where μ_i is the signal mean under H_i and ξ is an additive noise of known distribution with zero mean and variance σ^2. In the study, we focus on three noise distributions including the Gaussian, Laplacian, and the logistic distributions where, respectively, the pdfs of the last two are given by

$$f_{\text{Laplace}}(y) = \frac{1}{2\vartheta}e^{-|y-\mu|/\vartheta}, \quad \vartheta = \frac{1}{\sqrt{2}}\sigma \tag{33}$$

$$f_{\text{Logistic}}(y) = \frac{e^{-(y-\mu)/\rho}}{\rho\left[1 + e^{-(y-\mu)/\rho}\right]^2}, \quad \rho = \frac{\sqrt{3}}{\pi}\sigma. \tag{34}$$

While the use of the Gaussian and the Laplacian models is justified in many practical scenarios [3], the logistic distribution is included here for illustration purposes. These assumptions are introduced to facilitate the numerical studies but are in no way necessary for the proposed approximation. The approximation technique presented in this paper can be applied to other types of local detectors and other observation models.

We adopt the quantization rule in (1) and assume that the observations follow the shift in mean model described by (32). Unless otherwise specified, we assume $\mu_1 = -\mu_0 = \mu = 1$ while the variance σ^2 of the measurement noise depends on $\text{SNR} = 10\log_{10}\mu^2/\sigma^2$.

A. Lattice Versus Nonlattice Conditions

Here, we demonstrate the difference between lattice and nonlattice conditions and evaluate the accuracy of both forms of the saddlepoint approximation. We work with the shift-in-mean
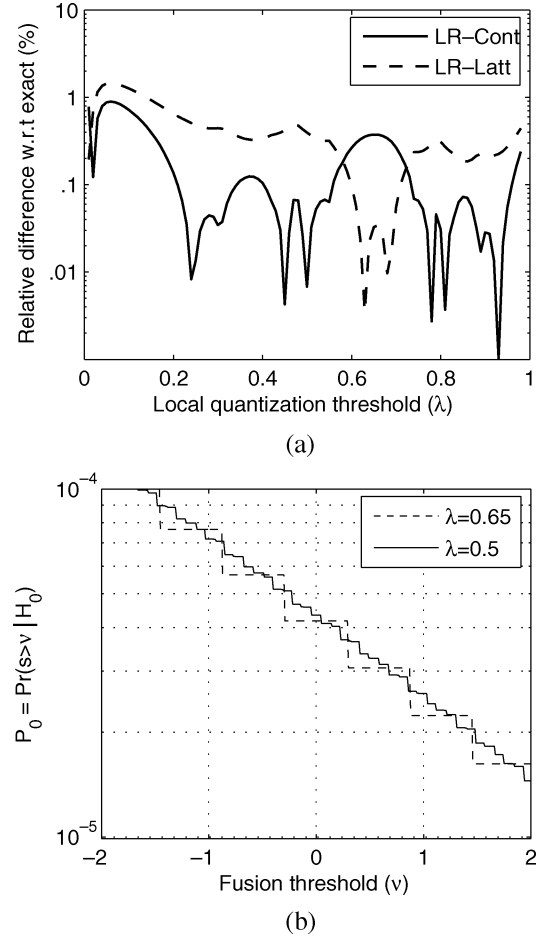

(a)

(b)

Fig. 2. (a) Accuracy of the LR-Cont and the LR-Latt approximations at different local quantization thresholds. The fusion threshold is fixed at $v = 0$. (b) Exact false alarm probability P_0 at different fusion thresholds under lattice $(\lambda = 0.65)$ and nonlattice $(\lambda = 0.5)$ conditions.

model of (32) and assume that the noise is Laplacian and the signal-to-noise ratio (SNR) is -10 dB. We consider a network of $N = 100$ quaternary sensors $(b = 2$ bits/sensor) with a symmetric local decision threshold vector $\lambda = (-\lambda, 0, \lambda)$. The value of λ changes in the range from 0 to 1 in order to produce examples where ℓ becomes lattice distributed. In implementing the LR-Latt approximation using $g_i(\theta)$ given by (31), we take the span of the lattice to be $\delta_\Delta = L_2 - L_1$. Fig. 2(a) compares the relative error of the LR-Cont and LR-Latt approximations. The figure demonstrates the high accuracy of the LR-Cont approximation for most of the range, where the relative error is below 1%, often about 0.01%, except around $\lambda = 0.65$. When $\lambda = 0.65$, the relative error for the LR-Latt approximation dips well below the relative error for the LR-Cont approximation. This is because for $\lambda = .65$ the values of $L_m, m = 1, \ldots, M$ belong to a lattice with span $\delta_\Delta \simeq 0.58$. This is further illustrated in Fig. 2(b), where the exact probability $\Pr(s > v|H_0)$ is plotted for two values of the local detection threshold: $\lambda \simeq 0.65$ (corresponding to the lattice case); and $\lambda = 0.5$. The network size is still fixed at $N = 100$ while the fusion threshold v is varied from -2 to 2. The plot clearly illustrates the regular wide jumps for the lattice case. In contrast, when $\lambda = 0.5$, the jumps become irregular and closely spaced. As the number of sensors
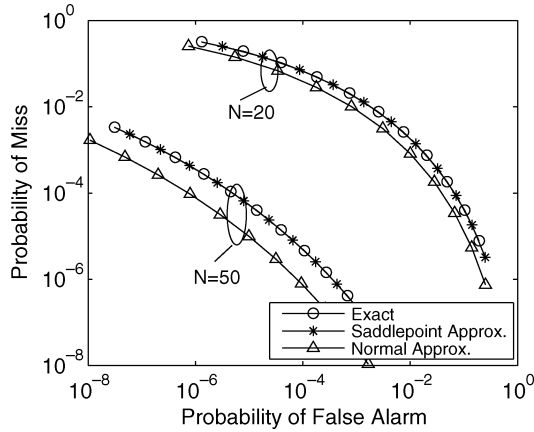
Fig. 3. ROC for networks of $N = 20$ and $N = 50$ sensors at $\mathrm{SNR} = -5$ dB.

N is increased the jumps become even closer (not shown here due to space limitation).

B. Receiver Operating Characteristics (ROC)

We now illustrate the use of the saddlepoint approximation to compute the receiver operating characteristic (ROC). In this case we need to compute the probability of false alarm and the probability of detection, not their asymptotic exponential decay rates. We consider a network of N sensors, where each sensor has a $b = 3$ bit local detector, i.e., $M = 8$ quantization levels. The local detection thresholds are fixed arbitrarily at $\lambda = (-2, -1, 0, 1, 2)$. The noise is assumed to be logistic and the $\mathrm{SNR} = -5$ dB. Fig. 3 shows the ROC of two networks of $N = 20$ and $N = 50$ sensors obtained by computing P_0 and P_1 at different fusion thresholds v. The error probabilities are computed using the exact, the normal approximation, and the saddlepoint approximation. Asymptotic measures can not be used here since they do not compute the error probabilities P_0 and P_1, which are necessary for computing the ROC curve.

We discuss first the normal approximation. There is a significant relative difference between the normal approximation and the exact ROC curves, which approaches one order of magnitude for the $N = 50$ sensor network. This difference is smaller for the $N = 20$ network, an observation that will be examined further in the following subsection. On the other hand, Fig. 3 shows a nearly perfect agreement between the exact ROC curves and the saddlepoint approximation for both networks $N = 50$ and $N = 20$. We emphasize that, for example, when $N = 50$, evaluating each point of the direct ROC curve requires computing a sum of about 2.6×10^8 terms, while the saddlepoint approximation requires roughly 500 operations regardless of N.

C. Bayesian Detection

We now illustrate how the saddlepoint approximation can be used to compute the average probability of error, which is the metric of interest in Bayes detection. We consider a setup similar to that described in Section V-B, with equal priors $\pi_0 = \pi_1 = 1/2$. To design the $\min P_e$ detector or study its performance, we need a speedy and accurate way of computing the probability of error P_e. The average probability of error $P_e = \pi_0 P_0 + \pi_1 P_1$ is
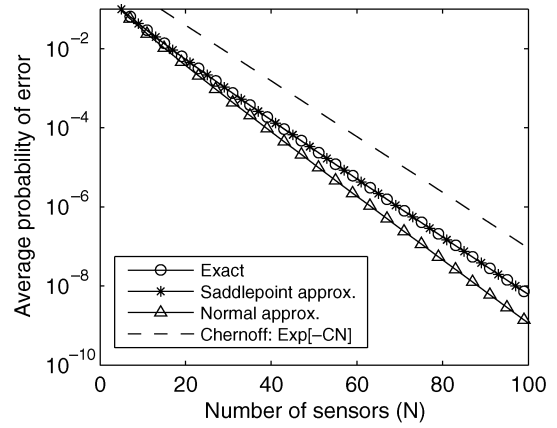
(a)
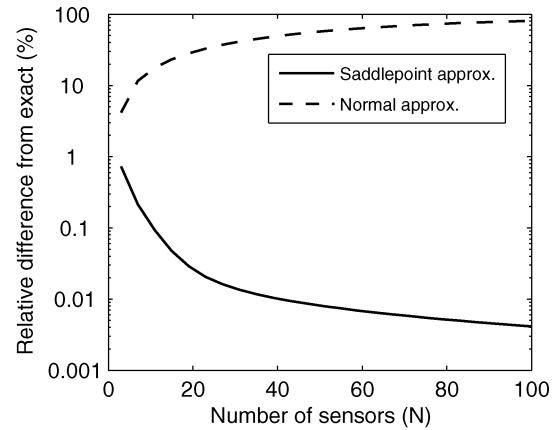
(b)

Fig. 4. (a) Average probability of error for a network of N sensors with 3-bit quantizers. (b) Relative differences between the exact and approximate values of P_e (i.e., $|\widetilde{P}_e - P_e|/P_e \times 100$).

approximated using different approximation methods and compared with the exact value as the number of sensors N grows. The saddlepoint approximation agrees with the exact P_e, the corresponding plots fall on top of each other, as seen in Fig. 4(a), while the normal approximation gets worse for higher values of N. The *relative differences* ($|\widetilde{P}_e - P_e|/P_e \times 100$) between the probability of error computed by the saddlepoint formulas, or the normal approximation, and the exact value of the probability of error are shown in Fig. 4(b). While the saddlepoint approximation leads to relative errors that rapidly fall below 0.1%, the relative errors in the normal approximation approach 100%. The accuracy of the normal approximation becomes worse as the number of sensors is increased. On the contrary, the accuracy of the saddlepoint approximation improves when the number of sensors is increased and it performs well even for small networks.

Also, included in Fig. 4(a) is the e^{-CN} dashed line representing the Chernoff estimate of the error probability, where $C = 0.1621$ is the Chernoff distance computed from (13). There is about one order of magnitude difference between the Chernoff bound e^{-CN} and the exact error probability P_e. The slope of the exact P_e curve approaches that of the Chernoff bound when $N \rightarrow \infty$. This issue is further investigated in Fig. 5 where we compute the "true" exponential decay rate of
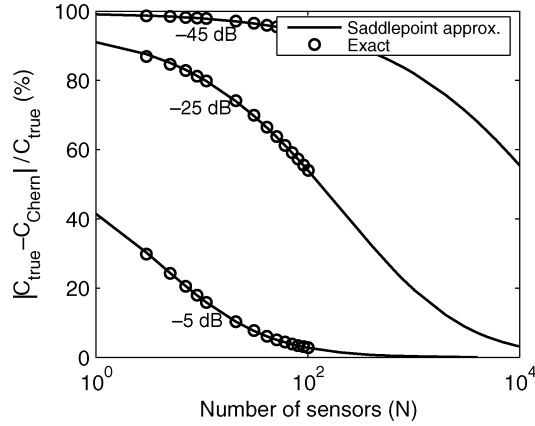
Fig. 5. Relative difference between the "true" decay rate C_{true} and the asymptotic Chernoff rate C_{Chern} (i.e., $|C_{\text{true}} - C_{\text{Chern}}|/C_{\text{true}} \times 100$). Markers indicate points computed exactly while the solid line is obtained using the saddlepoint approximation.

$P_e(N)$ and compare it to the asymptotic Chernoff rate in (13). Fig. 5 shows the relative difference between the "true" decay rate $C_{\text{true}}(N)$ and the asymptotic Chernoff rate C_{Chern} (i.e., $|C_{\text{true}}(N) - C_{\text{Chern}}|/C_{\text{true}}(N) \times 100$) at three different SNR values. The true decay rate $C_{\text{true}}(N)$ is obtained through numerical differentiation of $\log(P_e(N))$, where $P_e(N)$ is computed using both, the exact formula (circle markers; up to $N = 100$ only, due to complexity issues) and the saddlepoint approximation (solid line; for the whole N range since complexity is independent of N). From Fig. 5, it is clear that the convergence speed of the true decay rate towards the Chernoff asymptotic rate depends on the SNR. For higher SNR values, the Chernoff rate can accurately estimate the true decay rate even for networks with ten's of sensors. For low SNRs, however, the Chernoff rate is an accurate estimate of the true decay rate only for large networks with hundreds or even thousands of sensors. Moreover, while the Chernoff distance in (13) is known to be an asymptotically ($N \to \infty$) tight bound on the *exponential decay rate* of P_e, it fails to provide a reasonable approximation to the probability of error P_e itself as can be seen from Fig. 4(a).

D. NP Detection

Here we illustrate the application of the saddlepoint approximation to computing the probability of false alarm, P_0, and the probability of a miss, P_1, quantities of interest in designing NP detectors. Consider the setup described in Section V-B where we do not have knowledge of the priors (π_0 and π_1), and we want to solve the NP detection problem for a given size α and a given number of sensors N. By using numerical root finding techniques, we solve for the fusion threshold v such that $P_0(v) = \alpha$. The root finding algorithm is run multiple times with different initial values of v to find the global solution. We remark that, in all of the considered cases, for fixed α and N, the algorithm converges to the same solution. Fig. 6(a) shows the NP thresholds v as a function of the number of sensors N under two false alarm probabilities, 10^{-3} and 10^{-7}. For each false alarm probability, the three curves in Fig. 6(a) represent the solution of the NP problem using the exact, normal approximation, and saddlepoint approximation. The exact solution is only provided up to $N = 27$ due to complexity issues. (In the numerical algorithm,
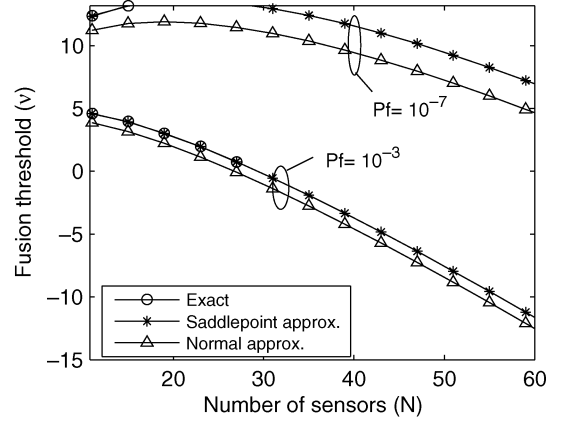
(a)
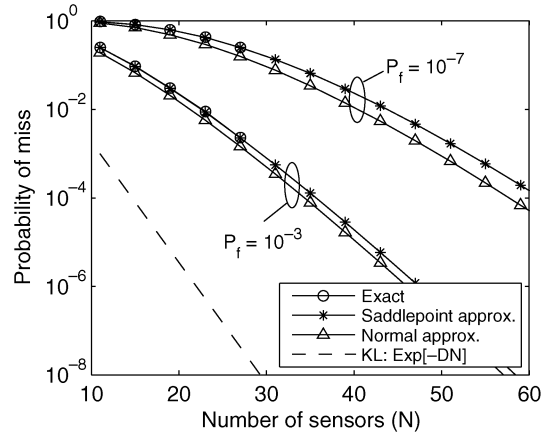
(b)

Fig. 6. (a) Fusion threshold v for the NP detection problem under different number of sensors N and false alarm constraints $P_0 < \alpha, \alpha \in \{10^{-3}, 10^{-7}\}$. (b) The corresponding minimum probability of miss (P_1).

P_0 needs to be computed about 100 times to reach the solution v. At $N = 27$, each exact evaluation of P_0 requires computing 5×10^6 terms.) The complexity of the saddlepoint and normal approximations is independent of N and, hence, they can be used for larger values of N. In Fig. 6(a) there is a significant difference between the threshold v obtained by using the exact error probabilities and values computed using the normal approximation, especially for lower values of the false alarm probability. On the other hand, the threshold obtained using the saddlepoint approximation coincides almost perfectly with the true optimum thresholds.

The corresponding minimum miss probability P_1 is shown in Fig. 6(b) where we also include a dashed line representing e^{-DN}, where $D = 0.6285$ is the KL distance computed from (14). The KL distance provides the asymptotic exponential *rate of decay* of P_1 when $N \to \infty$ and the false alarm probability $\alpha \to 0$. Fig. 6(b) illustrates the high accuracy of the saddlepoint approximation in the context of NP problems.

E. Optimizing the Local Detectors Thresholds

In previous examples, the local detectors thresholds at the sensors are fixed arbitrarily. Now, we consider optimizing these thresholds with respect to the error probability. This problem has been considered in the past in a variety of contexts using
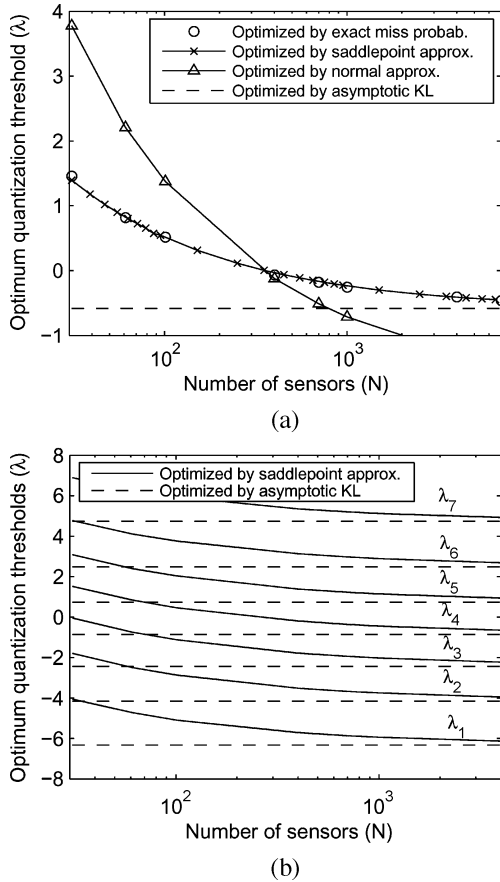
Fig. 7. (a) The local quantization threshold optimized with respect to the miss probability when the sensors are binary ($b = 1$). (b) When the sensors use $M = 8$-ary detectors (b = 3, 7-thresholds).

different optimization criteria (see, e.g., [26] and [4, Sec. 3.4]). Here we optimize the thresholds with respect to the saddlepoint approximation of the error performance. We consider a NP problem with a probability of false alarm $P_f = 10^{-6}$. The measurement noise is assumed to be Gaussian and the SNR $= -10$ dB. Notice that even though the measurements are Gaussian, the decisions of the local detectors, which are used in the global fusion, are not Gaussian. These quantized measurements are discrete and, hence, approximating the distribution of their LLR sum is not a trivial task as it might appear at first sight.

Binary Sensors: We first consider a network of binary ($b = 1$) sensors with a common quantization threshold λ. The quantization threshold λ and the fusion threshold v are optimized numerically to minimize the miss probability P_1 subject to the constraint that $P_0 < 10^{-6}$ using the exact method of computing the miss probability as well as its normal and saddlepoint approximations. In optimizing the approximate miss probability, we run the algorithm multiple times with different initial conditions to find the global optimum. In all of the cases that we considered, the algorithm converges to the same solution. On the other hand, in addition to the linear complexity growth with N, the exact error probability is discrete and, hence, it is much more difficult to optimize than its continuous approximations. In this case, we evaluate the exact error probability at fine displacements ($\Delta\lambda = 0.001$) to find the threshold that yields the lowest miss probability. Fig. 7(a) shows the optimization results; it is

clear from these that the saddlepoint-optimized thresholds are very close to those optimized using the exact method. On the other hand, thresholds obtained by optimizing the normal approximation are considerably different from the optimal thresholds. It can also be seen that the saddlepoint-optimized thresholds approach the KL-optimized threshold as $N \to \infty$ [dashed line in Fig. 7(a)]: $\lambda_{\text{KL}} \simeq -0.586$); they come within 20% of the asymptotic threshold when the number of sensors N is larger than 7000.

$b = 3$-bit Sensors: In Fig. 7(b), we consider a similar example as in the previous paragraph, but with $M = 8$-ary ($b = 3$ bits/sensor) local detectors; we assume that these local detectors all have common thresholds $\lambda = (\lambda_1, \ldots, \lambda_7)$. The thresholds are optimized with respect to the saddlepoint approximation of the miss probability and are shown in Fig. 7(b). The dashed lines represent the values of the seven thresholds when optimized using the asymptotic KL distance. Note that, in this case where $M = 8$, optimization could not be carried out with respect to the exact miss probability due to its extremely high computational complexity.

The examples in Fig. 7 show that the local detectors designed using the normal approximation may be significantly different from the optimal ones. Also, designing the local detectors using the asymptotic decay rate may require a very large number of sensors, in the examples provided $N > 7000$, to yield the correct threshold values. In contrast, the thresholds designed using the saddlepoint approximation are indistinguishable from the correct ones.

VI. EXTENSIONS

A. Nonidentical Measurements and Nonidentical Local Detectors

So far, we have only considered scenarios where the measurements are identically distributed across sensors and we assumed identical local detectors. Here, we demonstrate the application of the saddlepoint approximation to approximate error probabilities when the observations are nonidentically distributed and the local detectors are nonidentical. Consider a network of J groups of sensors. Sensors within the same group use identical local detectors and their measurements are identically distributed. But across groups, the number of sensors N_j, the measurement conditional densities $f_{y_j}(y|H_i)$, the local detectors, and the number of quantization levels M_j, $j = 1, \cdots, J$, are possibly different. The sensors from all J groups communicate their local decisions to a single fusion center, which makes the final binary decision as in Fig. 1. Saddlepoint techniques can be applied to compute the error probabilities accurately. The key point is to use the saddlepoint approximation of densities (22) to approximate the density $f_{s_j,i}(x) = f_{s_j}(x|H_i)$ of the LLR sum $s_j = \sum_{n=1}^{N_j} \ell_{j,n}$, $j = 1, \ldots, J$, of each group under H_i, $i = 0,1$, where $\ell_{j,n} \in \{L_{j,1}, \ldots, L_{j,M_j}\}$ is the LLR of the decision of the nth sensor of the jth group. The values of the local LLRs at the jth group are given by $L_{j,m} = \log[Q_{j,m1}/Q_{j,m0}]$, where $Q_{j,mi}$ is the probability of m-decisions at the output of sensors of the jth group under H_i. The density $f_{s,i}(x)$ of the global LLR $s = \sum_{j=1}^{J} s_j$ is estimated by convolving all the J

densities $\tilde{f}_{s_j,i}(x)$, $j = 1, \ldots, J$. The probability of error is finally approximated by integrating the tails of the approximate density.

We illustrate the technique by computing the global false alarm probability $P_0 = \Pr(s > v|H_0)$ in a network composed of $J = 2$ groups, where $s = s_1 + s_2$ is the global LLR sum at the fusion center. Applying the independence assumption, we evaluate the following double integration numerically

$$P_0 \simeq \int_{\max\{v, a_2+x\}}^{b_2+x} \int_{a_1}^{b_1} \tilde{f}_{s_1,0}\left[x, \hat{\theta}_{1,0}(x)\right] \\ \times \tilde{f}_{s_2,0}\left[z - x, \hat{\theta}_{2,0}(z-x)\right] dx\,dz \quad (35)$$

where

$$\tilde{f}_{s_j,i}(x, \theta) = \frac{e^{K_{j,i}(\theta) - x\theta}}{\sqrt{2\pi K_{j,i}''(\theta)}}$$

and $K_{j,i}(\theta)$ is the cumulant generating function of s_j under H_i, i.e.,

$$K_{j,i}(\theta) = N \log \sum_{m=1}^{M_j} Q_{j,mi} e^{\theta L_{j,m}}$$

and $\hat{\theta}_{j,i}(x)$ is the saddlepoint obtained by solving $K_{j,i}'(\hat{\theta}_{j,i}) = x$. The limits of integration rely on the support intervals of s_1 and s_2 and are given by $a_j = N_j L_{\min}^{(j)}$ and $b_j = N_j L_{\max}^{(j)}$, where $L_{\min}^{(j)} = \min_n(L_{j,m})$ and $L_{\max}^{(j)} = \max_m(L_{j,m})$.

At each step of the numerical integration in (35), one has to solve for two saddlepoints $\hat{\theta}_{1,0}(x)$ and $\hat{\theta}_{2,0}(z-x)$. This can be simplified by performing a change of variables $x = K_{1,0}'(\theta)$. Since $dx = K_{1,0}''(\theta)d\theta$, (35) can be rewritten as

$$P_0 \simeq \int_{\max\{v, a_2+K_{1,0}'(\theta)\}}^{b_2+K_{1,0}'(\theta)} \int_{-\infty}^{\infty} K_{1,0}''(\theta)\tilde{f}_{s_1,0}(x, \theta) \\ \times \tilde{f}_{s_2,0}\left(z - K_{1,0}'(\theta), \hat{\theta}_{2,0}\left[z - K_{1,0}'(\theta)\right]\right) d\theta\,dz \quad (36)$$

which requires computing only one saddlepoint $\hat{\theta}_{2,0}$ (by solving $K_{2,0}'(\hat{\theta}_{2,0}) = z - K_{1,0}'(\theta)$) at each integration step.

We demonstrate the accuracy of the saddlepoint point method through the following example. We consider two groups of sensors with the following parameters:

Group 1: $N = 20$ quaternary sensors, i.e., with $b = 2$ bit local detectors at each sensor, $\lambda = (-1, 0, 1)$, Gaussian noise, $\mu_1 = -\mu_0 = 0.1$, $\sigma^2 = 1$.

Group 2: $N = 40$ binary sensors, i.e., with $b = 1$ bit local detectors at each sensor, $\lambda = (0)$, Laplacian noise, $\mu_1 = -\mu_0 = 0.05$, $\sigma^2 = 1$.

All sensors in each group send their b-bit local decision to a single fusion center that computes the global LLR s and then makes the final decision u_0 by comparing s against the fusion threshold v. Using the saddlepoint approximation in (36), we compute the global error probabilities P_0 and P_1 for different
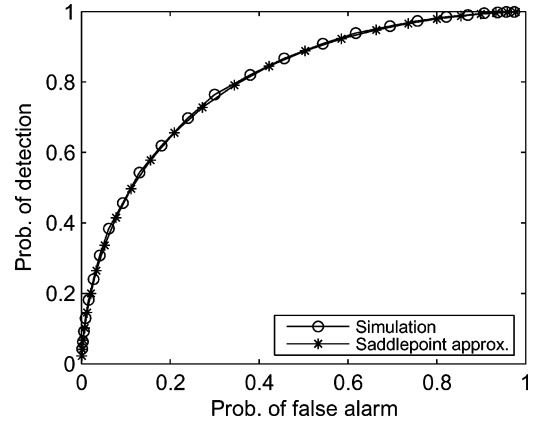

Fig. 8. ROC of a network of nonidentical sensors with nonidentically distributed measurements.

values of the fusion threshold v, from which we get the ROC in Fig. 8. Fig. 8 also includes the ROC curve obtained with 10^6 Monte Carlo runs. The two plots show a very good agreement of the ROC curve derived by Monte Carlo and its estimate provided by the saddlepoint approximation.

B. Imperfect Communication Channels

Up to this point, we have assumed that the communication links between the sensors and the fusion center are noiseless, so that the local decisions are delivered to the fusion center without errors. Now, we extend the saddlepoint techniques to practical cases involving noisy communication links between the sensors and the fusion center. We model the communication links between the sensors and the fusion center as independent discrete memoryless channels (DMC). The DMC is characterized by a stochastic channel transition matrix $[h_{ij}]$, where h_{ij}, $i, j = 1, \ldots, M$ is the probability that a local decision $u_n = j$ of one of the sensors is received incorrectly as $w_n = i$ at the fusion center, i.e., $h_{ij} = \Pr(w_n = i | u_n = j)$. The fusion center makes its decision based on the received noisy M-ary messages $w_n \in \{1, \ldots, M\}$, $n = 1, \ldots, N$. The communication links of all sensors are assumed to be identical and independent of each other.

Saddlepoint techniques can be easily extended to this case since the noisy decisions w_n received at the fusion center are i.i.d. The probability of receiving a message $w_n = m$ under hypothesis H_i is given by

$$\tilde{Q}_{mi} = \Pr(w_n = m | H_i) \\ = \sum_{k=1}^{M} \Pr(w_n = m | u_n = k) \Pr(u_n = k | H_i) \\ = \sum_{k=1}^{M} h_{mk} Q_{ki}$$

where $Q_{ki} = \Pr(u_n = k | H_i)$. The LLR of the noisy message w_n of the nth sensor is

$$\tilde{\ell}_n = \log\left[\Pr(w_n | H_1) / \Pr(w_n | H_0)\right] \in \{\tilde{L}_1, \ldots, \tilde{L}_M\}$$

where $\tilde{L}_m = \log(\tilde{Q}_{m1}/\tilde{Q}_{m0})$, $m = 1, \ldots, M$. The fusion center makes its decision based on the global LLR statistic

$$s = \sum_{n=1}^{N} \tilde{\ell}_n \underset{0}{\overset{1}{\gtrless}} v.$$

It is straightforward to see that the same saddlepoint approximation formulas for densities and distributions presented in Section IV can be used here too, except that the conditional probabilities Q_{mi}, $m = 1, \ldots, M$, $i = 0,1$, of the "error-free" local decisions u_n and the LLR values L_m should now replaced by their "noisy" counterparts $\tilde{Q}_{mi}$ and $\tilde{L}_m$.

To illustrate the application of the saddlepoint approximation under noisy communication channels, we consider a Bayesian detection problem where the priors are assumed equal, i.e., $\pi_0 = \pi_1 = 1/2$ ($v = 0$). A network of 50 binary sensors collects noisy measurements corrupted with Laplacian noise. Note that this distribution describes the measurement noise, not the communication links. The local detectors all use the same threshold $\lambda = (0)$. The local binary decisions $u_n \in \{0, 1\}$ are transmitted through independent binary symmetric channels (BSC) with transition probabilities $h_{ij} = \Pr(w_n = i | u_n = j) = \varepsilon$ if $i \neq j$ and $h_{ij} = 1 - \varepsilon$ if $i = j$, where ε represents the error rate of the *communication links*. Fig. 9 shows the probability of error P_e at the fusion center computed using the saddlepoint approximation compared to that obtained by exact evaluation. Notice that, in this example, where $b = 1$ ($M = 2$), the computational complexity of the exact method grows linearly with the number of sensors N. The exact method becomes much more complicated when $b > 1$. The complexity of the saddlepoint approximation is independent of N regardless of the number of quantization bits b. We address the simple case of $b = 1$ here to get insight on the effect of communication bit-errors on the global decision fusion. As expected, the reliability of the global decisions at the fusion center deteriorates when the bit error rate of the communication links between the sensors and the fusion center increases. More interestingly, Fig. 9 shows that the fusion performance under unreliable communication links can be close to that under error-free communication links even when the bit-error-rate of the communication links is relatively high. For instance, when $\mathrm{SNR} = -2$ dB, the fusion probability of error is about 10^{-8} under error-free communication links. It can be seen from Fig. 9 that the fusion performance remains close to 10^{-8} even when the bit error rate of the communication links is as high as 10^{-3}. This may be attributed to the inherent coding capabilities resulting from the fusion of multiple independent decisions. This issue is the subject of a future study. The main emphasis here is to illustrate the application of the saddlepoint approximation under unreliable communication links. Fig. 9 illustrates the good agreement between the values of the probability of error calculated by the saddlepoint approximation and their exact values.

VII. CONCLUSION

The paper presents a large deviation theory method, the saddlepoint approximation, to compute the detection probabilities—e.g., average probability of error, probability of false
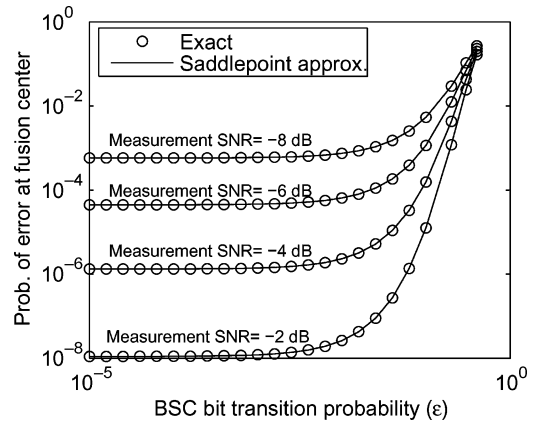

Fig. 9. Probability of decision error at the fusion center $\Pr(\widehat{H} \neq H)$ when the local decisions of the sensors are sent to the fusion center through a binary symmetric channel with a bit-transition rate of ε.

alarm, and probability of detection—in distributed detection in sensor networks. The saddlepoint technique is highly accurate and simple to compute, providing an efficient method to perform various design and performance assessment tasks such as computing ROCs, designing fusion rules and local detectors thresholds, and computing error probabilities, regardless of the network size. In all experimental studies, the results obtained with the saddlepoint approximation practically coincide with the exact ones (when available), in contrast with the corresponding results obtained with the normal approximation or using the asymptotic exponential decay rate of the error probabilities. We demonstrated the application of the saddlepoint approximation in practical situations when the observations and local detectors are different from sensor to sensor and when the communication among sensors and the fusion center is through imperfect, noisy communication links.

REFERENCES

[1] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification and tracking of targets," *IEEE Signal Process. Mag.*, vol. 19, pp. 17–29, Mar. 2002.

[2] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors: part I—Fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, Jan. 1997.

[3] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, pp. 64–79, Jan. 1997.

[4] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1996.

[5] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, pp. 98–101, Jul. 1981.

[6] W. A. Hashlamoun and P. K. Varshney, "An approach to the design of distributed Bayesian detection structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, pp. 1206–1211, Sep./Oct. 1991.

[7] ——, "Near-optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. 44, pp. 294–297, Mar. 1996.

[8] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. 25, pp. 479–484, May 1977.

[9] H. Delic, P. Papantoni-Kazakos, and D. Kazakos, "Fundamental structures and asymptotic performance criteria in decentralized binary hypothesis testing," *IEEE Trans. Commun.*, vol. 43, pp. 32–43, Jan. 1995.

[10] J.-S. Leu and A. Papamarcou, "Asymptotic optimality of likelihood ratio threshold tests in decentralized detection," *IEEE Trans. Inf. Theory*, vol. 45, pp. 572–585, Mar. 1999.

[11] 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.

[12] B. Chen, R. Jiang, T. Kasetkasem, and P. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3454–3458, Dec. 2004.

[13] C. K. Sestok, M. R. Said, and A. V. Oppenheim, "Randomized data selection in detection with applications to distributed signal processing," *Proc. IEEE*, vol. 91, pp. 1184–1198, Aug. 2003.

[14] R. Niu, P. Varshney, M. H. Moore, and D. Klamer, "Distributed fusion in a wireless sensor network with a large number of sensors," presented at the 7th Int. Conf. Information Fusion, Stockholm, Sweden, Jun. 2004.

[15] J. L. Jensen, *Saddlepoint Approximations*. New York: Oxford Univ. Press, 1995.

[16] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr., Signals, Syst.*, vol. 1, no. 2, pp. 167–182, 1988.

[17] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Inf. Theory*, vol. 35, pp. 995–1000, Sep. 1989.

[18] C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-25, pp. 761–765, Sep. 1989.

[19] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, no. 4, pp. 550–558, Apr. 1993.

[20] P. K. Willett and D. J. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inf. Theory*, vol. 38, pp. 355–361, Mar. 1992.

[21] J. Kolassa, *Series Approximation Methods in Statistics*. New York: Springer-Verlag, 1994.

[22] J. G. Booth, P. Hall, and A. T. Wood, "On the validity of Edgeworth and saddlepoint approximations," *J. Multivar. Anal.*, vol. 51, pp. 121–138, Oct. 1994.

[23] N. Reid, "Saddlepoint methods and statistical inference," *Statist. Sci.*, vol. 3, pp. 213–227, May 1988.

[24] H. E. Daniels, "Tail probability approximations," *Int. Statist. Rev.*, vol. 55, pp. 37–48, Apr. 1987.

[25] G. R. Terrell, "A stabilized Lugannani-Rice formula," presented at the Symp. Interface: Computing Sci. Statist., Salt Lake City, UT, Mar. 2003.

[26] H. V. Poor and J. B. Thomas, "Optimum quantization for local decisions based on independent samples," *J. Franklin Inst.*, vol. 303, pp. 549–561, Jun. 1977.

Saeed A. Aldosari (S'02–M'06) received the B.S. and M.S. degrees in electrical engineering from King Saud University, Riyadh, Saudi Arabia, in 1994 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 2005.

He is currently an Assistant Professor with the Electrical Engineering Department, King Saud University. His research interests include distributed signal processing, sensor networks, and multiinput multioutput wireless communication systems.

José M. F. Moura (S'71–M'75–SM'90–F'94) received the electrical engineering degree in 1969 from the Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and the D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1973 and 1975, respectively.

He is a Professor of Electrical and Computer Engineering and of BioMedical Engineering at Carnegie Mellon University, Pittsburgh, PA, where he is a founding codirector of the Center for Sensed Critical Infrastructures Research (CenSCIR). During the 2006–2007 academic year, he is a visiting Professor of EECS at MIT, on sabbatical leave from Carnegie Mellon University. He was on the faculty at IST (1975–1984) and has held visiting faculty appointments with MIT (1984–1986 and 1999–2000) and as a research scholar with the University of Southern California, Los Angeles (summers of 1978–1981). His research interests include statistical signal, image and bioimaging, and video processing, algebraic signal processing, and digital communications. He has published more than 300 technical journal and conference papers, is the coeditor of two books, holds six patents on image and video processing, and digital communications with the U.S. Patent Office, and has given numerous invited seminars at U.S. and European Universities and industrial and government Laboratories.

Dr. Moura has served the IEEE Signal Processing Society (SPS) in several capacities, including President Elect (2006–2007), Vice President for Publications, and member of the Board of Governors (2000–2002), Editor-in-Chief for the IEEE TRANSACTIONS IN SIGNAL PROCESSING (1975–1999), interim Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS (December 2001–May 2002), founding member of the Bioimaging and Signal Processing (BISP) Technical Committee, and a member of several other Technical Committees. He was Vice-President for Publications for the IEEE Sensors Council (2000–2002) and is or was on the Editorial Board of several journals, including the IEEE PROCEEDINGS, the IEEE SIGNAL PROCESSING MAGAZINE, and the ACM TRANSACTIONS ON SENSOR NETWORKS. He chaired the IEEE TAB Transactions Committee (2002–2003) that joins the more than 80 Editors-in-Chief of the IEEE TRANSACTIONS and served on the IEEE TAB Periodicals Review Committee (2002–2005). He is on the Steering Committees of the International Symposium on BioImaging (ISBI) and of the International Conference on Information Processing and Sensor Networks (IPSN), and has been on the program committee of over 30 conferences and workshops. He was on the IEEE Press Board (1991–1995). He is a Fellow of the American Association for the Advancement of Science, and a corresponding member of the Academy of Sciences of Portugal (Section of Sciences). He was awarded the 2003 IEEE Signal Processing Society meritorious service award and the IEEE Millenium Medal in 2000. He is affiliated with several IEEE societies, Sigma Xi, AMS, AAAS, IMS, and SIAM.