Decentralized Detection in Sensor Networks

Jean-François Chamberland, *Student Member, IEEE*, and Venugopal V. Veeravalli, *Senior Member, IEEE*

***Abstract*—In this paper, we investigate a binary decentralized detection problem in which a network of wireless sensors provides relevant information about the state of nature to a fusion center. Each sensor transmits its data over a multiple access channel. Upon reception of the information, the fusion center attempts to accurately reconstruct the state of nature. We consider the scenario where the sensor network is constrained by the capacity of the wireless channel over which the sensors are transmitting, and we study the structure of an optimal sensor configuration. For the problem of detecting deterministic signals in additive Gaussian noise, we show that having a set of identical binary sensors is asymptotically optimal, as the number of observations per sensor goes to infinity. Thus, the gain offered by having more sensors exceeds the benefits of getting detailed information from each sensor. A thorough analysis of the Gaussian case is presented along with some extensions to other observation distributions.**

***Index Terms*—Bayesian estimation, decentralized detection, sensor network, wireless sensors.**

I. Introduction

THE ESTIMATION of a random variable based on noisy observations is a standard problem in statistics. In this work, we investigate the related scenario where information about a random variable is made available to a fusion center by a set of geographically separated sensors. Each sensor receives a sequence of observations about the state of nature H and transmits a summary of its information over a wireless multiple access channel. Based on the received data, the fusion center produces an estimate of the state of nature. We focus our attention on the special case of binary hypothesis testing, where H takes on one of two possible values, where the observations across sensors are independent and identically distributed (i.i.d.) conditioned on H, and where the observation process at each sensor conditioned on H is a sequence of i.i.d. random variables.

If the structure of the information supplied by each sensor is predetermined, the fusion center faces a classical hypothesis testing problem. The probability of estimation error is then minimized by the maximum *a posteriori* detector. Alternatively, one can consider the problem of deciding what type of information each sensor should send to the fusion center. Providing some answers to the latter question will be the aim of this paper. We begin by mentioning that several different variants of this problem have been studied in the past. Notably, the class of decentralized detection problems where each sensor must select one of D possible messages has received much attention. In this setting, which was originally introduced by Tenney and Sandell [1], the goal is to find what message should be sent by which sensor and when. See Tsitsiklis [2] and the references contained therein for an elaborate treatment of the decentralized detection problem. More recently, the problem of decentralized detection with correlated observations has also been addressed (see, e.g., [3] and [4]).

In essence, having each sensor select one of D possible messages upper bounds the amount of information available at the fusion center. Indeed, the quantity of information relayed to the fusion center by a network of L sensors, each sending one of D possible messages, does not exceed $L\lceil \log_2 D \rceil$ bits per unit time. In the standard decentralized problem formulation, the number of sensors L and the number of distinct messages D are fixed beforehand. A more natural approach in context of wireless sensor networks is to constrain the capacity of the multiple access channel available to the sensors. For instance, a multiple access channel may only be able to carry R bits of information per unit time. Thus, the new design problem becomes selecting L and D_ℓ, where D_ℓ is the number of messages admissible to sensor S_ℓ, to minimize the probability of error at the fusion center, subject to the capacity constraint

$$\sum_{\ell=1}^{L} \lceil \log_2 (D_\ell) \rceil \leq R. \tag{1}$$

In the remainder of this paper, we consider the detection problem where the sensor network is limited by the capacity of the wireless channel over which the sensors are transmitting. In Section II, we introduce a mathematical framework for the study of decentralized detection in capacity constrained wireless sensor networks. In Section III, we review briefly some useful concepts from statistics, and we establish that, under certain conditions, the design problem admits a very simple solution. That is, we find sufficient conditions for which having R identical binary sensors minimizes the probability of error at the fusion center. We also show that these conditions are fulfilled whenever observations have Gaussian or exponential distributions, although they do not hold for arbitrary distributions. Section IV contains an alternative formulation of the Neyman–Person type for the detection problem at hand. Finally, we give our conclusions in the last section.

II. Statement of the Problem

Let H be a random variable drawn from the binary alphabet $\{H_0, H_1\}$ with prior probabilities π_0 and π_1, respectively. Cou-

Manuscript received February 19, 2002; revised September 24, 2002. This work was supported by the National Science Foundation under the CAREER/PECASE award CCR-0049089. The associate editor coordinating the review of this paper and approving it for publication was Prof. Xiaodong Wang.

The authors are with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois, Urbana, IL 61801 USA (e-mail: chmbrlnd@uiuc.edu; vvv@uiuc.edu).

Digital Object Identifier 10.1109/TSP.2002.806982

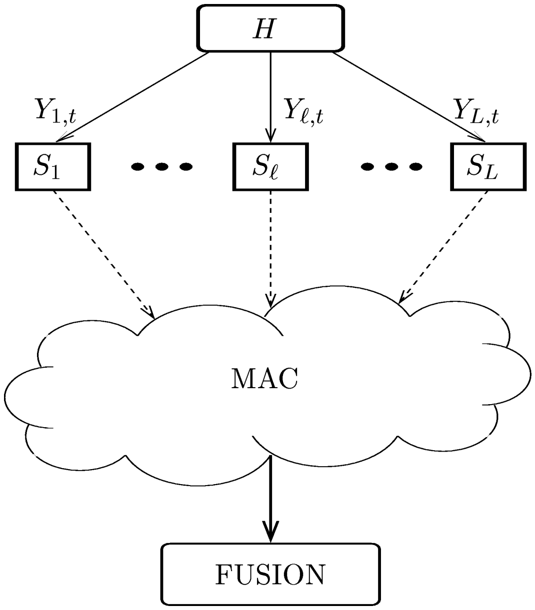


Fig. 1. Block diagram of a wireless sensor network where sensors transmit information to the fusion center over a multiple access channel.

pled with this random variable is a network of wireless sensors, with each sensor receiving a sequence of observations

$$\{Y_{\ell,t}: t = 1, 2, \ldots, T\} \tag{2}$$

about the value of H. We assume that the random variables $Y_{\ell,t}$ are i.i.d., given H, with conditional distribution $p_{Y|H}(\cdot|H_i)$. At discrete time t, the sensors are required to send a summary $U_{\ell,t} = \gamma_\ell(Y_{\ell,t})$ of their own observation to a fusion center. Information is transmitted over a multiple access channel. The information-theoretic capacity of a Gaussian multiple access channel is governed by bandwidth, power, and noise spectral density. More generally, the admissible rate of a practical system with a simple encoding scheme may depend on bandwidth, power, noise density, and maximum tolerable bit error rate at the output of the decoder. Specifying these quantities is equivalent to fixing the admissible rate of the multiple access channel. In this paper, we disregard the specifics of these parameters and assume a joint constraint on the information rates of the sensors. Furthermore, we neglect communication errors in the transmitted bits. In other words, we assume that the network of sensors can transmit reliably at a maximum rate of R bits per unit time. Upon reception of the data, the fusion center produces an estimate $\hat{H}$ of the state of nature H. This setting is illustrated in Fig. 1. Our goal is to design an admissible strategy that minimizes the probability of error $P_e = P\{\hat{H} \neq H\}$.

Definition 1: An admissible strategy, which is denoted by γ, consists of an integer L and a set of decision rules $\gamma_\ell: \mathcal{Y} \rightarrow \{1, 2, \ldots, D_\ell\}$ such that

$$\sum_{\ell=1}^{L} \lceil \log_2 D_\ell \rceil \leq R. \tag{3}$$

At time t, sensor S_ℓ evaluates message $u_{\ell,t} = \gamma_\ell(y_{\ell,t})$, which is subsequently forwarded to the fusion center. We write $\Gamma(R)$ to denote the set of all admissible strategies corresponding to a multiple access channel with capacity R.

Remark 1: We assume that the set $\mathcal{Y}$ is endowed with a σ-field $\mathcal{F}$ and that the decision rules γ_ℓ are measurable functions with respect to $\mathcal{F}$. Furthermore, we assume that the channel capacity R is an integer.

In this work, we are primarily concerned with the asymptotic regime where the observation interval goes to infinity $(T \rightarrow \infty)$. For any reasonable transmission strategy, the associated probability of error at the fusion center goes to zero exponentially fast as T grows unbounded. Thus, a suitable way to compare transmission schemes is through the error exponent measure

$$C(\gamma) = -\lim_{T\rightarrow\infty} \frac{1}{T} \log P_e^{(T)}(\gamma) \tag{4}$$

where $P_e^{(T)}(\gamma)$ is the probability of error at the fusion center associated with strategy γ when a maximum *a posteriori* detector is used. The error exponent is also known as the Chernoff information. For a multiple access channel that is able to carry R bits of information per unit time, we can pose the design problem formally as follows.

Problem 1: Find an admissible strategy $\gamma \in \Gamma(R)$ that maximizes the error exponent

$$C(\gamma) = -\lim_{T\rightarrow\infty} \frac{1}{T} \log P_e^{(T)}(\gamma). \tag{5}$$

III. Basic Concepts and Results

We begin our analysis with a concise review of some basic concepts and properties of Bayesian statistics related to the problem at hand. Consider an arbitrary admissible strategy

$$\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_L). \tag{6}$$

We denote the space of received information corresponding to this strategy by

$$\Upsilon = \{1, 2, \ldots, D_1\} \times \{1, 2, \ldots, D_2\} \times \cdots \times \{1, 2, \ldots, D_L\} \tag{7}$$

so that

$$(\gamma_1(y_1), \gamma_2(y_2), \ldots, \gamma_L(y_L)) \in \Upsilon \tag{8}$$

for all observation vectors $(y_1, y_2, \ldots, y_L) \in \mathcal{Y}^L$. In general, the maximum *a posteriori* detector is known to minimize the probability of estimation error at the fusion center. For a finite observation interval T, it may be impractical (or impossible) to compute this probability of error exactly. We can, however, evaluate the error exponent corresponding to strategy γ by using Chernoff's theorem, which we state without proof (see, e.g., [5]).

Let $p_{\underline{U}|H}(\cdot|H_0)$ and $p_{\underline{U}|H}(\cdot|H_1)$ represent the conditional probability mass functions on Υ, given hypotheses H_0 and H_1, respectively. Mathematically, for $\underline{u} = (u_1, u_2, \ldots, u_L)$ and $i \in \{0, 1\}$, we have

$$p_{\underline{U}|H}(\underline{u}|H_i) = P_i\{\underline{y}: (\gamma_1(y_1), \gamma_2(y_1), \ldots, \gamma_L(y_L)) = \underline{u}\} = \prod_{\ell=1}^{L} P_i\{\gamma_\ell^{-1}(u_\ell)\} \quad (9)$$

where $P_i\{A\}$ represents the probability of event A under hypothesis H_i, and

$$\gamma_\ell^{-1}(u_\ell) = \{y: \gamma_\ell(y) = u_\ell\}.$$

We underline briefly that (9) exploits conditional independence across sensors.

Theorem 1 (Chernoff): Suppose $\gamma \in \Gamma(R)$ is given. The best achievable exponent in the probability of error at the fusion center is given by

$$C(\gamma) = -\min_{0 \le s \le 1} \log \left[\sum_{\underline{u} \in \Upsilon} (p_{\underline{U}|H}(\underline{u}|H_0))^s (p_{\underline{U}|H}(\underline{u}|H_1))^{1-s} \right]. \quad (10)$$

In light of Theorem 1, we can rewrite our original problem as follows.

Problem 2: Find an admissible strategy $\gamma \in \Gamma(R)$ such that the Chernoff information

$$C(\gamma) = -\min_{0 \le s \le 1} \log \left[\sum_{\underline{u} \in \Upsilon} (p_{\underline{U}|H}(\underline{u}|H_0))^s (p_{\underline{U}|H}(\underline{u}|H_1))^{1-s} \right] \quad (11)$$

is maximized.

As stated, Problem 2 is difficult to solve. Even when assignment vector $(D_1, D_2, \ldots, D_L)$ is fixed *a priori*, the problem of finding optimal decision rules $\gamma_1, \gamma_2, \ldots, \gamma_L$ is, in most cases, hard (see, e.g., [2]). In the remainder of this section, we derive a set of conditions under which Problem 2 simplifies greatly. In particular, we find sufficient conditions for which having R sensors, each sending one bit of information, is optimal. First, we establish the following useful result, where we upper bound the contribution of a single sensor to the Chernoff information.

Proposition 1: For strategy γ, the contribution $C_{S_\ell}(\gamma)$ of sensor S_ℓ to the Chernoff information $C(\gamma)$ is bounded above by the Chernoff information C^* contained in one observation Y, i.e.,

$$C_{S_\ell}(\gamma) \le C^* \triangleq -\min_{0 \le s \le 1} \log \left[\int_{\mathcal{Y}} (p_{Y|H}(y|H_0))^s \cdot (p_{Y|H}(y|H_1))^{1-s} \, dy \right]. \quad (12)$$

Proof: Without loss of generality, we consider the contribution of sensor S_1. The Chernoff information for strategy $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_L)$ is given by

$$\begin{aligned} C(\gamma) &= -\min_{0 \le s \le 1} \log \left[\sum_{\underline{u} \in \Upsilon} (p_{\underline{U}|H}(\underline{u}|H_0))^s (p_{\underline{U}|H}(\underline{u}|H_1))^{1-s} \right] \\ &= -\log \left[\prod_{\ell=1}^{L} \left(\sum_{u_\ell=1}^{D_\ell} (P_0\{\gamma_\ell^{-1}(u_\ell)\})^{s^*} (P_1\{\gamma_\ell^{-1}(u_\ell)\})^{1-s^*} \right) \right] \\ &= -\sum_{\ell=1}^{L} \log \left[\sum_{u_\ell=1}^{D_\ell} (P_0\{\gamma_\ell^{-1}(u_\ell)\})^{s^*} (P_1\{\gamma_\ell^{-1}(u_\ell)\})^{1-s^*} \right] \\ &= -\log \left[\sum_{u_1=1}^{D_1} (P_0\{\gamma_1^{-1}(u_1)\})^{s^*} (P_1\{\gamma_1^{-1}(u_1)\})^{1-s^*} \right] \\ &\quad - \sum_{\ell=2}^{L} \log \left[\sum_{u_\ell=1}^{D_\ell} (P_0\{\gamma_\ell^{-1}(u_\ell)\})^{s^*} (P_1\{\gamma_\ell^{-1}(u_\ell)\})^{1-s^*} \right] \end{aligned} \quad (13)$$

where s^* is the value of s that maximizes the Chernoff information $C(\gamma)$. It is then clear that the contribution of sensor S_ℓ to the Chernoff information $C(\gamma)$ is no greater than

$$-\min_{0 \le s \le 1} \log \left[\sum_{u_\ell=1}^{D_\ell} (P_0\{\gamma_\ell^{-1}(u_\ell)\})^{s} (P_1\{\gamma_\ell^{-1}(u_\ell)\})^{1-s} \right] \quad (14)$$

which in turn is upper bounded by the Chernoff information contained in one observation Y. ■

In words, Proposition 1 asserts that the contribution of a single sensor to the total Chernoff information cannot exceed the information contained in each observation. Based on this proposition, we derive a sufficient condition for which having R binary sensors is optimal.

Define $C_1(\gamma_\ell)$ to be the Chernoff information corresponding to a single sensor with decision rule γ_ℓ, i.e.,

$$C_1(\gamma_\ell) = -\min_{0 \le s \le 1} \log \left[\sum_{u=1}^{D_\ell} (P_0\{\gamma_\ell^{-1}(u)\})^{s} (P_1\{\gamma_\ell^{-1}(u)\})^{1-s} \right] \quad (15)$$

and let Γ_{b} be the set of binary functions on the observation space $\mathcal{Y}$.

Proposition 2: Suppose there exists a binary function $\tilde{\gamma}_{\mathrm{b}} \in \Gamma_{\mathrm{b}}$ such that

$$C_1(\tilde{\gamma}_{\mathrm{b}}) \ge \frac{C^*}{2} \quad (16)$$

then having R identical sensors, each sending one bit of information, is optimal.

Proof: Let rate R and strategy $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_L) \in \Gamma(R)$ be given. To prove the claim, we construct an admissible binary strategy $\gamma' \in \Gamma(R)$ such that $C(\gamma') \ge C(\gamma)$. We begin by dividing the collection of decision rules $\{\gamma_1, \gamma_2, \ldots, \gamma_L\}$

into two sets; a first set contains all the binary functions, whereas the other is composed of the remaining decision rules. Formally, we define I_{b} to be the set of integers for which the function γ_ℓ is a binary decision rule

$$I_{\mathrm{b}} = \{\ell\colon 1 \le \ell \le L,\ \gamma_\ell \in \Gamma_{\mathrm{b}}\}. \tag{17}$$

Similarly, we let $I_{\mathrm{nb}} = \{1, 2, \ldots, L\} - I_{\mathrm{b}}$. We choose a binary decision rule $\breve{\gamma}_{\mathrm{b}} \in \Gamma_{\mathrm{b}}$ such that

$$C_1(\breve{\gamma}_{\mathrm{b}}) \ge \max\left\{\max_{\ell \in I_{\mathrm{b}}}\{C_1(\gamma_\ell)\},\ \frac{C^*}{2}\right\}. \tag{18}$$

Such a function $\breve{\gamma}_{\mathrm{b}}$ always exists since by assumption $\tilde{\gamma}_{\mathrm{b}} \in \Gamma_{\mathrm{b}}$ and

$$C_1(\tilde{\gamma}_{\mathrm{b}}) \ge \frac{C^*}{2}. \tag{19}$$

Notice that $\ell \in I_{\mathrm{nb}}$ implies that $D_\ell > 2$, which in turn yields $\lceil \log_2 D_\ell \rceil \ge 2$. Thus, we can replace each sensor with index in I_{nb} by two binary sensors without exceeding the capacity (R bits per unit time) of the multiple access channel. We then consider the alternative scheme γ' in which we replace every sensor with index in I_{b} by a binary sensor with decision rule $\breve{\gamma}_{\mathrm{b}}$ and every sensor with index in I_{nb} by two binary sensors with decision rule $\breve{\gamma}_{\mathrm{b}}$. By construction, γ' is an admissible strategy. Furthermore, this new scheme outperforms the original strategy γ. Indeed

$$\begin{aligned}
&C(\gamma') \\
&= (|I_{\mathrm{b}}| + 2|I_{\mathrm{nb}}|)C_1(\breve{\gamma}_{\mathrm{b}}) \ge |I_{\mathrm{b}}|C_1(\breve{\gamma}_{\mathrm{b}}) + |I_{\mathrm{nb}}|C^* \\
&\ge \sum_{\ell=1}^{L}\left[-\min_{0\le s\le 1}\log\left[\sum_{u_\ell=1}^{D_\ell}(P_0\{\gamma_\ell^{-1}(u_\ell)\})^s(P_1(\gamma_\ell^{-1}(u_\ell)\})^{1-s}\right]\right] \\
&\ge -\min_{0\le s\le 1}\log\left[\sum_{\underline{u}\in\Upsilon}\left(\prod_{\ell=1}^{L}(P_0\{\gamma_\ell^{-1}(u_\ell)\})^s(P_1\{\gamma_\ell^{-1}(u_\ell)\})^{1-s}\right)\right] \\
&= C(\gamma).
\end{aligned} \tag{20}$$

Implicit to this proof is the fact that observations are i.i.d. across sensors, conditioned on H. We also note that for a fixed decision rule $\breve{\gamma}_{\mathrm{b}}$, the Chernoff information at the fusion center is monotonically increasing in the number of sensors. We can therefore improve performance by augmenting the number of sensors in γ' until the rate constraint R is met with equality. The strategy γ being arbitrary, we conclude that having R identical sensors, each sending one bit of information, is optimal. ∎

Our attempt to simplify Problem 2 is in vain if the conditions of Proposition 2 are never satisfied. In the following examples, we show that these requirements are indeed fulfilled for the problem of detecting deterministic signals in Gaussian noise and for the problem of detecting a fluctuating signal in Gaussian noise using a square-law detector. In such cases, having R identical binary sensors is optimal. In proving these assertions, we repetitively use the Bhattacharyya coefficient as a lower bound for the Chernoff information. The Bhattacharyya coefficient $B_1(\gamma_\ell)$ corresponding to a single sensor with decision rule γ_ℓ is given by

$$\begin{aligned}
B_1(\gamma_\ell) &\triangleq -\log\left[\sum_{u=1}^{D_\ell}(P_0\{\gamma^{-1}(u)\})^{1/2}(P_1\{\gamma^{-1}(u)\})^{1/2}\right] \\
&\le -\min_{0\le s\le 1}\log\left[\sum_{u=1}^{D_\ell}(P_0\{\gamma^{-1}(u)\})^{s}(P_1\{\gamma^{-1}(u)\})^{1-s}\right] \\
&= C_1(\gamma).
\end{aligned} \tag{21}$$

The appeal of the Bhattacharyya coefficient obviously lies in its greater simplicity. Examples are next.

A. Gaussian Observations

For a binary signal in Gaussian noise, we consider the conditional distributions given by

$$p_{Y|H}(y|H_0) \sim \mathcal{N}(-m, \sigma^2) \tag{22}$$

$$p_{Y|H}(y|H_1) \sim \mathcal{N}(m, \sigma^2) \tag{23}$$

where $\mathcal{N}(m, \sigma^2)$ denotes a Gaussian distribution with mean m and variance σ^2. We remark that the results of this section can easily be extended to Gaussian observations with arbitrary means. However, because allowing for arbitrary means renders notation complex and provides no further insight, we adopt the symmetric case where $\mathrm{E}[Y|H_1] = -\mathrm{E}[Y|H_0] = m > 0$. We initiate our analysis of the Gaussian case by finding the maximum contribution from each sensor to the total Chernoff information.

Lemma 1: For observations with Gaussian distributions, the contribution of a single sensor to the Chernoff information $C(\gamma)$ is less than or equal to $C^* = m^2/2\sigma^2$.

To demonstrate that having R binary sensors is optimal, we establish the condition of Proposition 2, namely, that there exists a binary decision rule $\tilde{\gamma}_{\mathrm{b}} \in \Gamma_{\mathrm{b}}$ such that $C_1(\tilde{\gamma}_{\mathrm{b}}) \ge C^*/2$. As a potential strategy candidate, we consider the binary threshold function $\tilde{\gamma}_{\mathrm{b}}(y) = \chi_{[0,\infty)}(y)$, where χ_A is the indicator function of set A

$$\chi_A(y) = \begin{cases} 1, & y \in A \\ 0, & y \notin A. \end{cases} \tag{24}$$

We lower bound $C_1(\tilde{\gamma}_{\mathrm{b}})$ by computing the Bhattacharyya coefficient associated with the binary decision rule $\tilde{\gamma}_{\mathrm{b}}$.

Lemma 2: Let $\tilde{\gamma}_{\mathrm{b}}$ be the binary decision rule defined by $\tilde{\gamma}_{\mathrm{b}}(y) = \chi_{[0,\infty)}(y)$. The Bhattacharyya coefficient corresponding to $\tilde{\gamma}_{\mathrm{b}}$ is equal to

$$B_1(\tilde{\gamma}_{\mathrm{b}}) = -\frac{1}{2}\log\left[4\mathcal{Q}\left(-\frac{m}{\sigma}\right)\mathcal{Q}\left(\frac{m}{\sigma}\right)\right] \tag{25}$$

where $\mathcal{Q}$ is the complementary Gaussian cumulative distribution function

$$\mathcal{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}\, e^{-(\xi^2/2)}\, d\xi. \tag{26}$$

Proof: This result is evident from the definition of the Bhattacharyya coefficient

$$
\begin{aligned}
B_1(\tilde{\gamma}_{\mathrm{b}}) &= -\log\left[\sum_{u=1}^{2}(P_0\{\gamma_{\mathrm{b}}^{-1}(u)\})^{1/2}(P_1\{\gamma_{\mathrm{b}}^{-1}(u)\})^{1/2}\right] \\
&= -\log\left[\left(\mathcal{Q}\left(-\frac{m}{\sigma}\right)\right)^{1/2}\left(\mathcal{Q}\left(\frac{m}{\sigma}\right)\right)^{1/2} + \left(\mathcal{Q}\left(\frac{m}{\sigma}\right)\right)^{1/2}\left(\mathcal{Q}\left(-\frac{m}{\sigma}\right)\right)^{1/2}\right] \\
&= -\frac{1}{2}\log\left[4\mathcal{Q}\left(-\frac{m}{\sigma}\right)\mathcal{Q}\left(\frac{m}{\sigma}\right)\right]. \qquad (27)
\end{aligned}
$$

This completes the proof. ■

Lemmas 1 and 2, together with Proposition 2, imply the following theorem.

Theorem 2: For a binary signal in additive Gaussian noise, having R identical sensors, each sending one bit of information, is optimal.

Proof: Again, we let $\tilde{\gamma}_{\mathrm{b}}$ denote the binary decision rule defined by $\tilde{\gamma}_{\mathrm{b}}(y) = \chi_{[0,\infty)}(y)$. By Proposition 2, it is sufficient to show that the inequality $C_1(\tilde{\gamma}_{\mathrm{b}}) \geq C^*/2$ holds true to prove the theorem. We note that

$$
\begin{aligned}
C_1(\tilde{\gamma}_{\mathrm{b}}) - \frac{C^*}{2} &\geq B_1(\tilde{\gamma}_{\mathrm{b}}) - \frac{C^*}{2} \\
&= -\frac{1}{2}\log\left[4\mathcal{Q}\left(-\frac{m}{\sigma}\right)\mathcal{Q}\left(\frac{m}{\sigma}\right)\right] - \frac{m^2}{4\sigma^2} \\
&= -\frac{1}{2}\log\left[4\mathcal{Q}\left(-\frac{m}{\sigma}\right)\mathcal{Q}\left(\frac{m}{\sigma}\right)\exp\left(\frac{m^2}{2\sigma^2}\right)\right]. \qquad (28)
\end{aligned}
$$

Thus, we need only show that the inequality

$$\mathcal{Q}(-x)\mathcal{Q}(x) \leq \tfrac{1}{4}\, e^{-(x^2/2)} \qquad (29)$$

is valid for every $x \in [0, \infty)$. This is accomplished by a simple change of variables in the integral definition of the $\mathcal{Q}$ function

$$
\begin{aligned}
\mathcal{Q}(-x)\mathcal{Q}(x) &= \frac{1}{4} - \int_0^x\int_0^x \frac{1}{2\pi}\, e^{-((\xi^2+\zeta^2)/2)}\, d\xi\, d\zeta \\
&\leq \frac{1}{4} - \int_0^x\int_0^{\pi/2} \frac{1}{2\pi}\, e^{-(\rho^2/2)}\rho\, d\theta\, d\rho \\
&= \tfrac{1}{4}\, e^{-(x^2/2)}. \qquad (30)
\end{aligned}
$$

This establishes the desired result. ■

B. Exponential Observations

The strategy proposed in Theorem 2 possesses a simple structure that is well suited for both implementation and analysis. A natural question to ask, then, is whether having R identical binary sensors is also optimal for non-Gaussian observations. In this section, we provide a partial answer to this question and show that the results of the previous section can be duplicated for signals with exponential distributions. Exponentially distributed observations occur, for instance, when one attempts to detect the presence of a fluctuating complex Gaussian signal in additive white Gaussian noise from the data available at the output of a preprocessing scheme consisting of a match filter and square-law detector. The conditional distributions for exponential observations are given by

$$p_{Y|H}(y|H_0) = \alpha_0 e^{-\alpha_0 y} \qquad (31)$$

$$p_{Y|H}(y|H_1) = \alpha_1 e^{-\alpha_1 y} \qquad (32)$$

where $y \in [0, \infty)$. Without loss of generality, we assume that $0 < \alpha_0 < \alpha_1$. For convenience, we define the ratio $a = \alpha_0/\alpha_1$, with $0 < a < 1$.

Lemma 3: For observations with exponential distributions, the contribution of a single sensor to the Chernoff information $C(\gamma)$ is less than or equal to

$$C^* = -\log\left[-\frac{\log a}{1-a}\exp\left(\frac{\log a}{1-a}+1\right)\right]. \qquad (33)$$

Proof: By Proposition 1, we know that the contribution of a single sensor to $C(\gamma)$ is at most the Chernoff information contained in one observation Y. We obtain (33) by computing the latter quantity explicitly from the optimization problem

$$
\begin{aligned}
C^* &= -\min_{0\leq s\leq 1}\log\left[\int_{\mathcal{Y}}(p_{Y|H}(y|H_0))^s(p_{Y|H}(y|H_1))^{1-s}\,dy\right] \\
&= -\log\left[\min_{0\leq s\leq 1}\int_0^\infty \alpha_0^s e^{-s\alpha_0 y}\alpha_1^{1-s}e^{-(1-s)\alpha_1 y}\,dy\right]. \qquad (34)
\end{aligned}
$$

First, we evaluate the integral part

$$
\begin{aligned}
&\int_{\mathcal{Y}}(p_{Y|H}(y|H_0))^s(p_{Y|H}(y|H_1))^{1-s}\,dy \\
&= \alpha_0^s\alpha_1^{1-s}\int_0^\infty e^{-(s\alpha_0+(1-s)\alpha_1 y)}\,dy = \frac{a^s}{1-s(1-a)}. \qquad (35)
\end{aligned}
$$

We determine the extreme values of equation (35) by differentiating it with respect to s,

$$\frac{d}{ds}\left[\frac{a^s}{1-s(1-a)}\right] = \frac{a^s\log a}{1-s(1-a)} + \frac{a^s(1-a)}{(1-s(1-a))^2}. \qquad (36)$$

This yields a unique critical point at

$$s^* = \frac{1}{1-a} + \frac{1}{\log a} = \frac{1-a+\log a}{(1-a)\log a}. \qquad (37)$$

We observe that $0 \leq s^* \leq 1$, as desired. A second derivative test insures that s^* is a local minimum and a straightforward evaluation of equation (35) at the end points $s \in \{0, 1\}$ confirms that s^* is a global minimum over the set $0 \leq s \leq 1$. As a consequence, the Chernoff information contained in one observation reduces to

$$C^* = -\log\left[-\frac{\log a}{1-a}\exp\left(\frac{\log a}{1-a}+1\right)\right] \qquad (38)$$

and the lemma is verified. ■

We turn to the problem of estimating the amount of information provided to the fusion center by each binary sensor. Again, we seek a decision rule $\tilde{\gamma}_{\mathrm{b}} \in \Gamma_{\mathrm{b}}$ such that $B_1(\tilde{\gamma}_{\mathrm{b}}) \geq C^*/2$. For the exponential case, we select the binary decision rule $\tilde{\gamma}_{\mathrm{b}}(y) = \chi_{[\tau,\infty)}(y)$, where threshold τ is given by $\tau = (\log a)/(\alpha_0 - \alpha_1)$.

Lemma 4: Let $\tilde{\gamma}_{\mathrm{b}}$ be the binary threshold function defined by $\tilde{\gamma}_{\mathrm{b}}(y) = \chi_{[\tau, \infty)}(y)$ with threshold at $\tau = (\log a)/(\alpha_0 - \alpha_1)$. The Bhattacharyya coefficient corresponding to $\tilde{\gamma}_{\mathrm{b}}$ is equal to

$$B_1(\tilde{\gamma}_{\mathrm{b}}) = -\log\left[\left(1 - \exp\left(\frac{a \log a}{1-a}\right)\right)^{1/2} \cdot \left(1 - \exp\left(\frac{\log a}{1-a}\right)\right)^{1/2} + \exp\left(\frac{1+a}{2(1-a)} \log a\right)\right]. \quad (39)$$

Proof: The decision rule $\tilde{\gamma}_{\mathrm{b}}$ yields the following quantities:

$$P_0\{\tilde{\gamma}_{\mathrm{b}}^{-1}([0, \tau))\} = 1 - e^{-\alpha_0 \tau} \quad (40)$$

$$P_1\{\tilde{\gamma}_{\mathrm{b}}^{-1}([0, \tau))\} = 1 - e^{-\alpha_1 \tau} \quad (41)$$

$$P_0\{\tilde{\gamma}_{\mathrm{b}}^{-1}([\tau, \infty))\} = e^{-\alpha_0 \tau} \quad (42)$$

$$P_1\{\tilde{\gamma}_{\mathrm{b}}^{-1}([\tau, \infty))\} = e^{-\alpha_1 \tau} \quad (43)$$

which we use to compute the Bhattacharyya coefficient $B_1(\tilde{\gamma}_{\mathrm{b}})$

$$\begin{aligned} B_1(\tilde{\gamma}_{\mathrm{b}}) &= -\log\left[\sum_{u=1}^{2} (P_0\{\gamma_{\mathrm{b}}^{-1}(u)\})^{1/2} (P_1\{\gamma_{\mathrm{b}}^{-1}(u)\})^{1/2}\right] \\ &= -\log\left[(1 - e^{-\alpha_0 \tau})^{1/2} (1 - e^{-\alpha_1 \tau})^{1/2} + (e^{-\alpha_0 \tau} e^{-\alpha_1 \tau})^{1/2}\right] \\ &= -\log\left[\left(1 - \exp\left(\frac{a \log a}{1-a}\right)\right)^{1/2} \left(1 - \exp\left(\frac{\log a}{1-a}\right)\right)^{1/2} + \exp\left(\frac{1+a}{2(1-a)} \log a\right)\right]. \end{aligned} \quad (44)$$

The last equality is precisely the statement of the lemma. ■

Lemmas 3 and 4 are preliminary steps in establishing the following theorem.

Theorem 3: For a binary hypothesis test with exponential observations, having R identical sensors, each sending one bit of information, is optimal.

Much like in the proof of Theorem 2, we validate Theorem 3 by showing that the inequality $B_1(\tilde{\gamma}_{\mathrm{b}}) \geq C^*/2$ holds for all $a \in (0, 1)$. However, in the case of exponential observations, proving this fact requires a substantial effort. A comparative graph of $B_1(\tilde{\gamma}_{\mathrm{b}})$ and $C^*/2$ as functions of a is presented in Fig. 2. See a formal proof of Theorem 3 in the Appendix.

C. Counterexamples

At this point, we may be lured into believing that having R identical binary sensors is always optimal. Unfortunately, this is not true. In this section, we provide examples for which using R binary sensors is suboptimal. Proposition 2 states sufficient conditions for which having binary sensors is optimal; not fulfilling these conditions does not imply that binary sensors are suboptimal. To obtain counterexamples, we need to develop a set of requirements that insures the suboptimality of binary sensors. This is accomplished in the following proposition.

Proposition 3: Suppose the maximum transmission rate R is an even integer. If there exists a quaternary function $\tilde{\gamma}_{\mathrm{q}}$ such that

$$\frac{C_1(\tilde{\gamma}_{\mathrm{q}})}{2} > \sup_{\gamma_{\mathrm{b}} \in \Gamma_{\mathrm{b}}} \{C_1(\gamma_{\mathrm{b}})\} \quad (45)$$

then having R binary sensors is a suboptimal strategy.

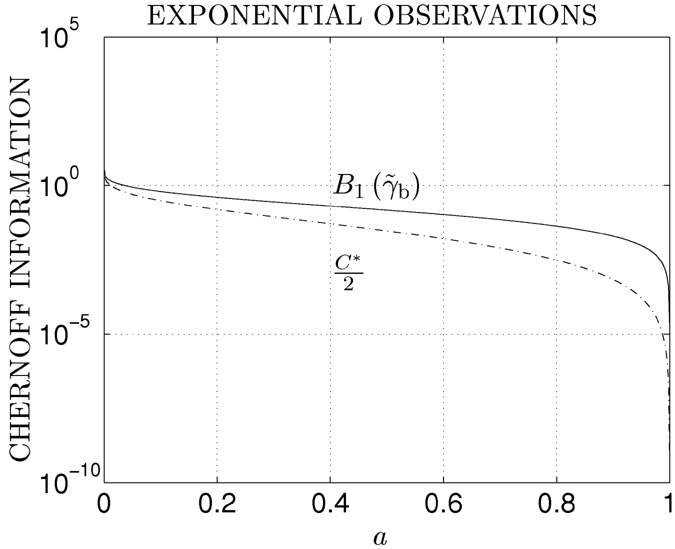


Fig. 2. Bhattacharyya coefficient corresponding to binary decision rule $\tilde{\gamma}_{\mathrm{b}}$ versus half of the Chernoff information contained in one observation.

Proof: Let admissible strategy $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_L) \in \Gamma(R)$ be given, and assume that $\gamma_\ell \in \Gamma_{\mathrm{b}}$ for all ℓ such that $1 \leq \ell \leq L$. We prove the claim by constructing an admissible strategy $\gamma' \in \Gamma(R)$ such that $C(\gamma') > C(\gamma)$. Consider the alternate strategy $\gamma' \in \Gamma(R)$ composed of $R/2$ quaternary sensors with decision rule $\tilde{\gamma}_{\mathrm{q}}$. These two strategies are related as follows:

$$\begin{aligned} C(\gamma) &= -\min_{0 \leq s \leq 1} \log\left[\sum_{\underline{u} \in \Upsilon} \left(\prod_{\ell=1}^{L} (P_0\{\gamma_\ell^{-1}(u_\ell)\})^s \cdot (P_1\{\gamma_\ell^{-1}(u_\ell)\})^{1-s}\right)\right] \\ &\leq \sum_{\ell=1}^{L} \left[-\min_{0 \leq s \leq 1} \log\left[\sum_{u_\ell=1}^{2} (P_0\{\gamma_\ell^{-1}(u_\ell)\})^s \cdot (P_1\{\gamma_\ell^{-1}(u_\ell)\})^{1-s}\right]\right] \\ &\leq R \sup_{\gamma_{\mathrm{b}} \in \Gamma_{\mathrm{b}}} \{C_1(\gamma_{\mathrm{b}})\} < R\frac{C_1(\tilde{\gamma}_{\mathrm{q}})}{2} = C(\gamma'). \end{aligned} \quad (46)$$

That is, strategy γ' outperforms any admissible binary strategy γ. Hence, having R sensors, each sending one bit of information, is suboptimal. ■

To construct counterexamples, we consider the simple scenario where $\mathcal{Y} = \{1, 2, 3, 4\}$. That is, there are only four possible observations at each sensor. The conditional probability mass functions are assigned values

$$p_{Y|H}(y|H_0) = [0.002 \quad p \quad 0.798 - p \quad 0.2] \quad (47)$$

$$p_{Y|H}(y|H_1) = [0.2 \quad 0.798 - p \quad p \quad 0.002] \quad (48)$$

where p is a free parameter $(0.008 \leq p \leq 0.399)$. In this example, only the binary sensors need to quantize their observations. The quaternary sensors simply retransmit the value of each observation to the fusion center. Without loss of optimality, we can assume that the binary decision rules are deterministic threshold rules on the likelihood ratio of the observation space

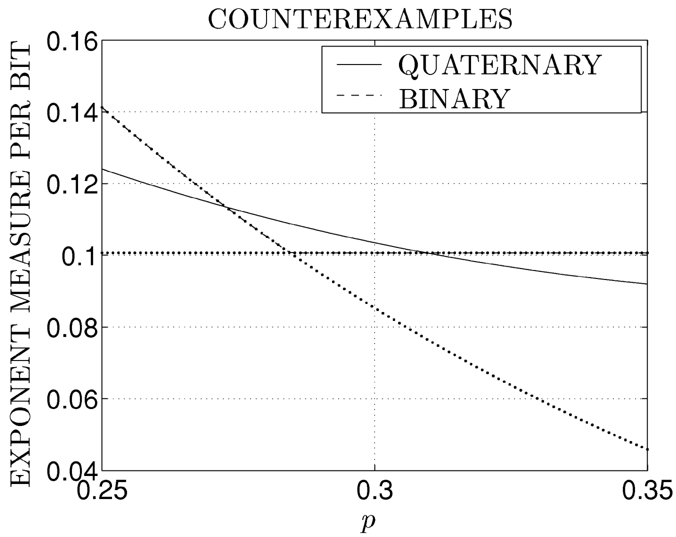


Fig. 3. Comparative plot of the performance per transmission bit of a quaternary sensor versus binary sensors. For parameter values near $p = 0.29$, the quaternary sensor outperforms binary sensors.

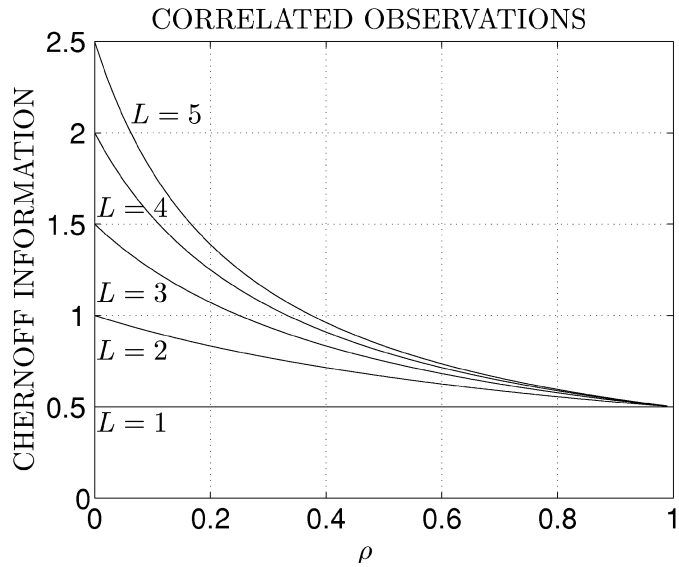


Fig. 4. Chernoff information captured by L sensors as a function of correlation coefficient ρ.

$\mathcal{Y}$. Indeed, in maximizing Chernoff information, there always exists a deterministic likelihood ratio quantizer that is optimal [6]. Because of the symmetry in the problem, we need only consider two binary threshold rules, namely, $\gamma_{\rm b}(y) = \chi_{\{1\}}(y)$ and $\gamma_{\rm b}(y) = \chi_{\{1,\,2\}}(y)$. We write $\gamma_{\rm b}^*$ for the optimal binary decision rule. Fig. 3 provides a comparative plot of the quaternary function $C_1(\gamma_{\rm q})/2$ and the binary function $C_1(\gamma_{\rm b}^*)$ for various values of the free parameter p. Obviously, for values of p near 0.29, the quaternary decision rule performs better than the best binary decision rule. By Proposition 3, this implies that binary strategies are suboptimal for the corresponding conditional distributions (R even).

In general, it is not easy to create good counterexamples. Binary policies seem to be optimal for most probability distributions. Whenever they are not, the performance loss inherent to using the best binary policy appears negligible. This section serves to illustrate the limitations of our results. In practice, the simplicity of binary sensors may prevail over a small hit in performance.

D. Correlated Observations

Throughout, we have assumed that observations are independent. This assumption is reasonable if the limited accuracy of the sensors is responsible for noisy observations. However, if the observed signal is stochastic in nature or if the sensors are subject to external noise, this assumption may fail. In general, decentralized detection with dependent observations is a difficult topic (see, e.g., [3]).

To illustrate how correlation affects our previous results, we study the specific case of a binary signal in equicorrelated Gaussian noise. In this scenario, the conditional distributions $p_{\underline{Y}|H_0}(\cdot|H_i)$ are given by

$$p_{\underline{Y}|H_0}(\underline{y}) \sim \mathcal{N}(-\underline{m},\, \Sigma) \tag{49}$$

$$p_{\underline{Y}|H_1}(\underline{y}) \sim \mathcal{N}(\underline{m},\, \Sigma) \tag{50}$$

where the covariance matrix Σ is of the form

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}. \tag{51}$$

The Chernoff information contained in L observations is equal to

$$C^* = \frac{\underline{m}^T \Sigma^{-1} \underline{m}}{2}. \tag{52}$$

Fig. 4 shows the amount of information contained in L observations for signal energy $\underline{m} = \underline{1} \triangleq (1,\, 1,\, \ldots,\, 1)^T$ and unit noise variance. As the correlation coefficient ρ goes to one, the amount of information contained in L observations approaches the amount of information contained in one observation. Hence, in the limit, having one sensor sending R bits of information is optimal. This suggests that correlation in the observations favors having fewer sensors sending multiple bits, or having nonidentical sensors, rather than employing a set of identical binary sensors.

We complement this remark with a simple example. In Fig. 5, the performance of two identical binary sensors with threshold at zero is plotted against that of a single quaternary sensor. In this example, the signal energy is set to $\underline{m} = \underline{1}$ and the noise variance to unity. For illustrative purposes, the Chernoff information corresponding to a single binary sensor with threshold at zero also appears on the graph. Not surprisingly, the Chernoff information provided by the two binary sensors with threshold at zero decreases to the information supplied by a single such sensor as ρ increases to one. It is interesting to note however that the quaternary sensor outperforms the two binary sensors for correlation coefficient $\rho > 0.65$. Thus, it should not be assumed that having R identical binary sensors is optimal for an arbitrary correlation matrix.

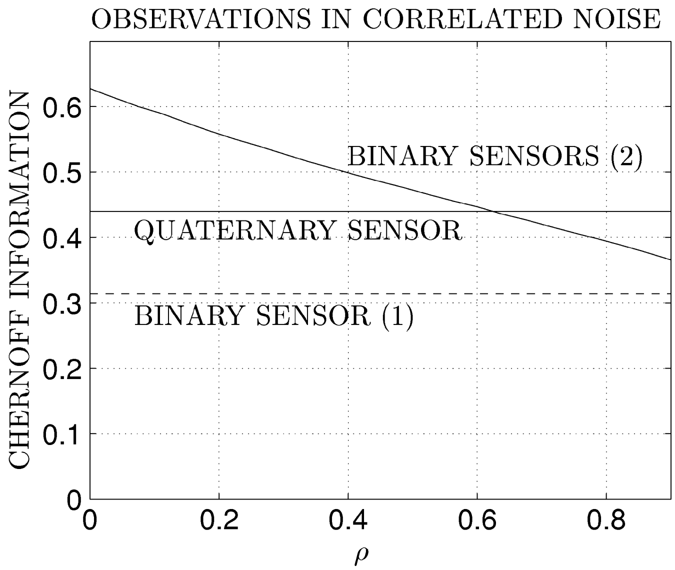


Fig. 5. Comparative plot of the performance of one quaternary sensor versus two binary sensors in correlated Gaussian noise.

IV. Neyman–Pearson Problem

This section presents an alternative formulation of the Neyman–Pearson type for the problem studied in the preceding sections. We consider the optimization problem where one of the probabilities of error is fixed and we wish to minimize the second probability of error. In this case, the best achievable exponent in the probability of error is given by the relative entropy (Kullback–Leibler distance). We state this standard result, known as Stein's lemma, without proof (see, e.g., [5]).

Theorem 4 (Stein's Lemma): Suppose $\gamma \in \Gamma(R)$ is given, and assume that

$$D(p_{\underline{U}|H}(\cdot|H_0) \,\|\, p_{\underline{U}|H}(\cdot|H_1)) < \infty \tag{53}$$

where $D(p \,\|\, q)$ represents the relative entropy of probability distribution q with respect to true distribution p. Let $A_n \subseteq \Upsilon^n$ be an acceptance region for hypothesis H_0, and let the probabilities of error be

$$\alpha_n = P_0^n(A_n^c), \qquad \beta_n = P_1^n(A_n). \tag{54}$$

Furthermore, for $0 < \epsilon < 1/2$, define

$$\beta_n^\epsilon = \min_{A_n \subseteq \Upsilon^n,\, \alpha_n < \epsilon} \beta_n. \tag{55}$$

Then

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \beta_n^\epsilon = -D(p_{\underline{U}|H}(\cdot|H_0) \,\|\, p_{\underline{U}|H}(\cdot|H_1)). \tag{56}$$

The relative entropy between joint distributions of independent random variables being additive, we can upper bound the contribution of a single sensor to

$$D(p_{\underline{U}|H}(\cdot|H_0) \,\|\, p_{\underline{U}|H}(\cdot|H_1)) \tag{57}$$

by the relative entropy corresponding to one observation. We make this statement precise in the following proposition.

Proposition 4: For strategy γ, the contribution $D_{S_\ell}(\gamma)$ of sensor S_ℓ to the relative entropy

$$D(p_{\underline{U}|H}(\cdot|H_0) \,\|\, p_{\underline{U}|H}(\cdot|H_1)) \tag{58}$$

is bounded above by the relative entropy corresponding to a single observation Y,

$$D(p_{Y|H}(\cdot|H_0) \,\|\, p_{Y|H}(\cdot|H_1)). \tag{59}$$

Proof: Without loss of generality, we consider the contribution of sensor S_1. The relative entropy for strategy $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_L)$ is given by

$$\begin{aligned} &D(p_{\underline{U}|H}(\cdot|H_0) \,\|\, p_{\underline{U}|H}(\cdot|H_1)) \\ &\quad= \sum_{\underline{u} \in \Upsilon} p_{\underline{U}|H}(\underline{u}|H_0) \log \left[\frac{p_{\underline{U}|H}(\underline{u}|H_0)}{p_{\underline{U}|H}(\underline{u}|H_1)}\right] \\ &\quad= \sum_{\underline{u} \in \Upsilon} \left(\prod_{\ell=1}^{L} P_0\{\gamma_\ell^{-1}(u_\ell)\} \log \left[\frac{\prod_{\ell=1}^{L} P_0\{\gamma_\ell^{-1}(u_\ell)\}}{\prod_{\ell=1}^{L} P_1\{\gamma_\ell^{-1}(u_\ell)\}} \right] \right) \\ &\quad= \sum_{u_1=1}^{D_1} P_0\{\gamma_1^{-1}(u_1)\} \log \left[\frac{P_0\{\gamma_1^{-1}(u_1)\}}{P_1\{\gamma_1^{-1}(u_1)\}}\right] \\ &\qquad+ \sum_{u' \in \Upsilon'} \left(\prod_{\ell=2}^{L} P_0\{\gamma_\ell^{-1}(u_\ell)\} \log \left[\frac{\prod_{\ell=2}^{L} P_0\{\gamma_\ell^{-1}(u_\ell)\}}{\prod_{\ell=2}^{L} P_1\{\gamma_\ell^{-1}(u_\ell)\}} \right] \right) \end{aligned} \tag{60}$$

where $u' = (u_2, u_3, \ldots, u_L)$, and Υ' is the product space

$$\Upsilon' = \{1, 2, \ldots, D_2\} \times \{1, 2, \ldots, D_3\} \times \cdots \times \{1, 2, \ldots, D_L\}. \tag{61}$$

The contribution of sensor S_1 to the relative entropy $D(p_{\underline{U}|H}(\cdot|H_0) \,\|\, p_{\underline{U}|H}(\cdot|H_1))$ is therefore no greater than

$$\sum_{u_1=1}^{D_1} P_0\{\gamma_1^{-1}(u_1)\} \log \left[\frac{P_0\{\gamma_1^{-1}(u_1)\}}{P_1\{\gamma_1^{-1}(u_1)\}}\right] \tag{62}$$

which in turn is upper bounded by the relative entropy contained in one observation Y. ■

With an upper bound on the contribution of each sensor to the relative entropy of the joint observation distributions, we proceed to establish a result analog to Proposition 2.

Proposition 5: If there exists a binary function $\tilde{\gamma}_{\mathrm{b}} \in \Gamma_{\mathrm{b}}$ such that

$$\sum_{u=1}^{2} P_0\{\tilde{\gamma}_{\mathrm{b}}^{-1}(u)\} \log \left[\frac{P_0\{\tilde{\gamma}_{\mathrm{b}}^{-1}(u)\}}{P_1\{\tilde{\gamma}_{\mathrm{b}}^{-1}(u)\}}\right] \geq \frac{D(p_{Y|H}(\cdot|H_0) \,\|\, p_{Y|H}(\cdot|H_1))}{2} \tag{63}$$

then having R identical sensors, each sending one bit of information, is optimal.

Proof: We only provide a sketch for the proof of Proposition 5 since it parallels closely the proof of Proposition 2. Let rate R and strategy $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_L) \in \Gamma(R)$ be given. We define I_{b} to be the set of integers for which the function γ_ℓ is a binary decision rule. Similarly, we let $I_{\rm nb} = \{1, 2, \ldots, L\} - I_{\rm b}$. We select a binary decision rule $\breve{\gamma}_{\rm b} \in \Gamma_{\rm b}$ such that

$$\sum_{u=1}^{2} P_0\{\breve{\gamma}_{\rm b}^{-1}(u)\} \log\left[\frac{P_0\{\breve{\gamma}_{\rm b}^{-1}(u)\}}{P_1\{\breve{\gamma}_{\rm b}^{-1}(u)\}}\right] \geq \max\left\{\sum_{u=1}^{2} P_0\{\tilde{\gamma}_{\rm b}^{-1}(u)\} \log\left[\frac{P_0\{\tilde{\gamma}_{\rm b}^{-1}(u)\}}{P_1\{\tilde{\gamma}_{\rm b}^{-1}(u)\}}\right], \max_{\ell \in I_{\rm b}}\left\{\sum_{u=1}^{2} P_0\{\gamma_\ell^{-1}(u)\} \log\left[\frac{P_0\{\gamma_\ell^{-1}(u)\}}{P_1\{\gamma_\ell^{-1}(u)\}}\right]\right\}\right\}. \tag{64}$$

We replace every sensor with index in $I_{\rm b}$ by a binary sensor with decision rule $\breve{\gamma}_{\rm b}$ and every sensor with index in $I_{\rm nb}$ by two binary sensors with decision rule $\breve{\gamma}_{\rm b}$. By construction, this admissible scheme outperforms the original strategy γ. Hence, having R identical sensors, each sending one bit of information, is optimal. ■

A. Gaussian Observations

We immediately turn to an example to show that Proposition 5 applies to concrete problems. As before, we consider the problem of detecting deterministic signals in Gaussian noise. For such signals, we employ the conditional distributions $p_{Y|H}(\cdot|H_i)$ given by

$$p_{Y|H}(y|H_0) \sim \mathcal{N}(-m, \sigma^2) \tag{65}$$

$$p_{Y|H}(y|H_1) \sim \mathcal{N}(m, \sigma^2) \tag{66}$$

where m is a positive real number.

Lemma 5: The relative entropy corresponding to Gaussian distributions as defined above is equal to

$$D(p_{Y|H}(\cdot|H_0)\,\|\,p_{Y|H}(\cdot|H_1)) = \frac{2m^2}{\sigma^2}.$$

To show that having R binary sensors is optimal for the Neyman–Pearson problem formulation with Gaussian observations, we establish the condition of Proposition 5. As a strategy candidate, we use the binary threshold function $\tilde{\gamma}_{\rm b}(y) = \chi_{[-m,\infty)}(y)$. We emphasize that this threshold function differs from the one employed in Section III. The dissimilarity arises from the difference between the Bayesian problem formulation and the Neyman–Pearson problem formulation.

Lemma 6: Let $\tilde{\gamma}_{\rm b}$ be the binary decision rule defined by $\tilde{\gamma}_{\rm b}(y) = \chi_{[-m,\infty)}(y)$. The relative entropy corresponding to $\tilde{\gamma}_{\rm b}$ is equal to

$$D(p_{U|H}(\cdot|H_0)\,\|\,p_{U|H}(\cdot|H_1)) = -\frac{1}{2}\log\left[4\mathcal{Q}\left(-\frac{2m}{\sigma}\right)\mathcal{Q}\left(\frac{2m}{\sigma}\right)\right]. \tag{67}$$

Proof: We compute the relative entropy corresponding to $\tilde{\gamma}_{\rm b}$ as follows:

$$\begin{aligned} D(p_{U|H}(\cdot|H_0)\,\|\,p_{U|H}(\cdot|H_1)) &= \sum_{u=1}^{2} P_0\{\tilde{\gamma}_{\rm b}^{-1}(u)\} \log\left[\frac{P_0\{\tilde{\gamma}_{\rm b}^{-1}(u)\}}{P_1\{\tilde{\gamma}_{\rm b}^{-1}(u)\}}\right] \\ &= \frac{1}{2}\log\left[\frac{1}{2\mathcal{Q}\left(\frac{2m}{\sigma}\right)}\right] + \frac{1}{2}\log\left[\frac{1}{2\mathcal{Q}\left(-\frac{2m}{\sigma}\right)}\right] \\ &= -\frac{1}{2}\log\left[4\mathcal{Q}\left(-\frac{2m}{\sigma}\right)\mathcal{Q}\left(\frac{2m}{\sigma}\right)\right] \end{aligned} \tag{68}$$

which possesses the desired form. ■

Lemmas 5 and 6, along with Proposition 5, yield the desired theorem.

Theorem 5: For a binary signal in additive Gaussian noise, having R identical sensors, each sending one bit of information, is optimal.

Proof: Let $\tilde{\gamma}_{\rm b}(y)$ denote the binary decision rule defined by $\tilde{\gamma}_{\rm b}(y) = \chi_{[-m,\infty)}(y)$. By Proposition 5, we need only show that the inequality

$$\sum_{u=1}^{2} P_0\{\tilde{\gamma}_{\rm b}^{-1}(u)\} \log\left[\frac{P_0\{\tilde{\gamma}_{\rm b}^{-1}(u)\}}{P_1\{\tilde{\gamma}_{\rm b}^{-1}(u)\}}\right] \geq \frac{D(p_{Y|H}(\cdot|H_0)\,\|\,p_{Y|H}(\cdot|H_1))}{2} \tag{69}$$

holds true to prove the claim. This is manifest from Lemma 5, Lemma 6, and (29). ■

V. Conclusions and Discussion

We considered a decentralized detection problem in which a network of wireless sensors provides relevant information about the state of nature to a fusion center. We addressed the specific case where the sensor network is constrained by the capacity of the multiple access channel over which the wireless sensors are transmitting and where observations are independent and identically distributed. Our primary focus was on minimizing the Chernoff information, which is equivalent to minimizing the error exponent associated with decisions taken at the fusion center. For Gaussian and exponential observations, having R identical binary sensors was found to be optimal in the asymptotic regime where the observation interval goes to infinity ($T \to \infty$). In other words, the gain offered by having more sensors outperforms the benefits of getting detailed information from each sensor whenever the number of observations per sensor is large.

We demonstrated, through counterexamples, that this property cannot be generalized to arbitrary observation distributions. In particular, there exist distributions for which the performance of quaternary sensors exceeds that of binary sensors. Moreover, having identical binary sensors may not be optimal when observations are dependent across sensors. Indeed, dependence across sensors may favor having fewer sensors sending multiple bits, or having nonidentical sensors over employing a set of identical binary sensors.

Finally, we showed that a similar analysis can be performed for an alternative problem formulation of the Neyman–Pearson type, where one of the probabilities of error is fixed, and the second probability of error is minimized. Again, in this case, we found that having R identical binary sensors is optimal for independent Gaussian observations.

The optimality of wireless sensor networks with identical binary sensors is encouraging. Such networks are easily implementable, amenable to analysis, and provide robustness to the system through redundancy. Avenues of further research include a more in depth analysis of the detection problem with dependent observations as well as extending the problem formulation to composite hypothesis testing.

Appendix
Proof of Theorem 3

This section is devoted to the proof of Theorem 3. As mentioned in Section III-B, it is sufficient to establish that there exists a binary policy $\tilde{\gamma}_{\mathrm{b}} \in \Gamma_{\mathrm{b}}$ such that $B_1(\tilde{\gamma}_{\mathrm{b}}) \geq C^*/2$ to prove the theorem. We consider the binary decision rule $\tilde{\gamma}_{\mathrm{b}}(y) = \chi_{[\tau, \infty)}(y)$ with threshold at $\tau = (\log a)/(\alpha_0 - \alpha_1)$. For convenience, we introduce the concise notation

$$g(a) \triangleq -\frac{\log a}{1-a}. \tag{70}$$

Recalling the results of Lemmas 3 and 4, we have

$$\begin{aligned} B_1(\tilde{\gamma}_{\mathrm{b}}) - \frac{C^*}{2} &= -\log\left[\left(1-e^{-ag(a)}\right)^{1/2}\left(1-e^{-g(a)}\right)^{1/2} + e^{-((1+a)g(a)/2)}\right] \\ &\quad + \log\left[(g(a))^{1/2}e^{-(g(a)/2)}e^{1/2}\right] \\ &= -\log\left[\left(1-e^{-ag(a)}\right)^{1/2}\left(e^{g(a)}-1\right)^{1/2}(g(a))^{-(1/2)}e^{-(1/2)} \right. \\ &\qquad \left. + e^{-(ag(a)/2)}(g(a))^{-(1/2)}e^{-(1/2)}\right]. \end{aligned} \tag{71}$$

We define the function $f(a)$ by

$$\begin{aligned} f(a) \triangleq \left(1-e^{-ag(a)}\right)^{1/2}\left(e^{g(a)}-1\right)^{1/2}(g(a))^{-(1/2)} \\ + e^{-(ag(a)/2)}(g(a))^{-(1/2)} \end{aligned} \tag{72}$$

and proceed to show that $f(a) \leq e^{1/2}$ for all $a \in (0, 1)$, which is equivalent to proving that $B_1(\tilde{\gamma}_{\mathrm{b}}) \geq C^*/2$. First, we state a few straightforward preliminary results.

Lemma 7: For element $x \in (0, 1)$

$$\frac{1}{x} \leq \coth x. \tag{73}$$

Lemma 8: For element $a \in (0, 1)$

$$\frac{1-e^{-ag(a)}}{1-e^{-g(a)}} \leq -\frac{a \log a}{1-a}. \tag{74}$$

Based on Lemmas 7 and 8, we can show that $(d/da)f(a) \geq 0$ for every $a \in (0, 1)$ by differentiating $f(a)$ explicitly and by making appropriate substitutions. Thus, the function $f(a)$ is monotonically increasing on the interval $(0, 1)$. Taking its limit as a approaches one, we obtain

$$\begin{aligned} \lim_{a \uparrow 1} f(a) = \lim_{a \uparrow 1}\left[\left(1-e^{-ag(a)}\right)^{1/2}\left(e^{g(a)}-1\right)^{1/2}(g(a))^{-(1/2)} \right. \\ \left. + \left(e^{-ag(a)}\right)^{1/2}(g(a))^{-(1/2)}\right] = e^{1/2}. \end{aligned} \tag{75}$$

That is, the inequality $f(a) \leq e^{1/2}$ holds for every $a \in (0, 1)$, as desired. This completes the proof of Theorem 3.

References

[1] R. R. Tenney and N. R. Sandell, Jr., "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, pp. 501–510, 1981.

[2] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statist. Signal Process.*, vol. 2, pp. 297–344, 1993.

[3] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad and ugly: Distributed detection of a known signal in dependent Gaussian noise," *IEEE Trans. Signal Processing*, vol. 48, pp. 3266–3279, Dec. 2000.

[4] Y. Zhu, R. S. Blum, Z.-Q. Luo, and K. M. Wong, "Unexpected properties and optimum-distributed sensor detectors for dependent observation cases," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 62–72, Jan. 2000.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[6] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, pp. 550–558, April 1993.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

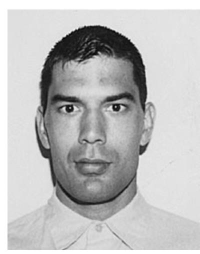
[8] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer, 1994.

[9] R. Durrett, *Probability: Theory and Examples*, 2nd ed. New York: Duxbury, 1996.

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

[11] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests, Part I: Asymptotic optimality," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2448–2462, Nov. 1999.

[12] ——, "Multihypothesis sequential probability ratio tests, Part II: Accurate asymptotic expansions for the expected sample size," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1366–1383, July 2000.



Jean-François Chamberland (S'98) received the B.Eng. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 1998 and the M.S. degree from the School of Electrical Engineering, Cornell University, Ithaca, NY, in 2000. He is currently pursuing the Ph.D. degree at the University of Illinois at Urbana-Champaign.

His research interests include communication systems and detection and estimation theory.

Venugopal V. Veeravalli (S'86–M'92–SM'98) received the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the B.Tech. degree in 1985 from the Indian Institute of Technology, Bombay, India (Silver Medal Honors), all in electrical engineering.

He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently an Associate Professor with the Department of Electrical and Computer Engineering and a Research Associate Professor in the Coordinated Science Laboratory. During 1996 to 2000, he was an Assistant Professor at Cornell University, Ithaca, NY. His research interests include mobile and wireless communications, detection and estimation theory, and information theory.

Dr. Veeravalli is currently an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY. Among the awards he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award in 1996, the National Science Foundation CAREER Award in 1998, the Presidential Early Career Award for Scientists and Engineers in 1999, and the Michael Tien Excellence in Teaching Award from the College of Engineering, Cornell University, in 1999.