Technical Research Report

S Y S T E M S
R E S E A R C H
C E N T E R

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

Distributed Detection with Feedback

by H.M.H. Shalaby and A. Papamarcou

TR 90-64

DISTRIBUTED DETECTION WITH FEEDBACK

Hossam M. H. Shalaby and Adrian Papamarcou

Electrical Engineering Department and Systems Research Center

University of Maryland, College Park, MD 20742

ABSTRACT

We investigate the effects of feedback on a decentralized detection system consisting of N sensors and a data fusion center. It is assumed that observations are independent and identically distributed across sensors, and that each sensor uses a randomized scheme for compressing its observations into a fixed number of quantization levels. We consider two variations on this setup. One entails the transmission of sensor data to the fusion center in two stages, and the broadcast of feedback information from the center to the sensors after the first stage. The other variation involves information exchange between sensors prior to transmission to the fusion center; this exchange is effected through a feedback decision center, which processes binary data from the sensors and thereafter broadcasts a single feedback bit back to the sensors. We show that under the Neyman-Pearson criterion, only the latter type of feedback yields an improvement on the asymptotic performance of the system (as $N \to \infty$), and we derive the associated error exponents. We also demonstrate that deterministic compression schemes are asymptotically as powerful as randomized ones.

DISTRIBUTED DETECTION WITH FEEDBACK

Hossam M. H. Shalaby and Adrian Papamarcou

Electrical Engineering Department and Systems Research Center

University of Maryland, College Park, MD 20742

1. Introduction.

We consider a binary hypothesis testing problem (H_0 versus H_1). The system in general consists of N identical sensors $\{S_i\}_{i=1}^N$. Each sensor observes a random variable X_i and transmits a compressed version of its information to a fusion center. The fusion center then decides which hypothesis is true based on the classical Neyman-Pearson criterion.

Tenney and Sandell [1] were the first to introduce this kind of problem. They considered two sensors (N=2), and designed, for a fixed fusion decision rule, the optimal local decision rules of the sensors on the basis of the minimum cost function criterion. Much related work has since appeared in literature [2-9]. Of special interest to our work here is [7], where a similar problem with multiple hypotheses was considered. The sensors and the fusion center used deterministic rules, and the optimization was based on minimizing the overall probability of error. The error exponent of this probability was evaluated as $N \to \infty$.

A modified version of the above system is as follows. Each sensor may transmit an encoded version of its observations to a *feedback decision center* (which may be the fusion center itself). This feedback decision center then broadcasts a compressed version of the received information to all the sensors (and perhaps also to the fusion center). Each sensor uses this feedback information to update its decision and transmit a new local decision to the fusion center, which will then decide on the true hypothesis. This system is useful in situations where the channel between the sensors and the fusion center is of restricted capacity and the transmission entails considerable cost.

2. Problem Definition and Preliminaries.

We have two hypotheses (H_0, H_1) and N sensors $\{S_i\}_{i=1}^N$. Each sensor S_i observes a random variable X_i that takes values from a finite set $\mathcal{X}$. We assume that $X_1, \cdots, X_N$ are identically distributed and independent under both hypotheses. The distributions of X_i under H_0 and H_1 are P_X and Q_X, respectively. We will consider four different systems for processing the data collected by the sensors. Before describing these systems we will introduce some notation.

General Notation. P_{X^N} and Q_{X^N} denote the distribution of the vector X^N under the null hypothesis and the alternative hypothesis, respectively. If the components $\{X_i\}$ of X^N are independent and P_{X_i} (resp. Q_{X_i}) denotes the distribution of X_i under H_0 (resp. H_1), then for all $x^N \in \mathcal{X}^N$

$$P_{X^N}(x^N) = \prod_{i=1}^{N} P_{X_i}(x_i), \qquad Q_{X^N}(x^N) = \prod_{i=1}^{N} Q_{X_i}(x_i) \ .$$

If, moreover, $P_{X_i} = P_X$ for all i, $1 \leq i \leq N$, then we may use P_X^N instead of P_{X^N}; similarly for Q. Also, as usual, $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$.

Problem (P1). Let $U_i \in \mathcal{U}$ be the decision of sensor S_i, where, $\mathcal{U}$ is any set of cardinality at most $|\mathcal{X}|$. To generate the random variable U_i, each sensor uses a behavioral rule, [10], which can be represented by a conditional distribution $\Delta_{U_i|X_i}$ on $\mathcal{U} \times \mathcal{X}$. No feedback is involved in this system, hence, as soon as the fusion center C collects the local decisions of each sensor $\{U_i\}$, it declares that hypothesis H_0 is true if U^N lies in some acceptance region $\mathcal{A}_N \subset \mathcal{U}^N$. We would like to choose the decision rules $\{\Delta_{U_i|X_i}\}$ and the acceptance region so as to minimize the type II error, $Q_{U^N}(\mathcal{A}_N)$, subject to the constraint

$$P_{U^N}(\mathcal{A}_N^c) \leq \epsilon$$

on type I error, where $\epsilon \in (0,1)$. For all $u^N \in \mathcal{U}^N$ P_{U^N} is defined as

$$P_{U^N}(u^N) = \prod_{i=1}^{N} P_{U_i}(u_i) \ ,$$

where

$$P_{U_i}(u_i) = \sum_x \Delta_{U_i|X_i}(u_i|x) P_X(x)$$

or simply $P_{U_i} = \Delta_{U_i|X_i} P_X$. $Q_{U^N}(u^N)$ is defined in a similar manner. We denote this minimum probability of type II error by $\beta_N^{(1)}(|\mathcal{U}|, \epsilon)$, i.e.,

$$\beta_N^{(1)}(|\mathcal{U}|, \epsilon) \stackrel{\text{def}}{=} \inf_{\{\Delta_{U_i|X_i}\}, \mathcal{A}_N \subset \mathcal{U}^N} \{Q_{U^N}(\mathcal{A}_N) : P_{U^N}(\mathcal{A}_N) \geq 1 - \epsilon\} \ .$$

We are interested in the asymptotic behavior of $\beta_N^{(1)}(|\mathcal{U}|, \epsilon)$ as $N \to \infty$. The resulting error exponent is given by

$$\theta^{(1)}(|\mathcal{U}|, \epsilon) \stackrel{\text{def}}{=} -\lim_N \frac{1}{N} \log \beta_N^{(1)}(|\mathcal{U}|, \epsilon) \ ,$$

provided the limit exists.

Problem (P2). Here feedback is involved and we assume that $\mathcal{U}$ is a binary set (each sensor transmits just one bit to C before receiving the feedback information). Once C receives U^N, it generates a feedback bit V. This bit is "0" if U^N lies in some region $\mathcal{C}_N \subset \mathcal{U}^N$ called the *feedback acceptance region*, and "1" otherwise. The center broadcasts the bit V to all sensors. Each sensor S_i generates two binary random variables Y_i, Z_i (that take values in the same binary set $\mathcal{W}$) according to distributions $\Delta_{Y_i|U_iX_i}, \Delta_{Z_i|U_iX_i}$ and uses the feedback bit to decide which of Y_i, Z_i to transmit to the fusion center as the second information bit W_i:

$$W_i \stackrel{\text{def}}{=} Y_i I_{[V=0]} + Z_i I_{[V=1]} \ , \tag{2.1}$$

where $I_{[\cdot]}$ denotes the indicator function. Since V depends on the observations of all sensors, the W_i's are in general dependent. The center C collects these decisions $\{W_i\}$ and uses those along with U^N to declare that H_0 is true if $U^N W^N$ lies in an acceptance region $\mathcal{A}_N$

which is a subset of $\mathcal{U}^N \times \mathcal{W}^N$. Our aim is to choose the decision rules $\{\Delta_{U_i Y_i|X_i}, \Delta_{U_i Z_i|X_i}\}$, the feedback acceptance region $\mathcal{C}_N$, and the acceptance region $\mathcal{A}_N$ so as to minimize the type II error, $Q_{U^N W^N}(\mathcal{A}_N)$, subject to the constraint

$$P_{U^N W^N}(\mathcal{A}_N^c) \leq \epsilon$$

on the type I error. Here, by (2.1), for all $(u^N, w^N) \in \mathcal{U}^N \times \mathcal{W}^N$, we can write

$$P_{U^N W^N}(u^N, w^N) = \begin{cases} P_{U^N Y^N}(u^N, w^N), & \text{if } u^N \in \mathcal{C}_N; \\ P_{U^N Z^N}(u^N, w^N), & \text{otherwise,} \end{cases} \tag{2.2}$$

where

$$P_{U^N Y^N}(u^N, w^N) = \prod_{i=1}^{N} P_{U_i Y_i}(u_i, w_i), \quad P_{U_i Y_i}(u_i, w_i) = \sum_x \Delta_{U_i Y_i|X_i}(u_i, w_i|x) P_X(x) ,$$

and

$$P_{U^N Z^N}(u^N, w^N) = \prod_{i=1}^{N} P_{U_i Z_i}(u_i, w_i), \quad P_{U_i Z_i}(u_i, w_i) = \sum_x \Delta_{U_i Z_i|X_i}(u_i, w_i|x) P_X(x) .$$

Hence, by (2.2),

$$P_{U^N W^N}(\mathcal{A}_N) = P_{U^N Y^N}(\mathcal{A}_N \cap (\mathcal{C}_N \times \mathcal{W}^N)) + P_{U^N Z^N}(\mathcal{A}_N \cap (\mathcal{C}_N^c \times \mathcal{W}^N)) .$$

$Q_{U^N W^N}$ is evaluated in a similar way. The optimal type II error is defined as follows

$$\beta_N^{(2)}(\epsilon) \stackrel{\text{def}}{=} \inf_{\substack{\{\Delta_{U_i Y_i|X_i}, \Delta_{U_i Z_i|X_i}\}, \\ \mathcal{C}_N \subset \mathcal{U}^N, \mathcal{A}_N \subset \mathcal{U}^N \times \mathcal{W}^N}} \{Q_{U^N W^N}(\mathcal{A}_N) \colon P_{U^N W^N}(\mathcal{A}_N) \geq 1 - \epsilon\} .$$

The corresponding error exponent is given by

$$\theta^{(2)}(\epsilon) \stackrel{\text{def}}{=} -\lim_N \frac{1}{N} \log \beta_N^{(2)}(\epsilon) .$$

Problem (P3). This system differs from (P2) in that here we have two different centers: the feedback decision center C_1 and the fusion center C_2. The sensors transmit their first decisions $\{U_i\}$ to C_1, which broadcasts a feedback decision bit V (generated in exactly the same manner as in (P2)) to all sensors and to C_2. The sensors then transmit their second decisions $\{W_i\}$ (generated as in (P2)) to C_2. C_2 uses W^N along with V to declare the final decision. Therefore, the acceptance region $\mathcal{A}_N \subset \mathcal{U}^N \times \mathcal{W}^N$ can be written as the disjoint union

$$\mathcal{A}_N = (\mathcal{C}_N \times \mathcal{F}_N) \cup (\mathcal{C}_N^c \times \mathcal{E}_N) \ ,$$

where $\mathcal{C}_N \subset \mathcal{U}^N$ and $\mathcal{F}_N, \mathcal{E}_N \subset \mathcal{W}^N$. With the aid of (2.2), we can write

$$P_{U^N W^N}(\mathcal{A}_N) = P_{U^N Y^N}(\mathcal{C}_N \times \mathcal{F}_N) + P_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{E}_N) \ .$$

The corresponding optimal type II error is thus

$$\beta_N^{(3)}(\epsilon) \stackrel{\text{def}}{=} \inf_{\substack{\{\Delta_{U_i Y_i | X_i}, \Delta_{U_i Z_i | X_i}\}, \\ \mathcal{C}_N \subset \mathcal{U}^N, \mathcal{F}_N, \mathcal{E}_N \subset \mathcal{W}^N}} \{Q_{U^N W^N}(\mathcal{A}_N) \colon \mathcal{A}_N = (\mathcal{C}_N \times \mathcal{F}_N) \cup (\mathcal{C}_N^c \times \mathcal{E}_N),\ P_{U^N W^N}(\mathcal{A}_N) \geq 1-\epsilon\}$$

and the error exponent is

$$\theta^{(3)}(\epsilon) \stackrel{\text{def}}{=} -\lim_N \frac{1}{N} \log \beta_N^{(3)}(\epsilon) \ .$$

Problem (P4). This system differs from P3 in that the two centers C_1, C_2 do not communicate. Thus, C_2 uses W^N only to determine the true hypothesis. The statment of the problem is summarized as follows. Let the acceptance region set by C_2 be $\mathcal{A}_N$ (it should be emphasized that this acceptance region is a subset of $\mathcal{W}^N$, not $\mathcal{U}^N$ as in (P1)). The type I and type II errors are given by

$$P_{W^N}(\mathcal{A}_N) = P_{U^N Y^N}(\mathcal{C}_N \times \mathcal{A}_N) + P_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{A}_N)$$

and

$$Q_{W^N}(\mathcal{A}_N) = Q_{U^N Y^N}(\mathcal{C}_N \times \mathcal{A}_N) + Q_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{A}_N) \ ,$$

respectively. The optimal type II error is thus

$$\beta_N^{(4)}(\epsilon) \stackrel{\text{def}}{=} \inf_{\substack{\{\Delta_{U_i Y_i | X_i}, \Delta_{U_i Z_i | X_i}\}, \\ \mathcal{C}_N \subset \mathcal{U}^N, \mathcal{A}_N \subset \mathcal{W}^N}} \{Q_{W^N}(\mathcal{A}_N) \colon P_{W^N}(\mathcal{A}_N) \geq 1 - \epsilon\}$$

and the error exponent is given by

$$\theta^{(4)}(\epsilon) \stackrel{\text{def}}{=} -\lim_N \frac{1}{N} \log \beta_N^{(4)}(\epsilon) .$$

In this report we are interested in comparing the error exponent in (P1) (when $|\mathcal{U}| = 4$) with that in (P2), the error exponent in (P3) with that in (P4), and the error exponent in (P1) (when $|\mathcal{U}| = 2$) with that in (P4). This comparison will measure the effect of a feedback bit in the performance of the system.

In Section 3 we will evaluate the error exponents of the above systems under the assumption that all sensors use the same behavioral decision rule. In Section 4 we will discuss and evaluate the error exponents in problems (P1), (P2) assuming that the sensors are not restricted to use the same rule. It turns out that system (P2) performs similarly to (P1)($|\mathcal{U}| = 4$). This means that the feedback bit does not lead to an improvement in the error exponent in (P2), but it is certainly useful in reducing the complexity of the system. On the other hand (P4) outperforms (P1)($|\mathcal{U}| = 2$) in general, hence this type of information exchange between sensors is useful. We will see as well that (P3) does not outperform (P4), i.e., C_2 does not need to know the feedback bit transmitted by C_1.

We will show also that the sensors can employ only deterministic decision rules without loss of asymptotic optimality.

Typical sequences. Our proofs rely on the concept of a typical sequence, as developed in [11]. Here we cite some basic definitions and facts on typical sequences.

The *type* of a sequence $x^N \in \mathcal{X}^N$ is the distribution λ_x on $\mathcal{X}$ defined by the relationship

$$(\forall a \in \mathcal{X}) \qquad \lambda_x(a) \stackrel{\text{def}}{=} \frac{1}{N} n(a|x^N),$$

where $n(a|x^N)$ is the number of terms in x^N equal to a. The set of all types of sequences in $\mathcal{X}^N$, namely $\{\lambda_x : x^N \in \mathcal{X}^N\}$, will be denoted by $\mathcal{P}_N(\mathcal{X})$.

Given a type $\hat{P}_X \in \mathcal{P}_N(\mathcal{X})$, we will denote by $\hat{T}_X^N$ the set of sequences $x^N \in \mathcal{X}^N$ of type $\hat{P}_X$:

$$\hat{T}_X^N \stackrel{\text{def}}{=} \{x^N \in \mathcal{X}^N : \lambda_x = \hat{P}_X\} .$$

Also, for an arbitrary distribution $\tilde{P}_X$ on $\mathcal{X}$ and a constant $\eta > 0$, we will denote by $\tilde{T}_{X,\eta}^N$ the set of $(\tilde{P}_X, \eta)$-*typical* sequences in $\mathcal{X}^N$. A sequence x^N is $(\tilde{P}_X, \eta)$-typical if $|\lambda_x(a) - \tilde{P}_X(a)| \leq \eta$ for every letter $a \in \mathcal{X}$ and, in addition, $\lambda_x(a) = 0$ for every a such that $\tilde{P}_X(a) = 0$. Thus, if $||\cdot||$ denotes the sup norm and $\ll$ denotes absolute continuity, we have

$$\tilde{T}_{X,\eta}^n \stackrel{\text{def}}{=} \{x^N \in \mathcal{X}^N\text{: } ||\lambda_x - \tilde{P}_X|| \leq \eta,\ \lambda_x \ll \tilde{P}_X\} .$$

In the same manner, we will denote by $T_{X,\eta}^N$ the set of (P_X, η)-typical sequences in $\mathcal{X}^N$.

The proofs of Lemmas 2.1, 2.2 appear in [11].

LEMMA 2.1. *For any $\hat{P}_X$ in $\mathcal{P}_N(\mathcal{X})$, any distribution Q_X on $\mathcal{X}$*

$$(N+1)^{-|\mathcal{X}|} \exp[NH(\hat{P}_X)] \ \leq\ |\hat{T}_X^N| \ \leq\ \exp[NH(\hat{P}_X)] ,$$

and

$$(N+1)^{-|\mathcal{X}|} \ \exp[-ND(\hat{P}_X||Q_X)] \ \leq\ Q_X^N(\hat{T}_X^N) \ \leq\ \exp[-ND(\hat{P}_X||Q_X)] ,$$

where $H(\cdot)$, $D(\cdot||\cdot)$ denote the informational entropy and divergence, respectively.

LEMMA 2.2. *For any distributions P_X, Q_X on $\mathcal{X}$, and $\eta > 0$,*

$$P_X^N(T_{X,\eta}^N) \geq 1 - \frac{|\mathcal{X}|}{4N\eta^2} ,$$

$$Q_X^N(T_{X,\eta}^N) \leq \exp[-N(D(P_X||Q_X) - \delta_N - \nu(\eta))] ,$$

where $\delta_N = \frac{|\mathcal{X}|\log(N+1)}{N} \to 0$, and $\nu(\eta) \to 0$ as $\eta \to 0$.

One can easily modify the above exposition to accommodate pairs $(x^N, y^N) \in \mathcal{X}^N \times \mathcal{Y}^N$ by reverting to their representation in $(\mathcal{X} \times \mathcal{Y})^N$. Thus the type of (x^N, y^N) is the distribution λ_{xy} on $\mathcal{X} \times \mathcal{Y}$ such that

$$\lambda_{xy}(a,b) = \frac{1}{N}\Big|\{i : (x_i, y_i) = (a,b)\}\Big| ,$$

and the sets $\mathcal{P}_N(\mathcal{X} \times \mathcal{Y})$, as well as $\hat{T}^N_{XY} \subset \mathcal{X}^N \times \mathcal{Y}^N$ and $\tilde{T}^N_{XY,\eta} \subset \mathcal{X}^N \times \mathcal{Y}^N$, are defined accordingly.

We will need the following lemma.

LEMMA 2.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be any binary sets. Fix $\rho > 0$, $\delta \in (0,1)$. Then there exists a sequence*

$$\nu_N = \nu_N(\rho, \delta, |\mathcal{X}|, |\mathcal{Y}|) \to 0$$

such that for every P_{XY}, $Q_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, $C \in \mathcal{X}^N$, $F \in \mathcal{Y}^N$ that satisfy

$$\min_{x,y: Q_{XY}(x,y) > 0} Q_{XY}(x,y) > \rho ,$$

$$D(P_{XY} || Q_{XY}) < \infty ,$$

and

$$P^N_{XY}(C \times F) \geq \delta ,$$

the following is true:

$$Q^N_{XY}(C \times F) \geq \exp[-N(d(P_X, P_Y || Q_{XY}) + \nu_N)] ,$$

where

$$d(P_X, P_Y || Q_{XY}) \stackrel{\text{def}}{=} \min_{\substack{\tilde{P}_{XY} \\ \tilde{P}_X = P_X,\ \tilde{P}_Y = P_Y}} D(\tilde{P}_{XY} || Q_{XY}) .$$

PROOF. *Case 1.* $Q_{XY} > 0$. The statement is implicitly proven in [12], Section 3, and is also true for arbitrary $\mathcal{X}, \mathcal{Y}$.

Case 2. Q_{XY} has zeros. The binary assumption on the sets $\mathcal{X}$ and $\mathcal{Y}$ is critical here. Let the distribution $\tilde{P}_{XY}$ achieve the minimum in $d(P_X, P_Y||Q_{XY})$, then $\tilde{P}_{XY} \ll Q_{XY}$, otherwise the divergence is infinity. The constraints $\tilde{P}_X = P_X$, $\tilde{P}_Y = P_Y$ force $\tilde{P}_{XY}$ to be identical to P_{XY}, i.e., here $d(P_X, P_Y||Q_{XY}) = D(P_{XY}||Q_{XY})$. Stein's Lemma [13] ensures the existence of a sequence $\lambda_N \to 0$ such that

$$Q_{XY}^N(C \times F) \ \geq \ \exp[-N(D(P_{XY}||Q_{XY}) + \lambda_N)] \ . \qquad \triangle$$

In the following sections we will omit the superscript N from T, as N will be essentially constant.

3. The Main Results.

THEOREM 3.1. *The error exponent for (P1), assuming all sensors use the same decision rule and $|\mathcal{U}| \leq |\mathcal{X}|$, is given by*

$$\theta^{(1)}(|\mathcal{U}|, \epsilon) = \sup_{\substack{\Delta_{U|X}: \\ P_U = P_X \Delta_{U|X},\ Q_U = Q_X \Delta_{U|X}}} D(P_U||Q_U)$$

for all $\epsilon \in (0,1)$.

PROOF. *Direct part.* If $|\mathcal{U}| = |\mathcal{X}|$, then the error exponent is given by Stein's Lemma [13]. Let $\mathcal{U}$ be a set of cardinality not greater than $|\mathcal{X}|$. We assume all sensors use the same decision rule, i.e., $\Delta_{U_i|X_i} = \Delta_{U|X}$, for all $1 \leq i \leq N$. We thus have for all $u \in \mathcal{U}$,

$$P_{U_i}(u) = P_U(u) = \sum_x \Delta_{U|X}(u|x) P_X(x), \qquad Q_{U_i}(u) = Q_U(u) = \sum_x \Delta_{U|X}(u|x) Q_X(x) \ .$$

Set the acceptance region

$$\mathcal{A}_N = T_{U,\eta} \ ,$$

where $\eta > 0$ is arbitrary small. Then from Lemma 2.2,

$$P_{U^N}(\mathcal{A}_N) = P_U^N(\mathcal{A}_N) \geq 1 - \frac{|\mathcal{U}|}{4N\eta^2} \ ,$$

which is greater than $1-\epsilon$ if N is large enough. The type II error is upper bounded by $\exp[-N(D(P_U||Q_U)-\delta_N-\nu(\eta))]$, where $\delta_N\to 0$ and $\nu(\eta)\to 0$ as $\eta\to 0$ (cf. Lemma 2.2). Since the conditional distribution $\Delta_{U|X}$ is arbirary, we have

$$\beta_N^{(1)}(|\mathcal{U}|,\epsilon)\le \inf_{\Delta_{U|X}} \exp[-N(D(P_U||Q_U)-\delta_N-\nu(\eta))] \ .$$

By definition of the error exponent,

$$\theta^{(1)}(|\mathcal{U}|,\epsilon)\ge \sup_{\Delta_{U|X}} D(P_U||Q_U)-\nu(\eta) \ .$$

Since $\eta>0$ is arbitrary, the proof of the direct part is complete.

Converse part. Assume that all sensors use the same decision rule $\Delta_{U|X}$. Let $\mathcal{A}_N\subset \mathcal{U}^N$ be any acceptance region satisfying the constraint

$$P_{U^N}(\mathcal{A}_N)\ge 1-\epsilon \ .$$

Hence, from Stein's Lemma, for N large enough there exists a sequence $\lambda_N\to 0$, depending only on $|\mathcal{U}|$ and ϵ, such that

$$\begin{aligned} Q_{U^N}(\mathcal{A}_N) &\ge \exp[-N(D(P_U||Q_U)+\lambda_N)] \\ &\ge \exp[-N(\sup_{\Delta_{U|X}} D(P_U||Q_U)+\lambda_N)] \ , \end{aligned}$$

Since $\Delta_{U|X}$ was arbitrary, and $\mathcal{A}_N$ was any set satisfying the constraint on type I error, we obtain

$$\beta_N^{(1)}(|\mathcal{U}|,\epsilon)\ge \exp[-N(\sup_{\Delta_{U|X}} D(P_U||Q_U)+\lambda_N)] \ .$$

Hence

$$\theta^{(1)}(|\mathcal{U}|,\epsilon)\le \sup_{\Delta_{U|X}} D(P_U||Q_U) \ . \qquad \triangle$$

THEOREM 3.2. *The error exponent for (P2), assuming all sensors use the same pair of decision rules, is given by*

$$\theta^{(2)}(\epsilon)=\sup_{\substack{\Delta_{UY|X}:\\ P_{UY}=P_X\Delta_{UY|X},\ Q_{UY}=Q_X\Delta_{UY|X}}} D(P_{UY}||Q_{UY})$$

for all $\epsilon \in (0,1)$.

REMARK. One can see that if $|\mathcal{U}| = 4$ in Theorem 3.1, then the error exponents in (P1) and (P2) are exactly the same, so the feedback bit used by the fusion center actually does not convey essential information to the sensors. As a matter of fact, one can show that the same error exponent prevails for any *fixed* number of feedback bits.

SKETCH OF THE PROOF. *Direct part.* It is obvious that this system will do at least as well as the one with no feedback, hence $\theta^{(2)}(\epsilon) \geq \theta^{(1)}(\epsilon)$.

Converse part. Assume that all sensors use a pair of decision rules $\Delta_{UY|X}, \Delta_{UZ|X}$ Let $\mathcal{C}_N \subset \mathcal{U}^N$, $\mathcal{A}_N \subset \mathcal{U}^N \times \mathcal{W}^N$ be any regions satisfying the constraint

$$P_{U^N W^N}(\mathcal{A}_N) \;=\; P_{U^N Y^N}(\mathcal{A}_N \cap (\mathcal{C}_N \times \mathcal{W}^N)) + P_{U^N Z^N}(\mathcal{A}_N \cap (\mathcal{C}_N^c \times \mathcal{W}^N)) \;\geq 1-\epsilon\,.$$

Hence, either $P_{U^N Y^N}(\mathcal{A}_N \cap (\mathcal{C}_N \times \mathcal{W}^N)) \geq \frac{1-\epsilon}{2}$ or $P_{U^N Z^N}(\mathcal{A}_N \cap (\mathcal{C}_N^c \times \mathcal{W}^N)) \geq \frac{1-\epsilon}{2}$. Using the same method as in the proof of the converse part of Theorem 3.1, we get that

$$\begin{aligned}\theta^{(2)}(\epsilon) &\leq \{\sup_{\Delta_{UY|X}} D(P_{UY}||Q_{UY})\} \vee \{\sup_{\Delta_{UZ|X}} D(P_{UZ}||Q_{UZ})\} \\ &= \sup_{\Delta_{UY|X}} D(P_{UY}||Q_{UY})\,.\end{aligned}$$

$\triangle$

THEOREM 3.3. *The error exponents for (P3) and (P4), assuming all sensors use the same pair of decision rules and* $D(P_X||Q_X) < \infty$, *are given by*

$$\theta^{(3)}(\epsilon) = \theta^{(4)}(\epsilon) = \sup_{\substack{\Delta_{UY|X}: \\ P_{UY}=P_X\Delta_{UY|X},\; Q_{UY}=Q_X\Delta_{UY|X}}} d(P_U, P_Y||Q_{UY})$$

for all $\epsilon \in (0,1)$.

REMARK. In general, the error exponent in (P3) and (P4) is better than that in (P1) when $|\mathcal{U}| = 2$, because we always have $d(P_U, P_Y||Q_{UY}) \geq D(P_U||Q_U)$.

SKETCH OF THE PROOF. *Direct part.* By the problem statement, we have $\theta^{(3)}(\epsilon) \geq \theta^{(4)}(\epsilon)$. Hence it is enough to show the direct part for (P4) only. Pick an arbitrary pair of conditional distributions $\Delta_{UY|X}$, $\Delta_{UZ|X}$ as fixed decision rules to all sensors. Set $\mathcal{C}_N = T_{U,\eta}$, $\mathcal{A}_N = T_{Y,\eta}$, where $\eta > 0$ arbitrary. For the type I error in (P4), we can write

$$\begin{aligned}
P_{W^N}(\mathcal{A}_N^c) &= P_{U^N Y^N}(\mathcal{C}_N \times \mathcal{A}_N^c) + P_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{A}_N^c) \\
&\leq P_{Y^N}(\mathcal{A}_N^c) + P_{U^N}(\mathcal{C}_N^c) \\
&\leq \frac{|\mathcal{Y}|}{4N\eta^2} + \frac{|\mathcal{U}|}{4N\eta^2} \leq \epsilon
\end{aligned}$$

if N is large enough. The Type II error in (P4) is upper bounded as follows

$$\begin{aligned}
Q_{W^N}(\mathcal{A}_N) &= Q_{U^N Y^N}(\mathcal{C}_N \times \mathcal{A}_N) + Q_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{A}_N) \\
&\leq Q_{U^N Y^N}(T_{U,\eta} \times T_{Y,\eta}) + Q_{Z^N}(T_{Y,\eta}) \\
&\leq \exp[-N(d(P_U, P_Y || Q_{UY}) - \delta_N - \nu(\eta))] + Q_{Z^N}(T_{Y,\eta}) \ .
\end{aligned}$$

By a suitable choice of $\Delta_{Z|X}$ the last term in the above inequality can be equal to zero. Indeed, for any fixed y with $P_Y(y) > 0$ we can always choose a trivial decision rule $\Delta_{Z|X}$ such that for the above fixed y, $\Delta_{Z|X}(y|x) = 0$ for all $x \in \mathcal{X}$. It follows that $Q_Z(y) = 0$. Since $P_Y(y) > 0$, there exists an $1 \geq i(y^N) \geq N$ such that $y_{i(y^N)} = y$ for each sequence y^N in $T_{Y,\eta}$. Thus $Q_{Z^N}(T_{Y,\eta}) = 0$. This, together with the fact that the decision rules were arbitrary, yields

$$\theta^{(4)}(\epsilon) \geq \sup_{\Delta_{UY|X}} d(P_U, P_Y || Q_{UY}) - \nu(\eta) \ .$$

Converse part. It is enough to show the converse part for (P3) only because $\theta^{(3)}(\epsilon) \geq \theta^{(4)}(\epsilon)$. Assume that all sensors use a pair of decision rules $\Delta_{UY|X}, \Delta_{UZ|X}$. Note that $D(P_{UY}||Q_{UY}) < \infty$ and $D(P_{UZ}||Q_{UZ}) < \infty$ since $D(P_X||Q_X) < \infty$. Let $\mathcal{C}_N \subset \mathcal{U}^N$, $\mathcal{F}_N, \mathcal{E}_N \subset \mathcal{W}^N$, $\mathcal{A}_N = \mathcal{C}_N \times \mathcal{F}_N \cup \mathcal{C}_N^c \times \mathcal{E}_N$ be satisfying the type I error constraint

$$P_{U^N W^N}(\mathcal{A}_N) = P_{U^N Y^N}(\mathcal{C}_N \times \mathcal{F}_N) + P_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{E}_N) \geq 1 - \epsilon \ .$$

Hence, either $P_{U^N Y^N}(\mathcal{C}_N \times \mathcal{F}_N) \geq \frac{1-\epsilon}{2}$ or $P_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{E}_N) \geq \frac{1-\epsilon}{2}$. Using Lemma 2.3, we have either

$$Q_{U^N Y^N}(\mathcal{C}_N \times \mathcal{F}_N) \geq \exp[-N(d(P_U, P_Y || Q_{UY}) + \nu_N)]$$

$$\geq \exp[-N(\sup_{\Delta_{UY|X}} d(P_U, P_Y||Q_{UY}) + \nu_N)]$$

or

$$\begin{aligned} Q_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{E}_N) &\geq \exp[-N(d(P_U, P_Z||Q_{UZ}) + \nu_N)] \\ &\geq \exp[-N(\sup_{\Delta_{UY|X}} d(P_U, P_Y||Q_{UY}) + \nu_N)] \ . \end{aligned}$$

This yields

$$\begin{aligned} Q_{U^N W^N}(\mathcal{A}_N) &= Q_{U^N Y^N}(\mathcal{C}_N \times \mathcal{F}_N) + Q_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{E}_N) \\ &\geq Q_{U^N Y^N}(\mathcal{C}_N \times \mathcal{F}_N) \vee Q_{U^N Z^N}(\mathcal{C}_N^c \times \mathcal{E}_N) \\ &\geq \exp[-N(\sup_{\Delta_{UY|X}} d(P_U, P_Y||Q_{UY}) + \lambda_N)] \ , \end{aligned}$$

and therefore also

$$\theta^{(3)}(\epsilon) \leq \sup_{P_{UY|X}} d(P_U, P_Y||Q_{UY}) \ . \qquad \triangle$$

In what follows we will see that it is sufficient for all sensors to employ deterministic decision rules in order to achieve the above error exponents. We need to define $\Phi(\Delta_{U|X})$ and $\Psi(\Delta_{UY|X})$ as follows.

$$\Phi(\Delta_{U|X}) \stackrel{\text{def}}{=} D(P_U||Q_U) \ , \tag{3.1}$$

where

$$P_U(\cdot) = \sum_x \Delta_{U|X}(\cdot|x) P_X(x), \quad Q_U(\cdot) = \sum_x \Delta_{U|X}(\cdot|x) Q_X(x) \tag{3.2}$$

and

$$\Psi(\Delta_{UY|X}) \stackrel{\text{def}}{=} d(P_U, P_Y||Q_{UY}) \ , \tag{3.3}$$

where

$$P_U(\cdot) = \sum_{x,y} \Delta_{UY|X}(\cdot, y|x) P_X(x), \quad P_Y(\cdot) = \sum_{u,x} \Delta_{UY|X}(u, \cdot|x) P_X(x) \ , \tag{3.4}$$

$$Q_{UY}(\cdot, \cdot) = \sum_x \Delta_{UY|X}(\cdot, \cdot|x) Q_X(x) \ . \tag{3.5}$$

The following lemma asserts the convexity of $\Phi(\cdot)$ and $\Psi(\cdot)$.

LEMMA 3.1. *$\Phi(\Delta_{U|X})$ defined in (3.1) is a convex function in $\Delta_{U|X}$, and $\Psi(\Delta_{UY|X})$ defined in (3.3) is a convex function in $\Delta_{UY|X}$.*

PROOF. For $\alpha \in (0,1)$ and any two conditional distributions $\Delta_{U|X}, \tilde{\Delta}_{U|X}$, let P_U, Q_U be defined as in (3.2), and $\tilde{P}_U, \tilde{Q}_U$ be defined similarly. Then

$$\begin{aligned}
\Phi(\alpha\Delta_{U|X} + (1-\alpha)\tilde{\Delta}_{U|X}) &= D(\alpha P_U + (1-\alpha)\tilde{P}_U || \alpha Q_U + (1-\alpha)\tilde{Q}_U) \\
&\leq \alpha D(P_U||Q_U) + (1-\alpha)D(\tilde{P}_U||\tilde{Q}_U) \\
&= \alpha\Phi(\Delta_{U|X}) + (1-\alpha)\Phi(\tilde{\Delta}_{U|X}) ,
\end{aligned}$$

where we have made use of the convexity of the divergence. This proves that $\Phi(\Delta_{U|X})$ is a convex function in $\Delta_{U|X}$.

Now for any $\alpha \in (0,1)$ and any two conditional distributions $\Delta_{UY|X}, \tilde{\Delta}_{UY|X}$, let P_U, P_Y, Q_{UY} be defined as in (3.4), (3.5), and similarly for $\tilde{P}_U, \tilde{P}_Y, \tilde{Q}_{UY}$. Then

$$\begin{aligned}
&\alpha\Psi(\Delta_{UY|X}) + (1-\alpha)\Psi(\tilde{\Delta}_{UY|X}) \\
&\qquad = \alpha d(P_U, P_Y||Q_{UY}) + (1-\alpha)d(\tilde{P}_U, \tilde{P}_Y||\tilde{Q}_{UY}) \\
&\qquad = \alpha D(P_{UY}^{(1)}||Q_{UY}) + (1-\alpha)D(P_{UY}^{(2)}||\tilde{Q}_{UY}) ,
\end{aligned}$$

for some $P_{UY}^{(1)}$ having marginals P_U, P_Y, and $P_{UY}^{(2)}$ having marginals $\tilde{P}_U$, $\tilde{P}_Y$. By convexity of the divergence and definition of $d(\cdot,\cdot||\cdot)$, we obtain

$$\begin{aligned}
&\alpha\Psi(\Delta_{UY|X}) + (1-\alpha)\Psi(\tilde{\Delta}_{UY|X}) \\
&\qquad \geq D(\alpha P_{UY}^{(1)} + (1-\alpha)P_{UY}^{(2)}||\alpha Q_{UY} + (1-\alpha)\tilde{Q}_{UY}) \\
&\qquad \geq d(\alpha P_U + (1-\alpha)\tilde{P}_U, \alpha P_Y + (1-\alpha)\tilde{P}_Y||\alpha Q_{UY} + (1-\alpha)\tilde{Q}_{UY}) \\
&\qquad = \Psi(\alpha\Delta_{UY|X} + (1-\alpha)\tilde{\Delta}_{UY|X}) .
\end{aligned}$$

This proves that $\Psi(\Delta_{UY|X})$ is a convex function in $\Delta_{UY|X}$. $\triangle$

In what follows we assume that Π and Λ are partitions of $\mathcal{X}$. We denote by $\Pi \vee \Lambda$ the coarsest common refinement of Π and Λ. We use $P|_\Pi$ to denote the restriction of P_X on Π.

THEOREM 3.4. *Assume all sensors use the same decision rule.*

(i) For all $\epsilon \in (0,1)$, $|\mathcal{U}| \leq |\mathcal{X}|$, if Π is a partition of $\mathcal{X}$, then

$$\theta^{(1)}(|\mathcal{U}|, \epsilon) = \max_{\substack{\Pi: \\ P_U = P|_\Pi,\ Q_U = Q|_\Pi}} D(P_U || Q_U) .$$

(ii) If $\mathcal{U}$, $\mathcal{Y}$ are any binary sets and Π, Λ are partitions of $\mathcal{X}$, then for all $\epsilon \in (0,1)$

$$\theta^{(2)}(\epsilon) = \max_{\substack{\Pi,\Lambda: \\ P_{UY} = P|_{\Pi\vee\Lambda},\ Q_{UY} = Q|_{\Pi\vee\Lambda}}} D(P_{UY} || Q_{UY}) ,$$

(iii) If, in addition to (ii), $D(P_X || Q_X) < \infty$, then for all $\epsilon \in (0,1)$

$$\theta^{(3)}(\epsilon) = \theta^{(4)}(\epsilon) = \max_{\substack{\Pi,\Lambda: \\ P_{UY} = P|_{\Pi\vee\Lambda},\ Q_{UY} = Q|_{\Pi\vee\Lambda}}} d(P_U, P_Y || Q_{UY}) .$$

PROOF. From Theorem 3.1 and the definition of $\Phi(\cdot)$ we have

$$\theta^{(1)}(|\mathcal{U}|, \epsilon) = \sup_{\Delta_{U|X}} \Phi(\Delta_{U|X}) . \tag{3.6}$$

Observe, however, that any distribution $\Delta_{U|X}$ can be written as a convex combination of at most $|\mathcal{U}|^{|\mathcal{X}|}$ extremal distributions $\{\Delta_{U_i|X_i}\}$, which are such that $\Delta_{U_i|X_i}(u|x) = 0$ or 1, i.e., if $M = |\mathcal{U}|^{|\mathcal{X}|}$, then we can write

$$\Delta_{U|X} = \sum_{i=1}^{M} \alpha_i \Delta_{U_i|X_i} ,$$

where $\alpha \geq 0$, $\sum_{i=1}^{M} \alpha_i = 1$. Substituting in (3.6) and making use of the convexity of $\Phi(\cdot)$, we obtain

$$\begin{aligned} \theta^{(1)}(|\mathcal{U}|, \epsilon) &= \sup_{\Delta_{U|X}} \Phi(\Delta_{U|X}) = \sup_{\{\alpha_i\}_{i=1}^{M}} \Phi(\sum_{i=1}^{M} \alpha_i \Delta_{U_i|X_i}) \\ &\leq \sup_{\{\alpha_i\}_{i=1}^{M}} \sum_{i=1}^{M} \alpha_i \Phi(\Delta_{U_i|X_i}) \leq \max_{1\leq i\leq M} \Phi(\Delta_{U_i|X_i}) . \end{aligned} \tag{3.7}$$

The reverse inequality is obviously true. This proves the first statment of the theorem. The remaining statments can be proven in a similar way. $\triangle$

4. Extensions and Concluding Remarks.

In the previous section we considered the situation in which all sensors use the same behavioral decision rule, and showed that no loss of optimality resulted from using deterministic rules. In this section we consider a more general situation, in which the sensors are allowed to use different decision rules.

We show here that in this case the error exponents of (P1), (P2) will still be given by the corresponding expressions in Theorem 3.4, and thus the sensors can use the same deterministic rule without loss of optimality. We have the following theorem.

THEOREM 4.1. *The error exponents for (P1) and (P2), assuming* $D(P_X||Q_X) < \infty$ *but no further constraint on the local decision rules, are given by the corresponding expressions in Theorem 3.4.*

PROOF. *Direct part.* In (P1) let $\mathcal{U}$ be a set of cardinality not greater than $|\mathcal{X}|$. Fix any conditional distribution $\Delta_{U|X}$ and let all sensors use the same decision rule, i.e., $\Delta_{U_i|X_i} = \Delta_{U|X}$ for all $1 \leq i \leq N$. As in Theorem 3.1, the error exponent in (P1) is lower bounded by

$$\theta^{(1)}(|\mathcal{U}|, \epsilon) \geq \sup_{\Delta_{U|X}} D(P_U||Q_U) \geq \max_{\substack{\Pi: \\ P_U = P|_\Pi,\ Q_U = Q|_\Pi}} D(P_U||Q_U) .$$

Converse part. In (P1) assume that each sensor S_i uses an arbitrary behavioral decision rule $\Delta_{U_i|X_i}$. For all $u_i \in \mathcal{U}$, $i \in \{1, \cdots, N\}$, define

$$P_{U_i}(u_i) = \sum_x \Delta_{U_i|X_i}(u_i|x) P_X(x), \qquad Q_{U_i}(u_i) = \sum_x \Delta_{U_i|X_i}(u_i|x) Q_X(x) .$$

It is obvious that the P_{U_i}'s are independent and so are the Q_{U_i}'s. Let $\mathcal{A}_N \subset \mathcal{U}^N$ be any acceptance region satisfying the constraint

$$P_{U^N}(\mathcal{A}_N) \geq 1 - \epsilon .$$

Define, for $\eta > 0$ arbitrary, the set

$$T_\eta^N \stackrel{\text{def}}{=} \{u^N \in \mathcal{U}^N : \big| \log \frac{P_{U^N}(u^N)}{Q_{U^N}(u^N)} - \sum_{i=1}^N D(P_{U_i} \| Q_{U_i}) \big| \leq N\eta\} \; .$$

If $E_P(\cdot)$, $\mathrm{Var}_P(\cdot)$ denote expectation and variance under P, then

$$\begin{aligned} E_P \log \frac{P_{U^N}(U^N)}{Q_{U^N}(U^N)} &= \sum_{i=1}^N E_P \log \frac{P_{U_i}(U_i)}{Q_{U_i}(U_i)} \\ &= \sum_{i=1}^N D(P_{U_i} \| Q_{U_i}) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}_P \log \frac{P_{U^N}(U^N)}{Q_{U^N}(U^N)} &= \sum_{i=1}^N \mathrm{Var}_P \log \frac{P_{U_i}(U_i)}{Q_{U_i}(U_i)} \\ &\leq N \sup_{\Delta_{U|X}} \mathrm{Var}_P \log \frac{P_U(U)}{Q_U(U)} = N\sigma^2 \; , \end{aligned}$$

where $\sigma^2 = \sup_{\Delta_{U|X}} \mathrm{Var}_P \log \frac{P_U(U)}{Q_U(U)} < \infty$ (cf. Appendix A). We have from Chebychev's inequality that

$$\begin{aligned} P_{U^N}(T_\eta^c) &= P_{U^N}\{\big| \log \frac{P_{U^N}(U^N)}{Q_{U^N}(U^N)} - \sum_{i=1}^N D(P_{U_i} \| Q_{U_i}) \big| > N\eta\} \\ &\leq \frac{\mathrm{Var}_P \log \frac{P_{U^N}(U^N)}{Q_{U^N}(U^N)}}{N^2\eta^2} \leq \frac{\sigma^2}{N\eta^2} \; . \end{aligned}$$

This yields

$$P_{U^N}(\mathcal{A}_N \cap T_\eta) \geq \frac{1-\epsilon}{2}$$

for N large enough. We can estimate the type II error rate as follows:

$$\begin{aligned} Q_{U^N}(\mathcal{A}_N) \;\geq\; Q_{U^N}(\mathcal{A}_N \cap T_\eta) &= \sum_{u^N \in \mathcal{A}_N \cap T_\eta^N} Q_{U^N}(u^N) \\ &\geq \sum_{u^N \in \mathcal{A}_N \cap T_\eta^N} P_{U^N}(u^N) \exp\big[- \sum_{i=1}^N D(P_{U_i} \| Q_{U_i}) - N\eta \big] \\ &\geq \frac{1-\epsilon}{2} \exp\big[- \sum_{i=1}^N D(P_{U_i} \| Q_{U_i}) - N\eta \big] \; . \end{aligned}$$

It follows that

$$\theta^{(1)}(|\mathcal{U}|,\epsilon) \;\leq\; \frac{1}{N}\sum_{i=1}^{N} D(P_{U_i}||Q_{U_i}) + \eta \leq \sup_{\Delta_{U|X}} D(P_U||Q_U) + \eta \;.$$

Combining the direct and converse parts and applying Theorem 3.4 completes the proof of the first statement of the theorem. The proof of the other is omitted. △

APPENDIX A

We will show that $\sigma^2 < \infty$ if $P_X \ll Q_X$. Let

$$f(u) \stackrel{\text{def}}{=} \sum_{u\in\mathcal{U}} P_U(u)\log^2 \frac{P_U(u)}{Q_U(u)} = \sum_{u:P_U(u)>0} P_U(u)\log^2 \frac{P_U(u)}{Q_U(u)} , \tag{A.1}$$

where

$$P_U(\cdot) = \sum_x \Delta_{U|X}(\cdot|x)P_X(x) = \sum_{x:Q_X(x)>0} \Delta_{U|X}(\cdot|x)P_X(x) ,$$

$$Q_U(\cdot) = \sum_{x:Q_X(x)>0} \Delta_{U|X}(\cdot|x)Q_X(x) .$$

We have, for all $u \in \mathcal{U}$ with $P_U(u) > 0$,

$$P_U(u) \le \frac{P_U(u)}{Q_U(u)} \le \frac{\sum_{x:Q_X(x)>0} \Delta_{U|X}(u|x)}{\rho\sum_{x:Q_X(x)>0} \Delta_{U|X}(u|x)} = \frac{1}{\rho} ,$$

where $\rho \stackrel{\text{def}}{=} \min_{x:Q_X(x)>0} Q_X(x)$. Consequently,

$$\log P_U(u) \le \log \frac{P_U(u)}{Q_U(u)} \le \log \frac{1}{\rho}$$

and hence,

$$\log^2 \frac{P_U(u)}{Q_U(u)} \le \log^2 \frac{1}{\rho} \vee \log^2 P_U(u) .$$

Substituting in (A.1), we obtain

$$f(u) \;\le\; \sum_{u:P_U(u)>0} \left(P_U(u)\log^2 \frac{1}{\rho}\right) \vee \left(P_U(u)\log^2 P_U(u)\right) .$$

Using the fact that $0 \le t\log^2 t \le \log^2 e^{2/e}$ for all $0 \le t \le 1$, it follows that

$$\begin{aligned} f(u) \;&\le\; \sum_{u:P_U(u)>0} \log^2 \frac{1}{\rho} \vee \log^2 e^{2/e} \\ &\le\; |\mathcal{U}|\log^2 \left(\frac{1}{\rho} \vee e^{2/e}\right) . \end{aligned}$$

Hence,

$$\begin{aligned} \sigma^2 \;&=\; \sup_{\Delta_{U|X}} \left\{E_P \log^2 \frac{P_U(u)}{Q_U(u)} - \left(E_P \log \frac{P_U(u)}{Q_U(u)}\right)^2\right\} \\ &\le\; \sup_{\Delta_{U|X}} f(u) \;\le\; |\mathcal{U}|\log^2 \left(\frac{1}{\rho} \vee e^{2/e}\right) . \end{aligned}$$

$\triangle$

REFERENCES

[1] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerospace and Electronic Sys.*, vol. AES-17, pp. 501-510, July 1981.

[2] J. J. Chao and C. C. Lee, "A distributed detection scheme based on soft local decisions," *the 24th Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois,* Oct. 1-3, 1986.

[3] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerospace and Electronic Sys.*, vol. AES-22, pp. 98-101, jan. 1986.

[4] I. Y. Hoballah and P. K. Varshney, "Neyman-Pearson detection with distributed sensors," *In Proceedings of the 25th IEEE Conference on Decision and Control,* Athens, Greece, Dec. 1986, pp. 237-241.

[5] S. C. Thomopoulos, R. Viswanathan, and D. C. Bougoulias, "Optimal decision fusion in multiple sensor systems," *IEEE Trans. Aerospace and Electronic Sys.*, vol. AES-23, pp. 644-653, Sep. 1987.

[6] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Automt. Contr.*, vol. AC-30, pp. 440-446, May 1985.

[7] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals, and Systems*, vol. 1,2, pp. 167-182, 1988.

[8] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerospace and Electronic Sys.*, vol. AES-23, pp. 24-30, Jan. 1987.

[9] G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. Aerospace and Electronic Sys.*, vol. AES-26,

pp. 282-292, Mar. 1990.

[10] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach.* New York: Academic Press, 1967.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1982 and Budapest, Hungary: Akadémiai Kiadó, 1981.

[12] H. M. H. Shalaby and A. Papamarcou, "Error exponents for a class of multiterminal detection systems," Technical Report TR 89-63, Systems Research Center, Univ. of Maryland, College Park, 1989.

[13] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations," *Ann. Math. Stat.*, vol. 23, pp. 493-507, 1952.