

Performance Analysis of Optimal Blind Fusion of Bits

Jean-Pierre Delmas, *Senior Member, IEEE*, and Yann Meurisse

Abstract—This paper is devoted to a statistical performance analysis of blind estimation of bit error rates (BERs) of a bank of detectors, using empirical estimation algorithms that have appeared in the literature (by Dixit *et al.*). In particular, we prove that these blind estimators asymptotically (in the number of observed bits) achieve the accuracy obtained with perfect knowledge of the transmitted bits. We propose a maximum-likelihood solution which follows from the standard expectation-maximization (EM) algorithm, considered to be a reference algorithm. Finally, the optimal fusion rule is revisited and our theoretical results are compared to Monte Carlo simulations.

Index Terms—Fusion of bits, performance analysis, expectation-maximization (EM) algorithm.

I. INTRODUCTION

IN recent years, signal processing with distributed sensors and decentralized detection has been gaining importance in many applications. Most results (see, e.g., [1] and [2] and the references therein) on decentralized detection assume that each sensor node produces a finite-valued function of its observation, conditioned on the state 0 or 1 of a phenomenon H , which is conveyed over a noisy channel to the fusion center. In this paper, we consider the simplest case where a binary sensor decision is conveyed reliably to the fusion center.

This happens, for example, in cellular communications, where one often wishes to merge data from several distributed detectors located in base stations, to improve overall link performance. In this case, raw bit decisions from each detector are fused into a final bit decision. Naturally, this scheme is suboptimal with respect to the optimal combination of sufficient statistics issued from each of the detectors. Assuming perfect knowledge of the individual BERs of each detector, optimal fusion of bit estimates is a standard problem [3]. However, when this knowledge is missing, only recent empirical algorithms [4]–[6] have addressed this problem. Reference [4] introduces the problem of blind fusion of bit estimates, curiously not studied until now: It presents a very clever, simple, and intuitive algorithm to solve this problem. Then, a refinement of the previous algorithm and an iterative procedure that outperforms the previous algorithms for small numbers of observations are proposed in [5] and [6], respectively. In this contribution, we

examine the statistical performance of blindly estimating bit error rates (BERs).

The paper is organized as follows. After stating the problem and giving a statistical framework in Section II, we consider an expectation-maximization (EM) algorithm where the transmitted bits play the role of the missing data and we analyze the statistical performance of empirical algorithms based on the number of mutual agreements between the different receivers in Sections III and IV, respectively. Conditions where the optimal fusion detector reduces to the majority detector or to the best detector are given in Section V. Our theoretical results are compared to Monte Carlo simulations in Section VI. Finally Section VII summarizes our contribution.

II. PROBLEM STATEMENT

A. Model and Notations

The problem we consider here involves the simultaneous transmission of T binary signals $(b_t)_{t=1,\dots,T} \in \{0,1\}$ over K independent memoryless binary symmetric channels (BSC). It is assumed that each bit b_t may be randomly flipped by an independent error sequence $n_{k,t} \in \{0,1\}$, depending on the channel k

$$y_{k,t} = b_t \oplus n_{k,t}, \quad k = 1, \dots, K \quad \text{and} \quad t = 1, \dots, T$$

where $(y_{k,t})_{k=1,\dots,K,t=1,\dots,T} \in \{0,1\}$ are the observations, and \oplus denotes modulo-two addition. Hence, the received bit $y_{k,t}$ from the channel k at time t is equal to the input bit b_t if $n_{k,t} = 0$; otherwise $y_{k,t} \neq b_t$ and an error occurs. The transmitted bits $(b_t)_{t=1,\dots,T}$ are assumed to be independent and have equal prior probabilities. The K error probabilities are $\theta_k = P(n_{k,t} = 1)$ and are unknown parameters. Let Θ be the parameter $(\theta_1, \dots, \theta_K)$.¹

Based on the observation $Y \stackrel{\text{def}}{=} (y_{k,t})_{k=1,\dots,K,t=1,\dots,T}$, the problem consists (see Fig. 1) of detecting the bits $B \stackrel{\text{def}}{=} (b_t)_{t=1,\dots,T}$ without any knowledge of the separate error probabilities $(\theta_k)_{k=1,\dots,K}$ with or without previous estimates of the parameter Θ .

B. Statistical Framework

It is a fundamental rule in decision theory that the statistical framework must be carefully specified before examining optimality. As usual, classical and Bayesian approaches can be considered.

¹Note that alternative modeling of correlated binary observations $(y_{k,t})_{k=1,\dots,K}$ has been studied in literature (see, e.g., [7] and the references therein).

Manuscript received October 1, 2005; revised June 23, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athanasios P. Llavas.

The authors are with the GET/INT, Département CITI, UMR-CNRS 5157, Institut National des Télécommunications, 91011 Evry Cedex, France (e-mail: jean-pierre.delmas@int-evry.fr; yann.meurisse@int-evry.fr).

Digital Object Identifier 10.1109/TSP.2006.890818

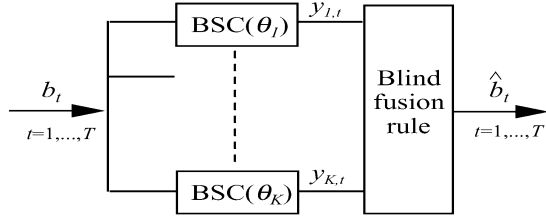


Fig. 1. Blind bit fusion.

In the classical approach, the parameter Θ is assumed to be a deterministic but unknown constant and we are faced with a joint parameter estimation and bit detection problem, also known as composite hypothesis testing in the statistical literature. The distributions $p(Y, B; \Theta)$ and its marginal $p(Y; \Theta)$ are well defined. Ideally, we would like to detect $(\hat{b}_t(Y))_{t=1, \dots, T}$ by minimizing the error probability $P[\hat{b}_t(Y) \neq b_t]_{t=1, \dots, T}$ for all values of $\Theta \in (0, (1/2)^K)$. This, however, is an unachievable goal because the probabilities $P[\hat{b}_t(Y) \neq b_t]_{t=1, \dots, T}$ depend in fact on Θ and consequently $(\hat{b}_t(Y))_{t=1, \dots, T}$ would depend on Θ . In these situations, several practical solutions, very similar to generalized-likelihood ratio tests, can be used. These solutions are composed of two consecutive maximizations. The most attractive one is the use of two maximum-likelihood estimates

$$\hat{\Theta}(Y) = \arg \left(\max_{\Theta} p(Y; \Theta) \right) \quad (2.1)$$

followed by

$$\hat{B}(Y) = \arg \left(\max_B p(Y, B; \hat{\Theta}(Y)) \right).$$

The first step can be solved using the EM algorithm (see Section III) and the second step gives the following detection rule, which will be commented on in Section V:

$$\sum_{k=1}^K (2y_{k,t} - 1) \ln \frac{1 - \hat{\theta}_k(Y)}{\hat{\theta}_k(Y)} \underset{\hat{b}_t(Y)=0}{\overset{\hat{b}_t(Y)=1}{>}} 0, \quad t = 1, \dots, T. \quad (2.2)$$

Very attractive empirical parameter estimators can replace the first step of the previous procedure, as addressed in Section IV.

In the Bayesian approach, the parameter Θ is assumed to be a random variable whose particular realization we must estimate. A prior distribution of Θ is assumed given. In this case, $p(Y, B)$ is a marginal issued from the joint distribution $p(Y, B, \Theta)$. Once a prior distribution has been chosen for Θ , optimal detection of the sequence $(b_t)_{t=1, \dots, T}$ can be considered. In particular, maximum *a posteriori* (MAP) estimators minimize the error probability $P[(\hat{b}_t(Y))_{t=1, \dots, T} \neq (b_t)_{t=1, \dots, T}]$. Because $(b_t)_{t=1, \dots, T}$ is uniformly distributed, the MAP estimator is given by

$$\hat{B}(Y) = \arg \left(\max_B p(Y/B) \right)$$

with $p(Y/B) = \int \dots \int p(Y/B, \Theta) p(\Theta) d\theta_1 \dots d\theta_K$ and $p(Y/B, \Theta) = \prod_{t=1}^T \prod_{k=1}^K \theta_k^{y_{k,t} \oplus b_t} (1 - \theta_k)^{1 - (y_{k,t} \oplus b_t)}$. If the

prior distribution of Θ is now assumed uniform in $(0, 1)$, a closed-form expression of $p(Y/B)$ can be derived

$$\begin{aligned} p(Y/B) &= \prod_{k=1}^K \int_0^1 \theta_k^{\sum_{t=1}^T y_{k,t} \oplus b_t} (1 - \theta_k)^{\sum_{t=1}^T (1 - y_{k,t} \oplus b_t)} d\theta_k \\ &= \prod_{k=1}^K \frac{\left(\sum_{t=1}^T y_{k,t} \oplus b_t \right)! \left(\sum_{t=1}^T (1 - y_{k,t} \oplus b_t) \right)!}{(T+1)!} \end{aligned}$$

and $(\hat{b}_t(Y))_{t=1, \dots, T}$ is the solution of the following intricate maximization:

$$\hat{B}(Y) = \arg \left(\max_B \prod_{k=1}^K \left(\sum_{t=1}^T y_{k,t} \oplus b_t \right)! \left(\sum_{t=1}^T (1 - y_{k,t} \oplus b_t) \right)! \right).$$

We note that the previous function takes the same value for the sequences $(b_t)_{t=1, \dots, T}$ and $(1 - b_t)_{t=1, \dots, T}$; this comes from the prior $(0, 1)^K$ uniform distribution.

If the MAP criterion is applied to the “parameter” (B, Θ) , a suboptimal detection procedure can be proposed. Since the joint distribution of (B, Θ) is uniform, we have

$$(\hat{B}(Y), \hat{\Theta}(Y)) = \arg \left(\max_{B, \Theta} p(Y/B, \Theta) \right). \quad (2.3)$$

Because

$$\begin{aligned} &\arg \left(\max_{\Theta} p(Y/B, \Theta) \right) \\ &= \arg \left(\max_{\Theta} \sum_{k=1}^K \left(\sum_{t=1}^T (y_{k,t} \oplus b_t) \ln \theta_k \right. \right. \\ &\quad \left. \left. + \sum_{t=1}^T (1 - (y_{k,t} \oplus b_t)) \ln (1 - \theta_k) \right) \right) \end{aligned}$$

and

$$\begin{aligned} &\arg \left(\max_B p(Y/B, \Theta) \right) \\ &= \arg \left(\max_B \sum_{t=1}^T \left(\sum_{k=1}^K (y_{k,t} \oplus b_t) \ln \theta_k \right. \right. \\ &\quad \left. \left. + \sum_{k=1}^K (1 - (y_{k,t} \oplus b_t)) \ln (1 - \theta_k) \right) \right) \end{aligned}$$

the maximization (2.3) can be solved with the following cyclic procedure, after an appropriate initialization at iteration i :

$$\sum_{k=1}^K (2y_{k,t} - 1) \ln \frac{1 - \hat{\theta}_k^{(i)}(Y)}{\hat{\theta}_k^{(i)}(Y)} \underset{\hat{b}_t^{(i+1)}(Y)=0}{\overset{\hat{b}_t^{(i+1)}(Y)=1}{>}} 0, \quad t = 1, \dots, T$$

and

$$\hat{b}_k^{(i+1)}(Y) = \frac{1}{T} \sum_{t=1}^T y_{k,t} \oplus \hat{b}_t^{(i+1)}(Y), \quad k = 1, \dots, K.$$

Within the Bayesian framework, we note that an optimal estimator $\hat{\Theta}$ always exists. However, this optimality is obtained “on average” (as different values of Θ are chosen). Of course, for a particular value of Θ , it may not perform as well as an estimator obtained in the classical framework. Consequently, in the sequel, only the classical approach will be considered.

III. MAXIMUM-LIKELIHOOD ESTIMATION OF THE ERROR PROBABILITIES

Due to the intricate expression of the likelihood (2.1), its direct maximization using standard optimization techniques is likely to be intractable. In these situations, a practical solution is to use the EM algorithm, an iterative method for finding maximum-likelihood estimates in problems involving incomplete data. In this problem, the incomplete data is the observation itself Y , and the complete data consists of $X = (Y, B)$ with $B \stackrel{\text{def}}{=} (b_t)_{t=1, \dots, T}$. The likelihood of X is referred to as the complete likelihood. The basic idea behind EM is to maximize the incomplete likelihood by iteratively maximizing the complete likelihood. The complete likelihood may be written as

$$p(X; \Theta) = \frac{1}{2^T} \prod_{t=1}^T \prod_{k=1}^K \theta_k^{y_{k,t} \oplus b_t} (1 - \theta_k)^{1 - (y_{k,t} \oplus b_t)}.$$

Each iteration of EM has two steps: an expectation step and a maximization step. The $(i+1)$ st E-step finds the conditional expectation of the complete data log-likelihood with respect to the conditional distribution of the missing data given the observation $Y = y$ and the current estimated parameter $\Theta^{(i)}$

$$\begin{aligned} \Theta &\mapsto Q(\Theta, \Theta^{(i)}) \\ &\stackrel{\text{def}}{=} \mathbb{E} \left(\ln \left(p(X; \Theta) / Y = y; \Theta^{(i)} \right) \right) \\ &\propto \sum_{t=1}^T \sum_{k=1}^K \mathbb{E} \left((y_{k,t} \oplus b_t) / Y = y; \Theta^{(i)} \right) \ln \theta_k \\ &\quad + \mathbb{E} \left((1 - (y_{k,t} \oplus b_t)) / Y = y; \Theta^{(i)} \right) \ln(1 - \theta_k) \\ &\propto \sum_{t=1}^T \sum_{k=1}^K P(y_{k,t} \oplus b_t = 1 / Y = y; \Theta^{(i)}) \ln \theta_k \\ &\quad + P(y_{k,t} \oplus b_t = 0 / Y = y; \Theta^{(i)}) \ln(1 - \theta_k) \end{aligned}$$

as a function of the unknown parameter Θ , given the values of the observation y and the parameter $\Theta^{(i)}$. Since the posterior probabilities $P(y_{k,t} \oplus b_t = 1 / Y = y; \Theta^{(i)})$ and $P(y_{k,t} \oplus b_t = 0 / Y = y; \Theta^{(i)})$ are given by

$$\begin{aligned} P(y_{k,t} \oplus b_t = 1 / Y = y; \Theta^{(i)}) &= \frac{a_{k,t}^{(i)}}{a_t^{(i)}} \\ P(y_{k,t} \oplus b_t = 0 / Y = y; \Theta^{(i)}) &= 1 - \frac{a_{k,t}^{(i)}}{a_t^{(i)}} \end{aligned}$$

with

$$a_{k,t}^{(i)} = \prod_{k'=1}^K \theta_{k'}^{(i) 1 - (y_{k',t} \oplus y_{k,t})} \left(1 - \theta_{k'}^{(i)} \right)^{y_{k',t} \oplus y_{k,t}}$$

and

$$\begin{aligned} a_t^{(i)} &= \prod_{k=1}^K \theta_k^{(i) 1 - y_{k,t}} \left(1 - \theta_k^{(i)} \right)^{y_{k,t}} \\ &\quad + \prod_{k=1}^K \theta_k^{(i) y_{k,t}} \left(1 - \theta_k^{(i)} \right)^{1 - y_{k,t}} \end{aligned}$$

the $(i+1)$ st M-step involves the following maximization:

$$\begin{aligned} \Theta^{(i+1)} &= \arg \left(\max_{\Theta} \sum_{t=1}^T \sum_{k=1}^K \left(\frac{a_{k,t}^{(i)}}{a_t^{(i)}} \ln \theta_k \right. \right. \\ &\quad \left. \left. + \left(1 - \frac{a_{k,t}^{(i)}}{a_t^{(i)}} \right) \ln(1 - \theta_k) \right) \right) \\ \Rightarrow \theta_k^{(i+1)} &= \frac{1}{T} \sum_{t=1}^T \frac{a_{k,t}^{(i)}}{a_t^{(i)}}, \quad k = 1, \dots, K. \end{aligned}$$

IV. BLIND EMPIRICAL FUSION RULE BY DIXIT *et al.*

A. Algorithms of Dixit *et al.*

To estimate the parameter Θ , Dixit *et al.* consider the following statistic $(c_t^{k,l})_{1 \leq k < l \leq K, t=1, \dots, T}$ issued from the observation Y :

$$c_t^{k,l} \stackrel{\text{def}}{=} 1 - (y_{k,t} \oplus y_{l,t}) = \begin{cases} 1, & \text{if } y_{k,t} = y_{l,t} \\ 0, & \text{if } y_{k,t} \neq y_{l,t}. \end{cases}$$

The idea behind these different numbers of mutual agreements between the different receivers $c_t^{k,l}$ is that $c_t^{k,l} = 1 - (n_{k,t} \oplus n_{l,t})$. Consequently this statistic does not depend on the bit sequence $(b_t)_{t=1, \dots, T}$, and each $c_t^{k,l}$ is Bernoulli distributed with parameter $\theta_{k,l} = (1 - \theta_k)(1 - \theta_l) + \theta_k \theta_l$.

From this statistic, Dixit *et al.* propose different methods of moments based on the $(K(K-1)/2)$ sample moments $\alpha_T^{k,l} = (1/T) \sum_{t=1}^T c_t^{k,l}$ that are consistent estimate of functions $\mathbb{E}(c_t^{k,l}) = \theta_{k,l}$ of the parameter Θ . All these methods are composed of two steps.

1) Empirical ordering of the channels ($\theta_1 \leq \theta_2 \leq \dots \leq \theta_K$).

This procedure is based on comparisons of the different outputs $y_{k,t}$ with the majority fusion rule detector (see [4]).

2) Resolution of a consistent system of K equations $((k, l) = (1, 2), (1, 3), \dots, (1, K), (2, 3))$ or an inconsistent overdetermined system of $(K(K-1)/2)$ ($1 \leq k < l \leq K$) obtained by replacing $\mathbb{E}(c_t^{k,l})$ by its sample estimate $\alpha_T^{k,l}$ and solving for Θ , where the unknown is the estimate $\hat{\Theta}$.

We note that first step is essential for the selection of the K equations and for the resolution of the overdetermined system.

Note that the random variables $(c_t^{k,l})_{1 \leq k < l \leq K}$ are not independent because $c_t^{k,l} = 1 - (c_t^{j,k} \oplus c_t^{j,l})$ for $1 \leq j < k < l \leq K$. Furthermore, given a set of consistent statistics $(c_t^{k,l})_{1 \leq k < l \leq K}$, Y can only take two complementary equally probable values; e.g., if $(c_t^{k,l} = 1)_{1 \leq k < l \leq K}$, $(y_{k,t} = 0)_{1 \leq k \leq K}$

or $(y_{k,t} = 1)_{1 \leq k \leq K}$. Consequently, these statistics $c_t^{k,l}$ are sufficient statistics for the parameter Θ .

In [4]–[6], it is claimed that these approaches are *optimal* and this assertion is supported by two points. The statistic $\{(c_t^{k,l})_{1 \leq k < l \leq K, t=1, \dots, T}\}$ carries the same information about Θ as the observation Y , and the random variables $c_t^{k,l}$ are independent. The reduced statistic $\{(\alpha_T^{k,l})_{1 \leq k < l \leq K}\}$, however, is not a sufficient statistic because the Neyman–Fisher factorization does not apply.

B. Statistical Analysis

After specifying the distribution of the statistic $\{(c_t^{k,l})_{1 \leq k < l \leq K, t=1, \dots, T}\}$, we give the asymptotic distribution of the statistic $\alpha_T^{k,l}$. This allows us to give the asymptotic distribution as well as closed-form expressions of the bias and variance of an estimate of Θ given by an arbitrary algorithm based on the sample moments $\alpha_T^{k,l}$. It is also possible to derive the asymptotic lower bound on the variance of an arbitrary consistent estimator based on these statistics and the associated Cramér–Rao bound.

Because $E(c_t^{k,l} c_t^{k',l'}) = P(n_{k,t} = n_{l,t} \text{ and } n_{k',t} = n_{l',t})$, it is straightforward to prove that the $(K(K-1)/2)$ -dimensional random variables $(c_t^{k,l})_{1 \leq k < l \leq K}$ for $t = 1, \dots, T$ are independent and composed of dependent Bernoulli random variables $c_t^{k,l}$ with parameter $\theta_{k,l}$ whose covariance matrix is \mathbf{C}_α given by (4.1) shown at the bottom of the page, with $\theta_{m,n,p} = (1 - \theta_m)(1 - \theta_n)(1 - \theta_p) + \theta_m \theta_n \theta_p$. Consequently, from the standard central limit theorem, the sequence of statistics $\alpha_T \stackrel{\text{def}}{=} (\alpha_T^{k,l})_{1 \leq k < l \leq K}$ is asymptotically Gaussian distributed

$$\sqrt{T}(\alpha_T - \alpha) \stackrel{\mathcal{L}}{\sim} \mathcal{N}(\mathbf{0}; \mathbf{C}_\alpha).$$

Here, α collects the associated $(\theta_{k,l})_{1 \leq k < l \leq K}$.

To consider the asymptotic performance of an arbitrary algorithm based on the statistic $(\alpha_T^{k,l})_{1 \leq k < l \leq K}$ or on a subset of these statistics, we adopt a functional approach that consists of recognizing that the whole process of constructing an estimate Θ_T of Θ is equivalent to defining a functional relation linking the estimate Θ_T to the sample statistic α_T from which it is inferred. This functional dependence is denoted $\Theta_T = \text{alg}[\alpha_T]$. Clearly, $\Theta = \text{alg}[\alpha]$, so the different algorithms $\text{alg}(\cdot)$ constitute distinct extensions of the mapping $\alpha \stackrel{\text{def}}{=} (\theta_{k,l})_{1 \leq k < l \leq K} \rightarrow \Theta$ to any statistics α_T . If \mathbf{J}_{alg} and $(\mathbf{H}_{\text{alg}}^k)_{k=1, \dots, K}$ denote, respectively, the Jacobian and the Hessian matrices associated with this mapping at point

²To emphasize that we are dealing with a sequence of estimates, we replace in this section the notation $\hat{\Theta}$ by Θ_T .

$\alpha \stackrel{\text{def}}{=} (\theta_{k,l})_{1 \leq k < l \leq K}$, the asymptotic bias and variance of an arbitrary algorithm are, respectively

$$\begin{aligned} \text{bias}(\Theta_T) &= \frac{1}{2T} \begin{bmatrix} \text{Tr} \left[\mathbf{H}_{\text{alg}}^1 \mathbf{C}_\alpha \right] \\ \vdots \\ \text{Tr} \left[\mathbf{H}_{\text{alg}}^K \mathbf{C}_\alpha \right] \end{bmatrix} + o\left(\frac{1}{T}\right), \\ E[(\Theta_T - \Theta)(\Theta_T - \Theta)^T] &= \frac{1}{T} \mathbf{J}_{\text{alg}} \mathbf{C}_\alpha \mathbf{J}_{\text{alg}}^T + o\left(\frac{1}{T}\right). \end{aligned}$$

And by the continuity theorem, Θ_T is asymptotically Gaussian distributed

$$\sqrt{T}(\Theta_T - \Theta) \stackrel{\mathcal{L}}{\sim} \mathcal{N}(\mathbf{0}; \mathbf{J}_{\text{alg}} \mathbf{C}_\alpha \mathbf{J}_{\text{alg}}^T)$$

where \mathbf{C}_α denotes the covariance matrix of the statistic $\alpha_T^{k,l}$ involved in the algorithm, deduced from (4.1).

For example, the algorithm proposed in [4], [5] consists in solving the following K consistent nonlinear equations

$$\begin{aligned} 1 - \theta_{1,T} - \theta_{2,T} + 2\theta_{1,T}\theta_{2,T} &= \alpha_T^{1,2} \\ 1 - \theta_{1,T} - \theta_{3,T} + 2\theta_{1,T}\theta_{3,T} &= \alpha_T^{1,3} \\ &\vdots \\ 1 - \theta_{1,T} - \theta_{K,T} + 2\theta_{1,T}\theta_{K,T} &= \alpha_T^{1,K} \\ 1 - \theta_{2,T} - \theta_{3,T} + 2\theta_{2,T}\theta_{3,T} &= \alpha_T^{2,3}. \end{aligned} \quad (4.2)$$

We note that the Jacobian and the Hessian matrices associated with a one-to-one mapping “alg” and its inverse mapping “alg⁻¹” are connected by the following relations (see, e.g., [10]):

$$\mathbf{J}_{\text{alg}} = (\mathbf{J}_{\text{alg}^{-1}})^{-1}$$

and

$$\begin{aligned} \begin{bmatrix} \mathbf{H}_{\text{alg}}^1 \\ \vdots \\ \mathbf{H}_{\text{alg}}^K \end{bmatrix} &= - \left((\mathbf{J}_{\text{alg}^{-1}})^{-1} \otimes \mathbf{I}_K \right) \\ &\quad \times \begin{bmatrix} (\mathbf{J}_{\text{alg}^{-1}})^{-T} \mathbf{H}_{\text{alg}^{-1}}^1 (\mathbf{J}_{\text{alg}^{-1}})^{-1} \\ \vdots \\ (\mathbf{J}_{\text{alg}^{-1}})^{-T} \mathbf{H}_{\text{alg}^{-1}}^K (\mathbf{J}_{\text{alg}^{-1}})^{-1} \end{bmatrix} \end{aligned}$$

in addition, it is straightforward to obtain the expressions of $\mathbf{J}_{\text{alg}^{-1}}$ and $\mathbf{H}_{\text{alg}^{-1}}^k$ from (4.2) as shown in the first equation at

$$[\mathbf{C}_\alpha]_{(k,l),(k',l')} \stackrel{\text{def}}{=} \text{Cov} \left(c_t^{k,l}, c_t^{k',l'} \right) = \begin{cases} 0 & \text{if } k, l, k' \text{ and } l' \text{ are all distinct,} \\ \theta_{k,l}(1 - \theta_{k,l}) & \text{if } k = k' \text{ and } l = l', \\ \theta_{m,n,p} - \theta_{k,l}\theta_{k',l'} & \text{if only two indices among} \\ & k, l, k', l' \text{ are equal,} \\ & (m, n, p \text{ are these distinct values}) \end{cases} \quad (4.1)$$

the bottom of the page. Consequently, \mathbf{J}_{alg} and $(\mathbf{H}_{\text{alg}}^k)_{k=1,\dots,K}$ associated with the algorithm proposed in [4] and [5] are deduced. With the expression of \mathbf{J}_{alg} , we prove the following property.

Property 1: The estimated parameters $\theta_{k,T}$ given by the Dixit *et al.* rule (4.2) have asymptotic variances $\text{var}(\theta_{k,T}) = (\theta_k + o(\Theta)/T)$. Furthermore, if $(\theta_k \ll 1)_{k=1,\dots,K}$, these estimates are asymptotically uncorrelated and have asymptotic variances invariant with respect to the choice of the K selected statistics $\alpha_T^{k,l}$ and $\text{var}(\theta_{k,T}) \approx (\theta_k/T)$.

As such, in these conditions, the blind estimation achieves the accuracy obtained with perfect knowledge of the transmitted bits given by $\text{var}(\theta_{k,T}) = (\theta_k(1 - \theta_k)/T) \approx (\theta_k/T)$. In practice, since the relative accuracy on the estimated parameters is given by $\sqrt{(\text{var}(\theta_{k,T})/\theta_k^2)} \approx (1/\sqrt{T\theta_k})$, the number T of bits observed must be roughly inversely proportional to the least bit error probability θ_k .

Proof: With $\theta_{k,l}(1 - \theta_{k,l}) = \theta_k + \theta_l + o(\Theta)$ (where $\lim_{\Theta \rightarrow 0}(o(\Theta)/|\Theta|) = 0$) and $\theta_{m,n,p} - \theta_{k,l}\theta_{l',l'} = \theta_i + o(\Theta)$ (where i is the common subscript among k, l, k', l'), we have the second equation shown at the bottom of the page. In the same way, the Jacobians \mathbf{J}_K associated with K channels are given by

$$\begin{aligned} \mathbf{J}_3 &= \left[\begin{pmatrix} -1 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & -1 \end{pmatrix} + \epsilon(\Theta) \right]^{-1} \\ &= \frac{1}{2} \begin{pmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{pmatrix} + \epsilon(\Theta) \end{aligned}$$

$$\begin{aligned} \mathbf{J}_k &= \left[\begin{pmatrix} -1 & -1 & 0 & \dots & \dots & 0 \\ -1 & 0 & -1 & 0 & \dots & \vdots \\ \vdots & 0 & \ddots & \ddots & \cdot & \vdots \\ \vdots & \vdots & \cdot & \ddots & \ddots & \vdots \\ -1 & 0 & \dots & \dots & 0 & -1 \\ 0 & -1 & -1 & \dots & \dots & 0 \end{pmatrix} + \epsilon(\Theta) \right]^{-1}, \\ &\text{for } k = 4, \dots, K \\ &= \frac{1}{2} \begin{pmatrix} -1 & -1 & 0 & \dots & 0 & 1 \\ -1 & 1 & 0 & \dots & \vdots & -1 \\ 1 & -1 & 0 & \dots & \vdots & \vdots \\ \vdots & 1 & -2 & 0 & \vdots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 & \vdots \\ 1 & 1 & 0 & 0 & -2 & -1 \end{pmatrix} + \epsilon(\Theta) \end{aligned}$$

with $\lim_{\Theta \rightarrow 0} \epsilon(\Theta) = 0$. Putting these values in $\mathbf{J}_{\text{alg}} \mathbf{C}_\alpha \mathbf{J}_{\text{alg}}^T$, we straightforwardly obtain after some tedious algebraic manipulations

$$\mathbf{J}_{\text{alg}} \mathbf{C}_\alpha \mathbf{J}_{\text{alg}}^T = \text{Diag}(\theta_1, \theta_2, \dots, \theta_K) + o(\Theta).$$

To assess the performance improvement obtained by an algorithm that uses all $(K(K - 1)/2)$ available statistics $\alpha_T^{k,l}$, we now consider the asymptotic lower bound on the variance of an arbitrary consistent estimator based on all these statistics. As with the Cramér–Rao bound, this bound can be used as a benchmark against which potential estimates are tested.

$$\begin{aligned} \mathbf{J}_{\text{alg}^{-1}} &= \begin{pmatrix} -1 + 2\theta_2 & -1 + 2\theta_1 & 0 & \dots & \dots & 0 \\ -1 + 2\theta_3 & 0 & -1 + 2\theta_1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 & \vdots \\ \vdots & \vdots & \cdot & \ddots & \ddots & 0 \\ -1 + 2\theta_K & 0 & \dots & \dots & 0 & -1 + 2\theta_1 \\ 0 & -1 + 2\theta_3 & -1 + 2\theta_2 & 0 & \dots & 0 \end{pmatrix} \\ [\mathbf{H}_{\text{alg}^{-1}}^k]_{(i,j)} &= \begin{cases} 2, & \text{for } (i,j) = (1, 1+k) \text{ and } (1+k, 1), k = 1, \dots, K-1 \\ 0, & \text{otherwise} \end{cases} \\ [\mathbf{H}_{\text{alg}^{-1}}^K]_{(i,j)} &= \begin{cases} 2, & \text{for } (i,j) = (2, 3) \text{ and } (3, 2) \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

$$\mathbf{C}_\alpha = \begin{pmatrix} \theta_1 + \theta_2 & \theta_1 & \dots & \dots & \theta_1 & \theta_2 \\ \theta_1 & \theta_1 + \theta_3 & \theta_1 & \ddots & \theta_1 & \theta_3 \\ \vdots & \theta_1 & \ddots & \ddots & \vdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \theta_1 & \vdots \\ \theta_1 & \dots & \dots & \theta_1 & \theta_1 + \theta_K & 0 \\ \theta_2 & \theta_3 & 0 & \dots & 0 & \theta_2 + \theta_3 \end{pmatrix} + o(\Theta).$$

From the works of [8] and [9], the asymptotic covariance \mathbf{C}_Θ of a consistent estimator of Θ is lower bounded by the symmetric positive-definite matrix $(\mathbf{D}^T \mathbf{C}_\alpha^{-1} \mathbf{D})^{-1}$, where \mathbf{D} is the $(K(K-1)/2) \times K$ matrix defined as $[\mathbf{D}]_{(i,j)} \stackrel{\text{def}}{=} (d\theta_{k,l}/d\theta_j)$, $1 \leq k < l \leq K$ (note that the particular ordering of the row of \mathbf{D} is irrelevant in the expression $(\mathbf{D}^T \mathbf{C}_\alpha^{-1} \mathbf{D})^{-1}$ if this order is consistent with the ordering of the terms of \mathbf{C}_α defined in (4.1)).

Finally, to assess the relative efficiency of the statistic $(\alpha_T^{k,l})_{1 \leq k < l \leq K}$, this asymptotically minimum variance lower bound is compared to the Cramér–Rao bound. We note that the Cramér–Rao bound appears to be prohibitive to compute because the distribution of $y_t = (y_{k,t})_{k=1,\dots,K}$ is a mixture of two K -dimensional Bernoulli distributions. We can use, however, a numerical expression derived from the Fisher information matrix, i.e.,

$$\text{CRB}(\Theta) = \frac{1}{T} (\mathbf{I}_F^{-1}) \quad \text{with}$$

$$(\mathbf{I}_F)_{k,l} = \sum_{y_t \in \{0,1\}^K} \left(\frac{\partial \ln p(y_t; \Theta)}{\partial \theta_k} \right) \times \left(\frac{\partial \ln p(y_t; \Theta)}{\partial \theta_l} \right) p(y_t; \Theta), \quad k, l = 1, \dots, K$$

where³ (see the equation at the bottom of the page).

V. OPTIMUM BIT FUSION RULE REVISITED

We revisit in this section the optimal fusion rule, for three purposes. First, we give closed-form expressions of the error probability given by the optimal fusion detector. Second, we give conditions where this optimal fusion detector reduces to the majority detector or to the best detector,⁴ and finally we study the sensitivity of the optimal fusion detector to errors in the estimate of the error probabilities of the different channels.

Assume here that $\theta_1 \leq \theta_2 \dots \leq \theta_K \leq (1/2)$. The optimal fusion rule (2.2) can be interpreted as a weighting of the outputs $y'_{k,t} \stackrel{\text{def}}{=} 2y_{k,t} - 1 \in \{-1, +1\}$ of the different channels according to their reliability because the weights $w_k \stackrel{\text{def}}{=} \ln(1 - \theta_k/\theta_k)$ are decreasing functions of θ_k . As for fixed-point binary representation of numbers, the most significant bits $y'_{1,t} \dots$ tend to impose their signs compared with the least significant bits $\dots y'_{K,t}$.

$$\sum_{k=1}^K w_k y'_{k,t} \begin{matrix} \hat{\delta}_t(Y) = 1 \\ > \\ < \\ \hat{\delta}_t(Y) = 0 \end{matrix} 0, \quad t = 1, \dots, T. \quad (5.1)$$

³We note that this score function appears to be prohibitive to compute at each iteration of a Fisher scoring algorithm.

⁴An illustrative example is given in [4, Fig. 6].

From this rule, the optimal error probability P_e^K is given by

$$P_e^K = \sum_{\mathbf{y}'^i: \mathbf{w}^T \mathbf{y}'^i < 0} \left(\prod_{k=1}^K \theta_k^{y_k^i} (1 - \theta_k)^{(1-y_k^i)} \right) + \frac{1}{2} \sum_{\mathbf{y}'^i: \mathbf{w}^T \mathbf{y}'^i = 0} \left(\prod_{k=1}^K \theta_k^{y_k^i} (1 - \theta_k)^{(1-y_k^i)} \right) \quad (5.2)$$

where $\mathbf{w} \stackrel{\text{def}}{=} (w_1, \dots, w_K)^T$, $\mathbf{y}'^i \stackrel{\text{def}}{=} (y_1^i, \dots, y_K^i)^T$, $i = 1, 2, \dots, 2^K$ is associated with $\mathbf{y}^i \in \{0, 1\}^K$. Consequently, the following property is proven.

Property: The optimal error probability P_e decreases or does not change if a channel is added. In particular it does not change if this added channel has an error probability equal to $(1/2)$.

Proof: If a channel is added, each “word” (y_1^i, \dots, y_K^i) is split into two words due to the weight w_{K+1} . Consequently, if $\theta_{K+1} = (1/2) \Leftrightarrow w_{K+1} = 0$, the optimal fusion rule and the error probability do not change. However, if $\theta_{K+1} < (1/2) \Leftrightarrow w_{K+1} > 0$, three cases may be considered. First, for the first K channels, there is a word (y_1^i, \dots, y_K^i) such that $\mathbf{w}^T \mathbf{y}_i = 0$. Consequently, the two words issued from this word introduce two terms $\mathbf{w}^T \mathbf{y}_i$ of opposite signs and consequently P_e^K decreases. Second, for the first K channels, suppose there is no word (y_1^i, \dots, y_K^i) such that $\mathbf{w}^T \mathbf{y}_i = 0$. If θ_{K+1} is very close to $(1/2) \Leftrightarrow w_{K+1} \approx 0$ such that the two words issued from each word (y_1^i, \dots, y_K^i) do not change the sign of $\mathbf{w}^T \mathbf{y}_i$, the optimal fusion rule and the error probability do not change. But if θ_{K+1} is not very close to $(1/2)$ such that among the two words issued from each word at least one term (y_1^i, \dots, y_K^i) changes the sign of $\mathbf{w}^T \mathbf{y}_i$, the optimal fusion rule changes and the error probability decreases. ■

The optimal error probability P_e^K for $K = 2, 3, 4$ given by (5.2) simplifies to

$$P_e^2 = \theta_1$$

$$P_e^3 = \inf (P_e^2, \theta_1 \theta_2 + \theta_1 \theta_3 + \theta_2 \theta_3 - 2\theta_1 \theta_2 \theta_3)$$

$$P_e^4 = \inf (P_e^3, \theta_1 (\theta_2 + \theta_3 + \theta_4 - \theta_2 \theta_3 - \theta_2 \theta_4 - \theta_3 \theta_4) + \theta_2 \theta_3 \theta_4)$$

if we suppose that $\sum_{k=1}^K w_k y_k^i \neq 0$ for all words \mathbf{y}_i . Unfortunately, for $K \geq 5$, such relations may not be proved by induction.

An interesting problem is to deduce conditions under which the optimal fusion rule reduces to the majority detector or to the best detector. If the different probabilities of error θ_k are equal, the optimal fusion rule reduces to the majority detector (see (5.1)) and intuitively if these probabilities of error θ_k are very close together, then this property remains true. On the other

$$\frac{\partial \ln p(y_t; \Theta)}{\partial \theta_l} = \frac{\prod_{1 \leq k \neq l \leq K} \theta_k^{1-y_{k,t}} (1 - \theta_k)^{y_{k,t}} (-1)^{y_{l,t}} + \prod_{1 \leq k \neq l \leq K} \theta_k^{y_{k,t}} (1 - \theta_k)^{1-y_{k,t}} (-1)^{1-y_{l,t}}}{\prod_{k=1}^K \theta_k^{1-y_{k,t}} (1 - \theta_k)^{y_{k,t}} + \prod_{k=1}^K \theta_k^{y_{k,t}} (1 - \theta_k)^{1-y_{k,t}}}$$

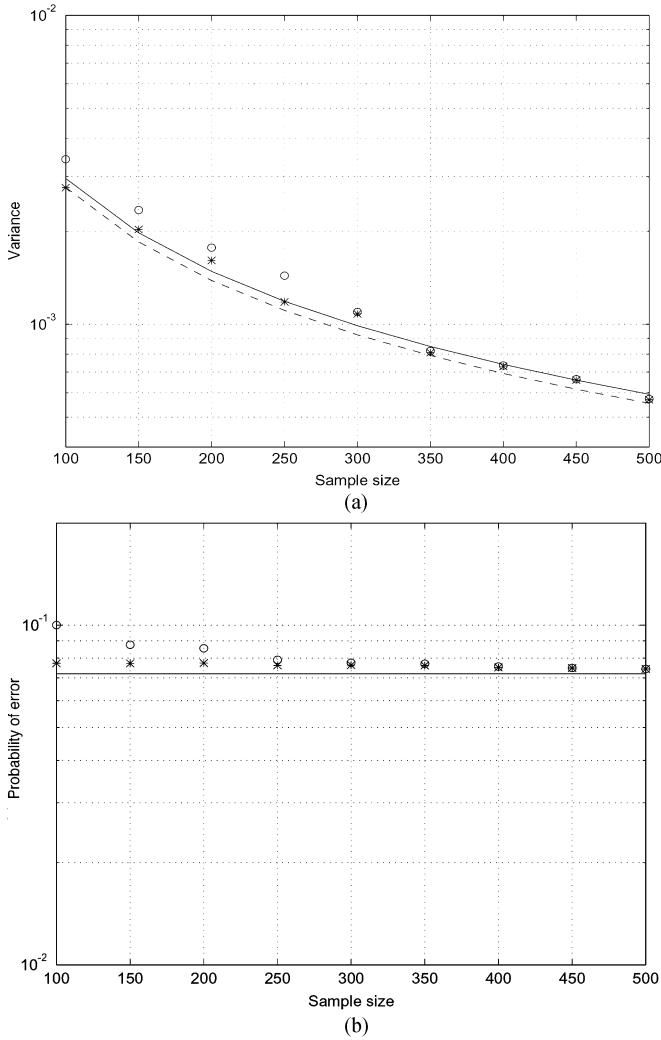


Fig. 2. (a) Theoretical (-) and empirical (o) variance given by the Dixit rule, CRB (-) and empirical (*) variance given by the EM algorithm of the parameter θ_1 as a function of the sample size. (b) Probability of error given by the “optimal” fusion (2.2) derived from channel BER estimated by the Dixit (4.2) (o) and the EM (*) algorithms, compared with the “optimal” fusion (-) derived from the exact values of $(\theta_k)_{k=1,\dots,K}$ as a function of the sample size.

hand, if the probability of error θ_1 of the best channel is very small compared to the probabilities of error of the other channels, then the optimal fusion rule reduces to the best detector. In the following, we try to quantify these notions.

From the optimal fusion rule (5.1), because $w_1 \geq \sum_{k=2}^K w_k \Leftrightarrow \text{sign}(y'_{1,t}) = \text{sign}(\sum_{k=1}^K w_k y'_{k,t})$, a necessary and sufficient condition for the optimal fusion rule to reduce to the best detector is

$$\frac{1 - \theta_1}{\theta_1} > \prod_{k=2}^K \frac{1 - \theta_k}{\theta_k}.$$

For K even, the optimal fusion rule never reduces to the majority rule, except if the error probabilities θ_k are equal. Indeed in the case of equality of 0 and 1, the majority detector

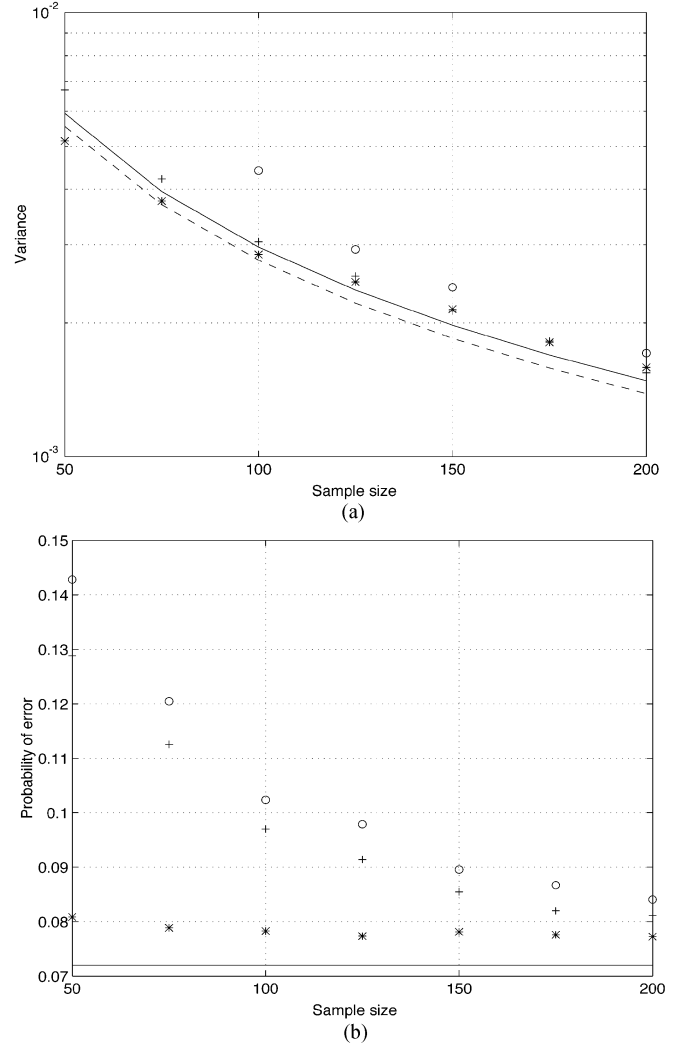


Fig. 3. (a) Theoretical (-) variance given by the Dixit rule, CRB (-) and empirical variances given by the Dixit rule (o) and algorithm [6] (+) and the EM algorithm (*) of the parameter θ_1 as a function of the sample size. (b) Probability of error given by the “optimal” fusion (2.2) derived from channel BER estimated by the Dixit rule (4.2) (o), [6] (+) and the EM (*) algorithms respectively, compared with the “optimal” fusion (-) derived from the exact values of $(\theta_k)_{k=1,\dots,K}$ as a function of the sample size.

draws lots for its detection if the error probabilities θ_k are very close together. For K odd, a necessary and sufficient condition for the optimal fusion rule to reduce to the majority detector seems to be very intricate in the general case. To draw conditions, consider the particular case where the error probabilities θ_K satisfy the condition $\theta_1 < \theta_2 = \dots = \theta_K$. Because $\text{sign}(\sum_{k=1}^K w_k y'_{k,t}) = \text{sign}((w_1/w_2)y'_{1,t} + \sum_{k=2}^K y'_{k,t})$, $\text{sign}(\sum_{k=1}^K w_k y'_{k,t}) = \text{sign}(\sum_{k=1}^K y'_{k,t}) \Leftrightarrow (w_1/w_2) < 2$. Consequently

$$\frac{1 - \theta_1}{\theta_1} < \left(\frac{1 - \theta_2}{\theta_2}\right)^2$$

is a necessary and sufficient condition for the optimal fusion rule to reduce to the majority detector.

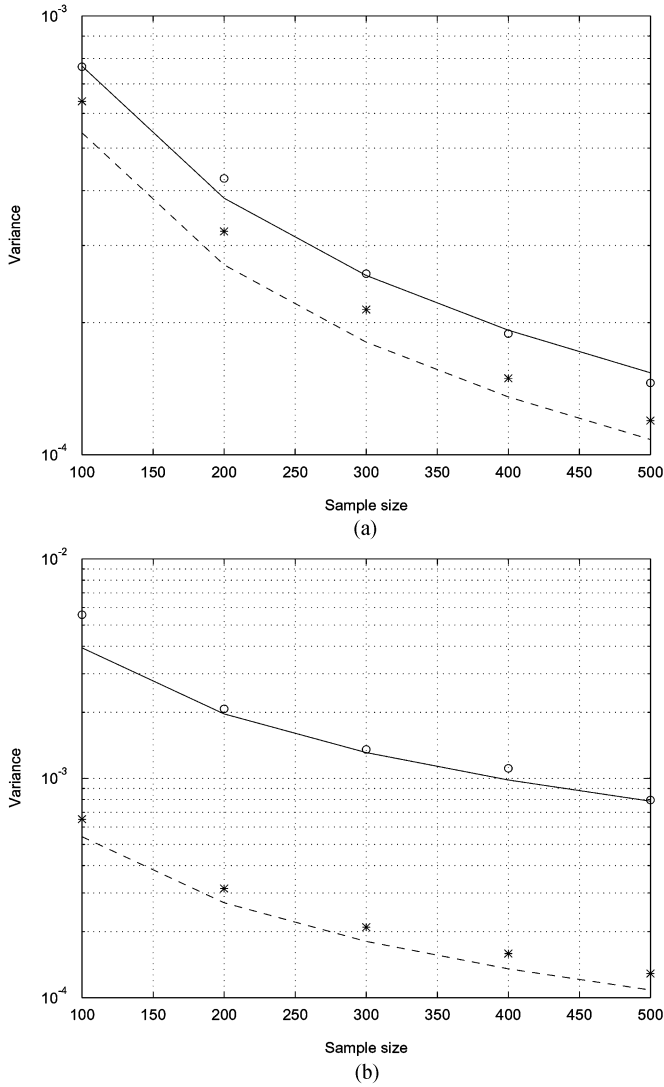


Fig. 4. Theoretical (-) and empirical (o) asymptotic variance given by the Dixit algorithm (4.2), CRB (-) and empirical (*) asymptotic variance given by the EM algorithm of the parameter θ_1 as a function of the sample size.

To study the sensitivity of the optimal fusion detector to errors in the estimate of the error probability of the different channels, we first note that the differential of w_k satisfies

$$d(w_k) = -\frac{d\theta_k}{\theta_k} \left(\frac{1}{1-\theta_k} \right) = -d\theta_k \left(\frac{1}{\theta_k(1-\theta_k)} \right).$$

Consequently, if the relative accuracy is considered, the sensitivity of the optimal fusion detector increases when the channel degrades and conversely if the absolute accuracy is considered, the sensitivity of the optimal fusion detector increases when the channel improves.

VI. ILLUSTRATIVE EXAMPLES

Our performance analysis is illustrated by two experiments. In each one, we plot the theoretical and empirical (averaged over 1000 runs) asymptotic variances of estimates of a probability of error of a particular channel given by different Dixit algorithms and by the EM algorithm. We also plot the resulting

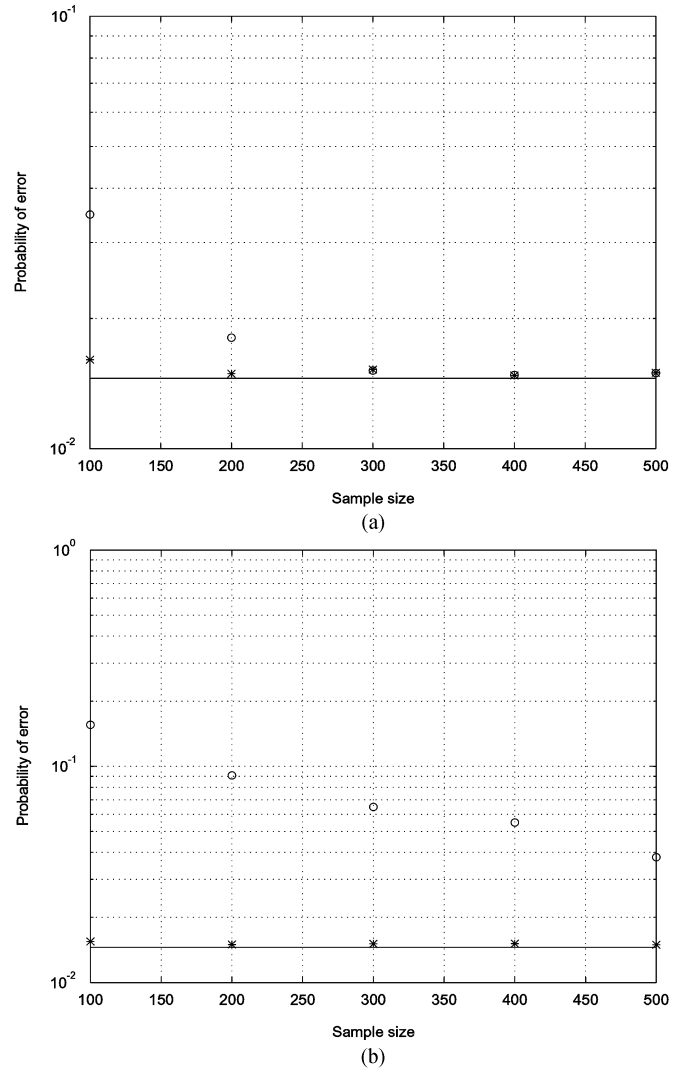


Fig. 5. Probability of error given by the “optimal” fusion (2.2) derived from channel BER estimated by the Dixit (4.2) (o), [6] (+), and the EM (*) algorithms respectively, compared with the “optimal” fusion (-) derived from the exact values of $(\theta_k)_{k=1,\dots,K}$ as a function of the sample size.

probability of error given by the “optimal” fusion (2.2) derived from channel $BER(\hat{\theta}_k(Y))_{k=1,\dots,K}$ estimated by these Dixit algorithms and by the EM algorithm.

The first experiment considers the case of three channels where $\Theta = (0.1; 0.2; 0.2)$. The EM algorithm is initialized by the following approximation of (4.2):

$$\begin{pmatrix} \theta_{1,0} \\ \theta_{2,0} \\ \theta_{3,0} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 - \alpha_T^{1,2} \\ 1 - \alpha_T^{1,3} \\ 1 - \alpha_T^{1,4} \end{pmatrix}$$

Fig. 2(a) and (b) compare the performance of the EM algorithm to the Dixit (4.2) rule. We see from Fig. 2(a), that our theoretic asymptotic results are valid over a large range of the number T of observed bits (from $T = 100$ and $T = 350$ for the EM and the Dixit algorithms (4.2) respectively). Naturally the EM algorithm outperforms the Dixit (4.2) rule but their performances are very close, except from a very small number of samples. We note that

the resulting probability of error given by the “optimal” fusion (2.2) derived from channel BER $(\hat{\theta}_k(Y))_{k=1,\dots,K}$ estimated by these algorithms is very close to this ”optimal” fusion derived from the exact values of $(\theta_k)_{k=1,\dots,K}$, except for a very small number of samples ($T < 300$).

Fig. 3(a) and (b) compares the performance of the EM algorithm to the Dixit (4.2) and [6] algorithms for a small number of samples. In contrast to the previous figure, the EM algorithm largely outperforms the Dixit algorithms especially for a very small number of samples. The algorithms [6] outperform the Dixit rule issued from (4.2) below 200 samples. Note that the Dixit (4.2) rule gives very bad estimates for $T \leq 125$, whereas the algorithms [6] keep on converging. We note that the resulting probability of error given by the “optimal” fusion (2.2) derived from channel BER $(\hat{\theta}_k(Y))_{k=1,\dots,K}$ estimated by all these algorithms is robust to the bad estimate of θ_1 given by the Dixit (4.2) rule.

The second experiment considers the case of four channels where $\Theta = (0.05; 0.1; 0.2; 0.2)$. To see the influence of the choice of the selected statistics $\alpha_T^{k,l}$ in the Dixit rule (4.2), Figs. 4 and 5 exhibit the case where the different channel BERs are arranged in ascending or descending values of θ_k respectively. In other words, the Dixit rule is based on $(\alpha_T^{1,2}, \alpha_T^{1,3}, \alpha_T^{1,4}, \alpha_T^{2,3})$ in Figs. 4(a) and 5(a) and $(\alpha_T^{4,3}, \alpha_T^{4,2}, \alpha_T^{4,1}, \alpha_T^{3,2})$ in Figs. 4(b) and 5(b). We see from Fig. 4 that our theoretic asymptotic results are valid over a large range of the number T of observed bits, but contrary to the case where $(\theta_k)_{k=1,\dots,K} \ll 1$ (see property 1), the Dixit rule is very sensitive to the choice of the selected statistics. Naturally, the EM algorithm (initialized by the following approximation of (4.2))

$$\begin{pmatrix} \theta_{1,0} \\ \theta_{2,0} \\ \theta_{3,0} \\ \theta_{4,0} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & -1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 - \alpha_T^{1,2} \\ 1 - \alpha_T^{1,3} \\ 1 - \alpha_T^{1,4} \\ 1 - \alpha_T^{2,3} \end{pmatrix}$$

outperforms the Dixit rule. Fig. 5 shows that the probability of error derived from channel BERs estimated by the EM algorithm is very close to the “optimal” fusion (2.2) derived from the exact values of $(\theta_k)_{k=1,\dots,K}$. However, this probability of error is very robust to bad estimate of $(\theta_k)_{k=1,\dots,K}$ given by Dixit rules for the second arrangement of $(\theta_k)_{k=1,\dots,K}$.

VII. CONCLUSION

In this paper, we have presented a comprehensive asymptotic statistical performance analysis of blind estimation of bit error rates (BERs) of a bank of detectors which have appeared in the literature (by Dixit *et al.*). We have proposed a maximum-likelihood solution obtained by the standard EM algorithm which can be considered as a reference algorithm. We have proven, in particular, that these blind estimators asymptotically (in the

number of observed bits) achieve the accuracy obtained with perfect knowledge of the transmitted bits. Finally, the optimal fusion rule is revisited and our theoretical results are compared with Monte Carlo simulations.

REFERENCES

- [1] J. F. Chamberland and V. V. Veeravalli, “Asymptotic results for decentralized detection in power constrained wireless sensor networks,” *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.
- [2] R. Viswanathan and P. K. Varshney, “Distributed detection with multiple sensors: Part I—Fundamentals,” *Proc. IEEE*, vol. 285, no. 1, pp. 54–63, Jan. 1997.
- [3] Z. Chair and P. K. Varshney, “Optimal data fusion in multiple sensor detection systems,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 98–101, Jan. 1986.
- [4] A. Dixit and G. C. Orsak, “Optimal empirical detector fusion,” presented at the IEEE Signal Processing Workshop, Hunt, TX, Oct. 15–18, 2000.
- [5] A. Dixit and G. C. Orsak, “Asymptotically optimal blind fusion of bit estimates,” presented at the IEEE Signal Processing Workshop, Callaway Gardens, GA, Oct. 13–16, 2002.
- [6] A. Dixit, S. C. Douglas, and G. C. Orsak, “Blind estimation of channel BERs in a multi-receiver network,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP) 2004*, Montreal, QC, Canada, May 17–21, 2004, vol. 2, pp. ii-305–ii-308.
- [7] R. L. Prentice, “Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors,” *J. Amer. Statist. Assoc.*, vol. 81, no. 394, Jun. 1986.
- [8] B. Porat and B. Friedlander, “Performance analysis of parameter estimation algorithms based on high-order moments,” in *Int. J. Adapt. Control, Signal Process.*, 1989, vol. 3, pp. 191–229.
- [9] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [10] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*, ser. Wiley Series in Probability and Statistics, rev. ed. New York: Wiley, 1999.



Jean Pierre Delmas (M’00–SM’06) was born in France in 1950. He received the Engineering degree from Ecole Centrale de Lyon, France, in 1973, the Certificat d’Etudes Supérieures from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1982, and the Habilitation à Diriger des Recherches (HDR) degree from the University of Paris, Orsay, France, in 2001.

Since 1980, he has been with the Institut National des Télécommunications, Evry, France, where he is presently a Professor in the CITI department and Deputy Director of the UMR CNRS 5157 (SAMOVAR) laboratory. His teaching and research interests are in the areas of statistical signal processing with application to communications and antenna array.

Mr. Delmas is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.



Yann Meurisse received the Engineer degree in telecommunications from the Institut National des Télécommunications (INT), Evry, France in 1983 and the Diplôme d’Etudes Approfondies in automatic control and signal processing and the Ph.D. degree from the University of Paris Sud, Orsay, France, in 1987 and 2002, respectively.

Since 1985, he has been with the INT, where he is currently an Assistant Professor in the Signal and Image Processing Department. His research interests are in antenna array processing.