On the Subexponential Decay of Detection Error Probabilities in Long Tandems

Wee Peng Tay, John N. Tsitsiklis, *Fellow, IEEE*, and Moe Z. Win, *Fellow, IEEE*

***Abstract*—We consider the problem of Bayesian decentralized binary hypothesis testing in a network of sensors arranged in a tandem. We show that the rate of error probability decay is always subexponential, establishing the validity of a long-standing conjecture. Under the additional assumption of bounded Kullback–Leibler (KL) divergences, we show that for all $d > 1/2$, the error probability is $\Omega(e^{-cn^d})$, where c is a positive constant. Furthermore, the bound $\Omega(e^{-c(\log n)^d})$, for all $d > 1$, holds under an additional mild condition on the distributions. This latter bound is shown to be tight.**

***Index Terms*—Decentralized detection, error exponent, serial network, tandem network.**

I. Introduction

Consider a tandem network, as shown in Fig. 1, with each sensor i observing an independent random variable X_i, which has marginal law $\mathbb{P}_j$ under hypothesis H_j, $j = 0, 1$. Sensor i sends a 1-bit message $Y_i = \gamma_i(Y_{i-1}, X_i)$ (Y_0 can be defined to be always 0) to sensor $i+1$. The transmission function γ_i used by sensor i is thus a function of the observed X_i and the received message Y_{i-1} from sensor $i-1$. We call the collection $(\gamma_1, \ldots, \gamma_n)$ a *strategy* for the n-sensor tandem network.

Let $\pi_j > 0$ be the prior probability of hypothesis H_j, and let $P_e(n) = \pi_0 \mathbb{P}_0(Y_n = 1) + \pi_1 \mathbb{P}_1(Y_n = 0)$ be the probability of error at sensor n, under some particular strategy. The goal of a system designer is to design a strategy so that the probability of error $P_e(n)$ is minimized. Let $P_e^*(n) = \inf P_e(n)$, where the infimum is taken over all possible strategies.

The problem of finding optimal strategies has been studied in [1]–[3], while the asymptotic performance of a long tandem network (i.e., $n \to \infty$) is considered in [2], [4]–[8] (some of these works do not restrict the message sent by each sensor to be binary). In the case of binary communications, [4] and [8] show that the error probability stays bounded away from zero iff $|\log \frac{d\mathbb{P}_1}{d\mathbb{P}_0}| \leq B$ almost surely, for some constant B. When the log-likelihood ratio is unbounded, numerical examples have indicated that the error probability goes to zero much slower than exponentially. This is to be contrasted with the case of a parallel configuration (all sensors send messages $\gamma_i(X_i)$ directly to a single fusion center), where the error probability decays exponentially fast with the number of sensors n [9]. This suggests that a tandem configuration performs worse than a parallel configuration, when n is large. It has been conjectured in [2], [8], [10], [11] that indeed, the rate of decay of the error probability is subexponential. However, a proof is not available. The goal of this correspondence is to prove this conjecture.

Manuscript received February 8, 2007; revised June 22, 2008. Current version published September 17, 2008. This work was supported, in part, by the National Science Foundation under Contracts ECCS-0701623, ECS-0426453, ANI-0335256 and ECS-0636519, and DoCoMo USA Labs. The material in this paper was presented in part at the 32nd International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, April 2007.

The authors are with the Laboratory for Information and Decision Systems, MIT, Cambridge MA 02139 USA (e-mail: wptay@mit.edu; jnt@mit.edu; moewin@mit.edu).

Communicated by L. Tong, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2008.929032

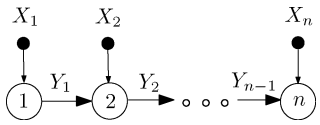

Fig. 1. A tandem network.

The study of tandem networks is of interest because their simple structure serves as a basis for the study of more complicated tree architectures [3], [8], [11], [12]. A tandem network can also be viewed as a representation of a single node with 1 bit of memory, making observations X_i at different time periods [4], [5]. Therefore, our results are also relevant to the sequential detection problem. For tractability reasons, the tandem networks we study in this correspondence are stylized approximations of practical networks, which typically allow multiple bits of communication between nodes, and have noisy communication links. Our analysis of this simplified network serves to give insights into the rate of error decay in practical tandem networks, and more general tree architectures.

We first note that there is a caveat to the subexponential decay conjecture: the probability measures $\mathbb{P}_0$ and $\mathbb{P}_1$ need to be equivalent, i.e., absolutely continuous with respect to (w.r.t.) each other. Indeed, if there exists a measurable set A such that $\mathbb{P}_0(A) > 0$ and $\mathbb{P}_1(A) = 0$, then an exponential decay rate can be achieved as follows: each sensor always declares 1 until some sensor m observes an $X_m \in A$, whereupon all sensors $i \geq m$ declare 0. For this reason, we assume throughout this correspondence that the measures $\mathbb{P}_0$ and $\mathbb{P}_1$ are equivalent. Under this assumption, we first show that the error probability decays subexponentially fast with the number of sensors n. When the error probability goes to zero, we would also like to quantify the best possible (subexponential) decay rate. In this spirit, we find lower bounds on the probability of error, under some further mild assumptions.

The rest of the correspondence is organized as follows. In Section II, we show that the error probability decays subexponentially. In Section III, we derive more detailed lower bounds on the error probabilities. In Section IV, we establish the tightness of one of our lower bounds. Finally, Section V contains concluding remarks.

II. Subexponential Decay

In this section, we show that the rate of decay of the error probability is always subexponential. Although the proof is simple, we have not been able to find it in the literature. Instead, all works on this topic, to our best knowledge, have only conjectured that the decay is subexponential, with numerical examples as supporting evidence.

We first state an elementary fact that we will make use of throughout this correspondence. A proof can be found in [13].

Lemma 1: Suppose that $\mathbb{P}$ and $\mathbb{Q}$ are two equivalent probability measures. If $A_1, A_2, \ldots$ is a sequence of measurable events such that $\mathbb{P}(A_n) \to 0$, as $n \to \infty$, then $\mathbb{Q}(A_n) \to 0$, as $n \to \infty$.

Let $L_i = \log \frac{d\mathbb{P}_1}{d\mathbb{P}_0}(X_i)$ be the log-likelihood ratio associated with the observation made by sensor i. From [1], [8], [10], [14], there is no loss in optimality if we require each sensor to form its messages by using a log-likelihood ratio quantizer (LLRQ), i.e., $Y_i = 0$ iff $L_i \leq t_{i,n}(y)$, where $t_{i,n}(y)$ is a threshold whose value depends on the message $Y_{i-1} = y$ received by sensor i. In the sequel, we will assume, without loss of optimality, that all sensors use an LLRQ. The next lemma follows easily from the existence results in [14], and Proposition 4.2 in [10]. A proof can be found in [13].

Lemma 2: There exists an optimal strategy under which each sensor uses an LLRQ, with thresholds that satisfy $t_{i,n}(1) \leq t_{i,n}(0)$ for all $i = 1, \ldots, n$.

Proposition 1: The rate of decay of the error probability in a tandem network is subexponential, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log P_e^*(n) = 0.$$

Proof: Suppose that $P_e^*(n) \to 0$ as $n \to \infty$, else the proposition is trivially true. Fix some n and consider an optimal strategy for the tandem network of length n. In view of Lemma 2, we can restrict to strategies in which $t_{i,n}(1) \leq t_{i,n}(0)$ for all i. We have

$$\begin{aligned}\mathbb{P}_0(Y_i = 1) &= \mathbb{P}_0\Big(L_i > t_{i,n}(0)\Big) \cdot \mathbb{P}_0(Y_{i-1} = 0) \\ &\quad + \mathbb{P}_0\Big(L_i > t_{i,n}(1)\Big) \cdot \mathbb{P}_0(Y_{i-1} = 1) \qquad (1)\\ \mathbb{P}_1(Y_i = 0) &= \mathbb{P}_1\Big(L_i \leq t_{i,n}(0)\Big) \cdot \mathbb{P}_1(Y_{i-1} = 0) \\ &\quad + \mathbb{P}_1\Big(L_i \leq t_{i,n}(1)\Big) \cdot \mathbb{P}_1(Y_{i-1} = 1). \qquad (2)\end{aligned}$$

From (1) and (2), with $i = n$, and applying Lemma 2, we have

$$\begin{aligned}P_e^*(n) &= \pi_0 \mathbb{P}_0(Y_n = 1) + \pi_1 \mathbb{P}_1(Y_n = 0) \\ &= \pi_0 \Big[\mathbb{P}_0\Big(L_n > t_{n,n}(0)\Big) \\ &\qquad + \mathbb{P}_0\Big(t_{n,n}(1) < L_n \leq t_{n,n}(0)\Big) \cdot \mathbb{P}_0(Y_{n-1} = 1)\Big] \\ &\quad + \pi_1 \Big[\mathbb{P}_1\Big(L_n \leq t_{n,n}(1)\Big) \\ &\qquad + \mathbb{P}_1\Big(t_{n,n}(1) < L_n \leq t_{n,n}(0)\Big) \cdot \mathbb{P}_1(Y_{n-1} = 0)\Big] \qquad (3)\\ &\geq \min_{j=0,1} \mathbb{P}_j\Big(t_{n,n}(1) < L_n \leq t_{n,n}(0)\Big) \cdot P_e^*(n-1). \qquad (4)\end{aligned}$$

From (3), in order to have $P_e^*(n) \to 0$ as $n \to \infty$, we must have $\mathbb{P}_0(L_n > t_{n,n}(0)) \to 0$ and $\mathbb{P}_1(L_n \leq t_{n,n}(1)) \to 0$, as $n \to \infty$. Because $\mathbb{P}_0$ and $\mathbb{P}_1$ are equivalent measures, from Lemma 1, we have $\mathbb{P}_1(L_n > t_{n,n}(0)) \to 0$ and $\mathbb{P}_0(L_n \leq t_{n,n}(1)) \to 0$, as $n \to \infty$. Hence, $\mathbb{P}_j(t_{n,n}(1) < L_n \leq t_{n,n}(0)) \to 1$ for $j = 0, 1$. Therefore, from (4), the error probability cannot decay exponentially fast. □

We have established that the decay of the error probability is subexponential. This confirms that the parallel configuration performs much better than the tandem configuration when n is large. In the next section, we investigate the best performance that a tandem configuration can possibly achieve.

III. Rate of Decay

In this section, we show that under the assumption of bounded Kullback–Leibler (KL) divergences, the error probability is $\Omega(e^{-cn^d})$, for some positive constant c and for all $d > 1/2$. Under some additional assumptions, the lower bound is improved to $\Omega(e^{-c(\log n)^d})$, for any $d > 1$. We rely on a sequence of comparisons of the tandem configuration with other tree configurations, whose performance can be quantified using methods similar to [12].

Our results involve the KL divergences, defined by

$$D_0 = \mathbb{E}_0\Big[\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big], \quad D_1 = \mathbb{E}_1\Big[\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big].$$

We assume that $-\infty < D_0 < 0 < D_1 < \infty$, throughout this section.

Let k and m be positive integers, and let $n = km$. Let us compare the following two networks: i) a tandem network, as in Fig. 1, with n sensors, where each sensor i obtains a single observation X_i; ii) a modified tandem network $T(k, m)$, as in Fig. 2, with k sensors, where each sensor v_i obtains m (conditionally) independent observations $X_{(i-1)m+1}, \ldots, X_{im}$, given either hypothesis. In both networks, a sensor sends a binary message to its recipient. It should be clear that when we keep the total number of observations $n = km$ the same in both networks, the network $T(k, m)$ can perform at least as well as the original one. Indeed, each sensor v_i in the modified network can emulate the behavior of m sensors in tandem in the original network.

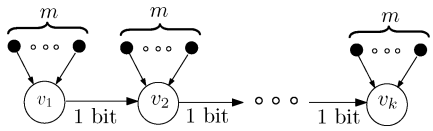

Fig. 2. A modified tandem network $T(k, m)$ that outperforms a tandem network with $n = km$ sensors.

Therefore, it suffices to establish a lower bound for the error probability in the network $T(k, m)$. Toward this goal, we will use some standard results in large deviations theory, notably Cramér's theorem [15], as stated in the lemma that follows. A proof is provided in [13].

Lemma 3: Suppose that $-\infty < D_0 < 0 < D_1 < \infty$, and that $X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.) under either hypothesis H_j, with marginal law $\mathbb{P}_j$. Let $S_m = \sum_{i=1}^m L_i$, and for $j = 0, 1$, let

$$\Lambda_j^*(t) = \sup_{\xi\in\mathbb{R}} \left\{\xi t - \log \mathbb{E}_j\Big[\Big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big)^{\xi}\Big]\right\}.$$

.

(a) For every $\epsilon > 0$, there exist $a \in (0, 1)$, $c > 0$, and $M \geq 1$, such that for all $m \geq M$

$$\begin{aligned}\mathbb{P}_0(S_m/m > D_1 + \epsilon) &\geq a e^{-mc} \\ \mathbb{P}_1(S_m/m \leq D_0 - \epsilon) &\geq a e^{-mc}.\end{aligned}$$

(b) Suppose that $\mathbb{E}_1\Big[\Big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big)^s\Big] < \infty$ for some $s > 0$. Then, there exists some $\epsilon > 0$, such that $\Lambda_1^*(D_1 + \epsilon) > 0$, and

$$\mathbb{P}_1(S_m/m \leq D_1 + \epsilon) \geq 1 - e^{-m\Lambda_1^*(D_1+\epsilon)}, \qquad \forall\, m \geq 1.$$

(c) Suppose that $\mathbb{E}_0\Big[\Big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big)^s\Big] < \infty$ for some $s < 0$. Then, there exists some $\epsilon > 0$, such that $\Lambda_0^*(D_0 - \epsilon) > 0$, and

$$\mathbb{P}_0(S_m/m > D_0 - \epsilon) \geq 1 - e^{-m\Lambda_0^*(D_0-\epsilon)}, \qquad \forall\, m \geq 1.$$

(d) For every $\epsilon > 0$, there exists some $M \geq 1$ such that

$$\mathbb{P}_1(S_m/m \leq D_1 + \epsilon) \geq 1/2, \qquad \forall\, m \geq M.$$

Moreover, if for some integer $r \geq 2$, $\mathbb{E}_1\Big[\Big|\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big|^r\Big] < \infty$, then there exists some $c_r > 0$ such that

$$\mathbb{P}_1(S_m/m \leq D_1 + \epsilon) \geq 1 - \frac{c_r}{m^{r/2}\epsilon^r}, \qquad \forall\, m \geq 1.$$

(e) For every $\epsilon > 0$, there exists some $M \geq 1$ such that

$$\mathbb{P}_0(S_m/m > D_0 - \epsilon) \geq 1/2, \qquad \forall\, m \geq M.$$

Moreover, if for some integer $r \geq 2$, $\mathbb{E}_0\Big[\Big|\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big|^r\Big] < \infty$, then there exists some $c_r > 0$ such that

$$\mathbb{P}_0(S_m/m > D_0 - \epsilon) \geq 1 - \frac{c_r}{m^{r/2}\epsilon^r}, \qquad \forall\, m \geq 1.$$

We now state our main result. Part (ii) of the following proposition is a general lower bound that always holds; part (i) is a stronger lower bound, under an additional assumption. Note that the condition in part (i) implies that $\mathbb{E}_j\left[\left|\log\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\right|^r\right]<\infty$ for all r, but the reverse implication is not always true.

Proposition 2: Suppose that $-\infty<D_0<0<D_1<\infty$.

(i) Suppose that there exists some $\epsilon'>0$ such that for all $s\in[-\epsilon',1+\epsilon']$, $\mathbb{E}_0\left[\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\right)^s\right]<\infty$. Then

$$\lim_{n\to\infty}\frac{1}{(\log n)^d}\log P_e^*(n)=0$$

for all $d>1$.

(ii) For all $d>1/2$, we have

$$\lim_{n\to\infty}\frac{1}{n^d}\log P_e^*(n)=0.$$

Furthermore, if for some integer $r\geq 2$, $\mathbb{E}_j\left[\left|\log\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\right|^r\right]<\infty$ for both $j=0,1$, then the above is true for all $d>1/(2+r/2)$.

Proof: Let us fix m and k, and an optimal strategy for the modified network $T(k,m)$. Let Y_{v_i} be the 1-bit message sent by sensor v_i, under that strategy. Let

$$S_{i,m}=\sum_{l=1}^{m}L_{(i-1)m+l} \tag{5}$$

which is the log-likelihood ratio of the observations obtained at sensor v_i. For the same reasons as in Lemma 2, an optimal strategy exists and can be taken to be an LLRQ, i.e., $Y_{v_i}=0$ iff $S_{i,m}/m\leq t_{i,m}(y)$, where $t_{i,m}(y)$ is a threshold whose value depends on the message y received by sensor v_i from sensor v_{i-1}. For the same reasons as in Lemma 2, we can assume that the optimal strategy is chosen such that $t_{i,m}(1)\leq t_{i,m}(0)$, for all $m\geq 1$, and for all $i\geq 1$.

Let $q_{0,i}=\mathbb{P}_0(Y_{v_i}=1)$ and $q_{1,i}=\mathbb{P}_1(Y_{v_i}=0)$ be the Type I and II error probabilities at sensor v_i. Suppose that the conditions in part (i) of the proposition hold. Let $\delta=\min\{\Lambda_0^*(D_0-\epsilon),\Lambda_1^*(D_1+\epsilon)\}$. From parts (ii) and (iii) of Lemma 3, there exists $\epsilon>0$ such that $\delta>0$. Let us fix such an ϵ, and let $a\in(0,1)$, $c>0$, and $M\geq 1$ be as in Lemma 3 (i). We first show a lower bound on the Type I and II error probabilities $q_{j,i}$.

Lemma 4: There exists some $\bar{M}$ such that for every $i\geq 1$, and every $m\geq\bar{M}$, either

$$q_{0,i}\geq\frac{a}{2}e^{-mc}(1-e^{-m\delta})^i \tag{6}$$

or

$$q_{1,i}\geq\frac{a}{2}e^{-mc}(1-e^{-m\delta})^i. \tag{7}$$

Proof: The proof proceeds by induction on i. When $i=1$, the result is an immediate consequence of Lemma 3 (i). Indeed, if the threshold t used by sensor v_1 satisfies $t\leq D_1$, then $q_{0,1}\geq ae^{-mc}$, and if $t\geq D_0$, then $q_{1,1}\geq ae^{-mc}$.

Assume now that $i>1$ and that the result holds for $i-1$. We will show that it also holds for i. Let $S_{i,m}$ be as defined in (5). We have for $i>1$

$$\begin{aligned}q_{0,i}&=(1-q_{0,i-1})\mathbb{P}_0(S_{i,m}/m>t_{i,m}(0))\\&\quad+q_{0,i-1}\mathbb{P}_0(S_{i,m}/m>t_{i,m}(1))\end{aligned} \tag{8}$$

$$\begin{aligned}q_{1,i}&=(1-q_{1,i-1})\mathbb{P}_1(S_{i,m}/m\leq t_{i,m}(1))\\&\quad+q_{1,i-1}\mathbb{P}_1(S_{i,m}/m\leq t_{i,m}(0)).\end{aligned} \tag{9}$$

We start by considering the case where $q_{0,i-1}<1/2$ and $q_{1,i-1}<1/2$. Suppose that $t_{i,m}(0)\leq D_1+\epsilon$. From (8) and Lemma 3 (i), we have for all $m\geq M$

$$\begin{aligned}q_{0,i}&\geq\frac{1}{2}\mathbb{P}_0(S_{i,m}/m>D_1+\epsilon)\\&\geq\frac{a}{2}e^{-mc}\\&\geq\frac{a}{2}e^{-mc}(1-e^{-m\delta})^i.\end{aligned}$$

Similarly, if $t_{i,m}(1)\geq D_0-\epsilon$, we have $q_{1,i}\geq ae^{-mc}(1-e^{-m\delta})^i/2$.

It remains to consider the case where $t_{i,m}(0)>D_1+\epsilon$ and $t_{i,m}(1)<D_0-\epsilon$. From (8) and Lemma 3 (iii), we obtain

$$\begin{aligned}q_{0,i}&\geq q_{0,i-1}\mathbb{P}_0(S_{i,m}/m>D_0-\epsilon)\\&\geq q_{0,i-1}(1-e^{-m\delta}).\end{aligned}$$

Similarly, from (9) and Lemma 3 (ii), we have

$$\begin{aligned}q_{1,i}&\geq q_{1,i-1}\mathbb{P}_1(S_{i,m}/m\leq D_1+\epsilon)\\&\geq q_{1,i-1}(1-e^{-m\delta}).\end{aligned}$$

Using the induction hypothesis, either (6) or (7) holds.

We next consider the case where $q_{0,i-1}\geq 1/2$ and $q_{1,i-1}<1/2$. If either

a) $t_{i,m}(1)\geq D_0-\epsilon$, or

b) $t_{i,m}(0)>D_1+\epsilon$ and $t_{i,m}(1)<D_0-\epsilon$,

we obtain, via the same argument as above, the desired conclusion. Suppose then that $t_{i,m}(0)\leq D_1+\epsilon$ and $t_{i,m}(1)<D_0-\epsilon$. From (8) and the Weak Law of Large Numbers (WLLN), we have for some $\bar{M}$ sufficiently large, and for all $m\geq\bar{M}$

$$\begin{aligned}q_{0,i}&\geq\frac{1}{2}\mathbb{P}_0(S_{i,m}/m>t_{i,m}(1))\\&\geq\frac{1}{2}\mathbb{P}_0(S_{i,m}/m>D_0-\epsilon)\geq\frac{1}{4}\end{aligned}$$

so that the claim holds trivially. The case $q_{0,i-1}<1/2$ and $q_{1,i-1}\geq 1/2$ is similar.

We finally consider the case where $q_{0,i-1}\geq 1/2$ and $q_{1,i-1}\geq 1/2$. If $t_{i,m}(1)\leq D_1$, then

$$q_{0,i}\geq\frac{1}{2}\mathbb{P}_0(S_{i,m}/m>D_1)\geq\frac{a}{2}e^{-mc}.$$

If, on the other hand, $t_{i,m}(1)>D_1$, then $t_{i,m}(0)\geq t_{i,m}(1)>D_1>D_0$, and

$$q_{1,i}\geq\frac{1}{2}\mathbb{P}_1(S_{i,m}/m\leq D_0)\geq\frac{a}{2}e^{-mc}.$$

This concludes the proof of the lemma. □

We return to the proof of part (i) of Proposition 2. Fix some $d>1$ and some $l\in(1/d,1)$. Let $k=k(m)=\exp(m^l)$. For a tandem network with n sensors, since $k(m)m=\exp(m^l)m$ is increasing in m, we have $\exp((m-1)^l)(m-1)<n\leq\exp(m^l)m$, for some m. Since the tree network $T(k(m),m)$ outperforms a tandem network with n sensors, we have

$$\begin{aligned}P_e^*(n)&\geq\pi_0q_{0,k(m)}+\pi_1q_{1,k(m)}\\&\geq\min\{\pi_0,\pi_1\}\frac{a}{2}e^{-mc}(1-e^{-m\delta})^{k(m)}\end{aligned} \tag{10}$$

where the last inequality follows from Lemma 4. Note that

$$\begin{aligned}&\frac{1}{(\log(k(m)m))^d}\log\left(e^{-mc}(1-e^{-m\delta})^{k(m)}\right)\\&=-\frac{mc}{(m^l+\log m)^d}+\frac{e^{m^l}}{(m^l+\log m)^d}\log\left(1-e^{-m\delta}\right)\\&=-\frac{mc}{(m^l+\log m)^d}+\frac{e^{m^l-m\delta}}{(m^l+\log m)^d}\log\left(1-e^{-m\delta}\right)^{e^{m\delta}}.\end{aligned} \tag{11}$$

Since $dl > 1$ and $l < 1$, the right-hand side (RHS) of (11) converges to 0 as $m \to \infty$. Moreover, since

$$1 \le \frac{\log(k(m)m)}{\log n} \le \frac{m^l + \log m}{(m-1)^l + \log(m-1)} \to 1$$

as $m \to \infty$, we have from (10)

$$\lim_{n\to\infty} \frac{1}{(\log n)^d} \log P_e^*(n) = 0$$

which proves part (i) of the proposition.

For part (ii), the argument is the same, except that we use parts (iv) and (v) of Lemma 3 (instead of parts (ii) and (iii)), and the inequalities (6) and (7) are replaced by

$$q_{0,i} \ge \frac{a}{2} e^{-mc} \frac{1}{2^i} \quad \text{and} \quad q_{1,i} \ge \frac{a}{2} e^{-mc} \frac{1}{2^i}$$

respectively, and we let $k = m^l$ where $l \in (1/d - 1, 1)$, for $1/2 < d < 1$. The conclusion when $\mathbb{E}_j\left[\left|\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\right|^r\right] < \infty$ for some integer $r \ge 2$ can be derived similarly. □

We note that our results also apply to the case of independent, but nonidentical distributions of the sensor observations, if we assume that no sensor can perform significantly better or worse than the others. To be more specific, we assume that there exists two pairs of distributions $(\mathbb{P}_0^l, \mathbb{P}_1^l)$ and $(\mathbb{P}_0^u, \mathbb{P}_1^u)$, such that for all sensors i, and for all $s \in [0,1]$, we have

$$\mathbb{E}_0\left[\left(\frac{\mathrm{d}\mathbb{P}_1^l}{\mathrm{d}\mathbb{P}_0^l}\right)^s\right] \le \mathbb{E}_0\left[\left(\frac{\mathrm{d}\mathbb{P}_1^i}{\mathrm{d}\mathbb{P}_0^i}\right)^s\right] \le \mathbb{E}_0\left[\left(\frac{\mathrm{d}\mathbb{P}_1^u}{\mathrm{d}\mathbb{P}_0^u}\right)^s\right]$$

where $\mathbb{P}_j^i$ is the distribution of X_i under hypothesis H_j. Then the bounds in Proposition 2 still hold. The above assumption is reasonable in most practical networks consisting of nodes that are similar in nature. It is violated when we have an infinite tandem network, with each node making raw observations that are qualitatively better than the previous one. However, we believe that such a scenario is uncommon in practice.

IV. Tightness

Part (i) of Proposition 2 translates to a bound of the form $\Omega(e^{-c(\log n)^d})$, for every $d > 1$. In this section, we show that this family of bounds is tight, in the sense that it cannot be extended to values of d less than one. This is accomplished by constructing an example in which the error probability is $O(e^{-c(\log n)^d})$, with $d = 1$, i.e., the error probability is of the order $O(n^{-c})$ for some $c > 0$.

Our example involves a Gaussian hypothesis testing problem. We assume that under H_j, X_1 is distributed according to a normal distribution with mean 0 and variance σ_j^2, where $0 < \sigma_0^2 < 1/2 < \sigma_1^2$. It is easy to check that the condition in part (i) of Proposition 2 is satisfied [13].

For each n, let $a_n = \sqrt{\log n}$. Consider the following suboptimal strategy [8]: for $i \ge 1$, let

$$\gamma_i(Y_{i-1}, X_i) = \begin{cases} 0, & \text{if } X_i^2 \le a_n^2 \text{ and } Y_{i-1} = 0 \\ 1, & \text{otherwise} \end{cases}$$

where $Y_0 = 0$. Thus, the decision at sensor n is $Y_n = 1$ iff we have $X_i^2 > a_n^2$ for some $i \le n$.

Proposition 3: With the above described strategy, the probability of error is $O(n^{-c})$, for some $c > 0$.

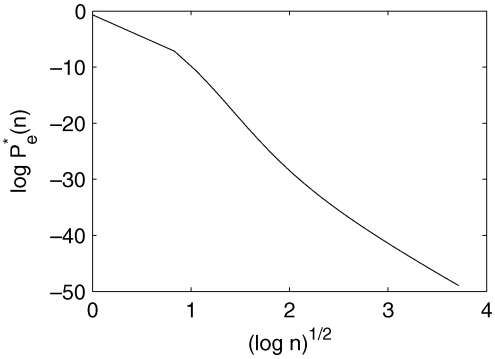

Fig. 3. A plot of the optimal error probability as a function of the number of sensors, for the problem of detecting the presence of a known signal in Gaussian noise. The optimal thresholds for the LLRQs at each sensor are given in [2]. For large n, the plot is almost linear.

Proof: Let $Q(\cdot)$ be the Gaussian complementary error function, i.e., $Q(x) = \mathbb{P}(Z \ge x)$, where Z is a standard normal random variable. We use the well-known bound $Q(x) \le \exp(-x^2/2)$ (see, e.g., [16]). The Type I error probability is given by

$$\begin{aligned} \mathbb{P}_0(Y_n = 1) &= \mathbb{P}_0(X_i^2 > a_n^2 \text{ for some } i) \\ &\le n\mathbb{P}_0(X_1^2 > a_n^2) \\ &= 2nQ\left(a_n/\sigma_0\right) \\ &\le 2ne^{-a_n^2/2\sigma_0^2} \\ &= 2n^{1-\frac{1}{2\sigma_0^2}} \end{aligned}$$

which is of the form $O(n^{-c})$, with $c > 0$.

The Type II error probability is

$$\begin{aligned} \mathbb{P}_1(Y_n = 0) &= \left(\mathbb{P}_1(X_1^2 \le a_n^2)\right)^n \\ &= \left(1 - \mathbb{P}_1(X_1^2 > a_n^2)\right)^n \\ &\le e^{-n\,\mathbb{P}_1(X_1^2 > a_n^2)}. \end{aligned} \tag{12}$$

From the lower bound $Q(x) \ge \frac{1}{x\sqrt{2\pi}}(1 - \frac{1}{x^2}) \exp(-x^2/2)$ (see [16]), we have

$$\begin{aligned} n\mathbb{P}_1(X_1^2 > a_n^2) &= 2nQ\left(a_n/\sigma_1\right) \\ &\ge \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma_1}{a_n}\left(1 - \frac{\sigma_1^2}{a_n^2}\right)\exp\left(-a_n^2/2\sigma_1^2\right)n \\ &= \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma_1}{\sqrt{\log n}}\left(1 - \frac{\sigma_1^2}{\log n}\right)n^{1-\frac{1}{2\sigma_1^2}} \\ &= \Omega(n^{d_1}) \end{aligned}$$

where $d_1 > 0$. From (12), we obtain that $\mathbb{P}_1(Y_n = 0) = O(\exp(-n^{d_1}))$. Hence, the overall error probability is dominated by the Type I error probability, and this strategy achieves a decay rate of n^{-c} for some positive constant c. □

We note that in most cases, the rate n^{-c} is not achievable. For example, suppose that under H_0, the distribution of X_1 is normal with mean $-\mu$ and variance 1, while under H_1, the distribution is normal with mean μ and variance 1. A numerical computation indicates that the optimal error probability decay is of the order $\exp(-c\sqrt{\log n})$ (see [2] and Fig. 3). Finding the exact decay rate analytically for particular pairs of distributions seems to be difficult because there is no closed-form

solution for the optimal thresholds used in the LLRQ decision rule at each sensor [8], except for distributions with certain symmetric properties [2].

V. Conclusion

In this correspondence, we have shown that, in Bayesian decentralized detection, using a long tandem of sensors, the rate of decay of the error probability is subexponential. We also provided lower bounds for the rate of error decay, under additional mild assumptions on the distributions.

In our model, we have assumed binary communication between sensors, and we have been concerned with a binary hypothesis testing problem. The question of whether k-valued messages (with $k > 2$) will result in a faster decay rate, or even an exponential decay rate, remains open. In the case of m-ary hypothesis testing using a tandem network where each sensor observation is a Bernoulli random variable, [6] shows that using $(m+1)$-valued messages is necessary and sufficient for the error probability to decrease to 0 as n increases. However, it is unknown what the decay rate is.

We finally note that under a Neyman–Pearson formulation, the picture is less complete. We are able to show the subexponential decay of the Type II error probability, but only for a myopic sensor strategy [13]. The case of general strategies is an interesting open problem.

References

[1] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, Jul. 1988.

[2] P. Swaszek, "On the performance of serial networks in distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 254–260, Jan. 1993.

[3] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part I—Tandem structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1044–1059, Sep. 1991.

[4] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, no. 3, pp. 828–835, 1969.

[5] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 765–782, 1970.

[6] J. Koplowitz, "Necessary and sufficient memory size for m-hypothesis testing," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 44–46, Jan. 1975.

[7] B. Chandrasekaran and C. C. Lam, "A finite-memory deterministic algorithm for the symmetric hypothesis testing problem," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 40–44, Jan. 1975.

[8] J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, May 1992.

[9] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr., Signals, Syst.*, vol. 1, pp. 167–182, 1988.

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

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

[12] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4155–4168, Sep. 2008.

[13] W. P. Tay, "Decentralized Detection in Resource-Limited Sensor Network Architectures," Ph.D. dissertation, MIT, Cambridge, MA, 2007.

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

[15] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer-Verlag, 1998.

[16] R. Durrett, *Probability: Theory and Examples*, 2nd ed. New York: Duxbury, 1995.

New Binary Sequences With Optimal Autocorrelation Magnitude

Nam Yul Yu, *Member, IEEE*, and Guang Gong, *Senior Member, IEEE*

***Abstract*—New binary sequences of period $N = 4(2^m - 1)$ for even $m \geq 4$ are found, where the sequences are described by a $4 \times (2^m - 1)$ array structure. The new sequences are almost balanced and have four-valued autocorrelation, i.e., $\{N, 0, \pm 4\}$, which is optimal with respect to autocorrelation magnitude. The complete autocorrelation distribution and the exact linear complexity of the sequences are mathematically derived. Finally, it is shown that the sequences are implemented by a combination of linear feedback shift registers and a simple logic.**

***Index Terms*—Binary sequences, interleaved sequences, linear complexity, optimal autocorrelation.**

I. Introduction

Binary pseudorandom sequences with optimal autocorrelation play important roles in many areas of communication and cryptography. In code-division multiple-access (CDMA) communication systems, the sequences are needed to acquire the accurate timing information of received signals. In cryptography, on the other hand, the sequences are employed to generate key streams in stream cipher encryptions. Therefore, lots of attention have been paid to binary sequences with optimal autocorrelation. More details on the sequences will be discussed in Section II.

For a binary sequence of period $N \equiv 0 \pmod 4$, the autocorrelation $\{N, 0, -4\}$ or $\{N, 0, 4\}$ is optimal in the sense that it has the two out-of-phase values with the smallest magnitudes [17]. If we allow three out-of-phase values with the smallest magnitudes, then the best autocorrelation should be $\{N, 0, \pm 4\}$, where the autocorrelation is *optimal* with respect to its magnitude. In practical applications, it has the same meaning as conventional optimal autocorrelation. Consequently, the autocorrelation of $\{N, 0, \pm 4\}$ is also considered as optimal in this correspondence.

In [13], Gong introduced the interleaved structure of sequences that is indeed a good method not only for understanding a sequence structure, but also for constructing new sequences of an interleaved form [13], [14]. In this correspondence, we show that binary sequences of period $4v$ with optimal autocorrelation shown in [1] can be represented by a $v \times 4$ interleaved structure. We also show that a binary product sequence [19] of period $4v$ with optimal autocorrelation can be represented as a $4 \times v$ interleaved structure. Inspired by these interpretations, we discover a new construction of binary sequences of period $N = 4(2^m - 1)$ with the four-valued autocorrelation $\{N, 0, \pm 4\}$ by the interleaved method. In details, we use a $4 \times (2^m - 1)$ interleaved structure defined by the perfect binary sequence of period 4 and a binary m-sequence of period $2^m - 1$. In the interleaved structure, a sequence defined over $\mathbb{Z}_4$ is used as a shift sequence. The new sequences are almost balanced, i.e., the difference between the numbers of zeros

Manuscript received August 17, 2006; revised August 22, 2006. Current version published September 17, 2008. This work was supported by NSERC SPG Grant. The material in this correspondence was presented in part at the IEEE Information Theory Workshop, Chengdu, China, October 2006.

N. Yu is with the Department of Electrical Engineering, Lakehead University, Thunder Bay, ON P7B 5E1 Canada (e-mail: nyu1@lakeheadu.ca).

G. Gong is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: ggong@calliope.uwaterloo.ca).

Communicated by A. Canteaut, Associate Editor for Complexity and Cryptography.

Digital Object Identifier 10.1109/TIT.2008.928999