ON THE SUB-EXPONENTIAL DECAY OF DETECTION ERROR PROBABILITIES IN LONG TANDEMS

Wee-Peng Tay, John N. Tsitsiklis, and Moe Z. Win

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

We consider the problem of decentralized binary hypothesis testing in a network of sensors arranged in a tandem. We show that the rate of error probability decay is always sub-exponential, establishing the validity of a long-standing conjecture. Under the additional assumption of bounded Kullback-Leibler 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, tandem, serial network, error exponent, tree network.

1. INTRODUCTION

Consider a tandem network, as shown in Figure 1, with n sensors, each sensor i observing a random variable X_i, taking values in $\mathcal{X}$. Under hypothesis H_j, $j = 0, 1$, X_i has marginal law $\mathbb{P}_j$, and all the X_i are independent. Sensor i is constrained to sending a 1-bit message Y_i to sensor $i+1$, of the form $Y_i = \gamma_i(Y_{i-1}, X_i)$ (Y_0 can be defined to be always 0), where $\gamma_i : \{0,1\} \times \mathcal{X} \mapsto \{0,1\}$. 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. 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.

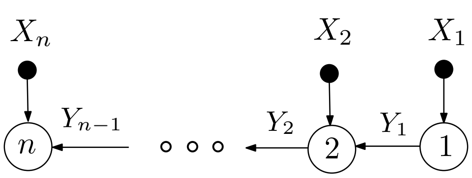

Fig. 1. A tandem network.

This research was supported, in part, by the National Science Foundation under contracts ECS-0426453, ANI-0335256 and ECS-0636519, the Charles Stark Draper Laboratory Reduced Complexity UWB Communication Techniques Program, an Office of Naval Research Young Investigator Award N00014-03-1-0489, and DoCoMo USA Labs.

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, 8] find necessary and sufficient conditions under which the error probability goes to zero in the limit of large n. To be specific, the error probability stays bounded away from zero iff there exists a $B < \infty$ such that $|\log \frac{d\mathbb{P}_1}{d\mathbb{P}_0}| \le B$ almost surely. 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 sub-exponential. However, a proof is not available. The goal of this paper is to prove this conjecture.

We first note that there is a caveat to the sub-exponential decay conjecture: the probability measures $\mathbb{P}_0$ and $\mathbb{P}_1$ need to be equivalent, i.e., absolutely continuous 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 a $X_m \in A$, whereupon all sensors $i \ge m$ declare 0. For this reason, we assume throughout the paper that the measures $\mathbb{P}_0$ and $\mathbb{P}_1$ are equivalent. Under this assumption, we show that

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

When the error probability goes to zero, we would also like to quantify the best possible (sub-exponential) decay rate. In this spirit, we find lower bounds on the probability of error, under the further assumption of bounded Kullback-Leibler (KL) divergences. In particular, we show that for any $d > 1/2$, and some positive constant c, the error probability is $\Omega(e^{-cn^d})$.[1] Under some further mild assumptions, which are valid in most practical cases of interest, we establish the bound $\Omega(e^{-c(\log n)^d})$ for all $d > 1$, and show that it is tight.

The rest of the paper is organized as follows. In Section 2, we show that the error probability decays sub-exponentially. In Section 3, we derive more detailed lower bounds on the error probabilities. In

[1] If f and g are nonnegative functions on the nonnegative integers, we write $f(n) = \Omega(g(n))$ if there exists a K such that $f(n) \ge Kg(n)$ for all n sufficiently large.

Section 4, we establish tightness of one of our lower bounds. Finally, Section 5 contains concluding remarks.

2. SUB-EXPONENTIAL DECAY

In this section we show that the rate of decay of the error probability is always sub-exponential. 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 sub-exponential, with numerical examples as supporting evidence.

Let $L_i = \log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(X_i)$ be the log-likelihood ratio associated with the observation made by sensor i. From [1, 8, 10, 12], 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., a rule of the form

$$Y_i = \begin{cases} 0, & \text{if } L_i \leq t_{i,n}(y), \\ 1, & \text{otherwise}, \end{cases} \tag{1}$$

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 a LLRQ. Moreover, the existence results from [12]), together with Proposition 4.2 in [10] gives us the following result.

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

In view of Lemma 1, we can restrict to strategies of the form

$$\gamma_i(Y_{i-1}, X_i) = \begin{cases} 0, & \text{if } L_i \leq t_{i,n}(1), \\ 1, & \text{if } L_i > t_{i,n}(0), \\ Y_{i-1}, & \text{otherwise}, \end{cases}$$

where $t_{i,n}(1) \leq t_{i,n}(0)$. Note that this is the type of strategies used in [4] to show that the error probability converges to zero.

Proposition 1. *The rate of decay of the error probability in a tandem network is sub-exponential, 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. We have, for all i,

$$\begin{aligned} \mathbb{P}_0(Y_i = 1) &= \mathbb{P}_0(L_i > t_{i,n}(0))\mathbb{P}_0(Y_{i-1} = 0) \\ &\quad + \mathbb{P}_0(L_i > t_{i,n}(1))\mathbb{P}_0(Y_{i-1} = 1) \qquad (2) \\ \mathbb{P}_1(Y_i = 0) &= \mathbb{P}_1(L_i \leq t_{i,n}(0))\mathbb{P}_1(Y_{i-1} = 0) \\ &\quad + \mathbb{P}_1(L_i \leq t_{i,n}(1))\mathbb{P}_1(Y_{i-1} = 1) \qquad (3) \end{aligned}$$

From (2) and (3), with $i = n$, and applying Lemma 1, we have

$$\begin{aligned} P_e^*(n) &= \pi_0\mathbb{P}_0(Y_n = 1) + \pi_1\mathbb{P}_1(Y_n = 0) \\ &= \pi_0\Big(\mathbb{P}_0(L_n > t_{n,n}(0)) + \mathbb{P}_0(t_{n,n}(1) < L_n \leq t_{n,n}(0)) \\ &\qquad \cdot \mathbb{P}_0(Y_{n-1} = 1)\Big) \\ &\quad + \pi_1\Big(\mathbb{P}_1(L_n \leq t_{n,n}(1)) + \mathbb{P}_1(t_{n,n}(1) < L_n \leq t_{n,n}(0)) \\ &\qquad \cdot \mathbb{P}_1(Y_{n-1} = 0)\Big) \qquad (4) \\ &\geq \min_{j=0,1} \mathbb{P}_j\big(t_{n,n}(1) < L_n \leq t_{n,n}(0)\big)P_e^*(n-1) \qquad (5) \end{aligned}$$

From (4), 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, we also 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 (5), the error probability cannot decay exponentially fast. □

We have established that the decay of the error probability is sub-exponential. This confirms that the parallel configuration performs much better than the tandem configuration when n is large. It now remains to investigate the best performance that a tandem configuration can possibly achieve. In the next section, we use a more elaborate technique to derive a lower bound for the error probability. Due to space limitations, some of the proofs are omitted, or only sketched; they can be found in [13].

3. RATE OF DECAY

In this section, we show that under the assumption of bounded 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 $d > 1$. The ideas in this section are inspired by the methods in [1] and [14]. In particular, we rely on a sequence of comparisons of the tandem configuration with other tree configurations, whose performance can be quantified using the methods of [14].

Our results involve the KL divergences, defined by

$$\begin{aligned} \bar{x}_0 &= \mathbb{E}_0\big[\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\big], \\ \bar{x}_1 &= \mathbb{E}_1\big[\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\big]. \end{aligned}$$

We assume that $-\infty < \bar{x}_0 < 0 < \bar{x}_1 < \infty$ 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 Figure 1, with n sensors, where each sensor i obtains a single observation X_i; (ii) a modified tandem network $T(k, m)$, as in Figure 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 successor. 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.

Therefore, it suffices to establish a lower bound for the error probability in the network $T(k, m)$. Towards this goal, we will use some standard results in Large Deviations Theory, notably Cramér's Theorem [15], as stated in the lemma below.

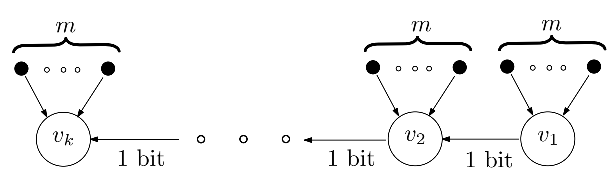

Fig. 2. A modified tandem network $T(k, m)$ that outperforms a tandem network with $n = km$ sensors.

Lemma 2. *Suppose that $-\infty < \bar{x}_0 < 0 < \bar{x}_1 < \infty$, and that $X_1, X_2, \ldots$ are 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 $\Lambda_j^*(t) = \sup_{\xi\in\mathbb{R}}\{\xi t - \log \mathbb{E}_j\big[\big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\big)^{\xi}\big]\}$.*

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

$$\mathbb{P}_0(S_m/m > \bar{x}_1 + \epsilon) \geq a e^{-mc},$$
$$\mathbb{P}_1(S_m/m \leq \bar{x}_0 - \epsilon) \geq a e^{-mc}.$$

(ii) 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 $\Lambda_1^*(\bar{x}_1 + \epsilon) > 0$, and

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

(iii) 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 $\Lambda_0^*(\bar{x}_0 - \epsilon) > 0$, and

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

(iv) Suppose that $\mathbb{E}_1\big[\big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\big)^s\big] = \infty$ for all $s > 0$. Then, for every $\epsilon > 0$, there exists some $M \geq 1$ such that for all $m \geq M$,

$$\mathbb{P}_1(S_m/m \leq \bar{x}_1 + \epsilon) \geq 1/2.$$

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

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

(v) Suppose that $\mathbb{E}_0\big[\big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\big)^s\big] = \infty$ for all $s < 0$. Then, for every $\epsilon > 0$, there exists some $M \geq 1$ such that for all $m \geq M$,

$$\mathbb{P}_0(S_m/m > \bar{x}_0 - \epsilon) \geq 1/2.$$

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

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

Proof. Omitted for brevity. □

We now state our main result. Note that the condition in part (i) of the proposition below implies that $\mathbb{E}_j\big[\big|\log \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\big|^r\big] < \infty$ for all r, but the reverse implication is not always true.

Proposition 2. *Suppose that $-\infty < \bar{x}_0 < 0 < \bar{x}_1 < \infty$.*

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

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

for all $d > 1$.

(ii) Suppose that $\mathbb{E}_0\big[\big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\big)^s\big] = \infty$ either for all $s > 1$ or for all $s < 0$. Then

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

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

Proof. (Outline) For part (i), we lower bound the error probability of a tandem network by the error probability of a modified network $T(k(m), m)$, where $k(m) = \exp(m^l)$, and $l \in (1/d, 1)$. Then, applying Lemma 2, we get the desired conclusion. A similar argument works for part (ii). □

4. TIGHTNESS

Part (i) of Proposition 2 gives 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 $0 < \sigma_0^2 < 1/2 < \sigma_1^2$. We first check that the condition in part (i) of Proposition 2 is satisfied. We have

$$\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(x) = \frac{\sigma_0}{\sigma_1} e^{-\frac{x^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)},$$

and (using the formula for the moment generating function of a χ^2 distribution),

$$\begin{aligned}\mathbb{E}_0\Big[\Big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\Big)^s\Big] &= \Big(\frac{\sigma_0}{\sigma_1}\Big)^s \mathbb{E}_0\big[e^{\frac{s}{2}\left(1-\sigma_0^2/\sigma_1^2\right)(X_1/\sigma_0)^2}\big] \\ &= \Big(\frac{\sigma_0}{\sigma_1}\Big)^s \Big(\frac{1}{1 - s\big(1-\sigma_0^2/\sigma_1^2\big)}\Big)^{1/2} < \infty,\end{aligned}$$

if $s < 1/(1-\sigma_0^2/\sigma_1^2)$. Hence, the condition in part (i) of Proposition 2 is satisfied.

Fix some n and let $a_n = \sqrt{\log n}$. We analyze the rate of decay of error probability of a particular sub-optimal strategy considered in [8], which is the following:

$$\gamma_1(X_1) = \begin{cases} 0, & \text{if } X_1^2 \leq a_n^2, \\ 1, & \text{otherwise,} \end{cases}$$

and for $i \geq 2$,

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

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

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

Proof. Omitted for brevity. □

We note that in most cases, the rate n^{-c} is not achievable. For example, consider the more common case of detecting the presence of a known signal in Gaussian noise: 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 Figure 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].

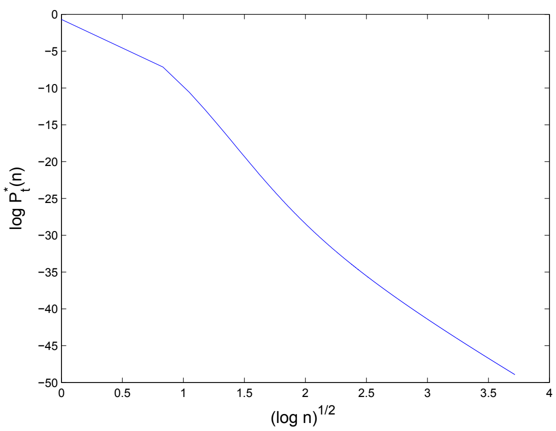

Fig. 3. A plot of the optimal error probability as a function of the number of sensors, for the problem of detecting the presence of a known signal in Gaussian noise. The optimal thresholds for the LL-RQs at each sensor are given in [2]. For large n, the plot is almost linear.

5. CONCLUSION

In this paper, we have shown that, in Bayesian decentralized detection, using a long tandem of sensors, the rate of decay of the error probability is sub-exponential. In order to obtain more precise bounds, we introduced a modified tandem network, which outperforms the original one, and used tools from Large Deviations Theory. Under the assumption of bounded KL divergences, we have shown that the error probability is $\Omega(e^{-cn^d})$, for all $d > 1/2$. Under the further assumption that the moments (under H_0) of order s of the likelihood ratio are finite for all s in an interval that contains $[0, 1]$ in its interior, we have shown that the lower bound can be improved to $\Omega(e^{-c(\log n)^d})$, for all $d > 1$, and that this latter bound is tight.

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. Nevertheless, we conjecture that the error decay rate is always sub-exponential.

6. 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, 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, 1993.

[3] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: part I- tandem structures," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 21, no. 5, pp. 1044–1059, 1991.

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

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

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

[7] B. Chandrasekaran and C. C. Lam, "A finite-memory deterministic algorithm for the symmetric hypothesis testing problem," *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 40–44, 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, 1992.

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

[10] ——, "Decentralized detection," *Advances in Statistical Signal Processing*, 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, pp. 54–63, 1997.

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

[13] W.-P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection probabilities in long tandems," Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Tech. Rep. LIDS-P-2746, Feb. 2007.

[14] ——, "Data fusion trees for detection: Does architecture matter?" Proc. 44th Allerton Annual Conference on Communication, Control, and Computing, 2006.

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