

A LARGE DEVIATION ANALYSIS OF DETECTION OVER MULTI-ACCESS CHANNELS WITH RANDOM NUMBER OF SENSORS

Animashree Anandkumar and Lang Tong

ABSTRACT

We consider the problem of distributed detection over a multi-access channel. Assuming a random number of sensors transmitting their observations using Type Based Multiple Access, we derive the detection performance using Large Deviations Principle as the mean number of sensors goes to infinity. We characterize the performance in terms of error exponents. We provide comparison with the case when the number of sensors is deterministic. We generalize this scheme to multiple collections, propose a Minimum Sum-Rate detector and characterize its error exponents.

1. INTRODUCTION

In the classical setting of distributed detection, sensors in the field sense certain physical phenomenon and transmit their observations to a fusion center, which makes decisions on the underlying phenomenon, with the transmissions assumed to be perfect. However, for large wireless networks this assumption is not valid. Moreover, bandwidth has to be allocated to sensors; transmissions must be made energy-efficient. Therefore, Medium Access Control becomes a crucial component. To this end, well-known deterministic scheduling schemes such as Time-Division Multiple Access (TDMA) may not be appropriate.

It is thus desirable to consider distributed scheduling techniques that facilitate the effective delivery of information from a random number of sensors to the fusion center. We couple the so-called Type-Based Multiple Access (TBMA) [1, 2] with a simple random access analogous to the ALOHA protocol. Referred to as Type-Based Random Access (TBRA), sensors transmit probabilistically using a set of orthogonal waveforms keyed to their measurement.

1.1. Related Work

The problem of classical distributed detection has been dealt in considerable detail [3]. Detection on wireless sensor networks [4, 5] assumed orthogonal schemes like TDMA, FDMA or CDMA. TBMA was proposed as a multi-access scheme by Mergen and Tong [6] and by Liu and Sayeed [2], independently.

In [6, 7, 2], it has been shown that, in the presence of conditionally i.i.d. data and identical channel gains, the asymptotic performance of TBMA is same as the case when fusion center has direct access to sensor observations. In [1, 8], an error exponent analysis of the TBMA scheme with i.i.d. non-zero mean channel gains and conditionally i.i.d. data was given. It was shown that the error probabilities decay exponentially with number of transmitting sensors. A Minimum Rate detector was proposed, which gives the best error

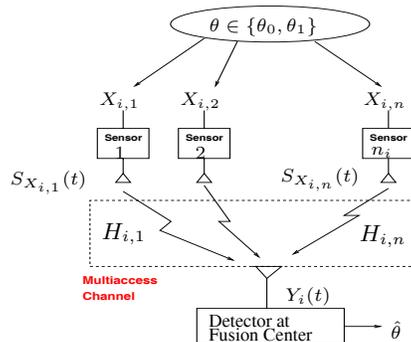


Fig. 1. Distributed Detection in Multiaccess, given $N_i = n$.

exponents for TBMA scheme. However, when channels are zero-mean fading, the Minimum Rate Detector was shown to give poor performance. In [9, 10] we present the analysis of TBRA involving multiple collections with zero mean fading and show the existence of an optimal activation strategy that maximizes the error exponents.

1.2. Summary of Results

We consider two scenarios involving Poisson number of transmitting sensors N : (a) Fusion Center has access to the realization of N . (b) Fusion Center does not know the realization of N . For both the cases, we prove that the normalized matched filter output at fusion center satisfies Large Deviations Principle as the mean number of sensors goes to infinity. We derive expressions for the rate functions in both the cases. We study the performance of Minimum Rate Detector in our setup and prove its asymptotic optimality in terms of error exponents. We give comparisons with the case when number of sensors is deterministic and demonstrate it with numerical examples. We also address issues such as scalability and computational requirement at the fusion center. We give a generalization of our scheme to multiple collections. We suggest a Minimum Sum-Rate detector and characterize its error exponents. Given the constraint of total expected transmissions, we analyze whether single or multiple collections is advantageous. When mean number of sensors per collection is large, there is no difference in performance.

2. SYSTEM MODEL

Sensor Observations and Hypothesis: In the i^{th} data collection, a sensor involved in the transmission¹, say sensor j , has measurement $X_{i,j} \in \{1, \dots, k\}$ quantized to k levels. We assume that $X_{i,j}$ are conditionally i.i.d across time and sensors with PMF \mathbf{p}_θ i.e.,

$$X_{i,j} \stackrel{\text{i.i.d.}}{\sim} \mathbf{p}_\theta = (p_\theta(1), \dots, p_\theta(k)), \quad \theta \in \{\theta_0, \theta_1\}.$$

We consider Simple Binary Hypothesis \mathcal{H}_i s.t,

$$\mathcal{H}_0 : \theta = \theta_0 \text{ vs. } \mathcal{H}_1 : \theta = \theta_1.$$

¹Without loss of generality, we will only consider those sensors involved in the transmission.

A.Anandkumar and L.Tong are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853. Email: {aa332@, ltong@ece.}cornell.edu.

This work is supported in part by the National Science Foundation under Contract CCR-0311055. Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011.

TBRA: In the i^{th} collection, node j may transmit probabilistically (possibly by flipping a coin). When it decides to transmit, it encodes $X_{i,j}$ to a certain waveform and transmits it over a multiaccess fading channel. As in TBMA, a set of k orthonormal waveforms $\{\phi_i(t), i = 1, \dots, k\}$ are used, each corresponding to a specific data value. Specifically, given energy constraint E per sensor transmission, the signal transmitted by sensor j in collection i is given by,

$$S_{i,j}(t) = \sqrt{E}\phi_{X_{i,j}}(t).$$

Channel Characteristics: We assume that the channels between the sensors and fusion center are flat fading with i.i.d non zero-mean channel gains² $H_{i,j} \in \mathbb{R}$ with mean μ_H . Noise $W_i(t)$ is AWGN process with p.s.d σ^2 . If N_i is the random number of sensors transmitting at collection i , the fusion center receives waveform,

$$Y_i(t) = \sum_{j=1}^{N_i} H_{i,j} S_{X_{i,j}}(t) + W_i(t). \quad (1)$$

Matched Filter Output: The matched filter output \mathbf{Y}_i is the inner product between the received signal $Y_i(t)$ and the orthogonal waveforms $\phi_1(t), \dots, \phi_k(t)$. Let $\mathbf{e}_1, \dots, \mathbf{e}_k$ be standard basis vectors and $\mathbf{W}_i \stackrel{iid}{\sim} \mathcal{N}(0, \frac{\sigma^2}{E}\mathbf{I})$. Then,

$$\begin{aligned} \mathbf{Y}_i &\triangleq \frac{1}{\sqrt{E}} \left[\langle Y_i(\cdot), \phi_1(\cdot) \rangle, \dots, \langle Y_i(\cdot), \phi_k(\cdot) \rangle \right]. \\ &= \sum_{j=1}^{N_i} H_{i,j} \mathbf{e}_{X_{i,j}} + \mathbf{W}_i. \end{aligned} \quad (2)$$

Poisson Number of sensors: We model the number of sensors involved in each transmission N_i as i.i.d Poisson with mean λ . There are several reasons to consider random access. The sensors may use a simple probabilistic wake-up strategy in which a sensor decides to participate in transmission based on a simple coin-flip. Another possibility is that the fusion center is a mobile access point, and it travels to different regions to collect data.

Error Exponents: For a given decision rule at the fusion center, let $\alpha \triangleq \mathbb{P}\{\mathcal{H}_0 \rightarrow \mathcal{H}_1\}$ and $\beta \triangleq \mathbb{P}\{\mathcal{H}_1 \rightarrow \mathcal{H}_0\}$ denote the Type-I/II error probabilities. Let η_1 and η_2 denote the error exponents.³

$$\eta_1 = -\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \alpha, \quad \eta_2 = -\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \beta. \quad (3)$$

Using the “worst exponent wins” rule, the exponent of error probability under Bayesian setting is given by $\min(\eta_1, \eta_2)$.

Decision Statistic and Type: In Information Theory, *type* or empirical distribution of a sequence is the relative proportion of occurrences of each symbol from a finite alphabet [11]. In the TBMA setup with fixed number of sensors *i.e.*, $N_i \equiv \lambda$, when noise and fading are absent ($H_{i,j} \equiv 1$ and $\sigma \equiv 0$),

$$\frac{\mathbf{Y}_i}{\lambda} = \frac{1}{\lambda} \sum_{j=1}^{\lambda} \mathbf{e}_{X_{i,j}}. \quad (4)$$

The normalized matched filter output is the scaled histogram or type of sensor observation. Hence we have the name Type Based Multiple Access.

²The results of this paper can be generalized to complex-valued channel gains with minor changes.

³Throughout the paper, the notation \log refers to natural logarithm.

3. MINIMUM RATE DETECTOR

We shall focus in this section and next section, on the single collection model. For ease of notation, we drop the time index i in (2).

$$\mathbf{Y} = \sum_{j=1}^N H_j \mathbf{e}_{X_j} + \mathbf{W}. \quad (5)$$

In this section, we recap results involving fixed number of sensors *i.e.*, $N \equiv \lambda$. We know that the optimal detector at fusion center has the form of a Likelihood Ratio detector. However, the computation of likelihood function for the normalized matched filter output is generally intractable for TBRA setup. Instead we study the asymptotic nature of random vector \mathbf{Y} in (5), using Large Deviations Principle (LDP). LDP characterizes the probability of large excursions of \mathbf{Y} from its “mean” behavior by quantifying the so-called rate function $I(\cdot)$ [12]. In essence for \mathbf{Y} satisfying LDP,⁴

$$\Pr\{\mathbf{Y} \in B_\epsilon(\mathbf{y})\} = e^{-\lambda(I(\mathbf{y}) + O(\epsilon)) + o(\lambda)} \doteq e^{-\lambda I(\mathbf{y})}. \quad (6)$$

where $\mathbf{y} \in \mathbb{R}^k$ and $B_\epsilon(\mathbf{y})$ is a open ball in \mathbb{R}^k centered at \mathbf{y} with radius $\epsilon > 0$. Note that $O(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ and $\frac{o(\lambda)}{\lambda} \rightarrow 0$ as $\lambda \rightarrow \infty$.

In other words for large λ , the probability that \mathbf{Y} turns out to be in the close vicinity of \mathbf{y} behaves as $e^{-\lambda I(\mathbf{y})}$. Let the rate function of \mathbf{Y} under hypothesis \mathcal{H}_i be $I_i(\cdot)$. Then the decision regions of Minimum Rate Detector are given by,

$$\Gamma_0 = \{\mathbf{y} \in \mathbb{R}^k : I_0(\mathbf{y}) \leq I_1(\mathbf{y})\}, \quad \Gamma_1 = \mathbb{R}^k \setminus \Gamma_0. \quad (7)$$

Thus the Minimum Rate Detector decides that \mathcal{H}_0 is true if the asymptotic likelihood under \mathcal{H}_0 is higher *i.e.*, $e^{-\lambda I_0(\mathbf{y})} \geq e^{-\lambda I_1(\mathbf{y})}$.

Theorem 1 [1, 8] *For fixed number of sensors λ , the normalized matched filter output $\frac{\mathbf{Y}}{\lambda}$ satisfies LDP as $\lambda \rightarrow \infty$ with rate functions $I_i(\cdot)$, under hypothesis \mathcal{H}_i .*

Then the Minimum Rate Detector maximizes $\min(\eta_1, \eta_2)$ amongst all detectors based on $\frac{\mathbf{Y}}{\lambda}$ and thus achieves the best exponent of error probability in the Bayesian setting for TBMA scheme. Its error exponents are given by,

$$\eta_1 = \inf_{x \in \Gamma_1} I_0(x), \quad \eta_2 = \inf_{x \in \Gamma_0} I_1(x). \quad (8)$$

4. SINGLE COLLECTION

Consider the following cases with Poisson number of sensors.

1. Fusion Center has access to the realization $N = n$.
2. Fusion Center does not know the realization $N = n$.

From Theorem 1, Minimum Rate Detector gives optimal error exponents when LDP holds. Thus the key issue is establishing LDP for the above two cases as the mean number of sensors goes to infinity. We also need to characterize their respective rate functions.

4.1. N is Random and Known

This is a reasonable assumption if the fusion center can get a good estimate of N (eg : identical channels and low noise variance). This scenario is intermediate between deterministic and random & unknown N . By studying this case, we can characterize the usefulness of knowing the realization of N . We first prove that LDP holds and give expressions for rate functions.

⁴“ \doteq ” is used in the same sense as in (6) throughout the paper

Theorem 2 (Characterization of Rate Function): $\frac{Y}{N}$ satisfies Large Deviations Principle for integral λ and $\lambda \rightarrow \infty$, with rate function $I_i^{rk}(\cdot)$ under hypothesis \mathcal{H}_i ,

$$I_i^{rk}(\mathbf{y}) = 1 - e^{-I_i^d(\mathbf{y})}.$$

where, $I_i^d(\cdot)$ is the rate function for deterministic number of sensors,

$$I_i^d(\mathbf{y}) = \sup_{\mathbf{t}} (\langle \mathbf{y}, \mathbf{t} \rangle - \log \phi_i(\mathbf{t})). \quad (9)$$

where $\phi_i(\mathbf{t}) = \mathbb{E}_{\theta_i}(e^{\langle \mathbf{t}, H_1 \mathbf{e}_{x_1} \rangle})$ is the moment generating function of a single “faded” observation.

Proof: We use the fact that, we have LDP w.r.t n , given $N = n$, with rate function $I_i^d(\cdot)$ i.e.,

$$\mathbb{P}\left(\frac{Y}{N} \in \mathcal{B}_\epsilon(\mathbf{y}) \mid N = n; \theta_i\right) \doteq e^{-nI_i^d(\mathbf{y})}.$$

We marginalize w.r.t N to obtain the likelihood function of $\frac{Y}{N}$. Using Stirling’s formula and the principle of logarithmic equivalence [13], we obtain the rate function. ■

Given $N = n$, the Minimum Rate Detector makes a decision based on $\frac{Y}{n}$ using the deterministic rate function $I_i^d(\cdot)$. Therefore the decision regions are the same as in deterministic case of n sensors. But to compute the error probabilities, we need to marginalize w.r.t N . We obtain the error exponents by substituting $I_i^{rk}(\cdot)$ in (8).

Since we can compute $\frac{Y}{n}$, TBRA delivers *type* of sensor observations in the absence of fading and noise, as in the deterministic case. However, we see that the rates of decay of likelihood functions and error probabilities are slower than in the deterministic case i.e., $I_i^{rk}(\cdot) \leq I_i^d(\cdot)$ and $\eta^{rk} \leq \eta^d$. Intuitively this is because of randomization in N which implies that for a given λ , every realization of N is not guaranteed to be large.

4.2. N is Random and Unknown

This scenario is easy to implement in practice. However, now the decision statistic is $\frac{Y}{\lambda}$ instead of $\frac{Y}{n}$. Thus even in the absence of noise and fading, TBRA does not deliver *type* of sensor observations. Intuitively we expect some performance degradation. However, we show that this degradation is only in the rate function and LDP still holds as $\lambda \rightarrow \infty$.

Theorem 3 (Characterization of Rate Function): $\frac{Y}{\lambda}$ satisfies Large Deviations Principle for integral λ and $\lambda \rightarrow \infty$, with rate function $I_i^{ru}(\cdot)$ under hypothesis \mathcal{H}_i ,

$$I_i^{ru}(\mathbf{y}) = \sup_{\mathbf{t} \in \mathbb{R}^k} (\langle \mathbf{y}, \mathbf{t} \rangle - \phi_i(\mathbf{t})) + 1 \quad (10)$$

where, $\phi_i(\mathbf{t}) = \mathbb{E}_{\theta_i}(e^{\langle \mathbf{t}, H_1 \mathbf{e}_{x_1} \rangle})$.

Proof: We compute the asymptotic log-moment generating function of $\frac{Y}{\lambda}$ and apply the Gärtner-Ellis Theorem [12]. ■

Computation of the Rate Function

The rate functions are unconstrained concave optimization problems in k -dimensions. They may not have closed form expressions, but can be evaluated numerically. This is much simpler than evaluating the actual likelihood function. We see that for random and unknown case, the optimization problem is separable as the rate function can be written as,

$$I_i^{ru}(\mathbf{y}) = \sup_{\mathbf{t}} \sum_{j=1}^k (y_j t_j - p_{\theta_i}(j) \phi_H(t_j)) + 1.$$

where $\phi_H(\cdot)$ is the MGF of fading distribution. Since the optimization is unconstrained, it can be reduced to k independent 1-dimensional optimizations. This is much simpler to compute than for the deterministic case, which is non-separable.

5. PERFORMANCE ANALYSIS

For different cases of N , the error exponents can be obtained by substituting their rate functions in (8). We give the expressions for rate functions in the table below.

Nature of N	Rate Function
Deterministic	$I_i^d(\mathbf{y}) = \sup_{\mathbf{t}} (\langle \mathbf{y}, \mathbf{t} \rangle - \log \phi_i(\mathbf{t}))$
Random & Known	$I_i^{rk}(\mathbf{y}) = 1 - e^{-I_i^d(\mathbf{y})}$
Random & Unknown	$I_i^{ru}(\mathbf{y}) = \sup_{\mathbf{t}} (\langle \mathbf{y}, \mathbf{t} \rangle - \phi_i(\mathbf{t}) + 1)$

We note that since σ^2 is not present in the rate functions, noise does not affect the error exponents. This is because the large deviation probability of the normalized noise term, $\tilde{W} \sim \mathcal{N}(0, \frac{\sigma^2}{En^2})$ decays super-exponentially in all the above cases.

Minimum Rate Detector is asymptotically optimal and gives the best error exponents for the TBRA scheme. This is because we have Large Deviations in all the above cases and assume the rate functions to be I-continuous in the interior of their domains.

Scalability: Consider the practical scenario of sensor duty-cycle. To increase λ we just need to increase the wake-up probability of sensors. Moreover on increasing λ , the energy distribution (battery power) is uniform among sensors, on the average. This is in contrast to the deterministic case, where to improve performance, new sensors have to be deployed in the field and the resulting energy distribution among the sensors is non-uniform.

Numerical Example : On-Off Channel

The channel gains are $H_i \sim \text{Bernoulli}\{0, 1\}$ with mean μ_H . The rate functions can be evaluated in closed form. From (Fig.2a),

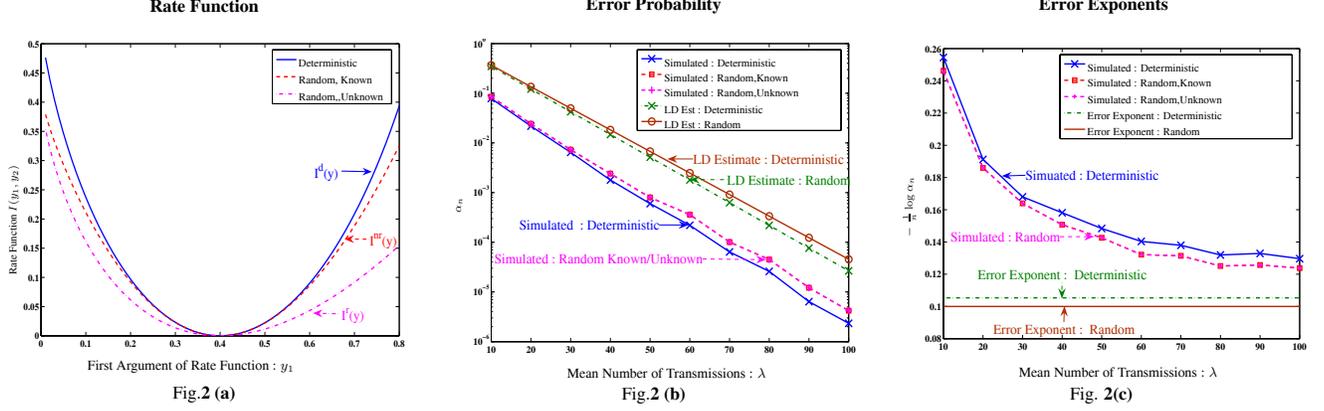
$$I_i^{ru}(\mathbf{y}) \leq I_i^{rk}(\mathbf{y}) \leq I_i^d(\mathbf{y}), \quad \therefore \eta^{ru} \leq \eta^{rk} \leq \eta^d.$$

The Type I/II error exponents are also evaluated in closed form and are attained on the boundary of Minimum Rate Detector. Additionally, we find that the error exponents for known and unknown N are equal i.e., $\eta^{rk} = \eta^{ru}$. We give simulation results in Fig.2b and Fig.2c. In Fig.2b, the Large Deviations (LD) estimate refers to $e^{-\lambda\eta}$. Since we neglect the $e^{\sigma(\lambda)}$ term, it is not accurate for small λ . There is a performance gap between the random and deterministic N at large λ as predicted. The curves for the known and unknown random N lie on each other. This agrees with the result $\eta^{rk} = \eta^{ru}$. Thus asymptotically there is no performance gain in knowing the realization of N for on-off channels. In Fig.2c as λ grows, the exponent in the simulations approaches the theoretical exponent. Thus the rate of decay of error probability agrees with theory for all cases.

6. MULTIPLE COLLECTIONS

Single collection scenario assumes that we can increase λ to any value to get the required performance. But there might be practical constraints in increasing λ , such as synchronization. Moreover increasing λ beyond a point results in more transmissions on the average. In a constrained area, this makes the fading gains correlated, thereby contradicting the i.i.d assumption. Thus we need to consider the general case where the fusion center collects more than one sample. We assume λ to be in *LD regime*. i.e., λ is large enough for the LD Estimate to be close to the actual error probability.

Let l be the number of collections. On the lines of Minimum Rate Detector, we define Minimum Sum-Rate Detector ($\hat{\theta}_{msd}$) by Decision Regions,



Simulation Values : $\mu_H = 0.5$, $\mathbf{p}_{\theta_0} = [0.8, 0.2]$, $\mathbf{p}_{\theta_1} = [0.2, 0.8]$, $\sigma^2 = 0.1$.

$$\Gamma_0 = \left(\sum_{j=1}^l (I_0(\mathbf{y}_j) - I_1(\mathbf{y}_j)) \leq 0 \right), \quad \Gamma_1 = \mathbb{R}^{km} \setminus \Gamma_0.$$

We can show that the Minimum Sum-Rate Detector gives optimal error exponents for l -collections in TBRA scheme. We express the error exponent for multiple collections, in terms of error exponent of a single collection, in the lemma below.

Lemma 1 Assume that the rate functions are continuous in the interior of their domain. Let η_s and $\eta_m(l)$ be the error exponents for a single collection and l collections respectively. Then,

$$\eta_m(l) = l\eta_s. \quad (11)$$

If the fusion center collects data using TBRA for l slots, the total average number of transmissions is $\rho = \lambda l$, which is also proportional to energy consumption. Fixing ρ , there is a spatio-temporal tradeoff between the average number of transmission per slot and the total number of slots in the data collection. For ρ in LD regime, we address this issue below.

1. If $\frac{\rho}{l}$ is also in LD regime : No difference in performance since the exponents are equal.

2. If $\frac{\rho}{l}$ is not in LD regime : There maybe performance degradation with l collections. This is because the $e^{-\rho(\frac{\rho}{l})}$ term is not small enough to be ignored and therefore the optimality of Minimum Sum-Rate detector for l -collections is not guaranteed.

7. CONCLUSIONS

In this paper we introduced TBRA, a random access protocol coupled with TBMA. This scheme has several attractive features. It is easy to deploy in practice. The access point does not need to monitor the state of sensors in the field. The bandwidth requirement is proportional to the number of local quantization levels of sensors, not to the number of sensors. It is easily scalable in λ . The rate functions are easier to evaluate and thus the decision rule at the fusion center requires less computational processing. However, we show that there is some performance degradation in terms of the rate of decay of error probability compared to having fixed number of sensors.

When there is a constraint in increasing the mean number of sensors per collection, we gave a scheme with multiple data collections

and showed linear improvement in the error exponents with collection size. In contrast to this fixed sample size scheme, we propose to search for sequential schemes where certain criteria have to be met and the size of the collection is a random variable.

The results in the paper depend crucially on the non-zero mean fading assumption. When channels are zero-mean fading, the Minimum Rate Detector gives poor performance. In [9, 10] we present an analysis of TBRA involving multiple collections with zero mean fading and show the existence of an optimal activation strategy (optimal λ) that maximizes the error exponents.

8. REFERENCES

- [1] G. Mergen, V. Naware, and L. Tong, "Asymptotic Detection Performance of Type-Based Multiple Access Over Multiaccess Fading Channels," submitted to IEEE Trans. on Signal Processing, May 2005.
- [2] Ke Liu and A. M. Sayed, "Optimal distributed detection strategies for wireless sensor networks,," in *42nd Annual Allerton Conf. on Commun., Control and Comp.*, Oct. 2004.
- [3] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, NY, 1997.
- [4] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE JSAC Special Issue on Wireless Sensor Networks*, 2004.
- [5] Biao Chen, Ruixiang Jiang, T. Kasetkasem, and P.K. Varshney, "Fusion of decisions transmitted over fading channels in wireless sensor networks,," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, Nov. 2002.
- [6] G. Mergen and L. Tong, "Estimation Over deterministic multiaccess channels,," in *Proceedings of the 42nd Allerton Conf. on Communications, Control, and Computing*, Monticello, IL, Sep.28-Oct.01 2003.
- [7] G. Mergen and L. Tong, "Type based estimation over multiaccess channels,," to appear in IEEE Transactions on Signal Processing, 2005.
- [8] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access in sensor networks,," in *Proceedings of the SPAWC 2005, IEEE's sixth international workshop on Signal Processing Advances in Wireless Communications.*, New York City, NY, June 2005.
- [9] A. Anandkumar and L. Tong, "Type-Based Random Access for Distributed Detection over Multiaccess Fading Channels,," *Submitted to IEEE Trans. Signal Proc.*, Dec. 2005.
- [10] A. Anandkumar and L. Tong, "Type-Based Random Access for Distributed Detection over Multiaccess Fading Channels,," in *Submitted to Proc. of CISS '06*, Princeton, NJ, March 2006.
- [11] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.
- [12] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, NY, 1998.
- [13] Frank den Hollander, *Large Deviations (Fields Institute Monographs, 14)*, American Mathematical Society, 2000.