Asymptotic Detection Performance of Type-Based Multiple Access Over Multiaccess Fading Channels

Gökhan Mergen, Vidyut Naware, and Lang Tong*

School of Electrical and Computer Engineering
384 Frank H.T. Rhodes Hall
Cornell University
Ithaca, NY 14853
Tel: (607) 255-3500. Fax: (607) 255-9072.
Email: ltong@ece.cornell.edu

Abstract

The problem of communicating sensor readings over a multiaccess channel for detecting a target using Type-Based Multiple Access (TBMA) is considered. TBMA is analyzed in a general framework by considering non-i.i.d. data and non-identical channel gains. An asymptotically optimal detector is proposed and its error-exponents for detection probabilities are characterized using tools from large deviations theory. In case of i.i.d. channel gains, it is shown that the performance of TBMA presents two distinct behaviors depending on whether the channel gains have zero mean. Numerical simulations are used to demonstrate that the error exponents provide reasonably accurate estimates of the performance of TBMA.

Index Terms

Sensor Networks, Distributed Detection, Error Exponents, Large Deviations, Types, Multiaccess Communication, Fading Channels.

EDICS: 2-DETC, 3-CNET.

---

* This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564, by the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011, and by the National Science Foundation under Contract CNS-0435190.

* Corresponding Author
I. INTRODUCTION

In this paper, we consider the medium access communication problem between sensor nodes and a fusion center. The fusion center is interested in detecting the presence of a target or estimating a parameter of the observed random field.

A. Modelling of Sensor Data

We consider a group of $n$ sensors trying to transmit their data to a fusion center over a multiaccess channel (MAC). In general, sensors observe real-valued data. For practical purposes, however, the observations are in general quantized before communication. In this paper, we do not deal with how the quantization is done, and assume that observation $X_i \in \{1, \cdots , k\}$ at sensor $i$ is already quantized to $k$ possible levels. In target detection, sensors may quantize their data to two levels indicating whether target is detected or not. In parameter estimation, on the other hand, $X_i$’s may model quantized measurements.

We use notation $\theta \in \mathbb{R}$ to denote the real parameter to be detected or estimated. The parameter can be discrete in case of target detection (i.e., $\theta \in \{0,1\}$ indicating the presence of target). In parameter estimation, $\theta$ may take continuous values representing the true state of the observed phenomenon.

Sensor data are statistically correlated, since nearby sensors tend to have correlated observations. A simple model for sensor data incorporates the conditionally i.i.d. assumption:

$$X_1, \cdots , X_n \overset{i.i.d.}{\sim} p_{\theta} \quad \text{given } \theta,$$

i.e., given the parameter $\theta$, the sensor data $X_1, \cdots , X_n$ are conditionally i.i.d. according to a probability mass function (pmf) $p_{\theta} = (p_{\theta}(1), \cdots , p_{\theta}(k))$. An interpretation of the conditionally i.i.d. assumption is that each sensor observes the same parameter $\theta$, but with i.i.d. observation noise.

While the conditionally i.i.d. assumption is applicable in some scenarios, it is restrictive in general, especially when observations of sensor nodes have varying degrees of reliability.\footnote{For example, in a target detection setting locations of sensor nodes play an important role. If the network is randomly deployed, some nodes may end up in unfavorable locations, which makes their data less reliable.} Furthermore, conditionally i.i.d. assumes that the sensed area is uniform, i.e., the parameter does
not vary in the observed area. In case the sensor observations come from a wide area with heterogenous parameter values, the model (1) needs to be generalized.

B. Type-Based Multiple Access

We deal with the transmission of sensor data $X_1, \cdots, X_n$ over a multiaccess channel. It is assumed sensor $i$ has channel gain $h_i \in \mathbb{R}$, which does not vary during the course of transmission. In this paper, we shall be primarily interested in the following scheme, which is called Type-Based Multiple Access (TBMA) [1], [2], [3], [4].

Let $s_1, \cdots, s_k$ be a set of $k$ predetermined orthonormal waveforms. In the TBMA scheme, sensor $i$ transmits the waveform $s_{X_i}$ corresponding to its observation $X_i$ with a certain energy $E$, i.e., it transmits $\sqrt{E} s_{X_i}$. Due to the additive nature of wireless medium, the fusion center receives

$$z = \sum_{i=1}^{n} h_i \sqrt{E} s_{X_i} + w, \quad (2)$$

where $w$ is white Gaussian channel noise with $\sigma^2/2$ power spectral density.$^3$

The motivation for TBMA arises from the special case that the sensor data are conditionally i.i.d. and sensor channel gains are identical (say, $h_i = 1, \forall i$). In this case, the received signal becomes

$$z = \sum_{j=1}^{k} \sqrt{EN_j} s_j + w, \quad (3)$$

where $N_j = \sum_{i=1}^{n} 1(X_i = j)$ is the number of sensors that observe symbol $j$. After matched filtering by $s_1, \cdots, s_k$, it can be seen that $z$ contains a noisy version of the histogram of sensor observations $(N_1, \cdots, N_k)$. The histogram scaled by the number of sensors, $N$, is also called the type in information theory [5], and hence the name TBMA. The basic idea in TBMA is to detect the target (or estimate the parameter) from the noisy type.

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

$^3$Eqn. (2) considers only one fusion center receiving data from its neighborhood, and it neglects the interference coming from other sensor-fusion center groups. In this paper, our objective is to analyze this so-called single-cell scenario thoroughly, which is a challenging problem by itself.
C. Related Work and Our Contribution

Estimation/detection over multiaccess channels has attracted considerable attention recently. TBMA has been proposed by the authors [1], [2] and by Liu and Sayeed [3], [4], independently. Works prior to TBMA (e.g.,[6], [7]) assumed that each sensor is allocated an orthogonal channel to transmit its observation as in TDMA, FDMA or CDMA.

Several asymptotic optimality properties of TBMA have been proved under the assumption of conditionally i.i.d. data and identical channel gains [2], [4], [8]. Since the histogram is a sufficient statistic for estimation/detection in the case of conditionally i.i.d. data, TBMA is a particularly good choice in that scenario. In parameter estimation, it has been shown in [1], [2] that the mean square error (MSE) with TBMA asymptotically achieves the Cramer-Rao Lower Bound as the number of sensors, $n$, go to infinity. Similarly, it has been shown in [4], [8] that TBMA achieves the best error exponent in target detection. The intuition behind these optimality results is that the effect of noise $w$ on $z$ (eqn. (3)) becomes negligible as $n \rightarrow \infty$. As a result, the asymptotic performance of TBMA is as if the fusion center has direct access to histogram, which is sufficient to get optimal performance. Thus, the main conclusion that can be drawn from [2], [4], [8] is that the asymptotic performance of TBMA is as if the fusion center has direct access to $X_1, \cdots, X_n$ in case of conditionally i.i.d. data and identical channel gains.

In other approaches such as TDMA, the bandwidth requirement grows linearly with $n$. In TBMA, however, the bandwidth requirement is independent of $n$—only $k$ orthogonal dimensions are needed. This implies that TBMA is significantly more bandwidth efficient than other orthogonal allocation methods, when the number of sensors, $n$, is large compared to $k$. This is likely to be the case in a target detection scenario with binary sensor observations ($k = 2$).

Wireless channels are subject to random fading, and different nodes may have different channels gains to the fusion center. In this paper, we first propose a detector that is asymptotically optimal in terms of providing the best error exponent in Bayesian hypothesis testing. Next, we provide an error exponent analysis of the TBMA scheme with i.i.d. random channel gains and conditionally i.i.d. data. This setup, despite being restricted to the i.i.d. situation, provides significant insights into the performance of TBMA. In particular, we identify two different regimes of operation. For the case that the channel gains $h_1, \cdots, h_n$ have non-zero mean, the detection error probabilities are shown to decay exponentially with $n$. On the other hand, if the
channel gains have zero mean, then the error probabilities may or may not go to zero as $n \to \infty$, depending on the statistics of sensor observations. Our results indicate that the performance of TBMA significantly depends on the channel characteristics, and to get the best performance TBMA should be used in channels with non-zero mean. Numerical simulations are provided to compare the performance of TBMA with other orthogonal allocation methods.

In a sensor network without transmit-power control, the sensor channel gains are generally not i.i.d. distributed, since some sensors may be closer to the fusion center than the others. Furthermore, the conditionally i.i.d. data assumption has its limitations as previously mentioned. In this paper, we also provide a general characterization of the detection error exponents for non-i.i.d. channel gains and data. For full generality, an abstract, large deviations theory framework (i.e., the Gärtner-Ellis Theorem) [9] is used.

Organization of the paper is as follows. In Section II, we review some results from large deviation theory in hypothesis testing and propose an asymptotic variant of the ML detector. In Section III, error exponents of TBMA with varying assumptions on data and channel statistics are provided and analyzed. In Section IV, some examples and simulation results are presented. Section V concludes the paper.

II. LARGE DEVIATIONS AND THE MINIMUM-RATE DETECTOR

In this section, we are interested in the hypothesis testing problem

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

Consider the TBMA scheme with received signal $z$. Upon reception of $z$, the fusion center decides on whether $\mathcal{H}_0$ or $\mathcal{H}_1$ is true. For a given decision rule at the fusion center, let $\alpha = \Pr\{\mathcal{H}_0 \to \mathcal{H}_1\}$ denote the probability that $\mathcal{H}_1$ is decided although $\mathcal{H}_0$ was true. Notation $\beta = \Pr\{\mathcal{H}_1 \to \mathcal{H}_0\}$ is defined analogously. The $\alpha$ and $\beta$ are generally called Type-I and Type-II error probabilities in literature [10]. We will use the notations $\alpha_n, \beta_n$, when the dependence on $n$ needs to be made explicit.
A. Large Deviations Principle

In this paper, we will be primarily interested in characterizing the error exponents\(^4\) (i.e., the rate of decay)
\[
- \lim_{n \to \infty} \frac{1}{n} \log \alpha_n, \quad - \lim_{n \to \infty} \frac{1}{n} \log \beta_n
\]
(5)
of the Type-I and Type-II error probabilities in various situations of interest (provided that the limits exist). Note that in a Bayesian setting with priors \(\{\pi_0, 1 - \pi_0\}\) on the hypotheses, the exponent for the probability of error is given by
\[
- \lim_{n \to \infty} \frac{1}{n} \log (\pi_0 \alpha_n + (1 - \pi_0) \beta_n),
\]
(6)
and it can be easily shown that [9]
\[
- \lim_{n \to \infty} \frac{1}{n} \log (\pi_0 \alpha_n + (1 - \pi_0) \beta_n) = \min \left\{ - \lim_{n \to \infty} \frac{1}{n} \log \alpha_n, - \lim_{n \to \infty} \frac{1}{n} \log \beta_n \right\}.
\]
(7)
Thus, in Bayesian hypothesis testing problems, the minimum of the Type-I and Type-II error exponents determines the overall rate of decay of the probability of error to zero.

A classic example is the case where the fusion center has complete access to \(\{X_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} p_{\theta_i}\) under hypothesis \(\theta = \theta_i\). Then, it is well known that the best achievable error exponent for Bayesian probability of error is achieved when the Type-I and Type-II error exponents are equal and is given by the so-called Chernoff information [5],
\[
C(p_{\theta_0}, p_{\theta_1}) = - \min_{0 \leq \lambda \leq 1} \log \left( \sum_{i=1}^k p_{\theta_0}^\lambda(i) p_{\theta_1}^{1-\lambda}(i) \right).
\]
(8)

In the TBMA scheme, the statistic used for hypothesis testing is the inner product (notation \(\langle \cdot, \cdot \rangle\)) between \(z\) and the waveforms \(s_1, \cdots, s_k\). Let
\[
y^{(n)} := \frac{1}{\sqrt{En}} [\langle z, s_1 \rangle \cdots \langle z, s_k \rangle]^T
\]
(9)
\[
= \frac{1}{n} \sum_{i=1}^n h_i e_{X_i} + \tilde{w},
\]
(10)
where \(e_1, \cdots, e_k\) are the standard basis vectors, and \(\tilde{w} \sim \mathcal{N}(0, \frac{\sigma^2}{En} I)\). In order to compute asymptotic error probabilities, we need to understand the asymptotics of the random vector \(y^{(n)}\). The theory of large deviations characterizes the probability of large excursions of \(y^{(n)}\) from its “mean” behavior by quantifying by the so-called rate function, which is defined below.

\(^4\)Throughout the paper, the notation \(\log\) refers to the natural logarithm.

DRAFT
**Definition** For a set \( B \subset \mathbb{R}^k \), let \( \text{int}(B) \) denote the interior of \( B \) and \( \text{cl}(B) \) denote the closure of \( B \). The sequence of random variables \( y^{(n)} \) for \( n = 1, 2, \cdots \) is said to satisfy the large deviations principle with rate function \( I \) if for any measurable set \( B \)

\[
- \inf_{x \in \text{int}(B)} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \Pr(y^{(n)} \in B) \leq \limsup_{n \to \infty} \frac{1}{n} \log \Pr(y^{(n)} \in B) \leq - \inf_{x \in \text{cl}(B)} I(x),
\]

where \( I : \mathbb{R}^k \to \mathbb{R}_+ \cup \{\infty\} \). The effective domain of the function \( I \) is defined as \( D_I = \{ x : I(x) < \infty \} \).

**Remark 1:** The sets of interest \( B \)'s in hypothesis testing mostly satisfy the so-called \( I \)-continuity property:

\[
\inf_{x \in \text{int}(B)} I(x) = \inf_{x \in \text{cl}(B)} I(x),
\]

which implies

\[
\lim_{n \to \infty} \frac{1}{n} \log \Pr(y^{(n)} \in B) = - \inf_{x \in B} I(x).
\]

The rate function \( I \) admits the following interpretation. For an integer \( k \) and \( r \in \mathbb{R}^k \), let \( B_\epsilon(r) \) be the open ball in \( \mathbb{R}^k \) centered at \( r \) with radius \( \epsilon > 0 \). If \( I \) is continuous in the interior of its domain \( D_I \), then it satisfies the condition in Remark 1 for all balls \( B = B_\epsilon(r) \) inside \( D_I \). The essence of the large deviations principle is that\(^5\)

\[
\Pr \{ y^{(n)} \in B_\epsilon(x) \} \doteq e^{-n(I(x) + o(\epsilon))}, \quad x \in \text{int}(D_I),
\]

where \( o(\epsilon) \) is a function that goes to zero as \( \epsilon \to 0 \). In other words, the likelihood of the event that \( y^{(n)} \) is in the close vicinity of \( x \) behaves as \( e^{-nI(x)} \).

**B. Minimum Rate Detector**

We will consider the detection error exponents with the ML detector. Given \( y^{(n)} \), the maximum-likelihood decision rule chooses the hypothesis under which the likelihood is maximum. The exact computation of the likelihood function of \( y^{(n)} \), however, is generally intractable in our setup. To alleviate the problem, we propose a variant of the ML detector as follows.

\(^5\)The "\( \doteq \)" notation in (12) means \( \lim_{n \to \infty} \frac{1}{n} \log \Pr \{ y^{(n)} \in B_\epsilon(x) \} = -n(I(x) + o(\epsilon)) \). This notation should be understood similarly in the rest of the paper.
Suppose $y^{(n)}$ satisfies LDP with rate function $I_i$ under hypothesis $\theta = \theta_i$. Recall that $e^{-nI_i(x)}$ can be viewed as the asymptotic probability of the event that $y^{(n)}$ turns out to be near $x$ under the hypothesis $\mathcal{H}_i$. Using this idea, we define the \textit{minimum-rate detector} as the decision rule with decision regions

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

(13)

Here, the detector decides that $\mathcal{H}_0$ is true if $e^{-nI_0(y^{(n)})} \geq e^{-nI_1(y^{(n)})}$ holds (i.e., the asymptotic likelihood of $y^{(n)}$ under $\mathcal{H}_0$ is higher). The decision region $\Gamma_1$ can be interpreted similarly. We expect the error exponents of this detector to be same as that of the exact ML detector.

Suppose we restrict ourselves to the class of detectors $C$ based on $y^{(n)}$ alone (and not on $n$ directly). Define a detector in $C$ to be \textit{max-min optimal} if it maximizes the minimum of the Type-I and Type-II error exponents amongst all detectors in $C$. Note that if a detector is max-min optimal, then it also has the best exponent for probability of error in a bayesian setting since the probability of error decays exponentially with the lower of the Type-I and Type-II exponents, as mentioned before. Next, we provide some results on the asymptotic performance of the minimum-rate detector.

\textit{Theorem 1:} Suppose that $y^{(n)}$ satisfies the LDP principle with rate functions $I_0$ and $I_1$ under hypothesis $\theta = \theta_0$ and $\theta = \theta_1$ respectively. Let $\Gamma_0$ (defined in (13)) be $I_1$-continuous and $\Gamma_1$ be $I_0$-continuous (eqn. (11)). Then,

\begin{enumerate}
  \item[i)] The error exponents of the minimum-rate detector are given by
  $$
  - \lim_{n \to \infty} \frac{1}{n} \log \alpha_n = \inf_{x \in \Gamma_1} I_0(x),
  $$
  $$
  - \lim_{n \to \infty} \frac{1}{n} \log \beta_n = \inf_{x \in \Gamma_0} I_1(x).
  $$
  (14)
  (15)
  \item[ii)] If the infimums in (14) and (15) are attained at the boundary $\partial \Gamma_1 = \{ x : I_0(x) = I_1(x) \}$, then the exponents of $\alpha_n$ and $\beta_n$ are the same and equal to
  $$
  \eta := \inf_{x \in \partial \Gamma_1} I_0(x),
  $$
  (16)
  and the minimum-rate detector is max-min optimal in $C$.
\end{enumerate}

\textit{Remark 2:} Note that for most fading channels of practical interest, the rate functions of $y^{(n)}$ are continuous in the interior of their domain. As a result, in such cases, the Type-I and Type-II error exponents are achieved on the boundary of the decision regions and they are equal.
Proof: See Appendix A.

III. TBMA IN MULTIACCESS FADING CHANNELS

In this section, we analyze the performance of the minimum-rate detector by characterizing the rate functions of $y^{(n)}$ under varying assumptions on the data and channel statistics. From here on, we suppress the superscript in $y^{(n)}$ for notational convenience.

A. Conditionally I.I.D. Data and Non-zero Mean I.I.D. Channels

Let $X_1, \cdots, X_n$ be conditionally i.i.d. with pmf $p_{\theta}$. Suppose that the channel gains $h_1, \cdots, h_n$ are i.i.d. and independent of $X_1, \cdots, X_n$. In this section, we will show that the error probabilities $\alpha_n, \beta_n$ decay exponentially with the network size $n$ under the condition that the channel gains have non-zero mean $h := \mathbb{E}(h_i)$ with the minimum-rate detector. In Section III-B, for the case that $\mathbb{E}(h_i) = 0$, we will argue that the error probabilities $\alpha_n, \beta_n$ may or may not go to zero simultaneously, depending on $p_{\theta_0}$ and $p_{\theta_1}$.

From the law of large numbers and Slutsky’s Theorem [11], it follows that in this case

$$y \rightarrow \mathbb{E}(h_i e^{X_i}) = h p_{\theta},$$

in probability as $n \rightarrow \infty$, where $p_{\theta}$ is viewed as a vector $[p_{\theta}(1) \cdots p_{\theta}(k)]^T$. Hence, for large $n$ one would expect to have $y \approx h p_{\theta_0}$ under hypothesis $\mathcal{H}_i$. Detection errors typically happen when $y$ is close neither to $h p_{\theta_0}$ nor to $h p_{\theta_1}$.

The next theorem characterizes the rate function of $y$. We adopt the following notations from [5]: $D(Q||P)$ denotes the relative entropy between the probability density functions (pdfs) $Q$ and $P$. For random variables $X$ and $Y$, $D(Y||X)$ denotes the relative entropy between the pdfs of $Y$ and $X$.

**Theorem 2:** Suppose that the moment generating function of $h_1, h_2, \cdots$ satisfies

$$\varphi(t) = \mathbb{E}e^{th_i} < \infty, \quad \forall t \in \mathbb{R}.$$  

(18)

Let $I_h : \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{\infty\}$ be the function defined as

$$I_h(r) = \inf_{\tilde{h} : \mathbb{E}(\tilde{h}) = r} D(\tilde{h}||h_i), \quad r \in \mathbb{R},$$

(19)
where the minimization is over real valued random variables \( \tilde{h} \). Similarly, for \( i = 0, 1 \) define 
\[
I_i : \mathbb{R}^k \to \mathbb{R}_+ \cup \{\infty\}
\]

\[
I_i(x) = \inf_{\tilde{p}} \{ D(\tilde{p}||p_{\theta_i}) + \sum_{j=1}^{k} \tilde{p}_j I_h(x_j/\tilde{p}_j) \},
\]

(20)

for each \( x \in \mathbb{R}^k \), where the minimization is over all probability vectors \( \tilde{p} \in \mathbb{R}^k \). Then, \( y \) satisfies the large deviations principle with the rate function \( I_i \) under hypothesis \( \theta = \theta_i \).

**Proof:** Given in Appendix B. The proof uses the Gärtner-Ellis Theorem presented in Section III-C.

**Remark 3:** The summation \( \frac{1}{n} \sum_{i=1}^{n} h_i e_{X_i} \) satisfies the large deviations principle with the same rate function \( I \). In other words, the existence of channel noise \( \tilde{w} \) does not affect the large deviations exponent. Intuitively speaking, the reason for this is that the large deviation probabilities of the noise \( \tilde{w} \sim \mathcal{N}(0, \sigma^2 \text{Ent} I) \) decays super-exponentially as \( e^{-n^2} \).

The result in (20) can be intuitively derived as follows. First, notice that \( \sum_{i=1}^{n} h_i e_{X_i} \) is the large deviations exponent of the sample mean of the channel gains, and \( D(\tilde{p}||p_\theta) \) is the large deviations exponent of the empirical measure \( \frac{1}{n} \sum_{i=1}^{n} e_{X_i} \). Moreover, for \( j = 1, \cdots, k \), \( \tilde{p}_j I_h(x_j/\tilde{p}_j) \) is the exponent of the event that \( n\tilde{p}_j \) nodes have sample channel mean near \( x_j/\tilde{p}_j \). Hence, \( D(\tilde{p}||p_\theta) + \sum_{j=1}^{k} \tilde{p}_j I_h(x_j/\tilde{p}_j) \) is the exponent of the event that the empirical measure of sensor data turns out to be near \( \tilde{p} \) and the summation \( \frac{1}{n} \sum_{i=1}^{n} h_i e_{X_i} \) is near \( x \). Following the dictum that “rare events happen in the most likely way of all unlikely ways” [9], we get (20).

**Corollary 1:** In the TBMA scheme, suppose that there is no fading \( \text{viz.}, h_i = 1 \text{ a.s. for all } i \). Then,

\[
I_i(x) = \begin{cases} 
D(x||p_{\theta_i}) & \text{if } \sum_{j=1}^{k} x_j = 1, x_j \geq 0 \forall j \\
\infty & \text{otherwise.}
\end{cases}
\]

(23)

*Eqns. (21), (22) can be obtained from the Chernoff and Sanov Theorems [9], respectively.*
Further, if $p_{\theta_0} \neq p_{\theta_1}$, then the Type-I and Type-II error exponents with the minimum-rate detector are the same and equal to the Chernoff information,

$$C(p_{\theta_0}, p_{\theta_1}) = -\min_{0 \leq \lambda \leq 1} \log \left( \sum_{i=1}^{k} p_{\theta_0}^\lambda(i), p_{\theta_1}^{1-\lambda}(i) \right).$$

(24)

In this case, TBMA with the proposed minimum rate detector is optimal in terms of Bayesian error exponent within the class of all detectors that are a function of the sensor data $X_1, X_2, \cdots, X_n$ and $n$ explicitly.

Proof: Equation (23) follows immediately from equations (19) and (20). Since the rate functions are continuous in the interior of their domains and convex, it follows that the infimums in equations (14) and (15) are attained at the boundary i.e., when $D(x||p_{\theta_0}) = D(x||p_{\theta_1})$ and hence the two error exponents coincide. Finally, the error exponents of the minimum-rate detector can be easily derived with Theorem 1 using Lagrange multipliers. Optimality of the minimum-rate detector in terms of providing the best bayesian error exponent is due to the properties of the Chernoff information [5].

Independently, Liu and Sayeed have shown the asymptotic optimality of TBMA with the ML detector in the no-fading scenario [4] using the Chernoff bounding techniques. The advantage of the Large Deviations framework compared to the Chernoff bound is that it can be easily generalized to the fading scenario as stated in Theorem 2. A disadvantage of the Large Deviations framework, though, is that it allows us to analyze only an asymptotic version of the ML detector—the minimum rate detector, but not the ML detector directly. In case of i.i.d. channels with non-zero mean we expect the error exponents of the ML detector to be same as that of the minimum-rate detector (as is the case with no-fading channels), although we do not have a proof for the same.

Next, we provide an alternative characterization of the rate function of $y$ as follows.

Theorem 3: Assume that (18) holds. Then, $I(x)$ can be equivalently expressed as

$$I(x) = \sup_{t \in \mathbb{R}^k} \left[ \sum_{j=1}^{k} x_j t_j - \sum_{j=1}^{k} p_{\theta}(j) \varphi(t_j) \right].$$

(25)

Furthermore, if $t^*$ attains the optimal value in (25), then

i) The $\tilde{p}^*$ attaining the minimum in (20) has pmf

$$\tilde{p}_{j}^* = \frac{p_j \varphi(t_j^*)}{\sum_{j=1}^{k} p_{j} \varphi(t_j^*)}.$$

(26)
ii) The $\tilde{h}^*$ attaining the minimum in (19) for $r = x_j/\tilde{p}_j^*$ has pdf $f(\tilde{h}^*)$ given by

$$f(\tilde{h}^*) = f(h)e^{\tilde{t}_j^h}/\varphi(t_j^*),$$

(27)

where $f(h)$ denotes the pdf of $h_i$.

**Proof:** Refer to Appendix C.

Equation (25) characterizes the rate function in terms of a maximization, and equation (20) characterizes it in terms of a constrained minimization. These two optimization problems always give the same answer, because they are *convex duals* of each other [12]. Through equations (26) and (27), Theorem 3 relates the optimal variables in the primal and dual characterizations of $I(x)$. In case (20) is used to evaluate $I(x)$, by using (26) and (27), one can obtain the optimal variables $\tilde{p}^*, \tilde{h}^*$ and characterize how “typical” realizations of rare events occur. Such characterizations, which are in part motivated by the Conditional Limit Theorem [5], are usually sought-after in large deviations theory.

**Remark 4:** Theorem 3 is useful for computational purposes as well. Since the rate function has two equivalent optimization forms, depending on the situation, computation of one might be preferable to the other. For example, if the dimension is small (e.g., $k = 2$), then one can do an exhaustive search over quantized $\tilde{p}$ to compute $I(x)$ via (20) since $I_h$ can be obtained in closed form for many channels of interest (such as Gaussian, binomial). On the other hand, if $k$ is large, then using (25) may be preferable since generic *unconstrained* optimization methods (such as steepest descent, Newton’s method) provide guaranteed convergence.

**B. Conditionally I.I.D. Data and Zero Mean I.I.D. Channels**

The proposed minimum-rate detector and the above result is useful only for the case that the channel gains have non-zero mean. If the channel mean is non-zero, then both error exponents in Theorem 1 are positive, *i.e.*, both $\alpha_n$ and $\beta_n$ converge to zero exponentially fast as $n \to \infty$. On the other hand, if the channel gains have zero mean, then one or more of the error exponents in (14), (15) are equal to zero, and the error probabilities may not even decay as $n \to \infty$. We will use the following theorem to understand the convergence behavior.

**Theorem 4:** If the channel gains $h_1, \ldots, h_n$ have zero-mean and variance $\sigma_h^2$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_i e^{X_i} \overset{d}{\to} \mathcal{N}(0, \sigma_h^2 \text{Diag}(p))$$

(28)
as $n \to \infty$, where $\overset{d}{\to}$ denotes the convergence in distribution, and $\text{Diag}(p_\theta)$ is a diagonal matrix with entries $p_\theta(1), \cdots, p_\theta(k)$.

**Proof:** Multivariate Central Limit Theorem [11] gives that
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_i e_{X_i} \overset{d}{\to} \mathcal{N}(0, \text{Cov}(h_i e_{X_i}))
\]
as $n \to \infty$. Eqn. (28) follows from the fact that $\text{Cov}(h_i e_{X_i}) = \sigma_i^2 \text{Diag}(p_\theta)$.

Let
\[
\tilde{y} := ny = \sum_{i=1}^{n} h_i e_{X_i} + n\tilde{w}.
\]

Intuitively speaking, Theorem 4 suggests that $\tilde{y}$ is asymptotically Gaussian distributed with covariance $n\sigma_i^2 \text{Diag}(p_\theta_0) + \sigma^2 E I$, under hypothesis $H_i$ i.e., the hypothesis testing problem can be equivalently viewed as a test between
\[
H_0 : \quad \tilde{y} \approx \mathcal{N}(0, n\sigma_i^2 \text{Diag}(p_\theta_0) + \sigma^2 E I),
\]
\[
H_1 : \quad \tilde{y} \approx \mathcal{N}(0, n\sigma_i^2 \text{Diag}(p_\theta_1) + \sigma^2 E I).
\]

This interpretation has significant implications. Possibilities for the error probability decay behavior are considered below.

i) If $p_\theta_i(j) > 0$ for all $i, j$, then the term $\frac{\sigma^2}{E} I$ in (29), (30) due to noise becomes negligible as $n \to \infty$. More precisely, what happens is that
\[
\frac{\tilde{y}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_i e_{X_i} + \sqrt{n}\tilde{w} \overset{d}{\to} \mathcal{N}(0, \sigma_i^2 \text{Diag}(p_\theta))
\]
as $n \to \infty$. The test between $\mathcal{N}(0, \sigma_i^2 \text{Diag}(p_\theta_0))$ and $\mathcal{N}(0, \sigma_i^2 \text{Diag}(p_\theta_1))$ for positive $p_\theta_i(j)$’s is a very nasty one; the Type-I and Type-II error probabilities of the ML detector are strictly positive. In general, eqn. (31) suggests that no decision rule can make the Type-I and Type-II error probabilities go to zero simultaneously as $n \to \infty$.

ii) If $p_\theta_0(j) = 0$, $p_\theta_1(j) > 0$ for some $j = j_0$ and $p_\theta_0(j) > 0$, $p_\theta_1(j) > 0$ for all $j \neq j_0$, then with the ML detector both Type-I and Type-II probabilities go to zero as $n \to \infty$. What enables the error probability decay is the $j_0$’th component of $\tilde{y}$; under $H_1$, the magnitude of the $j_0$’th component grows to infinity as $n \to \infty$ almost surely; under $H_0$, it stays constant. The other components of $\tilde{y}$ do not provide significant information in discriminating the hypotheses, similar to the case i) above.

Eqn. (31) follows from the fact that $\sqrt{n}\tilde{w}$ converges in probability to the zero vector. Eqn. (28) and the Slutsky’s Theorem [11] gives (31).
We will use the following lemma to analyze the error probabilities.

**Lemma 1:** Consider the hypothesis testing problem

\[ H_0' : r \sim \mathcal{N}(0, \sigma_1^2) \text{ vs. } H_1' : r \sim \mathcal{N}(0, \sigma_2^2), \]

where \( r \) is a real valued random variable. With the ML detector, the error probabilities scale as follows:

\[
\Pr\{H_0' \rightarrow H_1'\} = \sqrt{\frac{1}{\pi \sigma_2 \sqrt{\log(\sigma_2/\sigma_1)}}} \rightarrow 1,
\]

\[
\Pr\{H_1' \rightarrow H_0'\} = \sqrt{\frac{4 \sigma_1 \sqrt{\log(\sigma_2/\sigma_1)}}{\pi \sigma_2}} \rightarrow 1,
\]
as \( \sigma_2/\sigma_1 \rightarrow \infty \).

**Proof:** The proof relies on well known bounds for the \( Q \)-function. See Appendix D for details.

The test considering the \( j_0 \)’th component of \( \tilde{y} \) corresponds to \( \sigma_1^2 = \sigma^2/E \) and \( \sigma_2^2 = n\sigma_h^2 \theta(j_0) + \sigma^2/E \). Theorem 1 with this choice of \( \sigma_1^2 \), \( \sigma_2^2 \) and the preceding discussion suggests that the error probabilities \( \alpha_n, \beta_n \) of the ML detector based on \( \tilde{y} \) scale as \( \frac{1}{\sqrt{n \log n}}, \sqrt{\frac{\log n}{n}} \), respectively. In contrast with the case of non-zero mean channels, the decay rates in this case are not exponential. This decay behavior will also be confirmed via simulations in Section IV-C.

iii) If \( p_{\theta_0}(j) = 0, \ p_{\theta_1}(j) > 0 \) (or, the opposite) is true for more than one \( j \), then the error probabilities with the ML detector go to zero with a rate faster than \( \frac{1}{\sqrt{n \log n}}, \sqrt{\frac{\log n}{n}} \) given above. Even though the decay is faster, the error probabilities with the ML detector do not decay exponentially.

**C. Non-I.I.D. Data and Channels**

The sensor data need not be conditionally i.i.d. as elaborated in Section I-A. Moreover, the sensor channel gains need not be i.i.d., since some sensors may be closer than others to the collection agent. And, in general, sensor observations and channel gains may be dependent. In this section, we provide a generalization of the results of Section III-A for possibly dependent and not identically distributed data and channels.
Let \( \{(X_i, h_i)\}_{i=1}^{\infty} = \{(X_1, h_1), (X_2, h_2), \cdots \} \) be a sequence of \( \{1, \cdots, k\} \times \mathbb{R} \) valued random vectors. The distribution of \( \{(X_i, h_i)\}_{i=1}^{\infty} \) is determined by whether \( \mathcal{H}_0 \) or \( \mathcal{H}_1 \) is the correct hypothesis. Let
\[
\varphi(t) = \mathbb{E}e^{\langle t, y \rangle}, \quad t \in \mathbb{R}^k, n \in \mathbb{N}
\]
be the moment generating function of \( y \). The following theorem generalizes Theorem 2.

**Theorem 5:** (Gärtner-Ellis Theorem [13]) Suppose that
\[
\lim_{n \to \infty} \frac{1}{n} \log \varphi(nt) = \Lambda(t), \quad \forall t \in \mathbb{R}^k
\]
exists as an extended real number. If \( \Lambda \) is an essentially smooth, lower-semicontinuous function,\(^8\) then, \( y \) satisfies the Large Deviations Principle with rate function
\[
\Lambda^*(x) = \sup_{t \in \mathbb{R}^k} [\langle x, t \rangle - \Lambda(t)], \quad x \in \mathbb{R}^k.
\]
The function \( \Lambda^* \) is called the Legendre Transform of \( \Lambda \).

**Remark 5:** The above theorem does not require any independence, etc. assumption on the sequence \( \{(X_i, h_i)\}_{i=1}^{\infty} \). It only needs the existence of the asymptotic log moment generating function (mgf) (33) and its continuity, smoothness. Theorem 5 also characterizes the error exponents of the proposed minimum-rate detector.

We now present a specific non-i.i.d. case where the rate function can be calculated explicitly. Consider the standard TBMA setup with no fading \( \text{viz.} \),
\[
y = \frac{1}{n} \sum_{i=1}^{n} e^{X_i} + \tilde{w}
\]
where \( X_i \) is a binary random variable (\( \in \{1, 2\} \)) and \( \tilde{w} \sim \mathcal{N}(0, \sigma^2 I) \). As an example of the non i.i.d case, consider the following hypothesis testing problem:

- \( \mathcal{H}_0 : X_i \overset{i.i.d.}{\sim} \text{Bernoulli}(c) \).
- \( \mathcal{H}_1 : X_i \sim \text{Markov Chain with transition matrix } \Pi = \begin{bmatrix} 1 - a & a \\ b & 1 - b \end{bmatrix} \).

We assume that \( 0 < a, b, c < 1 \). As a result, the Markov chain is irreducible and aperiodic. We are interested in evaluating the performance of the minimum-rate detector proposed earlier with

\(^8\)See Dembo and Zeitouni [13] for precise meanings of these terms.
the TBMA scheme. We note the following:

i) Since we are interested in the asymptotic behavior as \( n \to \infty \), the initial distribution of the Markov chain is irrelevant.

ii) We do not expect TBMA to be optimal in this setup since in this case it is the empirical transition matrix and not the empirical distribution of the Markov chain which is sufficient for asymptotically optimal ML detection.

Next, we compute the rate functions for the minimum-rate detector under the two hypotheses.

It is obvious that the rate function under \( \mathcal{H}_0 \), \( I_0(x) \) is given by

\[
I_0(x) = \begin{cases} 
D([x_1, x_2]||[c, 1-c]), & \text{if } x_1 + x_2 = 1, \\
\infty, & \text{otherwise.}
\end{cases}
\]

In order to compute the rate function under hypothesis \( \mathcal{H}_1 \), we use a special form of the G"artner-Ellis Theorem [13] (Theorem 3.1.6, p. 76) for Markov chains. It follows then that the rate function \( I_1(x) \) is given by

\[
I_1(x) = \begin{cases} 
\sup_{u>0} \left[ \sum_{i=1}^{2} x_i \log \left( \frac{u_i}{u_{1i}} \right) \right], & \text{if } x_1 + x_2 = 1, \\
\infty, & \text{otherwise.}
\end{cases}
\] (34)

where we use the notation \( u > 0 \) to mean \( u_1 > 0, u_2 > 0 \).

**Lemma 2:** If \( x_1 + x_2 = 1 \), then

\[
I_1(x) = \left( x_1 \log(1 - a + \delta b) + x_2 \log(a\delta^{-1} + (1-b)) \right),
\]

where

\[
\delta = \frac{-ab(2x_1 - 1) + \sqrt{a^2b^2(2x_1 - 1)^2 + 4a(1-a)b(1-b)x_1x_2}}{2b(1-b)x_1}.
\]

**Proof:** From equation (34), it can be seen that \( I_1(x) \) corresponding to the case \( x_1 + x_2 = 1 \), is only a function of \( \delta = u_2/u_1 \). Optimizing over \( \delta > 0 \) gives the desired result. \( \blacksquare \)

**IV. EXAMPLES AND SIMULATION RESULTS**

**A. ON/OFF Channels**

Consider the ON/OFF channel, \( i.e., \ h_i \) is Bernoulli \( \{0, 1\} \) distributed with mean \( h \). For this scenario, \( I_h(\cdot) \) is the relative entropy function between two Bernoulli variables. Using the
Lagrange multipliers method, it is easy to get

\[ I(x) = XD(\bar{x} || p_\theta) + X \log \frac{X}{h} + (1 - X) \log \frac{1 - X}{1 - h}, \]

where \( X = h \sum_{j=1}^{k} x_j \) and \( \bar{x} = x / (\sum_{j=1}^{k} x_j) \). Using Theorem 1, the error exponent of \( \alpha_n \) is obtained as

\[ \eta = X^* C + X^* \log \frac{X^*}{h} + (1 - X^*) \log \frac{1 - X^*}{1 - h}, \]  

(35)

where \( C \) is the Chernoff information defined before (equation 8), and \( X^* = h / ((1 - h)e^C + h) \).

Simulation results for the hypotheses

\( \mathcal{H}_0 : X_i \overset{i.i.d.}{\sim} p_{\theta_0} = [0.8, 0.2], \quad \mathcal{H}_1 : X_i \overset{i.i.d.}{\sim} p_{\theta_1} = [0.2, 0.8] \)

(36)

are given in Fig. 1. The LD (Large Deviations) estimate refers to \( e^{-n\eta} \). SNR = \( E / \sigma^2 \) = 3dB. We compare TBMA with the following version of TDMA: Sensor nodes take turns to transmit their individual data in non-overlapping time intervals. More explicitly, if the \( i \)'th sensor observes 1, then it transmits the signal \( s_i = \sqrt{E} \) in its time-slot, otherwise it transmits \( s_i = -\sqrt{E} \) (i.e., the antipodal constellation is used). The fusion center receives \( h_i s_i + n_i, \quad i = 1, \cdots, N \), where \( n_i \) are i.i.d. \( \mathcal{N}(0, \sigma^2/2) \) channel noise. The ML detection is done based on \( h_i s_i + n_i, \quad i = 1, \cdots, N \), where the receiver is assumed to know the statistics of \( h_i \) only.

Some remarks are in order:

\( i) \) In the ON/OFF channel, the exact ML detector is computationally intractable. However, the proposed minimum-rate detector (13) is very easy to implement.

\( ii) \) Performance estimates provided by the large deviations theory are reasonably accurate for this scenario.

\( iii) \) The TBMA outperforms the TDMA scheme. In our simulations we observed that this is typically the case in channels with non-zero mean.

\( iv) \) The Chernoff information \( C \) is the optimal exponent obtained when the fusion center has direct access to \( X_1, \cdots, X_n \) [5]. From (35), it is seen that \( \eta \to C \) as \( h \to 1 \). In other words, the asymptotic performance of the TBMA scheme approaches the optimal one as the channel fading disappears. This is in accordance with the results in [2], [4], which show that the TBMA actually has the optimal error exponent, when there is no channel fading.
B. Plus/Minus-1 Channel

Suppose that the channel $h_i$ only takes values in $\{-1, +1\}$, and $h_1, \ldots, h_n$ are i.i.d. The rate function for this channel doesn’t have a closed form expression, but it can be computed numerically from Theorem 5. Figure 2 shows $\eta$ as a function of $\Pr\{h_i = 1\}$ for the binary observations model in (36) (SNR = $E/\sigma^2 = 0$dB). Two end points of this curve are interesting:

i) As $\Pr\{h_i = 1\}$ approaches 0.5, the $\eta$ approaches 0. This approves our intuition from Section III-B that the error probabilities in a zero-mean channel either do not decay to zero, or they decay sub-exponentially.

ii) As $\Pr\{h_i = 1\}$ approaches 1, the $\eta$ approaches the Chernoff information upper bound.

C. Gaussian Channels

Suppose that $h_1, \ldots, h_n$ are i.i.d. Gaussian with certain mean and variance. Depending on the mean, variance and $p_{\theta_i}$’s the Gaussian channels present a rich set of behaviors. In the Gaussian channel $h_i \sim \mathcal{N}(\mu, \sigma^2_h)$ with binary observations, the vector $y$ has a mixed Gaussian distribution,
and its likelihood function of $y$ can be computed from
\[
y \sim \sum_{i=0}^{n} \binom{n}{i} p_0^i (1-p_0)^n-i \mathcal{N} \left( \mu \begin{bmatrix} i \\ i \\ i \end{bmatrix}, \begin{bmatrix} i \sigma_h^2 + \sigma_x^2 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} i \\ i \\ (n-i) \sigma_h^2 + \sigma_x^2 \end{bmatrix} \right).
\]

Hence, the exact likelihood function can be computed unlike the situation, for example, in ON/OFF channels.

Figure 3 shows the error probability performance of TBMA and TDMA schemes when the channel has non-zero mean ($h_i \sim \mathcal{N}(1, 0.5)$, SNR = -20dB). Here, TBMA error probabilities decay exponentially, since the channel has non-zero mean. The large deviations estimate of $\alpha_n$ is rather coarse in this channel; the slope of the curve $e^{-\eta n}$ appears to be same as the slope of $\alpha_n$, however there is a non-vanishing gap between $e^{-\eta n}$ and $\alpha_n$ (this is because the large deviations estimate does not consider sub-exponential terms, and in particular the constants). The TBMA error performance again outperforms TDMA.

Figure 4 shows the error probabilities with TBMA in a channel with zero-mean ($h_i \sim \mathcal{N}(0, 0.5)$). As discussed in Section III-B, the error probabilities do not go to zero as $n \to \infty$ even when there is no channel noise and the exact ML detector is used at the fusion center. Hence, the
performance degradation is not due to the proposed minimum-rate detector. The performance with TDMA scheme is not shown in this figure, because TDMA still has exponential error probability decay, and its curve is significantly below that of TBMA.

Here, we see that while the TBMA is optimal in case of identical (and, non-fading) channels, its performance can be quite poor in fading channels with zero mean. The reason is that the sensor transmissions add non-coherently when the channel has zero-mean (i.e., they tend to cancel each other). Therefore, the fusion center can not get the type accurately. The only case for which the error probability decays is when some \( p_{\theta_0}(j) = 0 \) and \( p_{\theta_1}(j) \neq 0 \) or vice-versa.

Figure 5 shows the error probabilities with TBMA in another channel with zero-mean \( (h_i \sim \mathcal{N}(0,10), \sigma^2/E = 0.1) \) and \( p_{\theta_0} = [1 \ 0], \ p_{\theta_1} = [0.2 \ 0.8] \). The second curve is the estimate

\[
\sqrt{\frac{1}{\pi} \frac{\sigma_1}{\sigma_2 \sqrt{\log(\sigma_2/\sigma_1)}}}
\]

of \( \alpha_n \) provided by Theorem 1 \( (\sigma_1^2 = \sigma^2/E, \ \sigma_2^2 = n\sigma_k^2p_{\theta_i}(2) + \sigma^2/E) \). The error probability decay is due to the fact that the second entry of \( p_{\theta_0} \) is zero. This enables the fusion center to better distinguish between \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \), when there are more nodes transmitting. Here, it is
important to note that the decay is not exponential; it is actually slower. The estimate (37) of $\alpha_n$ is actually quite good even for small $n$.

D. Non-i.i.d. Data

We now present numerical examples for the detection of Markov process described in Section III-C. Figure 6 shows the decision regions with the minimum-rate detector for a specific case with $a = 0.2, b = 0.3, c = 0.4$. It can be seen that $\Gamma_0 \sim \{x : x_1 \in (0.15, 0.47)\}$ and $\Gamma_1 = \Gamma_0^c$. Since $y_1 + y_2 \neq 1$ almost surely, it appears as if the minimum-rate detector cannot distinguish between the two hypotheses. However, the detector based just on $y_1$ (by setting $y_2 = 1 - y_1$) can still be used.

By Theorem 1, it follows that the error exponent of both Type I and Type II error probabilities with the minimum-rate detector is $\sim 0.01$ for this example. Intuitively, the probability vector $[0.6, 0.4]$ is the limiting distribution of the Markov chain with transition matrix $\Pi_{a=0.2,b=0.3}$ and so, it is natural to expect an interval around the point $x_1 = 0.6$ to lie in $\Gamma_1$. Note also, that the “optimal” ML detector would require knowledge of the empirical transition matrix.
Figure 7 shows the simulated Type I error probability with the minimum-rate detector with $SNR = 0$ dB. It can be seen that the slope of the large deviations estimate is almost the same as that of the Type I error probability simulated with TBMA. We note in passing that if the stationary distribution of the Markov chain coincides with Bernoulli$(c)$, then TBMA cannot distinguish between the two hypothesis asymptotically.

V. Conclusions

In this paper, we have considered the transmission of sensor observations over a multiaccess fading channel for the purpose of detection. We analyzed the performance of the Type-Based Multiple Access using large deviations theory. An asymptotic version of the Maximum Likelihood detector is proposed, and its error exponents are characterized. For channels with zero mean, it is argued that the detection error probabilities may or may not decay to zero as the number of transmitters grow; even though the error probability may go to zero, it doesn’t go exponentially fast. Simulation results are presented to validate the theoretical findings and check...
the accuracy of the large deviations approximations.

For the detection scenarios considered, the TBMA is significantly more bandwidth efficient than the conventional approaches such as TDMA, FDMA, CDMA. Its detection error performance is also superior to these schemes depending on the fading statistics. One main reason why TBMA outperforms other orthogonal schemes is that the other schemes always have to contend with noise (the per-user noise does not diminish as the number of sensors grow), whereas the noise in TBMA becomes negligible asymptotically in non-zero mean channels. For other schemes, more users mean more number of dimensions—hence more noise, whereas the TBMA does not have this problem. The TBMA is also advantageous because of the fact that it tries to deliver a sufficient statistic and nothing more than that.

The advantage of the Large Deviations Theory presented is that it is fully applicable for general non i.i.d. data and channels with non-zero mean. The Large Deviations Theory framework also suggests a natural asymptotically-optimal detector (i.e., the minimum-rate detector) that works universally for all channels with non-zero mean. In situations where the exact maximum likelihood detector was computationally intractable, we saw that the minimum rate detector is
Fig. 7. Type I error probabilities for non i.i.d case, \(a = 0.2, b = 0.3, c = 0.4\).

very easy to implement. For computation of the rate functions for the i.i.d. scenario, we also provided a convex-dual characterization of the optimization problem that gives the rate function.

The performance of TBMA degrades in fading channels with zero-mean. The reason for that is that the sensor transmissions add non-coherently when the channel mean is zero. Consequently, they tend to cancel each other, and the fusion center does not get the type accurately. Mathematically, the received signal grows as \(cNp_\theta\) in non-zero mean channels, where \(c \neq 0\) is a constant, \(N\) is the number of sensors, and \(p_\theta \in \mathbb{R}^k\) is the pdf of observations represented as a vector. In zero mean channels, the signal grows as a Gaussian signal \(\mathcal{N}(0, cNp_\theta)\) because of non-coherent additions. Therefore, the effect of \(p_\theta\) on the received signal is seen as a second order effect, and the TBMA does not work well. Search for new schemes and improvements over TBMA for use in channels with zero-mean constitute an interesting area for research.

Another important issue that we have not fully investigated in this paper is the performance of TBMA for non-i.i.d. data and channels. While we characterized the rate function using the Gärtner-Ellis Theorem, the practical implications for non-i.i.d. data and channels need further study. Finding the non-i.i.d. cases where the exponent becomes zero and the error decay behavior
is sub-exponential also deserves further study.

APPENDIX

A. Proof of Theorem 1

i) is immediate from the definition of $I$-continuity.

For ii), note that $\inf_{x \in \partial \Gamma_1} I_0(x) = \inf_{x \in \partial \Gamma_1} I_1(x)$ and hence the two exponents are equal to (16).

Next, we prove the max-min optimality of the minimum-rate detector. Let the decision regions corresponding to any other detector in $C$ be $\Gamma_0^*, \Gamma_1^*$. Then, we have,

$$\max_{\Gamma_1^*} \left\{ \min \left\{ \inf_{x \in \Gamma_1^*} I_0(x), \inf_{x \in \Gamma_0^*} I_1(x) \right\} \right\} \geq \min \left\{ \inf_{x \in \Gamma_1} I_0(x), \inf_{x \in \Gamma_0} I_1(x) \right\}. \quad (38)$$

By assumption, $\exists x^* \in \Gamma_0^*$ such that

$$I_0(x^*) = I_1(x^*) = \min \left\{ \inf_{x \in \Gamma_1} I_0(x), \inf_{x \in \Gamma_0} I_1(x) \right\}.$$

Now, for any $\Gamma_1^*$, if $x^* \in \Gamma_1^*$,

$$\min \left\{ \inf_{x \in \Gamma_1^*} I_0(x), \inf_{x \in \Gamma_0^*} I_1(x) \right\} \leq \inf_{x \in \Gamma_1^*} I_0(x),$$

$$\leq I_0(x^*),$$

$$= I_1(x^*).$$

On the other hand, if $x^* \in \Gamma_0^*$, then

$$\min \left\{ \inf_{x \in \Gamma_1^*} I_0(x), \inf_{x \in \Gamma_0^*} I_1(x) \right\} \leq \inf_{x \in \Gamma_0^*} I_1(x),$$

$$\leq I_1(x^*).$$

and thus we have shown that

$$\max_{\Gamma_1^*} \left\{ \min \left\{ \inf_{x \in \Gamma_1^*} I_0(x), \inf_{x \in \Gamma_0^*} I_1(x) \right\} \right\} \leq \min \left\{ \inf_{x \in \Gamma_1} I_0(x), \inf_{x \in \Gamma_0} I_1(x) \right\}. \quad (39)$$

From (38) and (39), it follows that the minimum rate detector is max-min optimal.
B. Proof of Theorem 2

To obtain the result of Theorem 2 from the Gärtner-Ellis Theorem, we need to compute the asymptotic log moment generating function in (33). Since \((X_i, h_i)\)'s are i.i.d. and independent of the channel noise,

\[
\frac{1}{n} \log \varphi(nt) = \log \mathbb{E} e^{t h_i X_i} + \frac{1}{n} \log \mathbb{E} e^{(nt, \tilde{w})}.
\] (40)

Since \(\tilde{w} \sim \mathcal{N}(0, \frac{\sigma^2}{n^2} I)\), the second term \(\mathbb{E} e^{(nt, \tilde{w})}\) does not vary with \(n\); therefore,

\[
\frac{1}{n} \log \mathbb{E} e^{(nt, \tilde{w})} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad \forall t \in \mathbb{R}^k.
\] (41)

The limit (41) shows that the existence of noise does not affect the rate function. It follows from (41) that

\[
\Lambda(t) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \varphi(nt)
\]

is equal to the first term in (40), which can be expressed as

\[
\log \mathbb{E} e^{(t, h_i X_i)} = \log \sum_{j=1}^{k} p_\theta(j) \mathbb{E} e^{t_j h_i}.
\]

The function \(\mathbb{E} e^{t_j h_i}\) exists and is finite \(\forall t_j \in \mathbb{R}\) due to assumption (18). Moment generating functions are continuous and differentiable in the interior of their effective domain [9]. Hence, the function \(\Lambda(t)\) is finite, continuous and differentiable for all \(t \in \mathbb{R}^k\). This implies that the Gärtner-Ellis Theorem can be applied.

To establish Theorem 2, we only need to show that

\[
\Lambda^*(x) = \sup_{t \in \mathbb{R}^k} \left[ \sum_{j=1}^{k} x_j t_j - \log \sum_{j=1}^{k} p_\theta(j) \mathbb{E} e^{t_j h_i} \right]
\]

is equal to the \(I(x)\) in (20). First, observe that

\[
\sup_{t \in \mathbb{R}^k} \left[ \sum_{j=1}^{k} x_j t_j - \log \sum_{j=1}^{k} p_\theta(j) \mathbb{E} e^{t_j h_i} \right] = \sup_{t \in \mathbb{R}^k} \inf_{\tilde{p}} \left[ D(\tilde{p} || p) + \sum_{j=1}^{k} x_j t_j - \sum_{j=1}^{k} \tilde{p}_j \log \mathbb{E} e^{t_j h_i} \right],
\] (42)

where the equality can be seen by minimizing the objective functional in (42) with respect to the probability vector \(\tilde{p} \in \mathbb{R}^k\). The objective functional in (42) is convex with respect to
\(\tilde{p}\) and concave with respect to \(t\), therefore the minimization and maximization terms can be interchanged to get
\[
\inf_{\tilde{p}} \sup_{t \in \mathbb{R}^k} [D(\tilde{p}||p) + \sum_{j=1}^{k} x_j t_j - \sum_{j=1}^{k} \tilde{p}_j \log \mathbb{E}e^{t_j h_j}].
\]
(43)

In geometric terms, (42) and (43) are equal, because the solution lies at the saddle point of the objective functional. Take the supremum in (43) inside the functional to get
\[
\inf_{\tilde{p}} [D(\tilde{p}||p) + \sum_{j=1}^{k} \tilde{p}_j I_h(\frac{x_j}{\tilde{p}_j})],
\]
(44)

where the second equality can be seen by writing the minimization in the original definition of 
\(I_h(r) = \inf_{h \in \mathbb{R}^k} D(\tilde{h}||h)\) by using Lagrange multipliers. This establishes the equality \(I(x) = \Lambda^*(x)\).

**C. Proof of Theorem 3**

It follows directly from Appendix B that
\[
\inf_{\tilde{p}} \left[ D(\tilde{p}||p_\theta) + \sum_{j=1}^{k} \tilde{p}_j I_h(\frac{x_j}{\tilde{p}_j}) \right] = \sup_{t \in \mathbb{R}^k} \left[ \sum_{j=1}^{k} y_j t_j - \sum_{j=1}^{k} p_\theta(j) \varphi(t_j) \right].
\]
(45)

Next, suppose \(t^*\) is the optimal \(t\) in (25). Using equations (26) and (27) it can be shown that
\[
D(\tilde{p}^*||p_\theta) + \sum_{j=1}^{k} \tilde{p}_j^* I_h(\frac{x_j}{\tilde{p}_j^*}) = \sum_{j=1}^{k} y_j t_j^* - \sum_{j=1}^{k} p_\theta(j) \varphi(t_j^*),
\]
and thus it follows that \(\tilde{p}^*\) and \(\tilde{h}^*\) are the optimizing variables in (20). Actually the equations (26) and (27) can be seen to be Lagrange multipliers corresponding to the optimizations in equations (42) and (44) respectively.

**D. Proof of Lemma 1**

Let \(\sigma = \frac{\sigma_2}{\sigma_1}\). For the case that \(\sigma > 1\), it can be easily shown that the Type-I and Type-II error probabilities with the ML detector are given by
\[
\Pr\{\mathcal{H}_0 \rightarrow \mathcal{H}_1'\} = 2Q\left(\sigma \sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}\right).
\]
(46)
and
\[
\Pr\{\mathcal{H}'_1 \rightarrow \mathcal{H}'_0\} = Q\left(-\sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}\right) - Q\left(\sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}\right).
\] (47)

Hence, without loss of generality, we can consider \(\sigma \rightarrow \infty\). For the Type-I case, we use the inequality [14]
\[
\frac{1}{\sqrt{2\pi x}} \left(1 - \frac{1}{x^2}\right) e^{-x^2/2} < Q(x) < \frac{1}{\sqrt{2\pi x}} e^{-x^2/2} \quad x > 1.
\] (48)
Substituting \(x = \sigma \sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}\) in (48), dividing by \(\frac{1}{\sigma \sqrt{\pi \log \sigma}}\) and taking the limit as \(\sigma \rightarrow \infty\) gives
\[
\Pr\{\mathcal{H}'_0 \rightarrow \mathcal{H}'_1\} \rightarrow 1.
\]

For the Type-II case, we use the following inequalities [14].
\[
Q(x - y) - Q(x + y) \geq \frac{1}{2} - Q(2y) \quad -y \leq x \leq y,
\] (49)
\[
Q(x) \leq \frac{1}{2} e^{-\sqrt{2/\pi x}} \quad x \geq 0,
\] (50)
and
\[
(Q(x) - Q(y))^2 \leq \frac{1}{2\pi} (x - y)^2.
\] (51)
In equation (49), substituting \(x = 0, y = \sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}\) gives,
\[
\Pr\{\mathcal{H}'_1 \rightarrow \mathcal{H}'_0\} \geq \frac{1}{2} - Q\left(2 \sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}\right),
\]
\[
\geq \frac{1}{2} - \frac{1}{2} \exp\left(-\sqrt{\frac{2}{\pi} 2 \sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}}\right),
\] (52)
where we use (50) to get (52). Next, dividing both sides of (52) by \(\frac{2 \sqrt{\log \sigma}}{\sqrt{\pi \sigma}}\) and evaluating the limit of the right hand side as \(\sigma \rightarrow \infty\) gives,
\[
\lim_{\sigma \rightarrow \infty} \frac{\Pr\{\mathcal{H}'_1 \rightarrow \mathcal{H}'_0\}}{\sqrt{\frac{4 \sqrt{\log \sigma}}{\pi \sigma}}} \geq 1.
\]

For the reverse inequality, we use (51). In (51), choose \(-x = y = \sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}\). Then we have,
\[
\Pr\{\mathcal{H}'_1 \rightarrow \mathcal{H}'_0\} \leq \frac{2}{\sqrt{2\pi}} \sqrt{\frac{2 \log \sigma}{\sigma^2 - 1}}.
\] (53)
Dividing both sides by \(\frac{2 \sqrt{\log \sigma}}{\sqrt{\pi \sigma}}\), and taking the limit as \(\sigma \rightarrow \infty\) gives,
\[
\lim_{\sigma \rightarrow \infty} \frac{\Pr\{\mathcal{H}'_1 \rightarrow \mathcal{H}'_0\}}{\sqrt{\frac{4 \sqrt{\log \sigma}}{\pi \sigma}}} \leq 1,
\]
and we are done.
REFERENCES


