Type Based Estimation over Multiaccess Channels

Gökhan Mergen and Lang Tong†

***Abstract*— We study the problem of communicating sensor readings over a Gaussian multiaccess channel. We focus on the scenario that each sensor observes a single random variable, and transmits it using certain signaling in a shared channel. The objective is the design of channel waveforms (*i.e.,* the signal constellation) to facilitate the estimation of field parameters from the channel output. We propose a communication scheme in which sensors transmit according to the *type* of their observations—the Type Based Multiple Access (TBMA), and show that the TBMA is asymptotically optimal in the limit of large number of sensors if the sensor channel-gains are identical. In particular, we show that the TBMA together with a variant of the maximum-likelihood estimator achieves the Cramer-Rao bound asymptotically. We then extend the asymptotic analysis of TBMA to fading channels, and compare the performance of TBMA with other orthogonal allocation methods such as time-division multiple access (TDMA).**

***Index Terms*— Sensor Networks, Parameter Estimation, Types, Asymptotic Efficiency, Multiaccess Channel, Orthogonal Allocation, TDMA.**

I. Introduction

A. Context and Problem Setup

Main functions of wireless sensor networks include sensing of a physical phenomena, and the delivery of the sensed data. Since sensor data are correlated, the efficiency is improved by processing the data locally by a *fusion center* and then delivering compressed information. The fusion center can be a cluster-head in a hierarchical sensor network, or a mobile access point.

In this work we focus on the multiaccess part of sensor communication. How should the multiaccess be designed such that the sensor data are gathered by a fusion center most efficiently? The conventional approach mandates the data to be packetized and then transmitted according to a multiaccess protocol. This approach, however, ignores the fact that the sensor data are correlated and that the ultimate objective is the estimation of the field. In this paper we show that significant gains can be realized in estimation quality and in system resource consumption if the physical layer and the multiaccess are designed jointly for the purpose of estimation.

We consider the case that a group of n sensors observe conditionally independent and identically distributed (i.i.d.) data $X_1, \cdots, X_n$ (Fig. 1) given a parameter θ. For convenience, it is assumed that each $X_i \in \{1, \cdots, k\}$ is discrete[1] with the probability mass function (pmf) $p_\theta = (p_\theta(1), \cdots, p_\theta(k))$. The pmf belongs to a family $\{p_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}$ is the parameter space, and the objective is to estimate the parameter θ. Each sensor transmits a waveform s_{i,X_i} which depends on the node index i and the observation X_i (energy constraint $||s_{i,X_i}||^2 \leq E$ must be satisfied). The transmitted signals are received through a Gaussian multiaccess channel (MAC). The fusion center produces an estimate $\hat{\theta}$ of the parameter after reception. The objective is to design the channel waveforms and the estimator such that the mean squared error (MSE) $\mathbb{E}\{(\hat{\theta} - \theta)^2\}$ is minimized.

First, consider the ideal scenario that the fusion center has access to all X_i's directly. In this case a fundamental limit on estimation performance is given by the *Cramer-Rao bound* (CRB) [1]. That is, under regularity conditions on $\{p_\theta : \theta \in \Theta\}$, the MSE of any unbiased estimator $\hat{\theta}$ satisfies

$$\mathbb{E}\{(\hat{\theta} - \theta)^2\} \geq \frac{1}{nI(\theta)}, \quad (1)$$

where $I(\theta) = \sum_{i=1}^k \frac{(dp_\theta(i)/d\theta)^2}{p_\theta(i)}$ is the *Fisher information*[2] in observation X_i. The CRB is not always achievable for finite n, but there is a class of estimators, including the Maximum Likelihood (ML) estimator, achieving the CRB asymptotically, *i.e.,*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \frac{1}{I(\theta)}) \quad (2)$$

as $n \to \infty$[3]; such estimators are called *asymptotically efficient*. Of course, the fusion center having direct access to X_i's is an idealistic assumption due to the channel noise and the energy constraints, and this performance may not be achievable in the Gaussian MAC.

B. Type Based Multiple Access

In waveform design a crucial observation is that the estimator doesn't need to know the raw data $X_1, \cdots, X_n$ to achieve the best performance. Actually, if the nodes could deliver a *sufficient statistic* with their transmissions, then there is no loss of information. One such sufficient statistic is the *empirical measure* (*i.e.,* the type)

$$\tilde{p} = \frac{1}{n}(N_1, \cdots, N_k),$$

†Corresponding author

The authors are with the School of Electrical and Computer Engineering at Cornell University, Ithaca, NY. E-mail: {mergen,ltong}@ece.cornell.edu.
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 CCR-0330059.

[1] X_i should be viewed as *quantized* data. In this paper, we do not deal with continuous variables, or how the quantization is done.

[2] It is assumed that $I(\theta) < \infty \Leftrightarrow p_\theta(i) > 0, \forall i$.

[3] Notations $\xrightarrow{\mathrm{d}}$, $\xrightarrow{\mathrm{p}}$ denote the convergence in distribution and convergence in probability, respectively. Asymptotic efficiency at $\theta = \theta_0$ requires θ_0 to be *identifyable i.e.,* $\forall \epsilon > 0$, $\exists \delta > 0$ such that $||p_\theta - p_{\theta_0}|| < \delta$ for some $\theta \Rightarrow |\theta - \theta_0| < \epsilon$. This will be assumed throughout the paper.

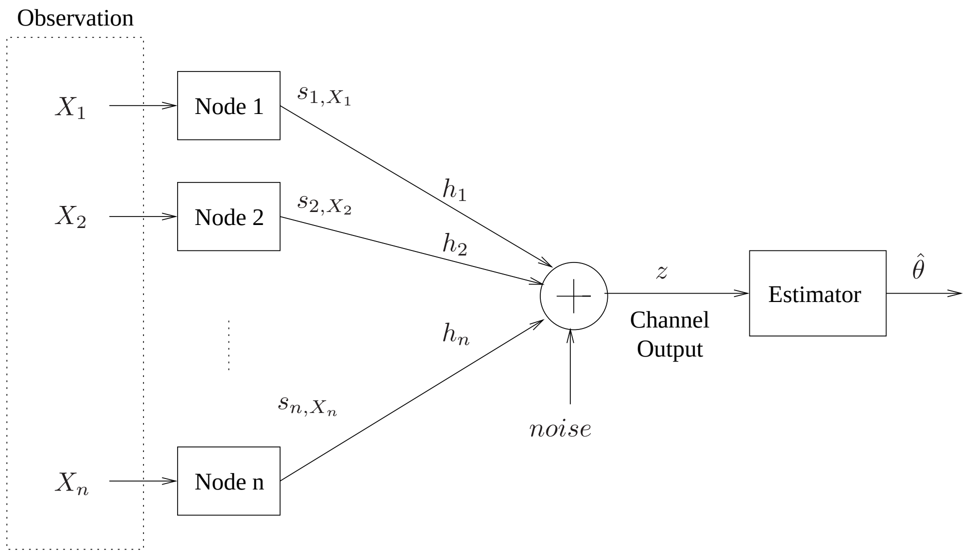

Fig. 1. Estimation over MAC setup.

where $N_j = \sum_{i=1}^{n} 1(X_i = j)$ is the number of nodes that observe j[4]. Sufficiency of $\tilde{p}$ motivates us to use the following scheme, which we shall call the *Type Based Multiple Access*. Let $u_1, \cdots, u_k$ be k orthonormal waveforms. Set $s_{i,X_i} = \sqrt{E} u_{X_i}$, *i.e.,* let every node observing j transmit u_j with energy E. The received signal at the fusion center is modelled as

$$z = \sum_{i=1}^{n} h_i \sqrt{E} u_{X_i} + v, \tag{3}$$

where v is the channel noise. This scheme is easier to understand when all h_i's are equal to 1. In this case (3) simplifies to

$$z = \sum_{j=1}^{k} \sqrt{E} N_j u_j + v. \tag{4}$$

After matched filtering by $u_1, \cdots, u_k$ and scaling by $1/\sqrt{E}n$, it is seen that the received signal contains a noisy version of the empirical measure.

C. Summary of Results

In this paper we provide an asymptotic performance analysis of the TBMA. Our main result states that the TBMA together with a variant of the ML estimator is *asymptotically efficient* if the channel gains from different nodes are identical. In other words, the asymptotic performance of TBMA is as if the fusion center has access to all X_i's directly.

In contrast, the asymptotic MSE of TDMA (time-division multiple access) is shown to scale as $\frac{1}{nJ(\theta)}$, where $J(\theta)$ is considerably smaller than $I(\theta)$ particularly at low SNR. In our model all orthogonal-allocation methods such as TDMA, FDMA and CDMA are mathematically identical. Therefore, our result implies that the TBMA outperforms all such methods with orthogonal allocation.

Another advantage of TBMA over TDMA is its bandwidth requirement; the TBMA uses k orthogonal dimensions irrespective of the number of users, whereas the bandwidth requirement of TDMA grows linearly with n. In a large network, $n \gg k$, this translates to significant bandwidth savings. The performances of TBMA and TDMA are compared by simulations for Bernoulli distributed data. It is observed that the asymptotic analysis provides accurate performance estimates for most of the considered cases even for finite n.

We next generalize the analysis of TBMA to symmetric fading channels ($h_1, \cdots, h_n$ are complex-valued i.i.d. random). For the case that the channel gain has *non-zero* mean $h = \mathbb{E}(h_i)$ and variance[5] $\sigma_h^2 = \text{Var}(\Re(\frac{h_i}{h}))$, it is shown that the estimation error with the proposed ML-variant estimator satisfies

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\text{d}} \mathcal{N}(0, \frac{1 + \sigma_h^2}{I(\theta)}).$$

In other words, fading incurs a loss of performance in MSE which depends on the value of σ_h^2. On the other hand, for circularly-symmetric fading with *zero* mean we obtain a starkly different result: the MSE does not go to zero even though $n \to \infty$. This is because transmissions from different nodes do not add up coherently (transmitted signals cancel each other).

Since band-pass channels are subject to phase uncertainty, zero-mean h_i is expected to be the norm rather than an exception. Thus, some form of transmitter channel side information (CSI) is needed for the TBMA to work in practice (to normalize the channel gains and to set the phases appropriately). By using CSI, the effective channel can be converted from zero-mean to non-zero mean. An estimate of transmitter CSI can be obtained in a distributed way from a pilot tone transmitted by the fusion center (see [2]). We analyze the performance of TBMA with transmitter CSI by simulations and analysis.

[4] $1(\cdot)$ is the *indicator function*, *i.e.,* $1(E) = 1$ if event E happens, and $1(E) = 0$ otherwise

[5] $\Re(\cdot)$, $\Im(\cdot)$ denote the real and imaginary parts of a complex number, respectively.

D. Related Work

Estimation over MAC problem has been previously considered in the context of information theory. Gastpar studied the scaling of distortion with respect to the number of sensors for the case that the sensor observations are noisy versions of a Gaussian source [3]. He showed that transmitting uncoded observations in the Gaussian MAC gives the best scaling law. Other relevant work includes source compression for detection or estimation under communication rate constraints (*e.g.,* [4]–[8]), and joint source-channel coding for the MAC (*e.g.,* [9]–[11]). The information theoretic approach considers the asymptote that the amount of source data/channel resources is large. On the other hand, our setup models the situation that very many sensors each with limited amount of data and finite energy access a common channel.

During the preparation of this work, Liu and Sayeed suggested communicating types [12], and independently proposed the TBMA scheme for distributed detection [13]. The effect of multiaccess protocol on the reconstruction MSE of correlated fields is studied in [14], [15]. Chamberland and Veeravalli [16] studied distributed detection under sum communication rate constraint. Distributed detection over *independent* noisy channels is considered in [17]–[19]. Quantization of observations for transmission over multiaccess channels with discrete input alphabet is investigated in [20]. Quantizer design for distributed estimation is studied in [21]–[23]. High-rate quantization for asymptotically optimal detection is studied in [24].

The organization of the paper is as follows. Section II analyzes the asymptotic performance of TBMA in deterministic and fading channels. Section III discusses the asymptotic performance of TDMA. Section IV gives numerical examples, and checks the validity of the asymptotic theory for finite n. Section V concludes the paper, and points out some further research directions. Proofs of the main theorems concerning the asymptotic efficiency of TBMA are provided in the appendix.

We use the notation p for p_{θ_0}, and $(p_1, \cdots, p_k)$ for $(p_{\theta_0}(1), \cdots, p_{\theta_0}(k))$, where θ_0 is the parameter the data comes from. All vectors, unless transposed, must be understood as column vectors. $(\cdot)'$ denotes the derivative with respect to θ.

II. Asymptotic Performance of TBMA

A. Identical Channels ($h_i = 1$)

Consider the received signal with the TBMA scheme in (4). We assume that the v is white proper-complex Gaussian noise with power $\sigma^2/2$ in each orthogonal dimension. Suppose that the fusion center processes the received signal to obtain

$$y := \frac{1}{\sqrt{En}} \Re\{((z|u_1), \cdots, (z|u_k))\}$$

where $(\cdot|\cdot)$ denotes the inner product. Here, we take the real component of the signal, since its imaginary component only contains noise. The signal y can be equivalently written as

$$y = \frac{1}{n}(N_1, \cdots, N_k) + (w_1, \cdots, w_k) =: \tilde{p} + w,$$

where $\tilde{p}$ is the empirical measure, and $w \sim \mathcal{N}(0, \frac{\sigma^2}{2En^2} I)$.

In the following lemma we characterize the asymptotics of $\tilde{p}$ and y, which actually turn out to be the same.

Lemma 1: $\tilde{p} \xrightarrow{\text{p}} p$ and $\sqrt{n}(\tilde{p} - p) \xrightarrow{\text{d}} \mathcal{N}(0, \Sigma)$ as $n \to \infty$, where

$$\begin{aligned} \Sigma &= \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_2p_k \\ \vdots & & & \\ -p_1p_k & & \cdots & p_k(1-p_k) \end{bmatrix} \\ &= \text{Diag}(p_1, \cdots, p_k) - pp^T. \end{aligned} \tag{5}$$

The same types of convergence hold true for y as well, *i.e.,* $y \xrightarrow{\text{p}} p$ and $\sqrt{n}(y - p) \xrightarrow{\text{d}} \mathcal{N}(0, \Sigma)$ as $n \to \infty$.

Remark 1: Intuitively, the lemma states that y is asymptotically Gaussian with mean p and covariance $\frac{1}{n}\Sigma$. To denote this, we use the notation

$$y \approx \mathcal{N}(p, \frac{1}{n}\Sigma)$$

for large n. This property will be instrumental in establishing the asymptotic efficiency of TBMA. The reason why the vector y and the empirical measure $\tilde{p}$ have the same asymptotics is that the noise term w has power decaying with $1/n^2$.

Proof: It is straightforward to check that the empirical measure has (scaled) multinomial distribution with mean p, and covariance $\frac{1}{n}\Sigma$. We have $\tilde{p} \xrightarrow{\text{p}} p$ from the law of large numbers, and $\sqrt{n}(\tilde{p} - p) \xrightarrow{\text{d}} \mathcal{N}(0, \Sigma)$ from the multivariate Central Limit Theorem [25, p. 385]. It follows from Slutky's Theorem (reviewed in Appendix B) that the addition of noise doesn't change this convergence behavior:

$$y = \tilde{p} + \frac{w}{n\sqrt{E}} \xrightarrow{\text{p}} p$$

since $\frac{w}{n\sqrt{E}} \xrightarrow{\text{p}} 0$. Similarly,

$$\sqrt{n}(y - p) = \sqrt{n}(\tilde{p} - p) + \frac{w}{\sqrt{nE}} \xrightarrow{\text{d}} \mathcal{N}(0, \Sigma).$$

□

Lemma 2: Let y be $\mathcal{N}(p, \frac{1}{n}\Sigma)$ distributed. Then,

i) The probability densiy function (pdf) of $y_1, \cdots, y_{k-1}$ is

$$f(y_1, \cdots, y_{k-1}) = \frac{1}{(2\pi/n)^{\frac{k-1}{2}}} \exp\left(-\frac{n\sum_{i=1}^k \frac{(p_i - y_i)^2}{p_i} + \log\prod_{i=1}^k p_i}{2}\right) \tag{6}$$

ii) $y_k = 1 - \sum_{i=1}^{k-1} y_i$.

Proof: To see (ii), notice that Σ is a singular matrix and $[1 \ \cdots \ 1]\Sigma[1 \ \cdots \ 1]^T = 0$. This implies that $\sum_{i=1}^k y_i$ has zero variance, *i.e.,* it is constant and is equal to its mean $\sum_{i=1}^k p_i = 1$.

The pdf of $(y_1, \cdots, y_{k-1})$ is

$$f(y_1,\cdots,y_{k-1}) = \frac{1}{(2\pi)^{\frac{k-1}{2}}\sqrt{|\Sigma_{(k-1)\times(k-1)}/n|}} \exp\left(-\frac{(y-p)^T(n\Sigma^{-1}_{(k-1)\times(k-1)})(y-p)}{2}\right)$$

$$= \frac{1}{(2\pi)^{\frac{k-1}{2}}\sqrt{n^{-(k-1)}}} \exp\left(-\frac{(y-p)^T(n\Sigma^{-1}_{(k-1)\times(k-1)})(y-p)+\log|\Sigma_{(k-1)\times(k-1)}|}{2}\right)$$

where $\Sigma_{(k-1)\times(k-1)}$ is the upper-left portion of the matrix Σ; similarly, $(y-p)$ should be understood as the first $k-1$ entries. Due to the special structure (5), the determinant can be obtained as

$$\begin{aligned}
&|\Sigma_{(k-1)\times(k-1)}| \\
&= |\mathrm{Diag}(p_1,\cdots,p_{k_1}) - pp^T| \\
&= |\mathrm{Diag}(p_1,\cdots,p_{k-1})|\cdot|I-\mathrm{Diag}(p_1^{-1},\cdots,p_{k-1}^{-1})pp^T| \\
&= |\mathrm{Diag}(p_1,\cdots,p_{k-1})|\cdot|1-p^T\mathrm{Diag}(p_1^{-1},\cdots,p_{k-1}^{-1})p| \quad (7)\\
&= \prod_{i=1}^{k} p_i, \quad (8)
\end{aligned}$$

where p again only refers to its first $k-1$ entries and the identity $|I-AB| = |I-BA|$ is used in (7). Similarly, the inverse matrix can be obtained from the Sherman-Morrison-Woodbury formula [26]:

$$\Sigma^{-1}_{(k-1)\times(k-1)} = D^{-1} + \frac{D^{-1}pp^TD^{-1}}{1-p^TD^{-1}p}, \tag{9}$$

where $D = \mathrm{Diag}(p_1,\cdots,p_{k-1})$. Direct evaluation gives

$$(y-p)^T\Sigma^{-1}_{(k-1)\times(k-1)}(y-p) = \sum_{i=1}^{k}\frac{(p_i-y_i)^2}{p_i}. \tag{10}$$

The lemma follows. □

The lemma gives the pdf of y when its distribution is *exactly* equal to $\mathcal{N}(p, \frac{1}{n}\Sigma)$, whereas for our y this is only asymptotically true. In establishing the asymptotic efficiency of TBMA one would like to consider the ML estimator based on y. However, the exact likelihood function of y has a complicated form, and the ML based on that doesn't seem tractable. This motivates us to consider the ML based on the asymptotic distribution of y, which amounts to maximizing the likelihood $f(y_1,\cdots,y_{k-1})$ in (6) with respect p. This is the same as minimizing the exponent. For large n, the second term $\log\prod_{i=1}^k p_i$ has a negligible effect on the minimization compared to the first one. Therefore, we propose the estimator $\hat\theta$ which minimizes

$$M(\theta) := \sum_{i=1}^{k}\frac{(p_\theta(i)-y_i)^2}{p_\theta(i)} \tag{11}$$

with respect to $\theta\in\Theta$. This can be viewed as an asymptotic version of the ML estimator, and as one would expect from the ML, it is asymptotically efficient.

Theorem 1: Consider a network with the TBMA scheme and the estimator $\hat\theta$ that minimizes $M(\theta)$ with respect to $\theta\in\Theta$. If the family $\{p_\theta : \theta\in\Theta\}$ satisfies certain regularity conditions,[6] then the estimator $\hat\theta$ is consistent and is asymptotically efficient, *i.e.,* $\hat\theta \xrightarrow{\mathrm{p}} \theta_0$ and

$$\sqrt{n}(\hat\theta-\theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \frac{1}{I(\theta_0)}) \tag{12}$$

as $n\to\infty$.

Proof: See Appendix B. □

Remark 2: Theorem 1 holds irrespective of the value of E as long as $E>0$. However, the magnitude of E determines the speed of convergence of MSE to the lower bound $\frac{1}{nI(\theta_0)}$. As it will be demonstrated later by simulations, the higher the SNR (*i.e.,* the higher the E), the faster the convergence.[7]

The intuition behind Theorem 1 is that the asymptotic behaviour of MSE is determined by the *gradient* of $\hat\theta(\cdot)$ with respect to y in the neighborhood $y\approx p$. That is, the Taylor expansion of $\hat\theta$ at $y=p$ gives

$$\hat\theta(y) = \hat\theta(p) + \nabla\hat\theta(p)^T(y-p) + \text{Higher order terms}, \tag{13}$$

where $\nabla\hat\theta(p) = (\frac{\partial\hat\theta}{\partial y_1},\cdots,\frac{\partial\hat\theta}{\partial y_k})\Big|_{y=p}$. Notice that $\hat\theta(p)=\theta_0$, therefore, the rest of the terms in the expansion constitute the estimation error. The dominant factor in the expansion is the linear term. Under the assumption that y has distribution $\mathcal{N}(p,\frac{1}{n}\Sigma)$, the linear term is also Gaussian with variance $\frac{1}{n}\nabla\hat\theta(p)^T\Sigma\nabla\hat\theta(p)$, which is equal to $\frac{1}{nI(\theta_0)}$. See Appendix A for a heuristic derivation of $\nabla\hat\theta(p)$ and the computation of $\frac{1}{n}\nabla\hat\theta(p)^T\Sigma\nabla\hat\theta(p)$.

B. Multiacess Fading Channel

This section analyzes the asymptotic performance of TBMA when $h_1,\cdots,h_n$ are i.i.d. random variables. We first focus on the case that the channel has *non-zero* mean $h := \mathbb{E}(h_i)$. Later, zero-mean circularly-symmetric h_i will be considered. It is assumed that the channel mean h is known at the receiver.

We use the notations $y := \Re\left\{\frac{z}{\sqrt{E}hn}\right\}$ and $u := \Im\left\{\frac{z}{\sqrt{E}hn}\right\}$. Also, let σ_h^2, α_h^2 be the variances of $\Re\{\frac{h_i}{h}\}$ and $\Im\{\frac{h_i}{h}\}$, respectively. The following lemma characterizes the asymptotic distribution of y and u.

Lemma 3: In a fading channel with non-zero mean,

$$y\to p \quad\text{and}\quad \sqrt{n}(y-p)\xrightarrow{\mathrm{d}}\mathcal{N}(0,\Sigma) \tag{14}$$

as $n\to\infty$, where

$$\Sigma = (1+\sigma_h^2)\mathrm{Diag}(p_1,\cdots,p_k) - pp^T. \tag{15}$$

Moreover,

$$u\to 0 \quad\text{and}\quad \sqrt{n}u\xrightarrow{\mathrm{d}}\mathcal{N}(0,\tilde\Sigma), \tag{16}$$

where $\tilde\Sigma = \alpha_h^2\mathrm{Diag}(p_1,\cdots,p_k)$.

Proof: Notice that $y = \bar e + \frac{w}{\sqrt{E}hn}$, where

$$e_i := \Re\{\frac{h_i}{h}\}(1(X_i=1),\cdots,1(X_i=k)) \text{ and } \bar e := \frac{1}{n}\sum_{i=1}^{n} e_i.$$

[6] Given in Appendix B.

[7] Actually, Theorem 1 depends on y only through its asymptotic statistics (cf. Lemma 1). Since, the asymptotic statistics of y are independent of the distribution of the channel noise, Theorem 1 holds true irrespective of the statistics of the noise *i.e.,* the noise need not even be Gaussian.

Observe that $\mathbb{E}(e_i) = p$ and $\mathbb{E}\{(e_i - p)(e_i - p)^T\} = [\Sigma_{jr}]$, where

$$\begin{aligned}\Sigma_{jj} &= \mathbb{E}(\Re\{\frac{h_1}{h}\}1(X_1 = j) - p_j)(\Re\{\frac{h_1}{h}\}1(X_1 = j) - p_j) \\ &= \mathbb{E}(\underbrace{(\Re\{\frac{h_1}{h}\} - p_j)^2}_{(\Re\{\frac{h_1}{h}\}-1+1-p_j)^2} p_j + (-p_j)^2(1-p_j)) \\ &= \mathbb{E}((\Re\{\frac{h_1}{h}\} - 1)^2 p_j + (1-p_j)^2 p_j + (-p_j)^2(1-p_j)) \\ &= p_j\sigma_h^2 + p_j(1-p_j).\end{aligned}$$

In general, it can be shown that

$$\begin{aligned}\Sigma_{jr} &= \mathbb{E}(\Re\{\frac{h_1}{h}\}1(X_1 = j) - p_j)(\Re\{\frac{h_1}{h}\}1(X_1 = r) - p_r) \\ &= \begin{cases} p_j\sigma_h^2 + p_j(1-p_j) & j = r \\ -p_j p_r & j \neq r. \end{cases}\end{aligned}$$

By the law of large numbers, $\bar{e} = \frac{1}{n}\sum_1^n e_i \xrightarrow{\text{p}} \mathbb{E}(e_i) = p$. By the multivariate central limit theorem, $\sqrt{n}(\bar{e} - p) \xrightarrow{\text{d}} \mathcal{N}(0, \Sigma)$. Eqn. (14) follows from Slutky's Theorem.

To obtain (16), consider

$$f_i := \Im\{\frac{h_i}{h}\}(1(X_i = 1), \cdots, 1(X_i = k)) \text{ and } \bar{f} := \frac{1}{n}\sum_{i=1}^{n} f_i.$$

Observe that

$$\mathbb{E}(\Im\{h_i/h\}) = \Im\{\mathbb{E}(h_i/h)\} = 0 \quad \Rightarrow \quad \mathbb{E}(f_i) = 0,$$

and $\mathbb{E}\{f_i f_i^T\} = \tilde{\Sigma}$. Thus, we get $\bar{f} \xrightarrow{\text{p}} 0$ and $\sqrt{n}\bar{f} \xrightarrow{\text{d}} \mathcal{N}(0, \tilde{\Sigma})$. Eqn. (16) follows from Slutky's Theorem. □

The above lemma indicates that the imaginary component $u \approx \mathcal{N}(0, \tilde{\Sigma})$ carries information about p only through its covariance. For large n, the information in $\mathcal{N}(0, \tilde{\Sigma})$ is of secondary importance compared to the information in $y \approx \mathcal{N}(p, \frac{1}{n}\Sigma)$. Therefore, we don't expect any loss in asymptotic performance from an estimator based only on y.

The difference between the fading and non-fading channels is that the Σ given in Lemma 3 is *invertible* when $\sigma_h^2 > 0$, which wasn't the case previously (eqn. 5). For invertible Σ, we have the following analogue of Lemma 2: If the distribution of y was *exactly* $\mathcal{N}(p, \frac{1}{n}\Sigma)$, then its pdf would be

$$f(y_1, \cdots, y_k) = \frac{1}{(2\pi/n)^{k/2}} \exp\left(-\frac{n(y-p)^T\Sigma^{-1}(y-p) + \log|\Sigma|}{2}\right). \quad (17)$$

In the fading channel, we define the estimator $\hat{\theta}$ as the $\theta \in \Theta$ minimizing

$$M(\theta) := (y - p_\theta)^T \Sigma_\theta^{-1}(y - p_\theta),$$

where Σ_θ is the covariance matrix (15) corresponding to θ. The following theorem characterizes the asymptotic performance of TBMA with this estimator.

Theorem 2: Consider the TBMA scheme over a fading MAC with non-zero mean. Under regularity conditions on $\{p_\theta : \theta \in \Theta\}$, the estimator $\hat{\theta}$ is consistent, and the scaled estimation error is asymptotically normal:

$$\sqrt{n}(\hat{\theta} - \theta_0)\xrightarrow{\text{d}}\mathcal{N}(0, \frac{1+\sigma_h^2}{I(\theta_0)}) \quad (18)$$

as $n\to\infty$.

Proof: See Appendix C. □

The theorem says that asymptotically the loss in performance due to fading is an increase in MSE by a factor of $(1 + \sigma_h^2)$.

Remark 3: In the proof of Theorem 2, we show the following more general result which may be of independent interest. Let y be a random vector with distribution a function of θ and n. For some $\{c_\theta \in \mathbb{R}^k : \theta \in \Theta\}$ and a set of invertible matrices $\{\Sigma_\theta : \theta \in \Theta\}$, suppose that

$$y\xrightarrow{\text{p}}c_\theta \quad \text{and} \quad \sqrt{n}(y - c_\theta)\xrightarrow{\text{d}}\mathcal{N}(0, \Sigma_\theta) \quad \text{as} \quad n\to\infty \quad (19)$$

for all $\theta \in \Theta$. Then, under some regularity conditions on $\{c_\theta\}$ and $\{\Sigma_\theta\}$, the estimator $\hat{\theta}$ minimizing $M(\theta) = (y - c_\theta)^T\Sigma_\theta^{-1}(y - c_\theta)$ is consistent and satisfies

$$\sqrt{n}(\hat{\theta} - \theta)\xrightarrow{\text{d}}\mathcal{N}(0, \frac{1}{(c'_\theta)^T\Sigma_\theta^{-1}c'_\theta}). \quad (20)$$

We next consider the case that the channel has circularly symmetric distribution, *i.e.,* $h_i \sim re^{j\rho}$ i.i.d. over i, where $\rho = \text{Uniform}[0, 2\pi]$ and r is a real valued random variable with $\mathbb{E}r^2 := \sigma_h^2$. Due to the uniform phase, circularly symmetric channels can be viewed as *channels with phase uncertainty*. The phase uncertainty exists in some channels naturally (*e.g.,* in Rayleigh fading). More generally, in band-pass wireless communications phase uncertainty is created by the phase difference between the modulator and demodulator clocks, and by the propagation delay [27]. It can be avoided only if the transmitters' and the receiver's clocks synchronize (possibly via transmitter CSI).

Lemma 4: Consider the TBMA scheme in a channel with circularly symmetric distribution. The scaled signal $y = \frac{z}{\sqrt{nE\sigma_h^2}}$ converges in distribution to $\mathcal{CN}(0, \text{Diag}(p_1, \cdots, p_k))$.

Proof: A complex random vector Y is called *proper* if $\mathbb{E}(YY^T) = 0$.[8] We need the following version of Central Limit Theorem (CLT): Let $Y_1, Y_2, \cdots$ be i.i.d. zero-mean proper complex random vectors. If $\mathbb{E}(Y_iY_i^H) = \Sigma$, then $\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i \xrightarrow{\text{d}} \mathcal{CN}(0, \Sigma)$ as $n\to\infty$. This result is straightforward to deduce from the original CLT [25], when one views complex random vectors in $\mathbb{C}^k$ as elements of $\mathbb{R}^{2k}$.

The $Y_i := \frac{1}{\sigma_h}h_i\delta_{X_i}$ are circularly symmetric ($\Rightarrow$ proper), and $\mathbb{E}(Y_iY_i^H) = \text{Diag}(p)$. Hence, the CLT implies $\frac{1}{\sqrt{n}\sigma_h}\sum_{i=1}^n h_i\delta_{X_i}\xrightarrow{\text{d}}\mathcal{CN}(0, \text{Diag}(p))$. The lemma follows from Slutky's Theorem □

Lemma 4 points out that TBMA does not work in a circularly symmetric channel. That is, $y\xrightarrow{\text{d}}\mathcal{CN}(0, \text{Diag}(p))$ and one expects the asymptotic MSE of any unbiased estimator to be $\geq \frac{1}{I(\theta)}$, where $I(\theta) < \infty$ is the Fisher information in

[8]This definition of properness is equivalent to the more conventional one which says that $Y = Y_R + iY_I$ is proper if $\mathbb{E}(Y_RY_R^T) = \mathbb{E}(Y_IY_I^T)$ and $\mathbb{E}(Y_RY_I^T) = -\mathbb{E}(Y_IY_R^T)$. To see this, notice that $\mathbb{E}(YY^T) = 0 \Leftrightarrow \mathbb{E}(Y_RY_R^T) - \mathbb{E}(Y_IY_I^T) + i[\mathbb{E}(Y_RY_I^T) + \mathbb{E}(Y_IY_R^T)] = 0$.

$\mathcal{CN}(0, \text{Diag}(p))$. This also implies that the MSE cannot go to zero even though $n \to \infty$. Notice that even the signal scalings are different for $\mathbb{E}(h_i) \neq 0$ and $\mathbb{E}(h_i) = 0$; the y was defined $\propto \frac{z}{n}$ for the case $\mathbb{E}(h_i) \neq 0$, while $y \propto \frac{z}{\sqrt{n}}$ in the above lemma.

C. Channel Side Information (CSI) at the Transmitter

In this section, we consider the fading channel with CSI available at the transmitter. It is assumed that every node knows *its own* channel state, and can control its transmission power and phase as a function of that. In practice, an estimate of channel gains (*i.e.,* the transmitter CSI) can be obtained in a distributed way from a pilot tone transmitted by the receiving node.

Suppose that the channel gain of user i is $h_i := r_i e^{j\rho_i}$, where $r_1, \cdots, r_n \in \mathbb{R}$ and $\rho_1, \cdots, \rho_n \in [0, 2\pi]$ are i.i.d. random variables (ρ_i's need not be Uniform$[0, 2\pi]$). With the help of transmitter CSI, nodes can control their transmissions to cancel the phase ρ_i and to normalize the gain r_i. That is, let the transmitted signal by the i'th node be $P(r_i)e^{-j\rho_i}\sqrt{E}\delta_{X_i}$ in TBMA, where $P(\cdot)$ is a *power control rule* satisfying the energy constraint $\mathbb{E}_{r_i}[P^2(r_i)] \leq 1$.

One possibility for power control is to invert the channel, *i.e.,* to set $P(r) = \frac{1}{r}$. Channel inversion effectively converts the fading channel into a non-fading one. Under such a rule, the results of Sec. II-A apply, and the asymptotic efficiency is achieved despite fading. For certain channels (*e.g.,* Rayleigh distributed r_i's), however, the channel inversion requires infinite energy ($\mathbb{E}_r[1/r^2] = \infty$). Because of this reason, we shall consider the following, more general class of power control rules:

$$P(r) = \begin{cases} \alpha/r, & r \in [\beta, \infty) \\ \gamma, & \text{otherwise} \end{cases}$$

where $\alpha, \beta, \gamma \in \mathbb{R}$ are constants independent of r. Under this scheme the network performance can be analyzed using the tools from Sec. II-B. That is, let $\tilde{h}_i := r_i P(r_i)$ be the *effective* channel gain of user i (*i.e.,* the channel gain seen by the receiver). Set $\tilde{h} := \mathbb{E}(\tilde{h}_i)$ and $\sigma_{\tilde{h}}^2 := \text{Var}(\tilde{h}_i)$. Notice that

$$\text{Var}(\Re\{\frac{\tilde{h}_i}{\tilde{h}}\}) = \text{Var}(\frac{\tilde{h}_i}{\tilde{h}}) = \mathbb{E}(\frac{\tilde{h}_i}{\tilde{h}} - 1)^2 = \frac{\sigma_{\tilde{h}}^2}{\tilde{h}^2}.$$

According to Theorem 2, the asymptotic MSE of TBMA with power control is $(1 + \frac{\sigma_{\tilde{h}}^2}{\tilde{h}^2})/nI(\theta)$.

Notice that $1 + \frac{\sigma_{\tilde{h}}^2}{\tilde{h}^2}$ is greater than or equal to 1, which corresponds to the case without fading (*i.e.,* the best case). A relevant question is how small can we make $1 + \frac{\sigma_{\tilde{h}}^2}{\tilde{h}^2}$ with the choice of α, β, γ, while satisfying the energy constraint $\mathbb{E}_r[P^2(r)] \leq 1$. The following lemma states that it can be made as close to 1 as possible by choosing α and β small enough.

Lemma 5: i) Consider a channel with the property $\mathbb{E}_r[1/r^2] < \infty$. Then, the parameters $\beta = 0$ and $\alpha = \frac{1}{\mathbb{E}_r[1/r^2]}$ satisfy $1 + \frac{\sigma_{\tilde{h}}^2}{\tilde{h}^2} = 1$.

ii) Suppose that $\mathbb{E}_r[1/r^2] = \infty$ and $\Pr\{r = 0\} = 0$. Let γ be equal to zero. For any given $\beta > 0$, choose α to satisfy $\mathbb{E}_r[P^2(r)] = 1$. Then, as $\beta \to 0$,

$$\alpha \to 0 \quad \text{and} \quad 1 + \frac{\sigma_{\tilde{h}}^2}{\tilde{h}^2} \to 1.$$

Proof: Part i) is obvious; only part ii) needs to be proved. Let $f(\cdot)$ be the pdf of r. For $\gamma = 0$, the power constraint is $\mathbb{E}_r[P^2(r)] = \int_\beta^\infty \frac{\alpha^2}{r^2} f(r) dr = 1$. Therefore, $\alpha \to 0$ as $\beta \to 0$. Furthermore, observe that

$$1 + \frac{\sigma_{\tilde{h}}^2}{\tilde{h}^2} = \frac{\mathbb{E}(\tilde{h}_i^2)}{\mathbb{E}^2(\tilde{h}_i)} = \frac{\alpha^2 \Pr\{r \in [\beta, \infty)\}}{(\alpha \Pr\{r \in [\beta, \infty)\})^2}$$

converges to $1/\Pr\{r \in (0, \infty)\} = 1$ as $\beta \to 0$. □

The above lemma says that with transmitter CSI the best asymptotic performance can still be achieved despite the existence of fading (as long as $\Pr\{r = 0\} = 0$). We saw in Sec. II-B that the phase uncertainty hinders the operation of TBMA. The results of this section, however, indicate that the phase uncertainty and, in general, fading issues can be resolved with the help of transmitter CSI.

Note that in the non-asymptotic regime (*i.e.,* finite n), the SNR of the received signal also makes a difference in MSE (the higher the received SNR, the closer the MSE to the Cramer-Rao Bound). The received SNR is proportional to the α and γ in the above power control scheme. Therefore, the α, γ minimizing the MSE for finite n need not be negligibly small as suggested by the above lemma.

III. Asymptotic Performance of TDMA

Packetization is a common practice in communication network design. The conventional layered architecture suggests the data to be mapped into a bit-stream, transmitted using some form of modulation, and then get received without collisions. In this section, the asymptotic performance of such an approach (the TDMA scheme) is considered.

For some $m \in \{1, 2, \cdots\}$, let $s_1, \cdots, s_k$ be vectors in $\mathbb{C}^m$ satisfying $||s_i||^2 \leq 1$ (these are viewed as points in a *constellation*). In TDMA users are allocated non-overlapping time slots of length m. Every node uses time shifted versions of the same set of waveforms in its own slot, *i.e.,* node i transmits vector $\sqrt{E}s_{X_i}$ in the i'th slot . Notice that the bandwidth requirement of TDMA linearly grows with number of users, whereas the TBMA uses k time units irrespective of the number users. We denote the i'th received TDMA packet by

$$z^{(i)} = h_i \sqrt{E} s_{X_i} + v^{(i)}, \tag{21}$$

where $v^{(i)} \sim \mathcal{CN}(0, \sigma^2 I_{m \times m})$.

For simplicity, first let's focus on the case that $h_i = 1$ for all i. The random vectors $z^{(1)}, \cdots, z^{(n)}$ are i.i.d. with *Gaussian mixture* density. That is, $z^{(i)}$ is distributed $\mathcal{CN}(\sqrt{E}s_j, \sigma^2 I)$ conditional on $X_i = j$, and the pdf of $z^{(i)}$ is

$$f(z^{(i)}) = \sum_{j=1}^{k} p_j \frac{1}{(\pi\sigma^2)^m} \exp\left(-\frac{||z^{(i)} - \sqrt{E}s_j||^2}{\sigma^2}\right). \tag{22}$$

From the CRB, the MSE of any unbiased estimator $\hat{\theta}$ based on $z^{(1)}, \cdots, z^{(n)}$ satisfies

$$\mathbb{E}\{(\hat{\theta} - \theta)\} \geq \frac{1}{nJ(\theta)},$$

where

$$J(\theta) = \mathbb{E}_{z^{(i)}}\left[\left(\frac{d\log f(z^{(i)})}{d\theta}\right)^2\right] \tag{23}$$

is the Fisher information in $z^{(i)}$. Moreover, the ML estimator based on $z^{(1)},\cdots,z^{(n)}$ achieves the MSE $\frac{1}{nJ(\theta)}$ asymptotically. Hence, the asymptotic performance of TDMA is determined by the Fisher information $J(\theta)$ in each TDMA packet.

An important problem is the choice of constellation. One ideally would like to maximize $J(\theta)$ in (23) with respect to $s_1,\cdots,s_k$ to get the best performance. This maximization does not appear tractable in general. For $k=2$, however, we have the following result.

Theorem 3: Let the channel $h_i = 1$ for all i. For $k=2$, the antipodal constellation $s_1 = 1,\ s_2 = -1$ maximizes $J(\theta)$ under the energy constraint $||s_i||^2 \le 1$.

Proof: Let s_1, s_2 be any constellation satisfying the energy constraint. Consider a received TDMA packet $z^{(i)} = \sqrt{E}s_{X_i} + v \in \mathbb{C}^m$. Invertible mappings of $z^{(i)}$ preserve the Fisher information. First, scale $z^{(i)}$ to get $\tilde{z}^{(i)} := s_{X_i} + v/\sqrt{E}$. Then, translate $\tilde{z}^{(i)}$ by $(s_1+s_2)/2$ to have the center of mass at the origin, *i.e.,* $\hat{s}_1 = -\hat{s}_2$, where

$$\tilde{z}^{(i)} \ \mapsto\ \tilde{z}^{(i)} - \frac{s_1+s_2}{2} = \underbrace{(s_{X_1} - \frac{s_1+s_2}{2})}_{:=\hat{s}_{X_i}} + \frac{v}{\sqrt{E}} := \hat{z}^{(i)}.$$

Let $2a := ||s_1 - s_2|| = ||\hat{s}_1 - \hat{s}_2||$. Multiply $\hat{z}^{(i)}$ by an orthogonal matrix U (*i.e.,* a rotation matrix) with the first row $\frac{s_1^* - s_2^*}{2a}$ ($(\cdot)^*$ denotes conjugation). Notice that $U\hat{s}_1 = (a, 0, \cdots, 0)^T$, $U\hat{s}_2 = (-a, 0, \cdots, 0)^T$ and $Uv \sim \mathcal{CN}(0, \sigma^2 I)$. Only the first component of $U\hat{z}^{(i)}$ conveys information (*i.e.,* it is a sufficient statistic), and the rest of $U\hat{z}^{(i)}$ can be dropped without any loss in Fisher information. Thus, we have just shown that without loss of generality we can focus on one dimensional, symmetric constellations. Scale the first component of $U\hat{z}^{(i)}$ by $\frac{1}{a}$ to get $\bar{z}^{(i)} := \bar{s}_{X_i} + \frac{\bar{v}}{a\sqrt{E}}$, where $\bar{s}_1 := 1,\ \bar{s}_2 := -1,\ \bar{v} \sim \mathcal{CN}(0, \sigma^2)$. Next, we will show that $a = 1$ maximizes the Fisher information in $\bar{z}^{(i)}$. Let $\bar{z}_*^{(i)}$ denote the $\bar{z}^{(i)}$ corresponding to $a = 1$. Observe that the $\bar{z}^{(i)}$ for any $a < 1$ can be viewed as a *degraded* version of $\bar{z}_*^{(i)}$. That is, $\bar{z}^{(i)} \sim \bar{z}_*^{(i)} + v_*$ for some $v_* \sim \mathcal{CN}(0, \frac{\sigma^2(1-a^2)}{Ea^2})$ independent of $\bar{z}_*^{(i)}$. A result analogous to the *data processing inequality* in information theory is that processing of $\bar{z}_*^{(i)}$ reduces the Fisher information (irrespective of whether the processing is random or deterministic) [28, p.138]. Hence, the theorem follows. □

For Bernoulli($\theta = 0.8$) distributed data and antipodal constellation, $J(\theta)$ and $I(\theta)$ are plotted in Fig. 2. Notice that the asymptotic MSE of TDMA, $\frac{1}{nJ(\theta)}$, is significantly higher than the asymptotic MSE of TBMA, $\frac{1}{nI(\theta)}$, at low SNR $:= \frac{E}{\sigma^2/2}$. It is also seen that in terms of asymptotic MSE the TDMA is as good as TBMA for SNR $\ge$ 20dB.

For $k = 8$, we evaluated $J(\theta)$ by Monte-Carlo integration for three types of constellation: BPSK, orthogonal s_i's and the simplex (orthogonal s_i's translated to have center of mass at the origin, and scaled to satisfy $||s_i|| = 1$). Fig. 3 shows $J(\theta)$ for Poisson (mean$=\theta = 1$) X_i truncated at $k = 8$. The simplex constellation is observed to be marginally better than the other two at all SNR values.

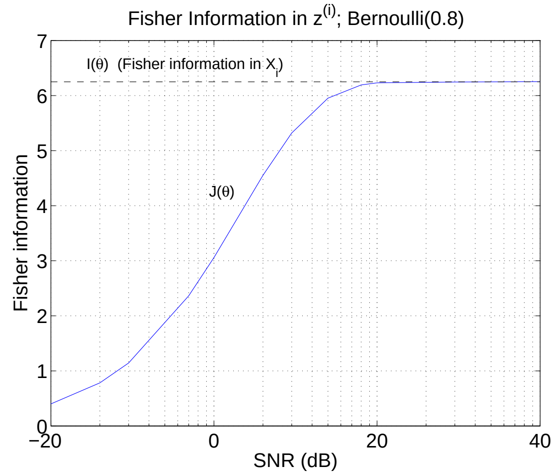

Fig. 2. Fisher Information in TDMA packets (Bernoulli data).

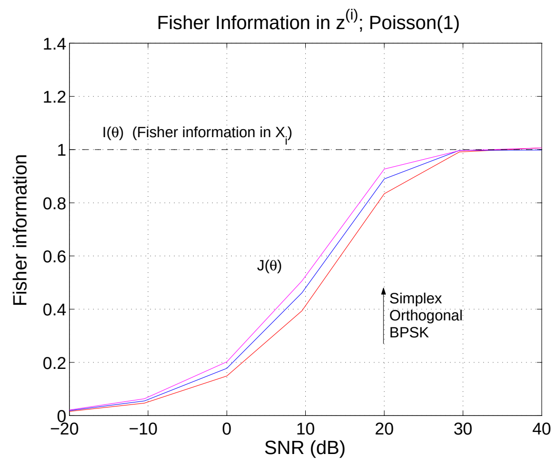

Fig. 3. Fisher Information in TDMA packets (Poisson data).

The discussion for the case that $h_i = 1$ can be extended to fading channels. When the channels $h_1,\cdots,h_n$ are i.i.d. random, the $z^{(i)}$ is distributed $\mathcal{CN}(h_i\sqrt{E}s_j, \sigma^2 I)$ conditional on $X_i = j$ and h_i. The pdf of $z^{(i)}$ can be expressed similar to (22), where this time the averaging is with respect to both X_i and h_i. Again, the asymptotic performance is determined by the Fisher information in $z^{(i)}$.

So far, we haven't considered any receiver CSI. In case the receiver has CSI, the estimation can be done based on $z^{(1)},\cdots,z^{(n)}$ and $h_1,\cdots,h_n$. The $(z^{(i)},h_i)$ are i.i.d. for different i, and the asymptotic performance is determined by the Fisher information $\hat{J}(\theta)$ in $(z^{(i)},h_i)$. Here, an important observation is that the pdf of $(z^{(i)},h_i)$ can be decomposed as $f(z^{(i)}|h_i)f(h_i)$, and

$$\hat{J}(\theta) \quad = \quad \mathbb{E}_{z^{(i)},h_i}\left[\left(\frac{d\log f(z^{(i)},h_i)}{d\theta}\right)^2\right]$$

$$= \mathbb{E}_{z^{(i)},h_i}\left[\left(\frac{d\log f(z^{(i)}|h_i)}{d\theta} + \underbrace{\frac{d\log f(h_i)}{d\theta}}_{=0}\right)^2\right]$$
$$= \mathbb{E}_{h_i}\mathbb{E}_{z^{(i)}|h_i}\left[\left(\frac{d\log f(z^{(i)}|h_i)}{d\theta}\right)^2\right]. \quad (24)$$

Hence, the Fisher information in fading channels with receiver CSI is same as the averaged Fisher information conditioned on the fading realization. For BPSK, orthogonal s_i and simplex, this means that $\hat{J}(\theta)$ can be obtained by averaging the values in Fig. 2 with respect to the SNR corresponding to the fading realization. The $\hat{J}(\theta)$ can also be viewed as an upper bound to the Fisher information without receiver CSI.

Finally, we would like to note that the above discussion can be extended to fading with transmitter CSI by changing h_i by $\tilde{h}_i$ as defined in Sec. II-C.

IV. Numerical Examples

A. Bernoulli Distributed Data (θ = mean)

Simulation results for $h_i = 1$ are given in Fig. 4. The curves are the following:

i) TDMA (antipodal constellation) with the ML estimator based on $z^{(1)}, \cdots, z^{(n)}$.
ii) Direct Access+ML: The hypothetical case that the estimator has access to X_i's directly.
iii) Asymptotic performance (of TBMA): MSE=$\frac{1}{nI(\theta)}$ predicted by our theory.
iv) The TBMA (SNR $= \frac{E}{\sigma^2/2} =$ 0dB) follows the expected asymptotic performance closely, whereas TBMA (SNR=-10dB) reaches the asymptotic limit only at large n.

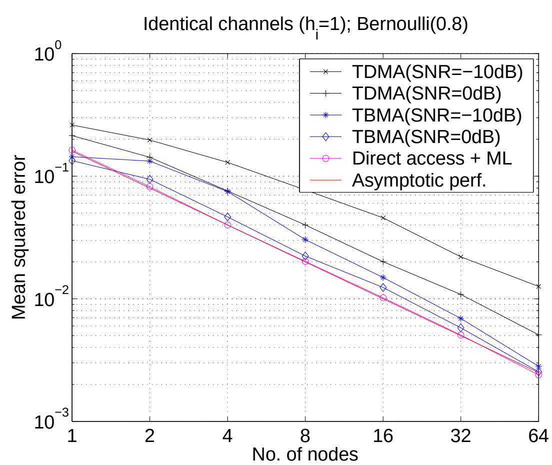

Fig. 4. MSE of TBMA, TDMA and direct access.

In the TDMA scheme the ML estimator based on $z^{(1)}, \cdots, z^{(n)}$ is not always computationally feasible. For $h_i = 1$, we were able to compute it because the likelihood has a simple, tractable form (eqn. 22). In case of fading, however, without the receiver CSI the likelihood involves an integration with respect to h_i. This generally makes its computation intractable, and motives us to find estimators other than the ML estimator. In Fig. 5 we plot the MSE of TDMA under three different estimators for $h_i = 1$:

i) Detect: A hard-decoder $\hat{X}_1, \cdots, \hat{X}_n$ (ML detector for channel symbols) followed by the estimator maximizing $\prod_{i=1}^n p_\theta(\hat{X}_i)$.
ii) Detect+MLE: Let q be the probability that $X_i \neq \hat{X}_i$. Because of channel errors, the distribution of $\hat{X}_i$ is $(\hat{p}_\theta(1), \hat{p}_\theta(2)) = ((1-q)p_\theta(1) + qp_\theta(2), (1-q)p_\theta(2) + qp_\theta(1))$. The estimator maximizes $\prod_{i=1}^n \hat{p}_\theta(\hat{X}_i)$ with respect to the θ.
iii) MLE: The ML estimator based on $z^{(1)}, \cdots, z^{(n)}$.

From Fig. 5 it is seen that these three estimators perform nearly the same when $h_i = 1$.

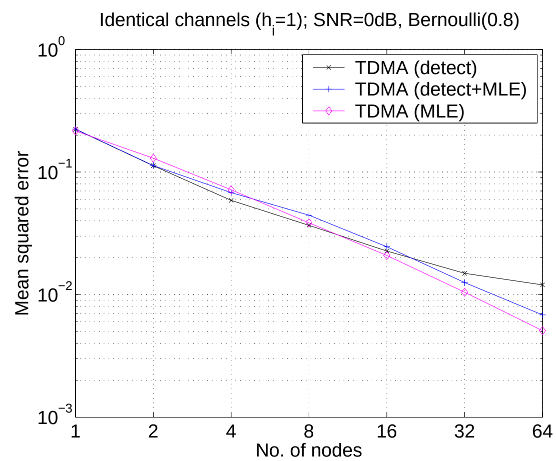

Fig. 5. Comparison of different estimators for TDMA

We next consider the case that the $h_i \sim \mathcal{CN}(0,1)$ are Rayleigh distributed (SNR = 0dB), and each transmitter has CSI for its own h_i. After trial and error on α, β, γ, we found that the following power control rule performs reasonably well:

$$P(r) = \begin{cases} 1/r, & r \in [P_0, \infty) \\ 1/P_0, & \text{otherwise} \end{cases}$$

where the parameter P_0 is chosen to meet the energy constraint $\mathbb{E}(P^2(r)) = 1 \Rightarrow P_0 \approx 0.89$. The transmitted signal is $P(r_i)e^{-j\rho_i}\sqrt{E}\delta_{X_i}$ in TBMA, and $P(r_i)e^{-j\rho_i}\sqrt{E}s_{X_i}$ in TDMA. The performance of these two schemes are given in Fig. 6. The figure also shows the asymptotic MSE of TBMA with power control predicted by the theory $(1 + \sigma_{\tilde{h}}^2/\tilde{h}^2 \approx 1.111)$. In TDMA, the previously mentioned detect+ML estimator is used, because the ML estimator based on $z^{(1)}, \cdots, z^{(n)}$ is not tractable.

In Fig. 7 the TDMA and TBMA schemes are compared in a Rayleigh fading channel without transmitter CSI (*i.e.,* no phase/power control). An advantage of Rayleigh channel is that we can express the pdf of the received signal compactly, and apply the ML estimator both for TBMA and TDMA. For example, in TBMA the distribution of z conditional on θ is

$$z \sim \sum_{i=0}^{n} \binom{n}{i} p_\theta^i(1) p_\theta^{n-1}(2)$$
$$\mathcal{CN}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix} iE+\sigma^2 & 0 \\ 0 & (n-i)E+\sigma^2 \end{bmatrix}\right).$$

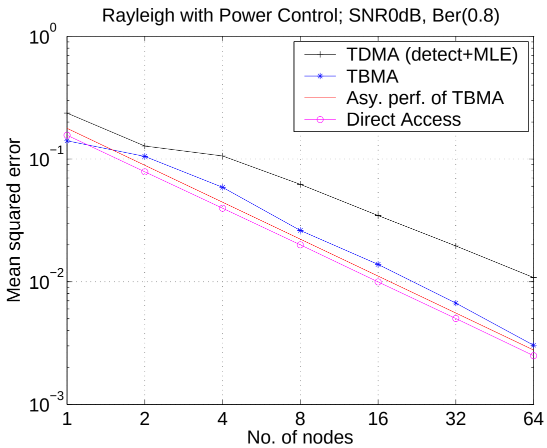

Fig. 6. Rayleigh fading with power control.

The orthogonal modulation is used in TDMA. As elaborated in Sec. II-B, the MSE of the TBMA method, even with the exact ML estimator, does not go to zero as $n\to\infty$. The MSE of TDMA, however, does go to zero.

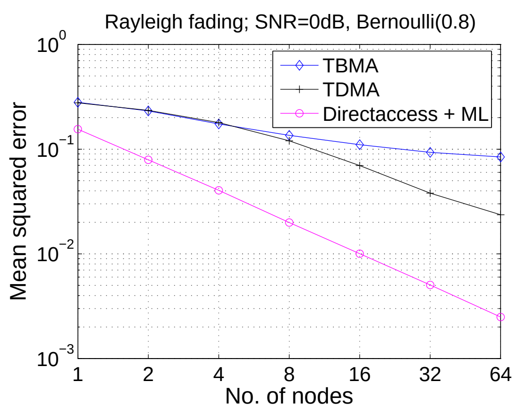

Fig. 7. Rayleigh fading without power control.

V. Conclusions

Communication for the purpose of estimation is a central issue in sensor networks. We studied the problem of parameter estimation for the case that a large number of sensors each with limited data and limited transmission energy access a common channel. We argued that with the use of TBMA significant gains can be realized in estimation quality and in system resource consumption compared to the conventional architecture allocating orthogonal channels to sensors. We characterized the asymptotic performance of TBMA, and observed that this characterization gives reasonably accurate performance estimates even for finite n. Note that the TBMA can also be used in a non-parametric setting for histogram estimation in case the family $\{p_\theta : \theta \in \Theta\}$ is unknown.

In this paper we have considered an individual power constraint for each sensor. This means that the total energy consumed goes to infinity as the network size grows. Our analysis can be extended to the network with *sum-power* constraint in a straightforward manner.[9] For example, if each node has power constraint E/n and all h_i's are equal to one, then the received signal satisfies

$$\Re[(z|u_1)\ \cdots\ (z|u_k)] = \sqrt{\frac{E}{n}}[N_1\ \cdots\ N_k] + w, \tag{25}$$

where $w \sim \mathcal{N}(0, \sigma^2 I/2)$. If we let y be $1/\sqrt{nE}$ times (25), then by following the steps of Lemma 1 we get

$$y \approx \mathcal{N}(p, \frac{\Sigma + \frac{\sigma^2}{2E}I}{n}), \qquad \Sigma = \text{Diag}(p) - pp^T,$$

as $n\to\infty$. Using the Remark after Theorem 2, it is seen that the MSE scales as

$$\frac{1}{n}\frac{1}{(p')^T(\Sigma + \frac{\sigma^2}{2E}I)^{-1}p'}.$$

In case of fading with non-zero mean, the same expression is obtained, where $\Sigma = (1+\sigma_h^2)\text{Diag}(p) - pp^T$. Surprisingly, despite the bounded sum-power, the MSE still goes to zero as $n\to\infty$. In this regime the asymptotic performance also depends on the noise power.

In their recent work [29], Liu and Sayeed showed that the asymptotic detection performance of TBMA is optimal when all h_i's are identical to deterministic and identical. In particular, they showed that the error exponents of TBMA in Bayesian hypothesis testing is as if the fusion center has direct access to data. In [30], we characterized the error exponents for the case that the sensor channels have i.i.d. fading for an asymptotic version of the ML detector.

Some notable future research directions are the following:

- As argued in Section II-B, the TBMA fails to deliver the empirical measure when the channel has zero mean. Transmitter CSI can be used to solve this problem. However, in certain cases (*e.g.*, in case of a mobile receiver) the channel may vary too fast to be tracked, and it may not be possible to obtain transmitter CSI. Strategies other than TBMA are needed for zero-mean fading.
- Another issue is the choice of the estimator for TBMA. The proposed estimator minimizing $M(\theta)$ is equivalent to ML estimator for large n, and is tractable. However, for finite n the problem of finding estimators better than the one proposed seems to deserve further attention.
- Optimal quantization of *continuous* variables is also important. One idea is to quantize such that the Fisher information in the quantized variable is maximized. However, the optimal number of quantization levels (possibly, as a function of n) and the structure of optimal quantizers are unknown.

Appendix

A. Derivation of the Gradient

The distribution family $\{p_\theta : \theta \in \Theta\}$ traces a curve in the k dimensional probability simplex (Fig. 8). For small perturbations,

$$\hat{\theta} = \theta_0 + e, \text{ where } |e| << 1,$$

[9] We would like thank to an anonymus reviewer who raised this question.

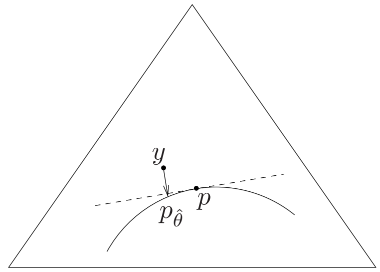

Fig. 8. Probability simplex and the parametric family $\{p_\theta : \theta \in \Theta\}$.

the $p_{\hat\theta}$ varies almost linearly along the tangent line:

$$p_{\hat\theta} \approx p + ep' \quad \text{and} \quad p'_{\hat\theta} \approx p' + ep'', \tag{26}$$

where $p' := p'_{\theta_0}$, $p'' := p''_{\theta_0}$. This property is very important for us, because e goes to zero as $y \to p$, and (26) becomes accurate. Instead of minimizing $M(\theta)$, one might as well solve for θ in $M'(\theta) = 0$.

$$\begin{aligned} M'(\theta) &= \sum_{i=1}^k \frac{2p'_\theta(i)(p_\theta(i) - y_i)}{p_\theta(i)} - \frac{p'_\theta(i)(p_\theta(i) - y_i)^2}{p_\theta(i)^2} \\ &\approx \sum_{i=1}^k \frac{2p'_\theta(i)(p_\theta(i) - y_i)}{p_\theta(i)}, \end{aligned} \tag{27}$$

where the approximation is because $(p_\theta(i) - y_i)^2$ is much smaller than $(p_\theta(i) - y_i)$ for $p_\theta \approx y$. Substituting (26), one gets

$$\begin{aligned} M'(\hat\theta) &\approx \sum_{i=1}^k \frac{2(p'_i + ep''_i)(p_i - y_i + ep'_i)}{p_i + ep'_i} \\ &\approx \sum_{i=1}^k \frac{2p'_i(p_i - y_i + ep'_i)}{p_i} = 0, \end{aligned} \tag{28}$$

where higher order terms with e^2 and $e(p_i - y_i)$ are neglected. From the last equation, e is obtained as

$$\begin{aligned} (\sum_{i=1}^k \frac{(p'_i)^2}{p_i})e &= \sum_{i=1}^k \frac{p'_i(y_i - p_i)}{p_i} \\ \Rightarrow \quad e &= \frac{1}{I(\theta_0)} \sum_{i=1}^k \frac{p'_i(y_i - p_i)}{p_i}. \end{aligned} \tag{29}$$

Here, e is the dominating term in the expansion (13), and we get $\nabla\hat\theta(p) = \frac{1}{I(\theta_0)}(p'_1/p_1, \cdots, p'_k/p_k)$. Under the assumption that $y \sim \mathcal{N}(p, \frac{1}{n}\Sigma)$, e is Gaussian with variance

$$\begin{aligned} \mathbb{E}(e^2) &= \frac{1}{nI^2(\theta_0)} \begin{bmatrix} \frac{p'_1}{p_1} & \cdots & \frac{p'_k}{p_k} \end{bmatrix} \\ &\qquad \left(\text{Diag}(p) - pp^T\right) \begin{bmatrix} \frac{p'_1}{p_1} & \cdots & \frac{p'_k}{p_k} \end{bmatrix}^T \\ &= \frac{1}{nI(\theta_0)}. \end{aligned}$$

B. Proof of Theorem 1

Let $\theta_0 \in \Theta$ be the parameter to be estimated. The estimator observes $y = \tilde p + \frac{w}{\sqrt{En}} \in \mathbb{R}^k$, where $\tilde p$ is the empirical measure and $w \sim \mathcal{N}(0, \sigma^2 I)$. Let $\hat\theta$ be the θ minimizing

$$M(\theta) = \sum_{i=1}^k \frac{(p_\theta(i) - y_i)^2}{p_\theta(i)}.$$

In this appendix, we prove the consistency and asymptotic efficiency of $\hat\theta$ (Theorems 4 and 7, respectively). The proof of Theorem 7 involves significant amount of derivation, which are given as Lemmas 6-9. The proof of the theorem can be read without reading the proofs of lemmas.

Let point $\theta_0 \in \Theta$ be identifiable in the following sense: $\forall\epsilon > 0$, $\exists\delta > 0$ such that

$$||p_\theta - p_{\theta_0}|| < \delta \text{ for some } \theta \in \Theta \quad \Rightarrow \quad |\theta - \theta_0| < \epsilon, \tag{30}$$

i.e., points close in the simplex are also close in Θ. Pictorially, $\forall\epsilon > 0$ one can draw a small enough ball around p_{θ_0} such that all p_θ within the ball satisfies $|\theta - \theta_0| < \epsilon$ (Fig. 9).

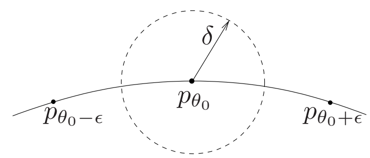

Fig. 9. Illustration of (30).

Theorem 4: If θ_0 satisfies (30), and $p_{\theta_0}(i) > 0$ $\forall i$, then the estimator $\hat\theta$ is consistent (*i.e.,* $\hat\theta \xrightarrow{\text{p}} \theta_0$).

Proof: Let B_δ denote the set of p_θ satisfying $||p_\theta - p_{\theta_0}|| < \delta$. We need to show that

$$\Pr\{|\hat\theta - \theta_0| > \epsilon\} \to 0 \tag{31}$$

for all $\epsilon > 0$. Fix $\epsilon > 0$, and assume that δ satisfies (30). We will argue that $\exists\delta' > 0$ such that $\delta' < \delta$ and

$$\text{if } y \in B_{\delta'} \quad \Rightarrow \quad \hat\theta \in B_\delta. \tag{32}$$

This is enough for (31), since $y \xrightarrow{\text{p}} p_{\theta_0}$ by Lemma 1 (which implies $\Pr\{y \in B_{\delta'}\} \to 1$).[10] To prove (32), it suffices to show that $\exists$ small enough $\delta' > 0$ such that $\forall\ y \in B_{\delta'}$

$$M(\theta_0) < M(\theta), \qquad \forall p_\theta \notin B_\delta.$$

Observe

$$\begin{aligned} M(\theta_0) &= \sum_{i=1}^k \frac{(p_{\theta_0}(i) - y_i)^2}{p_{\theta_0}(i)} \\ &\le \frac{1}{\min_{i=1,\cdots,k} p_{\theta_0}(i)} \sum_{i=1}^k (p_{\theta_0}(i) - y_i)^2 \\ &\le \frac{(\delta')^2}{\min_{i=1,\cdots,k} p_{\theta_0}(i)}, \end{aligned} \tag{33}$$

where the last step is because $y \in B_{\delta'}$. Also,

$$\begin{aligned} M(\theta) &= \sum_{i=1}^k \frac{(p_\theta(i) - y_i)^2}{p_\theta(i)} \\ &\ge \sum_{i=1}^k (p_\theta(i) - y_i)^2 \\ &\ge (\delta - \delta')^2, \end{aligned} \tag{34}$$

[10]It suffices to show the continuity of $\hat\theta$ at $y = p_{\theta_0}$ to establish the theorem. Eqn. 32 also proves that $\hat\theta$ is continous at $y = p_{\theta_0}$.

where the last step is because $y \in B_{\delta'}$, $p_\theta \notin B_\delta$. By choosing δ' small enough, the theorem follows. □

We state the following two theorems, which are standard in probability theory, without proofs (see [1], [25], [31]).

Theorem 5: Let $X, X_1, X_2, \cdots, Y, Y_1, Y_2, \cdots$ be random k-vectors defined on a probability space. Let $g : \mathbb{R}^k \rightarrow \mathbb{R}^l$ be a function continuous at every point of a set C such that $\Pr(X \in C) = 1$.
i) If $X_n \xrightarrow{\text{p}} X$, then $g(X_n) \xrightarrow{\text{p}} g(X)$.
ii) If $X_n \xrightarrow{\text{p}} X$ and $Y_n \xrightarrow{\text{p}} Y$, then $(X_n, Y_n) \xrightarrow{\text{p}} (X, Y)$.

Theorem 6: (Slutky's Theorem) Let $X, X_1, X_2, \cdots, Y, Y_1, Y_2, \cdots$ be random variables on a probability space. Suppose that $X_n \xrightarrow{\text{d}} X$ and $Y_n \xrightarrow{\text{p}} c$, where c is a real number. Then,
i) $X_n + Y_n \xrightarrow{\text{d}} X + c$
ii) $Y_n X_n \xrightarrow{\text{d}} cX$
iii) $X_n / Y_n \xrightarrow{\text{d}} X/c$ if $c \neq 0$.
The result i) holds if the random variables are changed by random k-vectors and $c \in \mathbb{R}^k$. Similarly, ii) is valid when one considers inner product of two random k-vectors.

Theorem 7: Suppose that θ_0 satisfies the conditions of Theorem 4 and the following
i) Θ contains an interval (a, b) including θ_0.
ii) $p_\theta : \Theta \rightarrow \mathbb{R}^k$ is three times continuously differentiable with respect to θ.
Then, the estimator $\hat{\theta}$ is asymptotically efficient, *i.e.,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\text{d}} \mathcal{N}(0, \frac{1}{I(\theta_0)}).$$

In the following we shall give a proof of Theorem 7 assuming that its conditions are satisfied. We adapt the proof technique in [32, Proposition IV.D.2]. Define the function $\Psi(\theta) = M(\theta)/2,\ \theta \in \Theta$. Since $\hat{\theta}$ is a minimizer of $M(\theta)$, and $M(\theta)$ is differentiable, one would expect to have $\Psi(\hat{\theta}) = 0$. The following lemma is based on this intuition.

Lemma 6: $\Pr(\Psi(\hat{\theta}) = 0) \rightarrow 1$ as $n \rightarrow \infty$.

Proof: For given y, the estimate $\hat{\theta}$ (= minimizer of $M(\theta)$) is either inside (a, b) and satisfies

$$\Psi(\hat{\theta}) = \frac{M'(\hat{\theta})}{2} = 0, \tag{35}$$

or is outside (a, b), in which case (35) may not hold since $M(\theta)$ may be minimized at a boundary point of Θ. Consistency of $\hat{\theta}$ implies $\Pr\{\hat{\theta} \in (a, b)\} \rightarrow 1$. The lemma follows. □

For the case $\Psi(\hat{\theta}) = 0$, one can expand the function Ψ around θ_0 using Taylor's Theorem:

$$0 = \Psi(\hat{\theta}) = \Psi(\theta_0) + (\hat{\theta} - \theta_0)\Psi'(\theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 \Psi''(\bar{\theta}), \tag{36}$$

where $\bar{\theta}$ is between θ_0 and $\hat{\theta}$. Eqn. (36) defines $\bar{\theta}$ only when $\Psi(\hat{\theta}) = 0$; define $\bar{\theta} := \theta_0$ when $\Psi(\hat{\theta}) \neq 0$. Rearranging (36) gives

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{-\sqrt{n}\Psi(\theta_0)}{\Psi'(\theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)\Psi''(\bar{\theta})} \qquad \text{when } \Psi(\hat{\theta}) = 0. \tag{37}$$

Lemma 7: $-\sqrt{n}\Psi(\theta_0) \xrightarrow{\text{d}} \mathcal{N}(0, I(\theta_0))$.

Proof: Recall that $M(\theta) = \sum \frac{(p_\theta(i) - y_i)^2}{p_\theta(i)}$. Direct calculation gives

$$\begin{aligned} M'(\theta) &= \sum_{i=1}^{k} \frac{2p'_\theta(i)[p_\theta(i) - y_i]p_\theta - p'_\theta(i)[p_\theta(i) - y_i]^2}{p^2_\theta(i)} \\ &= \sum_{i=1}^{k} \frac{p'_\theta(i)[p_\theta(i) - y_i][2p_\theta(i) - p_\theta(i) + y_i]}{p^2_\theta(i)} \\ &= \sum_{i=1}^{k} \frac{p'_\theta(i)[p_\theta(i)^2 - y_i^2]}{p^2_\theta(i)} \end{aligned} \tag{38}$$

Therefore,

$$-\sqrt{n}\Psi(\theta_0) = \frac{-\sqrt{n}M'(\theta_0)}{2} = -\sqrt{n}\sum_{i=1}^{k} \frac{p'_{\theta_0}(i)[p_{\theta_0}(i)^2 - y_i^2]}{2p^2_{\theta_0}(i)}. \tag{39}$$

The weak law of large numbers gives $y \xrightarrow{\text{p}} p_{\theta_0}$. Viewing $\frac{p'_{\theta_0}(i)(y_i + p_{\theta_0})}{2p^2_{\theta_0}(i)}$ as a continuous function of y, one gets

$$\frac{p'_{\theta_0}(i)(y_i + p_{\theta_0})}{2p^2_{\theta_0}(i)} \xrightarrow{\text{p}} \frac{p'_{\theta_0}(i)}{p_{\theta_0}(i)} := c_i, \tag{40}$$

from Theorem 5 i). Recall $\sqrt{n}(y - p_{\theta_0}) \xrightarrow{\text{d}} \mathcal{N}(0, \Sigma)$ from Lemma 1. Applying Theorem 6 ii) by setting $X_n = \sqrt{n}(y - p_{\theta_0})$, $Y_n =$[LHS of (40) as a vector] and $c = [c_1 \cdots c_k]^T$, we obtain $-\sqrt{n}\Psi(\theta_0) = X_n^T Y_n \xrightarrow{\text{d}} \mathcal{N}(0, c^T \Sigma c)$, where

$$\begin{aligned} c^T \Sigma c &= c^T D c - c^T p p^T c \\ &= c^T D c \\ &= \begin{bmatrix} \frac{p'_{\theta_0}(1)}{p_{\theta_0}(1)} \\ \vdots \\ \frac{p'_{\theta_0}(k)}{p_{\theta_0}(k)} \end{bmatrix}^T \begin{bmatrix} p_1 & & \\ & \ddots & \\ & & p_k \end{bmatrix} \begin{bmatrix} \frac{p'_{\theta_0}(1)}{p_{\theta_0}(1)} \\ \vdots \\ \frac{p'_{\theta_0}(k)}{p_{\theta_0}(k)} \end{bmatrix} \\ &= I(\theta_0). \end{aligned}$$

□

Lemma 8: $\Psi'(\theta_0) \xrightarrow{\text{p}} I(\theta_0)$

Proof: Differentiate (38) to get

$$\begin{aligned} \Psi'(\theta_0) = \sum_{i=1}^{k} \frac{1}{2p^4_{\theta_0}(i)} \ &\{p''_{\theta_0}(i)[p^2_{\theta_0}(i) - y_i^2] \\ &+ 2[p'_{\theta_0}(i)]^2 p_{\theta_0}(i)\}\ p^2_{\theta_0}(i) - 2[p'_{\theta_0}(i)]^2(p^2_{\theta_0}(i) - y_i^2)p_{\theta_0}(i) \\ = \sum_{i=1}^{k} &\frac{p''_{\theta_0}(i)p_{\theta_0}(i)[p_{\theta_0}(i)^2 - y_i^2] + 2[p'_{\theta_0}(i)]^2 y_i^2}{2p^3_{\theta_0}(i)}. \end{aligned}$$

Treating the whole expression as a continuous function of y, one can substitute y by its limit p_{θ_0} (Theorem 5 i) to get $\Psi'(\theta_0) \xrightarrow{\text{p}} \sum_{i=1}^{k} \frac{(p'_{\theta_0}(i))^2}{p_{\theta_0}(i)}$. □

Lemma 9: $\frac{1}{2}(\hat{\theta} - \theta_0)\Psi''(\bar{\theta}) \xrightarrow{\text{p}} 0$.

Proof: The estimate $\bar{\theta}$ is always between θ_0 and $\hat{\theta}$, and $\hat{\theta} \xrightarrow{\text{p}} \theta_0$. Therefore, $\bar{\theta} \xrightarrow{\text{p}} \theta_0$. The $\Psi''(\theta)$ is a continuous function of θ (we don't give an explicit expression, since it is not necessary). Therefore, $\Psi''(\bar{\theta}) \xrightarrow{\text{p}} \Psi''(\theta_0)$ by Theorem 5 i).

The estimator is consistent and $\frac{1}{2}(\hat{\theta} - \theta_0) \xrightarrow{\text{p}} 0$. Since both $\Psi''(\bar{\theta})$ and $\frac{1}{2}(\hat{\theta} - \theta_0)$ converge in probability, as a vector

$(\Psi''(\bar{\theta}), \frac{1}{2}(\hat{\theta}-\theta_0))$ they converge to $(\Psi''(\theta_0), 0)$ (Theorem 5 ii). Multiplication is a continuous function, and Theorem 5 i) gives the desired result. □

Corollary 1:

$$E_n := \frac{-\sqrt{n}\Psi(\theta_0)}{\Psi'(\theta_0)+\frac{1}{2}(\hat{\theta}-\theta_0)\Psi''(\bar{\theta})} \xrightarrow{\text{d}} \mathcal{N}(0, \frac{1}{I(\theta_0)}).$$

Proof: Immediate from Lemmas 7-9 and Theorems 5,6. □

Now, we can finish the proof of Theorem 7. For any $e \in \mathbb{R}$,

$$\begin{aligned}&\Pr(\sqrt{n}(\hat{\theta}-\theta_0) \le e,\ \hat{\theta}_n = 0)\\ &\quad\le \Pr(\sqrt{n}(\hat{\theta}-\theta_0) \le e)\\ &\quad\le \Pr(\sqrt{n}(\hat{\theta}-\theta_0) \le e,\ \hat{\theta}_n = 0) + \Pr(\hat{\theta} \neq 0)\end{aligned}$$

We wish to show that the term in the middle converges to cdf of $\mathcal{N}(0, 1/I(\theta_0))$. However, since $\Pr(\hat{\theta} \neq 0) \to 0$ by Lemma 6, it suffices to show that $\Pr(\sqrt{n}(\hat{\theta}-\theta_0) \le e,\ \hat{\theta} = 0)$ converges. Note that

$$\Pr(\sqrt{n}(\hat{\theta}-\theta_0) \le e,\ \hat{\theta} = 0) = \Pr(E_n \le e,\ \hat{\theta} = 0) \quad (41)$$

by (37), and

$$\begin{aligned}\Pr(E_n \le e,\ \hat{\theta}=0) - \Pr(\hat{\theta} \neq 0) &\le \Pr(E_n \le e,\ \hat{\theta}=0)\\ &\le \Pr(E_n \le e). \quad (42)\end{aligned}$$

Since $\Pr(E_n \le e)$ converges to the cdf of $\mathcal{N}(0, 1/I(\theta_0))$ (Corollary 1), the theorem follows.

C. Proof of Theorem 2

We will actually give a proof of the more general statement in Remark 3 (the connection between Theorem 2 and the remark is provided later by Lemma 10). The proof is a modification of the one given in previous Appendix. Let $\{c_\theta \in \mathbb{R}^k : \theta \in \Theta\}$ be a set (this can be visualized as *curve* in $\mathbb{R}^k$), and $\{\Sigma_\theta : \theta \in \Theta\}$ be a set of $k \times k$ invertible matrices. Suppose that $\hat{\theta}$ is the estimator minimizing $M(\theta) = (y - c_\theta)^T \Sigma_\theta^{-1}(y - c_\theta)$ with respect to $\theta \in \Theta$.

Let point $\theta_0 \in \Theta$ be identifiable in the following sense: $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$||c_\theta - c_{\theta_0}|| < \delta \text{ for some } \theta \in \Theta \quad \Rightarrow \quad |\theta - \theta_0| < \epsilon. \quad (43)$$

Theorem 8: Let Σ_θ be invertible $\forall \theta \in \Theta$, and $\sup_{\theta\in\Theta} ||\Sigma_\theta|| < \infty$, where $||\cdot||$ denotes the 2-norm of the matrix. If θ_0 satisfies (43), then the estimator $\hat{\theta}$ is consistent (*i.e.*, $\hat{\theta} \xrightarrow{\text{p}} \theta_0$).

Proof: Proof of Theorem 4 directly applies when p_θ is replaced by c_θ. Replace the bound (33) by

$$M(\theta_0) \le ||\Sigma_{\theta_0}^{-1}||\,||y - c_{\theta_0}||^2 \le ||\Sigma_{\theta_0}^{-1}||(\delta')^2, \quad (44)$$

and (34) by

$$\begin{aligned}M(\theta) &\ge [\text{smallest singular value of } \Sigma_\theta^{-1}]\ ||y - c_\theta||^2\\ &= \frac{1}{||\Sigma_\theta||}||y - c_\theta||^2\\ &\ge \frac{1}{\sup_{\theta\in\Theta}||\Sigma_\theta||}||y - c_\theta||^2\\ &\ge \frac{1}{\sup_{\theta\in\Theta}||\Sigma_\theta||}(\delta - \delta')^2. \quad (45)\end{aligned}$$

□

Theorem 9: Suppose that θ_0 satisfies the conditions of Theorem 8 and the following
i) Θ contains an interval (a, b) including θ_0.
ii) $c_\theta : \Theta \to \mathbb{R}^k$, $\Sigma_\theta^{-1} : \Theta \to \mathbb{R}^{k\times k}$ are three times continuously differentiable with respect to θ.
Then, the estimator $\hat{\theta}$ is asymptotically normal, and

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{\text{d}} \mathcal{N}(0, \frac{1}{(c'_{\theta_0})^T \Sigma_{\theta_0}^{-1} c'_{\theta_0}}).$$

Under the conditions of the theorem $M(\theta) = (y - c_\theta)^T \Sigma_\theta^{-1}(y - c_\theta)$ is three times continuously differentiable with respect to θ for all y. Define the function $\Psi(\theta) = \frac{M'(\theta)}{2}$ as before. Lemma 6 holds verbatim. One can expand $\Psi(\hat{\theta})$ using Taylor's Theorem as before, and establish

i) $-\sqrt{n}\Psi(\theta_0) \xrightarrow{\text{d}} \mathcal{N}(0, J(\theta_0))$
ii) $\Psi'(\theta_0) \xrightarrow{\text{p}} J(\theta_0)$
iii) $\frac{1}{2}(\hat{\theta}-\theta_0)\Psi''(\bar{\theta}) \xrightarrow{\text{p}} 0$

where $J(\theta) := (c'_\theta)^T \Sigma_\theta^{-1} c'_\theta$. Observe

$$M'(\theta) = -2(c'_\theta)^T \Sigma_\theta^{-1}(y - c_\theta) + (y - c_\theta)^T \left(\frac{d\Sigma_\theta^{-1}}{d\theta}\right)(y - c_\theta).$$

The left hand side term in $\frac{-\sqrt{n}M'(\theta_0)}{2}$ converges in distribution to $\mathcal{N}(0, J(\theta_0))$. And, from Slutky's Theorem the right hand side term converges to zero; this establishes i). One can differentiate M' further to get $2(c'_\theta)^T \Sigma_\theta^{-1} c'_\theta$ plus terms which all contain $y - c_\theta$. Consequently, ii) holds. iii) can be shown similarly. The proof of Theorem 9 is completed by following arguments identical to the one in the previous subsection.

As the following lemma asserts, Theorem 2 follows from Theorem 9 by substituting c_θ and Σ_θ corresponding to the asymptotic distribution of y in the fading channel (cf. Lemma 3).

Lemma 10: For the special case that $c_\theta = p_\theta$ and $\Sigma_\theta = (1 + \sigma_h^2)\text{Diag}(p_\theta) - p_\theta p_\theta^T$, we have

$$\frac{1}{(c'_\theta)^T \Sigma_\theta^{-1} c'_\theta} = \frac{1 + \sigma_h^2}{I(\theta)}. \quad (46)$$

Proof: Use Sherman-Morrison-Woodbury formula [26] to get

$$\Sigma_\theta^{-1} = D_\theta^{-1} + \frac{D_\theta^{-1} p_\theta p_\theta^T D_\theta^{-1}}{1 - p_\theta^T D_\theta^{-1} p_\theta},$$

where $D_\theta = (1 + \sigma_h^2)\text{Diag}(p_\theta)$. Substitute this into $1/(p'_\theta)^T \Sigma_\theta^{-1} p'_\theta$ to get the result. □

REFERENCES

[1] J. Shao, *Mathematical Statistics*. Springer Verlag, 1999.
[2] L. Tong, Q. Zhao, and S. Adireddy, “Sensor networks with mobile agents,” in *Proc. 2003 Military Communications Intl. Symp.*, Boston, MA, Oct. 2003.
[3] M. Gastpar and M. Vetterli, “Scaling laws for homogeneous sensor networks,” in *41st Annual Allerton Conf. on Commun., Control and Comp.*, Oct. 1-3 2003.
[4] T. S. Han and S. Amari, “Statistical inference under multiterminal data compression,” *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.
[5] R. Ahlswede and I. Csiszár, “Hypothesis testing with communication constraints,” *IEEE Trans. Inform. Theory*, vol. 32, no. 4, pp. 533–542, Jul. 1986.

[6] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887–902, May. 1996.

[7] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE Journal on Selected Areas in Communications: Special Issue on Sensor Networks*, 2004.

[8] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE J. Select. Areas Commun.*, 2004.

[9] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.

[10] J. Barros and S. D. Servetto, "The sensor reachback problem," Nov. 2003, submitted to the IEEE Trans. Inform. Theory.

[11] A. Murugan, P. Gopala, and H. E. Gamal, "Correlated sources and wireless channels: Cooperative source-channel coding," *IEEE J. Select. Areas Commun.*, 2004.

[12] K. Liu and A. M. Sayed, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," in *Proc. ICASSP'04 Conf.*, Montreal, Quebec, Canada, May 17-21 2004.

[13] ——, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," May 19 2004, ICASSP'04 presentation slides.

[14] M. Dong, L. Tong, and B. Sadler, "Impact of MAC design on signal field reconstruction in dense sensor networks," submitted to *IEEE Trans. on Signal Processing*, Nov. 2003.

[15] ——, "Information retrieval and processing in sensor networks: Deterministic scheduling vs. random access," submitted to *IEEE Trans. Inform. Theory*, Feb. 2004.

[16] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 407–416, February 2003.

[17] ——, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE JSAC Special Issue on Wireless Sensor Networks*, 2004.

[18] B. Chen, R. Jiang, T. Kasetkasem, and P. Varshney, "Fusion of decisions transmitted over fading channels in wireless sensor networks," in *in Proc.of the 36th Asilomar Conference on Signals, Systems and Computers*, Pacific Groove, CA, Nov. 2002, pp. 1184–1188.

[19] R. Niu, B. Chen, and P. Varshney, "Decision fusion rules in wireless sensor networks using fading channel statistics," in *Proceedings of CISS'2003*, Baltimore, MD, March 2003.

[20] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 34, no. 2, pp. 469–475, April 1998.

[21] W.-M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Trans. on Comm.*, vol. 41, no. 11, pp. 1602–1605, Nov. 1993.

[22] J. A. Gubner, "Distributed Estimation and Quantization," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.

[23] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *IEEE Trans. on Comm.*, vol. 48, no. 2, pp. 181–184, Feb. 2000.

[24] R. Gupta and A. Hero, "High rate vector quantization for detection," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1951–1969, Aug. 2003.

[25] P. Billingsley, *Probability and Measure 3rd. Ed.* New York, NY: Wiley Inter-Science, 1995.

[26] G. Golub and C. V. Loan, *Matrix Computations, 3rd Ed.* Baltimore, Maryland: The Johns Hopkins University Press, 1996.

[27] D. Tse and P. Viswanath, "EECS 224B: Fundamentals of Wireless Communications," lecture notes, University of California, Berkeley, Spring 2004.

[28] E. Lehmann and Casella, *Theory of Point Estimation, 2nd Ed.* New York, NY: Chapman & Hall, 1998.

[29] K. Liu and A. M. Sayeed, "Optimal distributed detection strategies for wireless sensor networks," in *42nd Annual Allerton Conf. on Commun., Control and Comp.*, Sep.29-Oct.1 2004.

[30] G. Mergen and L. Tong, "Sensor-fusion center communication over multiaccess fading channels," to appear in *ICASSP'05*.

[31] A. W. van der Vaart, *Asymptotic Statistics*. New York, NY: Cambridge University Press, 2000.

[32] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

Gökhan Mergen was born in 1978 in Ankara, Turkey. He received his B.S. degrees in Electrical and Electronics Engineering and Mathematics (double major) in 2000 from Middle East Technical University (METU), Turkey. He received his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Cornell University (the latter one in Jan. 2005). His industry experience includes work at the Bell Labs, Lucent Technologies, during the summer of 2003. He is currently employed with Qualcomm, Campbell CA. His research interests lie in the areas of communications theory, information theory, data networks and statistical signal processing.

Lang Tong is a Professor in the School of Electrical and Computer Engineering, Cornell University, Ithaca, New York. He received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and M.S. and Ph.D. degrees in electrical engineering in 1987 and 1991, respectively, from the University of Notre Dame, Notre Dame, Indiana. He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University in 1991. He was also the 2001 Cor Wit Visiting Professor at the Delft University of Technology.

Dr. Tong received the Young Investigator Award from the Office of Naval Research in 1996, the Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 Best Paper Award (with Min Dong) from the IEEE Signal Processing Society, and the 2005 Leonard G. Abraham Prize Paper Award (with P. Venkitasubramaniam, and S. Adireddy) from the IEEE Communications Society. His areas of interest include statistical signal processing, wireless communications, communication networks and sensor networks, and information theory.