Distributed Statistical Inference using Type Based Random Access over Multi-access Fading Channels

Animashree Anandkumar
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853.
Email: aa332@cornell.edu.

Lang Tong
School of Electrical and Computer Engineering
Cornell University,Ithaca, NY 14853.
Email: ltong@ece.cornell.edu.

Abstract— **The problem of distributed detection and estimation in a sensor network over a multiaccess fading channel is considered. A communication scheme known as the Type-Based Random Access (TBRA) is employed and its performance is characterized with respect to the mean transmission rate and the channel coherence index. For extreme values of channel coherence index *i.e.,* 0 and ∞, we give an optimal TBRA scheme which is essentially a sensor activation strategy that achieves the optimal allocation of transmission energy to spatial and temporal domains. For channels with zero coherence index, it is shown that there exists a finite optimal mean transmission rate maximizing performance. This optimal rate can be calculated numerically or estimated using the Gaussian approximation. On the other hand, for channels with infinite coherence index (*i.e.,* no fading) the optimal strategy is to allocate all the energy to the spatial domain. Numerical examples and simulations confirm our theory.**

Index Terms—**Signal Processing for Communications, distributed detection, distributed estimation, multi-sensor systems, performance analysis.**

I. INTRODUCTION

We consider the problem of distributed statistical inference over a multiaccess fading channel. Our focus is on the communication (or random access) aspect of the problem. The communication scheme we employ is the so-called Type-Based Random Access (TBRA) [1], [2], in which sensors transmit probabilistically using a set of orthogonal waveforms keyed to their measurement. Specifically, sensors with the same data value transmit using the same waveform on a multiaccess fading channel. The use of orthogonal waveforms eliminates interference among users with different data values and makes it possible to have coherent combining of transmissions in the absence of fading. However, we will see that the behavior is more complicated in presence of fading, since there is a possibility of cancelation among the signals. A schematic of our problem is illustrated in Fig. 1 with model definition given in Section II.

A. Summary of Main Results

Given the fixed local quantization rule at the sensors and the available set of orthogonal waveforms for transmission,
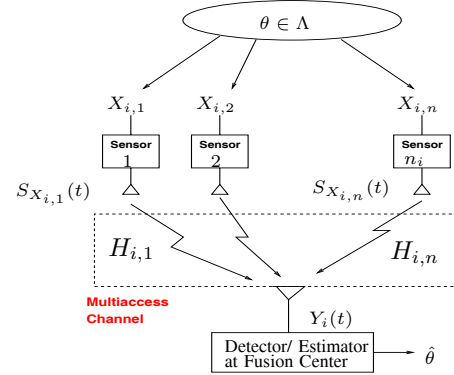
Fig. 1. Distributed Detection/ Estimation in Multiaccess.

the design of TBRA reduces to the optimal choice of the mean transmission rate λ by maximizing a performance metric $E(\lambda)$. Intuitively, if λ is too small, not enough sensors transmit, and performance suffers. On the other hand, if too many sensors transmit, the energy consumption is high. Since they transmit on a multiaccess channel, it is not obvious that sensors will not interfere with each other, resulting in poor performance.

The exact form of the performance metric $E(\lambda)$ depends on the type of detection (Neyman-Pearson or Bayesian) or estimation (Maximum likelihood or Bayesian). But it turns out that they all show similar dependence on the fading characteristics of the multiaccess channel. In particular, $E(\lambda)$ crucially depends on "channel coherence index" γ defined by

$$\gamma = \frac{|\mathbb{E}(H)|^2}{\text{Var}(H)}, \tag{1}$$

where H is the fading coefficient between a sensor and the fusion center. Intuitively, higher γ leads to better coherent addition of signals.

Illustrated in Fig. 2 are sketches of $E(\lambda)$ as functions of λ and γ. The shapes of these curves will be justified by the analytical and numerical results in Section III and Section IV. We see that for low coherence indices, there exists an optimal λ_* for which $E(\lambda)$ is maximized. This implies that there is an optimal sensor activation probability so that the average number of transmitting sensors is optimal. The intuition is
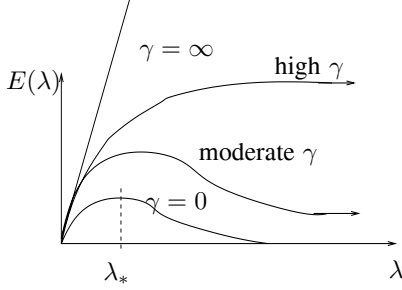
Fig. 2. Performance Metric $E(\lambda)$ as a function of λ for different channel coherence indices γ (see (1)).

that for fading channels with zero mean, $\gamma = 0$, sensors transmitting using the same waveform tend to cancel each other (in the mean), which is the reason that TBMA schemes involving a single data collection fail [3], [4]. A sharp contrast is the extreme case when the channel is deterministic without fading ($\gamma = \infty$). We show that there does not exist an optimizing λ, which means that the optimal strategy is to make all sensors transmitting at the same time in order to take advantage of channel coherency. This paper aims to provide insights into the optimal tradeoff.

We show the existence of optimal average transmission rate λ_* when channel coherence index is small. We also characterize of $E(\lambda)$ when λ is large. It is in fact the behavior of $E(\lambda)$ as $\lambda \to \infty$ that helps us to describe the shape of the curve in Fig. 2. By letting $\lambda \to \infty$, we employ a version of the Central Limit Theorem (CLT) that involves random summand, and the limiting distribution allows us to characterize $E(\infty)$ analytically. For large λ's, Gaussian approximation can be used to obtain estimates of $E(\lambda)$. Perhaps more importantly in practice, $\tilde{\lambda}_*$ obtained by optimizing the Gaussian approximation is close to the true optimal λ_*. We evaluate the accuracy of such an approximation via numerical evaluation and simulations.

B. Related Work

The problem of classical distributed detection has been investigated in considerable detail. [5]. In the context of power and bandwidth-constrained wireless sensor networks, large deviation techniques have been used for the optimal design of local quantization rules [6], [7]. We too use large deviation techniques, but for the design of multiaccess communications. Distributed detection in the presence of channel fading is considered in [8], [9] when each user has a dedicated channel to the fusion center. Distributed Estimation has been studied under information theoretic setup. See [10] for a survey.

The problem of of distributed statistical inference on multi-access channels are more recent [3], [4]. The transmission scheme used is the so-called Type Based Multiple Access (TBMA) proposed independently by Mergen and Tong [11], [3] and by Liu and Sayeed [4]. The positive result of TBMA is that, when there is no fading, the asymptotic performance of TBMA (as the number of sensors approaches infinity) is same as that when the fusion center has direct access to sensor

observations. The negative result, however, is that when the channel has zero mean fading, TBMA fails to be consistent. Furthermore, these results apply only when the number of sensors is deterministic. In [1], we proposed Type Based Random Access (TBRA) as a multiaccess scheme incorporating random number of sensors. We used large deviation approaches to compare detection performance of TBRA with TBMA for non-zero mean fading channels.

In [2], we employ TBRA in the context of distributed detection while allowing the fading channels to have zero mean. This scenario is relevant since in practice it may be difficult to synchronize transmissions among geographically distributed nodes to achieve phase coherency at the receiver. In this paper we will focus on the problem of distributed estimation and also unify the optimal strategies for both distributed detection and estimation under a common framework.

II. MODEL AND PROBLEM FORMULATION

A. TBRA: Transceiver and Sufficient Statistics

For the estimation problem, we assume a non-random real parameter θ lying in some set Λ. As illustrated in Fig. 1, the fusion center collects data in multiple time slots indexed by i. In each collection, there are N_i sensors involved in the transmission, where N_i is a random variable with mean λ and Probability Mass Function (PMF) $g(n, \lambda) \triangleq \Pr(N_i = n)$. We assume that the sequence N_i is IID.

In the ith data collection, a sensor involved in the transmission[1], say sensor j, has measurement $X_{i,j} \in \{1, \cdots, M\}$ quantized to M levels. We assume that the sensor data (X_{ij}) are conditionally IID across time and sensors with PMF $p_\theta(\cdot)$. In a vector notation,

$$X_{i,j} \overset{\text{i.i.d}}{\sim} \mathbf{p}_\theta = (p_\theta(1), \cdots, p_\theta(M)).$$

In the ith collection, the transmitter j encodes $X_{i,j}$ to a certain waveform and transmits it over a multiaccess fading channel. As in TBMA, a set of M ortho-normal waveforms $\{\phi_k(t), k = 1, \cdots, M\}$ are used, each corresponding to a specific data value. Specifically, if $\mathcal{E}$ is the energy of the transmission, then the signal transmitted by sensor j in collection i is $S_{i,j}(t) = \sqrt{\mathcal{E}}\phi_{X_{i,j}}(t)$.

At the fusion center, assuming that there is no inter-collection interference, the received signal is given by

$$Y_i(t) = \sum_{j=1}^{N_i} H_{i,j} S_{i,j}(t - \tau_{i,j}) + W_i(t), \quad i = 1, \cdots, l \quad (2)$$

where $\tau_{i,j}$ are the different delays experienced by sensors[2], the noise $W_i(t)$ is assumed to be zero-mean and white Gaussian with power density σ^2, and the channel fading coefficients[3]

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

[2]Due to lack of synchronization, the delays experienced by each sensor transmission in a data collection slot could be different.

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

$(H_{i,j} \in \Re)$ are IID with mean $\mu_H \triangleq \mathbb{E}(H_{i,j})$ and variance $\sigma_H^2 \triangleq \mathrm{Var}(H_{i,j})$. We define (transmitter) SNR by $\mathrm{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}$. Under the narrow-band signal assumption, the flat fading approximation which neglects the time dispersion is valid and therefore $S_{i,j}(t - \tau_{i,j}) \approx S_{i,j}(t)$. Therefore

$$Y_i(t) = \sum_{j=1}^{N_i} H_{i,j} S_{i,j}(t) + W_i(t), \quad i = 1, \cdots, l \quad (3)$$

The sufficient statistics $\{\mathbf{Y}_i\}$ are generated from the bank of filters matched to orthogonal basis $\{\phi_k(t)\}$. For i^{th} collection,

$$
\begin{aligned}
\mathbf{Y}_i & \triangleq \frac{1}{\sqrt{\mathcal{E}}} \Big[\langle Y_i(\cdot), \phi_1(\cdot) \rangle, \cdots, \langle Y_i(\cdot), \phi_M(\cdot) \rangle \Big] \\
& = \sum_{j=1}^{N_i} H_{i,j} \mathbf{e}_{X_{i,j}} + \mathbf{W}_i,
\end{aligned} \quad (4)
$$

where $\langle Y_i(\cdot), \phi_k(\cdot) \rangle$ is the output of the matched filter corresponding to $\phi_k(t)$, $\mathbf{e}_k$ the unit vector with non-zero entry at the kth position, and $\mathbf{W}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \frac{1}{\mathrm{SNR}} \mathbf{I})$.

To see the intuition behind the coherence index γ defined in (1), we write explicitly the mth entry of $\mathbf{Y}_i = [Y_{i,1}, \cdots, Y_{i,M}]^{\mathrm{T}}$

$$Y_{i,m} = \sum_{j=1}^{N_i} H_{i,j} 1_{\{X_{i,j}=m\}} + W_{i,m}, \quad (5)$$

where $1_{\mathcal{A}}$ is the event indicator function. The extreme case is when the channel is deterministic with $H_{i,j} = 1$ ($\gamma = \infty$). Transmissions from those sensors observing data value m add up coherently, and $Y_{i,m}$ is the number of sensors that observe data level m (plus noise), which gives rise to notion of type-based transmission[4]. On the other hand, when $\gamma = 0$, ($\mu_H = 0$), the transmissions add up non-coherently, and the mean of $\mathbf{Y}_i$ contains no information of the model.

B. Spatio-Temporal Tradeoff and Problem Formulation

The design of TBRA reduces to finding the optimal transmission rate λ that maximizes a performance metric $E(\lambda)$. If the fusion center collects data using TBRA for l slots, the total mean number of transmissions is $\rho \triangleq \lambda l$, which is also proportional to energy consumption. Fixing ρ, there is a spatio-temporal tradeoff between the mean number of transmission per slot and the total number of slots of data collection. Should energy be allocated mostly to simultaneous transmission by making λ large? Or should we rely on taking more data collections by choosing a large l. The optimal design of TBRA is to achieve optimal tradeoff between λ and l.

To this end, we need to characterize the mean square error (MSE) of maximum likelihood estimation (MLE). However, an explicit characterization is not tractable in our setup. We thus examine the case when the total average number of

---

[4]Given $X_{i,j} = x_{i,j}$, $N_i = n_i$ and the observation $\mathbf{Y}_i = \mathbf{y}_i$, in the absence of noise, the type of $x_{i,j}$ is $\frac{1}{n_i} \mathbf{y}_i$. [12], [13].

transmissions ρ is large. We will optimize TBRA through the asymptotic mean square error (MSE). We define the performance metric as

$$E(\lambda) \triangleq \lim_{\rho \to \infty} \frac{1}{\mathrm{Var}[\sqrt{\rho}(\hat{\theta}_{\mathrm{MLE}} - \theta)]}. \quad (6)$$

which is equivalent to saying that the asymptotic MSE decays as $\frac{1}{\rho}$. This will be justified in Section III. Maximizing the performance metric $E(\lambda)$ with respect to λ is equivalent to minimizing the asymptotic MSE.

For the detection problem, we consider two types of detectors: Bayesian and Neyman-Pearson. Let $P_e(\rho, \lambda)$ be the detection error probability (either the miss detection probability of Neyman-Pearson detector or the average of miss detection and false alarm probabilities in Bayesian setup). Here the performance metrics are the detection error exponents

$$E(\lambda) \triangleq - \lim_{\rho \to \infty} \frac{1}{\rho} \log P_e(\rho, \lambda), \quad (7)$$

We have used the same notation $E(\lambda)$ in (6) and (7), since we aim to unify the optimal strategies for detection and estimation. We will justify this in Section III. Ultimately we seek

$$\lambda_* = \arg \sup_{\lambda} E(\lambda). \quad (8)$$

Although $E(\lambda)$ can be evaluated numerically for a given statistical model and fading, it is of theoretical and practical significance to establish that λ_* is finite. To this end, we need to characterize $E(\lambda)$ as $\lambda \to \infty$ and $\lambda \to 0$.

III. Optimal Type-Based Random Access

A. Asymptotic MSE of ML estimation

Given $\mathbf{Y}_i$ IID, the ML estimator is given by[5],

$$\hat{\theta}_{\mathrm{MLE}} = \arg \max_{\theta \in \Lambda} \sum_{i=1}^{l} \log f_\theta(\mathbf{Y}_i). \quad (9)$$

In the following theorem we characterize $E(\lambda)$ in terms of Fisher Information, based on the asymptotic efficiency of MLE. The only trivial modification is a change in time index to the total mean number of transmissions ρ.

Theorem 1 (Asymptotic Efficiency): Given mean number of total transmissions ρ and mean transmission rate λ, let $\{\hat{\theta}_{\mathrm{MLE}}(\rho)\}_{\rho=1}^{\infty}$ be a consistent sequence of roots of likelihood equation. Under some regularity conditions, the ML estimator is asymptotically efficient and therefore we can define the performance metric $E^{\mathrm{MLE}}(\lambda, \theta)$ as

$$E^{\mathrm{MLE}}(\lambda, \theta) \triangleq \lim_{\rho \to \infty} \frac{1}{\mathrm{Var}[\sqrt{\rho}(\hat{\theta}_{\mathrm{MLE}}(\rho) - \theta)]} = \frac{I_\lambda(\theta)}{\lambda}, \quad (10)$$

---

[5]We assume that θ is identifiable.

where $I_\lambda(\theta) < \infty$ is the Fisher Information of a single data collection.

Proof: For asymptotic efficiency and related regularity conditions, see [14], pp. 183-184.

For the detection problem, the expressions for $E(\lambda)$ under Neyman Pearson and Bayesian Setup are given by

$$E^{\text{NPD}}(\lambda) \triangleq -\lim_{\rho \to \infty} \frac{1}{\rho} \log P_e^{\text{NP}}(\rho, \lambda) = \frac{D_\lambda(f_{\theta_0} || f_{\theta_1})}{\lambda},$$

$$E^{\text{BD}}(\lambda) \triangleq -\lim_{\rho \to \infty} \frac{1}{\rho} \log P_e^{\text{B}}(\rho, \lambda) = \frac{C_\lambda(f_{\theta_0}, f_{\theta_1})}{\lambda},$$

where $D_\lambda(f_{\theta_0} || f_{\theta_1})$ is the Kullback-Leibler distance and $C_\lambda(f_{\theta_0}, f_{\theta_1})$ the Chernoff information between the two hypotheses. [12], [13].

B. Optimal Transmission Rate

While Theorem 1 provides the basis for defining the performance metric $E(\lambda)$, it says little about its behavior as a function of λ, especially if there exists an optimizing λ. The following Theorem gives the results for the two extreme cases: $\gamma = 0$ and $\gamma = \infty$.

Theorem 2 (Existence of optimal λ): Given mean number of transmitting sensors λ, let $f_{\theta,\lambda}(\mathbf{y})$ be the probability density functions of matched filter output $\mathbf{Y}$ under θ. Assume that Theorem 1 holds and in addition assume that the density function $f_{\theta,\lambda}$ is differentiable with respect to λ. Also, assume that the PMF of N, $g(n, \lambda)$ is differentiable in λ and satisfies the following properties,

$$\lim_{\lambda \to 0} g(n, \lambda) = 1_{\{n=0\}}, \qquad (11)$$

$$\lim_{\lambda \to 0} \frac{d}{d\lambda} g(n, \lambda) = -a 1_{\{n=0\}} + a 1_{\{n=1\}}, \qquad (12)$$

where a is a constant and $1_{\mathcal{A}}$ is the event indicator function.

1) If the channel has zero-mean fading, *i.e.*, $\gamma = 0$. Then

$$\lim_{\lambda \to 0} E^{\text{MLE}}(\lambda, \theta) = \lim_{\lambda \to \infty} E^{\text{MLE}}(\lambda, \theta) = 0, \qquad (13)$$

which implies that there exists $0 < \lambda_*^{\text{MLE}} < \infty$ such that

$$\sup_\lambda E^{\text{MLE}}(\lambda, \theta) = \frac{1}{\lambda_*^{\text{MLE}}} I_{\lambda_*^{\text{MLE}}}(\theta). \qquad (14)$$

2) If channel is deterministic *i.e.*, $\sigma_H^2 = 0$ or $\gamma = \infty$, there does not exist optimizing λ that maximizes $E(\lambda)$ and

$$E^{\text{MLE}}(\lambda, \theta) = \Theta(\lambda) \qquad (15)$$

as $\lambda \to \infty$, where the notation Θ means that λ is an exponentially tight bound.[6]

Proof: See Appendix A. □

We have proven an analogous result for the detection problem in [2]. The conditions in (11) and (12) essentially state

[6]$\Theta(a(\lambda)) = \{b(\lambda) : 0 \le c_1 a(\lambda) \le b(\lambda) \le c_2 a(\lambda), \forall \lambda > \lambda_o\}$ for some $c_1, c_2, \lambda_o > 0$.

that, at low λ we either have none or utmost one transmission (*e.g.*, Poisson Distribution). Also in practice, the regularity conditions on the PDF $f_{\theta,\lambda}$ assumed in Theorem 2 are satisfied by many well behaved density functions.

Theorem 2 establishes general shape of $E(\lambda)$ as shown in Fig 2 for extreme values $\gamma = 0$ and $\gamma = \infty$. Note that the role of γ in $E(\lambda)$ is embedded through density function $f_{\theta,\lambda}(\mathbf{y})$, typically a continuous function of γ. Therefore, we can infer the behavior of $E(\lambda)$ for very small and very large γ.

C. Asymptotic Distribution and Gaussian Approximation

A key step in proving Theorem 2 is the investigation of $E(\lambda)$ when $\lambda \to \infty$. The idea is to use the continuity argument coupled by using a version of the Central Limit Theorem (CLT) involving random number of summands [15] to characterize $E(\lambda)$ as $\lambda \to \infty$.

We shall focus in this section on the single collection model, and evaluate the Fisher Information using the limiting distribution as $\lambda \to \infty$. For ease of notation, we drop the time index i in (4), and consider the model

$$\mathbf{Y} = \sum_{j=1}^N H_j \mathbf{e}_{X_j} + \mathbf{W}, \qquad (16)$$

where we have a random summand N with PMF $g(n, \lambda)$ and mean $\mathbb{E}(N) = \lambda$.

Theorem 3 (Asymptotic distribution of $\mathbf{Y}$): Assume that the channel gains H_j's are i.i.d distributed with mean μ_H and variance σ_H^2, and the number of sensors N is distributed according to PMF $g(n, \lambda)$ taking non-negative integral values with mean λ. Assume that

$$\frac{N}{\lambda} \xrightarrow{d} 1 > 0 \quad \text{as } \lambda \to \infty. \qquad (17)$$

When $p_\theta(j) > 0$ for all $j \in \{1, \ldots, M\}$, the shifted and scaled matched filter output has the asymptotic normal distribution according to

$$\frac{\mathbf{Y} - \lambda \mu_H \mathbf{p}_\theta}{\sqrt{\lambda}} \xrightarrow{d} \mathcal{N}(0, \sigma_H^2 \text{Diag}(\mathbf{p}_\theta)) \quad \text{as} \quad \lambda \to \infty, \quad (18)$$

where $\text{Diag}(\mathbf{p}_\theta) = \text{diag}(p_\theta(1), \cdots, p_\theta(M))$.

Proof : See [2].

The assumption that $\frac{N}{\lambda} \xrightarrow{d} 1$ is valid for many practical scenarios, (*e.g.*, the Poisson distribution and also the deterministic case). Since $\mathbf{Y}_i$ is asymptotically Gaussian, in large-λ regime, the detection or the estimation problem can be cast with

$$\mathbf{Y}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(\lambda \mu_H \mathbf{p}_\theta, \lambda \sigma_H^2 \text{Diag}(\mathbf{p}_\theta) + \frac{\sigma^2}{\mathcal{E}} \mathbf{I}\right). \qquad (19)$$

We now give the closed form expression for Fisher Information $\tilde{I}_\lambda(\theta)$ of the asymptotic Gaussian distribution (for a given θ) and specialize the results for coherent and non-coherent channels respectively. Similar results for detection error exponents can be found in [2].
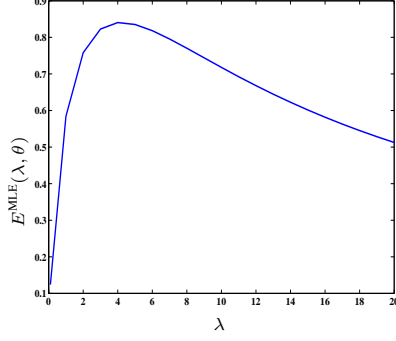
Fig. 3. Performance Metric vs. Transmission Rate. ($\gamma = 0$, SNR= 0db, $\sigma_H^2 = 1$, $\theta = 0.454$)
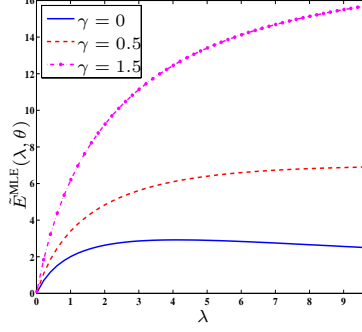
Fig. 4. Gaussian Approx. Performance Metric vs. Transmission Rate (SNR= 0db, $\sigma_H^2 = 1$, $\theta = 0.8$)
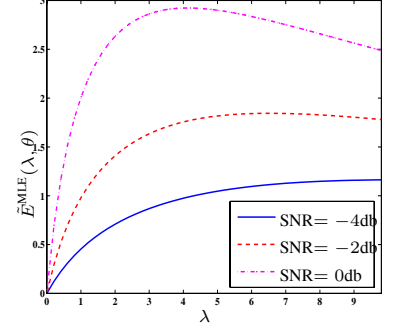
Fig. 5. Gaussian Approx. Performance Metric vs. Transmission Rate. ($\gamma = 0$, $\sigma_H^2 = 1$, $\theta = 0.8$)

Lemma 1 (Fisher Information): Let σ_H^2 be the channel variance, γ the channel coherence index, SNR $= \frac{\mathcal{E}}{\sigma^2}$ be the transmitter SNR. Denote $p'_\theta(j) \stackrel{\triangle}{=} \frac{\partial}{\partial \theta} p_\theta(j)$. The Fisher Information $\tilde{I}_{\lambda,\gamma}(\theta)$ is given by

$$
\begin{aligned}
\tilde{I}_{\lambda,\gamma}(\theta) &= 2\lambda^2 \, \text{SNR} \, \sigma_H^2 \Big[\sum_{j=1}^{M} \frac{\gamma \, p'_\theta(j)^2}{\lambda \sigma_H^2 \, \text{SNR} \, p_\theta(j) + 1} \\
&\quad + \sum_{j=1}^{M} \frac{\sigma_H^2 \text{SNR} \, p'_\theta(j)^2}{(\lambda \sigma_H^2 \, \text{SNR} \, p_\theta(j) + 1)^2} \Big].
\end{aligned}
\tag{20}
$$

Proof: By substituting in the expression for Fisher Information of Gaussian Distribution. □

We now provide expression for $E(\lambda)$ as $\lambda \to \infty$. We use this result to draw conclusions on the existence of optimal λ_*.

Theorem 4 (Limiting properties): The Fisher Information $\tilde{I}_{\lambda,\gamma}(\theta)$ is a monotonically increasing function of coherence index γ, average transmission rate λ and transmitter SNR. For finite γ, the normalized Fisher Information $\tilde{E}^{\text{MLE}}(\lambda, \theta)$ converges to a finite limit proportional to coherence index γ as $\lambda \to \infty$ given by

$$
\lim_{\lambda \to \infty} \tilde{E}^{\text{MLE}}(\lambda, \theta) = 2\gamma \sum_{j=1}^{M} \frac{p'_\theta(j)^2}{p_\theta(j)}.
\tag{21}
$$

We now investigate the case when the channel is perfectly coherent: $\mu_H = 1$ and $\sigma_H \to 0$ implying $\gamma \to \infty$.

Theorem 5 (Coherent channels): In the absence of fading

$$
\lim_{\sigma_H \to 0} \tilde{E}^{\text{MLE}}(\lambda, \theta) = 2\lambda \, \text{SNR} \sum_{j=1}^{M} p'_\theta(j)^2.
\tag{22}
$$

To contrast the perfectly coherent case, we examine the case when the channel is non-coherent, *i.e.,* $\mu_H = 0$ ($\gamma = 0$). Interestingly, the dependency of Fisher Information on the average transmission rate λ, SNR, and channel variance σ_H^2 can be summarized using a single parameter—the average receiver SNR,

$$
\chi \stackrel{\triangle}{=} \lambda \sigma_H^2 \text{SNR}.
\tag{23}
$$

Theorem 6 (Non-coherent channels): For the non-coherent channels ($\mu_H = 0$), the Fisher Information of the limiting distribution is a function of $\chi \stackrel{\triangle}{=} \lambda \sigma_H^2 \text{SNR}$ and satisfies the following properties:

1) $\tilde{I}_\chi$ is a monotonically increasing function of χ.
2) As $\chi \to \infty$, $\tilde{I}_\chi$ converges to a finite limit when $p_\theta(j) > 0$, $\forall j = 1, \ldots, M$.
3) Normalized function $\frac{\tilde{I}_\chi}{\chi}$ has unique a maximum.

The proofs for Theorem 4, 5 and 6 can be derived by evaluating (20). From a practical standpoint, the Gaussian approximation via CLT gives a computationally tractable way to approximate $E(\lambda)$ and therefore, the optimal λ_*. The accuracy of such approximation of course depends on the specific distributions, and we aim demonstrate it in Section IV.

IV. NUMERICAL RESULTS AND SIMULATIONS

In this section, we resort to numerical and simulation techniques to validate the theories developed in this paper. The channel fading is Gaussian $H_{i,j} \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(\mu_H, \sigma_H^2)$ and number of sensors involved in each transmission N_i is IID Poisson. We consider estimation of Bernoulli distributed data at the sensors with $\theta \in (0, 1)$ as the mean.

Since CLT is applicable only in large-λ regime, to draw conclusions for finite λ we numerically evaluated the Fisher Information. We found that the shapes are similar to detection error exponent plots [2]. We also found that the true $E^{\text{MLE}}(\lambda, \theta)$ and $\tilde{E}^{\text{MLE}}(\lambda, \theta)$ from Gaussian approximation have similar shapes and share the same trend with respect to λ, γ and SNR.

Fig. 3 shows the plot of true $E^{\text{MLE}}(\lambda, \theta)$ (without Gaussian Approximation) for a zero mean fading channel. The existence of an optimal λ_* is clearly seen in the plot. Fig. 4 and Fig. 5 show the behavior of $\tilde{E}^{\text{MLE}}(\lambda, \theta)$ vs. λ for different values of coherence index and SNR. Fig. 4 has varying channel coherence index γ with a fixed SNR and Fig. 5 has varying SNR for $\gamma = 0$. The existence of λ_* is evident for small γ.

V. Conclusions

In this paper, we focus on the communication aspect—random access in particular—of distributed statistical inference for large sensor networks. The main advance of this paper is the unification of optimal strategies for both detection and estimation under a common framework. By examining a number of extreme cases, we are able to obtain a general characterization of the performance metric for both detection and estimation as illustrated in Fig. 2. From a practical standpoint, using Gaussian approximation seems to give the correct insight into an optimal design. Such a characterization is a valuable guide, as a network designer pursues practical solutions.

We have left several important problems open. We have results on the existence of optimal λ_* only for extreme values of γ *i.e.,* 0 and ∞. Although we conjecture the existence of optimal λ_* for finite positive γ, with simulations confirming such a behavior, we have not proven it in this paper. We have also not dealt with the design of local quantization rule. A "cross-layer" optimization of local quantization, communications and global inference should be of interest.

Appendix

A. Proof of Theorem 2

For the PMF of N, $g(n, \lambda)$, applying Taylor's expansion for λ near zero, we have

$$g(0, \lambda) \approx 1 - \lambda a \quad \text{and} \quad g(1, \lambda) \approx \lambda a.$$

Define the conditional PDF of matched filter output $\mathbf{Y}$ given $N = 0, 1$ as

$$f_{\theta, \lambda}(\mathbf{y}|N=0) \stackrel{\Delta}{=} w(\mathbf{y}), \qquad f_{\theta, \lambda}(\mathbf{y}|N=1) \stackrel{\Delta}{=} h_\theta(\mathbf{y}),$$

where $w(\cdot)$ is the distribution of White Gaussian noise, independent of θ. Marginalizing over N, for small λ we have the PDF of $\mathbf{Y}$ as

$$f_{\theta, \lambda}(\mathbf{y}) \approx (1 - \lambda a)w(\mathbf{y}) + \lambda a h_\theta(\mathbf{y}).$$

Differentiating with respect to θ

$$\frac{\partial}{\partial \theta} f_{\theta, \lambda}(\mathbf{y}) \approx \lambda a \frac{\partial}{\partial \theta} h_\theta(\mathbf{y}).$$

From the definition of Fisher Information

$$E^{\text{MLE}}(\lambda, \theta) = \frac{I_\lambda(\theta)}{\lambda} = \frac{1}{\lambda} \int_{\mathbf{y}} \left(\frac{\partial}{\partial \theta} \log f_{\theta, \lambda}(\mathbf{y}) \right)^2 f_{\theta, \lambda}(\mathbf{y}) \mathbf{dy}.$$

Substituting for $f_{\theta, \lambda}(\mathbf{y})$

$$E^{\text{MLE}}(\lambda, \theta) = \int_{\mathbf{y}} \lambda a^2 \left(\frac{\partial}{\partial \theta} h_\theta(\mathbf{y}) \right)^2 \frac{\mathbf{dy}}{(1 - \lambda a)w(\mathbf{y}) + \lambda a h_\theta(\mathbf{y})}.$$

Since $f_{\theta, \lambda}(\mathbf{y})$ is a continuous function of λ and $\mathbf{y}$, $E^{\text{MLE}}(\lambda, \theta)$ is continuous in λ [16]. Therefore

$$\lim_{\lambda \to 0} E^{\text{MLE}}(\lambda, \theta) = 0.$$

For the case when $\lambda \to \infty$, again from the continuity of $f_{\theta, \lambda}(\mathbf{y})$ we have

$$\lim_{\lambda \to \infty} E^{\text{MLE}}(\lambda, \theta) = \lim_{\lambda \to \infty} \tilde{E}^{\text{MLE}}(\lambda, \theta).$$

Therefore we use the behavior of Fisher Information of asymptotic normal distribution. Refer Theorem 4, 5 and 6. $\square$

References

[1] A. Anandkumar and L. Tong, "A Large Deviation Analysis of Detection over Multi-Access Channels with Random Number of Sensors," in *Proc. of ICASSP'06*, Toulouse, France, May 2006.

[2] A. Anandkumar and L. Tong, "Type-Based Random Access for Distributed Detection over Multiaccess Fading Channels," *Submitted to IEEE Trans. Signal Proc.*, Dec. 2005.

[3] G. Mergen, V. Naware, and L. Tong, "Asymptotic Detection Performance of Type-Based Multiple Access Over Multiaccess Fading Channels," submitted to IEEE Trans. on Signal Processing, May 2005.

[4] Ke Liu and A. M. Sayed, "Optimal distributed detection strategies for wireless sensor networks,," in *42nd Annual Allerton Conf. on Commun., Control and Comp.*, Oct. 2004.

[5] R. Viswanathan and P.K.Varshney, "Distributed Detection with Multiple Sensors: Part I-Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[6] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE JSAC Special Issue on Wireless Sensor Networks*, 2004.

[7] S.A. Aldosari and J.M.F. Moura, "Detection in decentralized sensor networks," in *Proc. of ICASSP 04 Conf.*, Montreal, Canada.

[8] B. Chen, R. Jiang, T. Kasetkasem, and P.K. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Trans. on Signal Processing*, vol. 52, pp. 3454–3458, Dec. 2004.

[9] B. Chen and P.Willett, "On the Optimality of the Likelihood-Ratio Test for Local Sensor Decision Rules in the Presence of Nonideal Channels," *IEEE Trans. on Information Theory*, vol. 51, no. 2, pp. 693–700, Feb 2005.

[10] Te Sun Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[11] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 613–626, February 2006.

[12] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications, 2nd ed.*, Springer, NY, 1998.

[13] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

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

[15] P.Billingsley, *Probability and Measure*, vol. 3, Wiley Inter-Science, New York, NY, 1995.

[16] R.S.Strichartz, *The Way of Analysis*, vol. 2, Jones and Bartlett, Sudbury, MA, 2000.