Distributed Estimation via Random Access

Animashree Anandkumar, *Student Member, IEEE,*, Lang Tong†, *Fellow, IEEE* and Ananthram Swami *Senior Member, IEEE*

***Abstract*— The problem of distributed Bayesian estimation is considered in the context of a wireless sensor network. The Bayesian estimation performance is analyzed in terms of the expected Fisher information normalized by the transmission rate of the sensors. The sensors use a communication scheme known as the type-based random access (TBRA). Under a constraint on the expected transmission energy, an optimal spatio-temporal allocation scheme that maximizes the performance metric is characterized. It is shown that the metric is crucially dependent on the fading parameter known as the channel coherence index. For channels with low coherence indices, sensor transmissions tend to cancel each other, and there exists an optimal mean transmission rate that maximizes the performance metric. On the other hand, for channels with high coherence indices, there should be as many simultaneous transmissions as allowed by the network. The presence of a critical coherence index, where the change from one behavior to another occurs, is established.**

Index Terms—**Distributed inference, random access communications, sensor networks.**

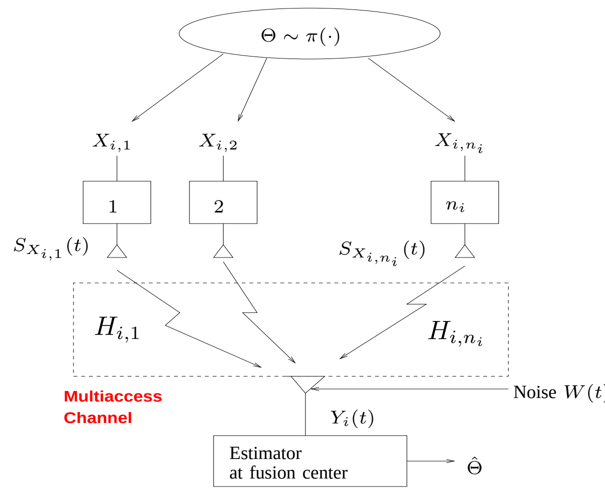

Fig. 1. Distributed Bayesian estimation in multi-access.

I. Introduction

We consider the distributed-estimation problem in the context of a wireless sensor network, when the number of reporting sensors is random. This may arise in large-scale wireless sensor networks, where random access may be the preferred medium access, as it does not require any centralized scheduling. Examples of random access include the ALOHA scheme, where sensors decide to transmit based on a simple coin-flip. Alternatively sensors may undertake a more sophisticated scheme and decide to transmit only significant data. Another scenario is when the fusion center is a mobile-access point and travels to different geographic locations, with nodes dispersed according to a point process. In this correspondence, we focus on the design of energy-optimal random-access schemes for distributed Bayesian estimation.

We employ the communication scheme known as type-based random access (TBRA), first proposed for distributed detection in [2], [3]. In TBRA, each sensor transmits probabilistically in a data collection, and the mean transmission rate λ (related to the probability of transmission) is the design parameter. The optimal TBRA can thus be obtained by maximizing an estimation performance metric (defined in section III-A) with respect to the mean transmission rate, under an energy constraint. We establish the existence of an optimal mean-transmission rate and its relationship with the channel fading characteristics.

Assuming constant energy for each sensor transmission, a constraint on the expected energy consumption translates to a constraint on the expected number of transmissions. Due to the presence of multi-access channel, we can have simultaneous sensor transmissions in a data-collection slot. A natural problem to consider is the optimal allocation of transmissions to spatial and temporal domains with the aim of maximizing a performance metric for estimation. Should energy be allocated to simultaneous transmissions, or should one collect more samples over time? In this correspondence, we illustrate the dependencies of the optimal TBRA allocation scheme on the channel fading characteristics.

There is extensive literature on distributed estimation. See [4] for a survey under the information-theoretic setup. Results on distributed estimation over multi-access channels are more recent. The type-based multiple access (TBMA) scheme was proposed in [5], [6]. TBMA, however, is only applicable when the mean of the fading is non-zero. In contrast, TBRA, originally proposed in [2], [3] for distributed detection, mitigates the canceling effects of zero-mean channel by choosing an optimal number of sensors to transmit and use multiple transmission slots. It is this crucial step that allows TBRA to provide consistent estimates, even when the fading channel has zero mean.

†Corresponding author.

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

A. Swami is with the Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: a.swami@ieee.org).

This work was supported in part through the collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011 and by the Army Research Office under Grant ARO-W911NF-06-1-0346. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

Part of this work has been presented in [1], [2]

II. Type-based Random Access

Distributed estimation via TBRA is illustrated in Fig.1, where we assume that a real random parameter Θ drawn from PDF $\pi(\theta)$ is to be estimated. 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 Poisson random variable with mean λ and probability-mass function (PMF) denoted by $g(n,\lambda)\stackrel{\Delta}{=}\Pr(N_i=n)$. We assume that the sequence N_i is IID.

In the i^{th} data collection, a sensor involved in the transmission[1], say sensor k, has quantized measurement $X_{i,k}\in\{1,\cdots,K\}$ (quantized to K levels). We assume that the sensor data $\{X_{i,k}\}$ are conditionally IID given θ, across time and sensors with PMF $p_\theta(j)\stackrel{\Delta}{=}\mathbb{P}[X_{i,k}=j|\Theta=\theta]$, for $j=1,\cdots,K$. In vector notation, the conditional PMF is given by

$$X_{i,k}|\theta \stackrel{\text{i.i.d}}{\sim} \mathbf{p}_\theta=(p_\theta(1),\cdots,p_\theta(K)).$$

In the i^{th} collection, the transmitter k encodes $X_{i,k}$ to a certain waveform and transmits it over a multi-access fading channel. As in TBMA, a set of K orthonormal waveforms $\{\phi_m(t), m=1,\cdots,K\}$ are used, each corresponding to a specific data value. Specifically, if $\mathcal{E}$ is the energy of one sensor transmission, then the signal transmitted by sensor k in collection i is $S_{i,k}(t)=\sqrt{\mathcal{E}}\phi_{X_{i,k}}(t)$.

The fading channel coefficients ($\tilde{H}_{i,k}\in\mathbb{C}$) are time-varying, IID across sensors and time. Assuming no inter-collection interference[2], the received complex-baseband signal after l data collections is

$$Y_i(t)=\sum_{k=1}^{N_i}\tilde{H}_{i,k}S_{i,k}(t-\tau_{i,k})+W_i(t),\quad i=1,\cdots,l, \quad (1)$$

where $\tau_{i,k}$ are the signal delays at the fusion center.

Under the narrow-band signal assumption, the flat-fading approximation which neglects the time dispersion in the signal is valid. Therefore, the delay is only through the carrier phase i.e., $S_{i,k}(t-\tau_{i,k})\approx S_{i,k}(t)\exp(-j2\pi f_c\tau_{i,k})$, where f_c is the carrier frequency. Defining a new fading statistic as $H_{i,k}\stackrel{\Delta}{=}\tilde{H}_{i,k}\exp(-j2\pi f_c\tau_{i,k})$ with mean $\mu_H\stackrel{\Delta}{=}\mathbb{E}(H_{i,k})$ and variance $\sigma_H^2\stackrel{\Delta}{=}\text{Cov}(H_{i,k})$, the received signal is thus given by

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

where we assume that $\{H_{i,k}\}$ are proper-complex Gaussian and unknown at the fusion center. The noise $W_i(t)$ is assumed to be zero-mean and complex white Gaussian with power density σ^2. We define the sensor signal to noise ratio by $\text{SNR}\stackrel{\Delta}{=}\frac{\mathcal{E}}{\sigma^2}$.

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

[2]Inter-collection interference can be removed by adding sufficient guard time between consecutive data collections.

We assume that the channel-state information is not known at the receiver. For the i^{th} collection, the bank of filters matched to orthogonal basis $\{\phi_k(t)\}$ generates

$$\begin{aligned}\mathbf{Y}_i &\stackrel{\Delta}{=} \frac{1}{\sqrt{\mathcal{E}}}\Big[\big\langle Y_i(\cdot),\phi_1(\cdot)\big\rangle,\cdots,\big\langle Y_i(\cdot),\phi_K(\cdot)\big\rangle\Big]\\ &= \sum_{k=1}^{N_i}H_{i,k}\mathbf{e}_{X_{i,k}}+\mathbf{W}_i,\end{aligned} \quad (3)$$

where $\big\langle Y_i(\cdot),\phi_k(\cdot)\big\rangle$ is the output of the matched filter corresponding to $\phi_k(t)$, $\mathbf{e}_k$ is the unit vector with non-zero entry at the k^{th} position, and $\mathbf{W}_i\stackrel{\text{i.i.d}}{\sim}\mathcal{N}(0,\frac{1}{\text{SNR}}\mathbf{I})$. Denote $\mathbf{Y}^l\stackrel{\Delta}{=}(\mathbf{Y}_1,\ldots,\mathbf{Y}_l)$. We define the sufficient statistics as

$$\mathbf{U}^l\stackrel{\Delta}{=}\frac{\mathbf{Y}^l}{\sqrt{\lambda^3}}, \quad (4)$$

where the normalization is needed to study the asymptotic behavior of the performance metric, defined in section III-A.

The design of TBRA is crucially dependent on the multi-access channel. We quantify this effect of the channel through a parameter known as the *channel-coherence index*,

$$\gamma=\frac{|\mathbb{E}(H)|^2}{\text{Cov}(H)}=\frac{|\mu_H|^2}{\sigma_H^2}, \quad (5)$$

where H is the effective fading coefficient between a sensor and the fusion center.

To see the intuition behind the coherence index γ defined in (5), we write explicitly the m^{th} entry of $\mathbf{Y}_i=[Y_{i,1},\cdots,Y_{i,K}]^{\text{T}}$

$$Y_{i,m}=\sum_{k=1}^{N_i}H_{i,k}1_{\{X_{i,k}=m\}}+W_{i,m}, \quad (6)$$

where $1_{\mathcal{A}}$ is the event-indicator function. The extreme case is when the channel is deterministic with $H_{i,k}\equiv 1$ ($\gamma=\infty$). 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 the notion of type-based transmission[3]. On the other hand, when $\gamma=0$, ($\mu_H=0$), the transmissions add up non-coherently, and the mean of received vector $\mathbb{E}[\mathbf{Y}_i]$ contains no information of the model.

Note that if the effective (or residual) channel phases $\arg(H_{i,k})$ are uniformly distributed, the channel is non-coherent ($\gamma=0$). Some degree of synchronization between the sensors and the fusion center is thus needed to attain a positive coherence index ($\gamma>0$). In practice, it is not possible to attain perfect coherence ($\gamma=\infty$).

[3]Given $X_{i,k}=x_{i,k}$, $N_i=n_i$ and the observation $\mathbf{Y}_i=\mathbf{y}_i$, in the absence of noise, the type of $x_{i,k}$ is $\frac{1}{n_i}\mathbf{y}_i$. [7], [8].

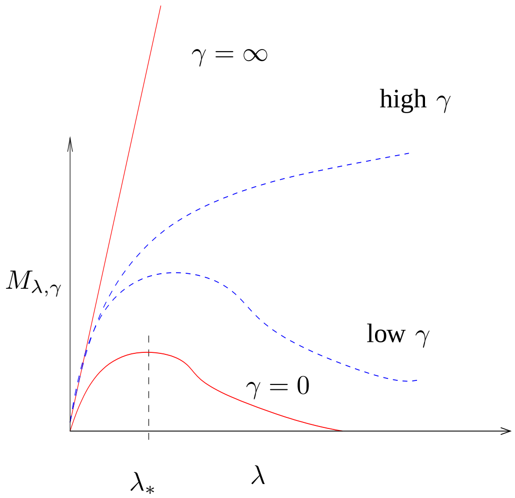

Fig. 2. Performance metric M_λ as a function of λ for different channel-coherence indices γ (see (5)).

III. Results on Optimal TBRA

A. Bayesian Cramér-Rao Bound

We define the performance metric for estimation as the normalized Bayesian Cramér-Rao lower bound (BCRB) [9]. Given expected number of transmissions ρ and mean transmission rate λ per data collection, let $\hat{\Theta}$ be a Bayesian estimator. Under some regularity conditions [9, p. 72], we have

$$\mathbb{E}(\hat{\Theta} - \Theta)^2 \geq \frac{1}{\frac{\rho}{\lambda}\mathbb{E}[I_\lambda(\Theta)] + A_\pi}, \tag{7}$$

with equality iff conditional PDF of $\mathbf{U}$, $f_U(\theta|\mathbf{u}^l)$, is Gaussian; and $I_\lambda(\theta)$ is the Fisher information of a single data collection of the sufficient statistic $\mathbf{U}$, for a given θ and A_π only depends on the PDF of Θ i.e., $\pi(\cdot)$. To obtain design guidelines, we define the normalized expected Fisher information, given by

$$M_\lambda \triangleq \frac{\mathbb{E}[I_\lambda(\Theta)]}{\lambda}, \tag{8}$$

where the expectation is taken over Θ. Maximizing the normalized Bayesian information with respect to λ, gives the least BCRB. In general, the BCRB is not achieved by the MMSE estimator. Note if we instead formulate θ as a deterministic parameter, then the optimal TBRA scheme would depend on θ. In addition to the regularity conditions for the existence of BCRB, we assume that the PDF $f_U(\mathbf{u}|\theta;\lambda)$ is differentiable up to second order ($\mathbb{C}^2$) in $\mathbf{y}, \theta$ and λ.

B. Optimal Transmission Rate

Having defined the performance metric, the design of optimal TBRA now reduces to finding an optimal transmission rate, with mean number of transmissions ρ fixed,

$$\lambda_* \triangleq \arg \sup_{\lambda \in \Re^+} M_\lambda. \tag{9}$$

Although the performance metric M_λ can be evaluated numerically for a given statistical model, it is of theoretical and practical significance to establish that λ_* is finite. This is because if λ_* is bounded, we need to design optimal sleeping strategies to limit interference. On the other hand, if λ_* is unbounded, the sensors simply need to transmit simultaneously to maximize performance.

The nature of λ_* is determined by the nature of interference between simultaneous transmissions, quantified by the channel coherence index γ in (5). In the following theorem, we establish the general shape of M_λ as shown in Fig.2, for extreme values $\gamma = 0$ (non-coherent channels) and $\gamma = \infty$ (perfectly coherent channels). Note that the role of γ in M_λ is embedded through joint PDF $f_U(\mathbf{u}, \theta; \lambda)$, which, we assume is a continuous function of γ. We therefore can infer the behavior of M_λ for very small and very large γ.

Theorem 1 (Existence of λ_*): Given the mean number of transmitting sensors λ, let $f_U(\mathbf{u}, \theta; \lambda)$ be the joint probability-density functions of the sufficient statistic $\mathbf{U}$ and Θ.

1) If the channel has zero-mean fading i.e., $\gamma = 0$ and if $p_\theta(k) > 0$ and $\frac{\partial}{\partial\theta}p_\theta(k) < \infty$ a.e, for each $k = 1, \ldots, K$. Then

$$\lim_{\lambda \to 0} M_\lambda = \lim_{\lambda \to \infty} M_\lambda = 0, \tag{10}$$

which implies that there exists $0 < \lambda_* < \infty$ such that

$$\sup_{\lambda} M_\lambda = \frac{1}{\lambda_*}\mathbb{E}[I_{\lambda_*}(\Theta)]. \tag{11}$$

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

$$M_\lambda = \Omega(\lambda) \tag{12}$$

as $\lambda \to \infty$, where the notation Ω means that λ is an exponentially tight bound.[4]

Proof: See appendix A. □

In the above theorem, we established the existence of a bounded optimal average-transmission rate λ_* for non-coherent channels ($\gamma = 0$). The intuition is that for these channels, 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 [5], [6]. A sharp contrast is the extreme case when the channel is perfectly coherent ($\gamma = \infty$). We establish that there does not exist an optimizing λ_*, which means that the optimal strategy is for all sensors to transmit at the same time, in order to take advantage of channel coherency.

IV. Gaussian Approximation

A key step in proving theorem 1 is the investigation of M_λ, as $\lambda \to \infty$. The idea is to use the continuity argument coupled with a version of the central limit theorem (CLT) involving random number of summands [10] to characterize M_λ as $\lambda \to \infty$.

[4] $\Omega(a(\lambda)) = \{b(\lambda) : 0 \leq c_1 a(\lambda) \leq b(\lambda) \leq c_2 a(\lambda), \forall \lambda > \lambda_o\}$ for some $c_1, c_2, \lambda_o > 0$.

We shall focus in this section on the single-collection model, and evaluate the Fisher information for a given θ using the limiting-conditional distribution as $\lambda \to \infty$. For ease of notation, we drop the time index i in (3), and consider the model

$$\mathbf{Y} = \sum_{k=1}^{N} H_k \mathbf{e}_{X_k} + \mathbf{W}, \tag{13}$$

where we have a random summand N with mean $\mathbb{E}(N) = \lambda$. When N is Poisson, for a given θ, the number of sensors transmitting a particular data level is independently Poisson by the property of marking.

Theorem 2 (Asymptotic distribution of $\mathbf{Y}$): Assume that the channel gains $\{H_m\}$ are IID with mean μ_H and variance σ_H^2, and the number of sensors N is Poisson. Then each entry of vector $\mathbf{Y}$ given θ is independent, and asymptotically Gaussian with

$$\frac{Y(k) - \lambda\mu_H p_\theta(k)}{\sigma_H\sqrt{\lambda p_\theta(k)}}\Big|\ \theta \xrightarrow{d} \mathcal{N}_c(0,1) \quad \text{as} \quad \lambda \to \infty. \tag{14}$$

Proof : See appendix B. □

Since $\mathbf{Y}_i$ is asymptotically Gaussian for a given θ, in the large-λ regime, the estimation problem can be approximated as follows: Estimate θ from a Gaussian random vector, which, given θ, is drawn from

$$\mathcal{N}_c\Big(\lambda\mu_H \mathbf{p}_\theta, \lambda\sigma_H^2 \mathrm{Diag}(\mathbf{p}_\theta) + \frac{\sigma^2}{\mathcal{E}}\mathbf{I}\Big). \tag{15}$$

We define the Gaussian metric as

$$\tilde{M}_\lambda \triangleq \frac{\mathbb{E}[\tilde{I}_\lambda(\Theta)]}{\lambda}, \tag{16}$$

where $\tilde{I}_\lambda$ is the Gaussian Fisher information. We now give the closed-form expression for $\tilde{M}_\lambda$ and specialize the results for coherent and non-coherent channels.

Lemma 1 (Gaussian metric): Let σ_H^2 be the channel variance, γ the channel-coherence index, $\mathrm{SNR} = \frac{\mathcal{E}}{\sigma^2}$ be the SNR per sensor. Denote $p'_\theta(k) \triangleq \frac{\partial}{\partial\theta} p_\theta(k)$. The Gaussian metric $\tilde{M}_\lambda$ is given by

$$\begin{aligned}\tilde{M}_\lambda \quad &= \quad 2\lambda\,\mathrm{SNR}\,\sigma_H^2 \mathbb{E}\Big[\sum_{k=1}^{K} \frac{\gamma\, p'_\Theta(k)^2}{\lambda\sigma_H^2\,\mathrm{SNR}\,p_\Theta(k) + 1} \\ &+ \quad \sum_{k=1}^{K} \frac{\sigma_H^2 \mathrm{SNR}\, p'_\Theta(k)^2}{(\lambda\sigma_H^2\,\mathrm{SNR}\,p_\Theta(k) + 1)^2}\Big].\end{aligned} \tag{17}$$

Proof: By substituting in the expression for Fisher information of Gaussian distribution for a given θ, and then taking the expectation. □

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

Theorem 3 (Limiting properties): The Fisher information $\tilde{I}_{\lambda,\gamma}(\theta)$ given θ, is a monotonically-increasing function of coherence index γ, average transmission rate λ and sensor SNR. Assume that $p_\theta(k) > 0$ and $p'_\theta(k) < \infty$ a.e, for each $k = 1, \dots, K$. For a fixed γ, the actual metric M_λ and the Gaussian metric $\tilde{M}_\lambda$ converge to the same finite limit, proportional to coherence index γ, as $\lambda \to \infty$, given by

$$\lim_{\lambda\to\infty} M_\lambda = \lim_{\lambda\to\infty} \tilde{M}_\lambda = 2\gamma \sum_{k=1}^{K} \mathbb{E}\Big[\frac{p'_\Theta(k)^2}{p_\Theta(k)}\Big]. \tag{18}$$

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

Theorem 4 (Perfectly Coherent Channels): In the absence of fading

$$\lim_{\sigma_H\to 0} \tilde{M}_\lambda = 2\lambda\,\mathrm{SNR} \sum_{k=1}^{K} \mathbb{E}[p'_\Theta(k)^2]. \tag{19}$$

Proof : Substituting $\sigma_H = 0$, we derive the expression by finding the Fisher information of $\mathcal{N}_c(\lambda\mathbf{p}_\theta, \sigma^2)$, given θ. □

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 given θ, can be summarized using a single parameter—the average receiver SNR for zero-mean fading,

$$\chi \triangleq \lambda\sigma_H^2 \mathrm{SNR}. \tag{20}$$

Theorem 5 (Non-coherent Channels): Assume that $p_\theta(k) > 0$ and $p'_\theta(k) < \infty$ a.e, for each $k = 1, \dots, K$. For non-coherent channels ($\mu_H = 0$), given θ, the Fisher information of the limiting distribution is a function of average receiver signal-to-noise ratio $\chi = \lambda\sigma_H^2\mathrm{SNR}$ and satisfies the following properties:

1) $\tilde{I}_\chi(\theta)$ is a monotonically-increasing function of χ.
2) As $\chi \to \infty$, $\tilde{I}_\chi(\theta)$ converges to a finite limit.
3) Normalized function $\frac{\tilde{I}_\chi}{\chi}(\theta)$ has a unique maximum and hence $\tilde{M}_\lambda$ has a unique maximum.

The proofs for theorem 3, 4 and 5 can be derived by evaluating (17). From a practical standpoint, the Gaussian approximation via CLT gives a computationally tractable way to approximate M_λ, and therefore, the optimal λ_*. The accuracy of such an approximation, of course, depends on the specific distributions, as we will demonstrate in section V.

A. Critical Coherence Index γ_*

In theorem 1, we have characterized the behavior of the metric $M_{\lambda,\gamma}$ and thereby the optimal transmission rate $\lambda_*(\gamma)$, for extreme values of the coherence index i.e., ($\gamma = 0$) and ($\gamma = \infty$). For finite positive γ, we expect smooth transition between these extreme behaviors, especially for well-behaved distributions. To study the nature of λ_*, it is crucial to characterize the slope of M_λ, since a negative slope at large-λ implies that λ_* is bounded. However, we can only numerically evaluate M_λ for finite λ.

If we impose an additional regularity condition that conditional PDF $f_\lambda(\mathbf{y}|\theta)$ is continuously differentiable to second

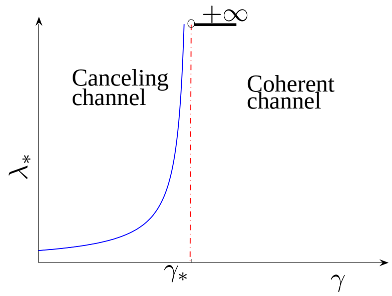

Fig. 3. Optimal transmission rate λ_* as a function of coherence index γ. Note the behavior around critical coherence index γ_*.

order, then the partial derivatives up to the second derivative are continuous [11]. Therefore,

$$\frac{\partial}{\partial\lambda}M_\lambda \to \frac{\partial}{\partial\lambda}\tilde{M}_\lambda, \quad \text{as} \quad \lambda \to \infty. \tag{21}$$

This condition is satisfied by well-behaved distributions. For the Poisson-Gaussian distribution, we can express the conditional PDF $f_\lambda(\mathbf{y}|\theta)$ as an infinite sum. On evaluating the limits, we find that it satisfies (21).

Therefore, at large-λ, we can reasonably approximate the slope of the actual metric by the slope of the Gaussian metric i.e.,

$$\frac{\partial}{\partial\lambda}M_\lambda \approx \frac{\partial}{\partial\lambda}\tilde{M}_\lambda. \tag{22}$$

Rewriting the Gaussian performance metric,

$$\tilde{M}_\lambda = 2\lambda\,\text{SNR}\,\sigma_H^2\mathbb{E}\Big[\sum_{k=1}^{K}\frac{\gamma\, p'_\Theta(k)^2}{\lambda\sigma_H^2\,\text{SNR}\,p_\Theta(k)+1} + \sum_{k=1}^{K}\frac{\sigma_H^2\text{SNR}\,p'_\Theta(k)^2}{(\lambda\sigma_H^2\,\text{SNR}\,p_\Theta(k)+1)^2}\Big], \tag{23}$$

we note that the two terms signify the opposing effects of coherence and cancelation respectively. This is because at large values of λ, the first terms approaches a constant, proportional to γ; whereas the second term decays to zero. Moreover, for all values of λ, the first term is increasing in λ and the second term is decreasing. Hence, if the first term dominates to such an extent that $\tilde{M}_\lambda$ is always increasing in λ, then the optimal λ_* is infinite. If the first term dominates for some value γ_*, then it dominates for all $\gamma > \gamma_*$. In the following theorem, we establish such a critical coherence index γ_*; signifying transition between these opposing effects.

Theorem 6: For the Gaussian metric $\tilde{M}_{\lambda,\gamma}$ given by (23), suppose the optimal transmission rate $\tilde{\lambda}_*(\gamma)$ is given by

$$\tilde{\lambda}_*(\gamma) \stackrel{\Delta}{=} \arg\sup_{\lambda\in\Re^+}\tilde{M}_{\lambda,\gamma}. \tag{24}$$

Then there exists a critical coherence index γ_* such that

$$\tilde{\lambda}_*(\gamma) = \infty\,, \forall\gamma > \gamma_*. \tag{25}$$

Additionally for $\gamma < \gamma_*$, the metric $\tilde{M}_\lambda$ is unimodal.

The critical coherence index γ_* given by,

$$\gamma_* = \sigma_H^2\text{SNR}. \tag{26}$$

Proof : We evaluate the sign of derivative of $\tilde{M}_\lambda$ with respect to λ. See appendix C for details. □

In the above theorem, we characterized the nature of optimal λ_* for finite positive γ. For well behaved distributions, the optimal $\lambda_*(\gamma)$ is a continuous function of γ (Fig.3). The critical coherence index γ_* divides the channels into two categories, viz.,

- coherent channels $(\gamma > \gamma_*)$: the optimal λ_* is unbounded, which implies that increasing the number of simultaneous transmissions always improves the performance metric.
- canceling channels $(\gamma < \gamma_*)$: λ_* is bounded and unique, which implies that increasing the number of simultaneous transmissions beyond a point degrades the performance metric.

Hence, for the canceling channels, we need to design sleeping strategies to limit interference. On the other hand for coherent channels, the sensors simply need to transmit simultaneously, in order to maximize performance.

V. 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 proper complex Gaussian $H_{i,k} \stackrel{\text{i.i.d}}{\sim} \mathcal{N}_c(\mu_H, \sigma_H^2)$ and number of sensors involved in each transmission N_i is IID Poisson. Θ is drawn from triangular distribution $\Delta(0.2, 0.8)$ with 0.2 and 0.8 as the end-points. We consider the estimation of Bernoulli-distributed data at the sensors with Θ as the mean i.e.,

$$\mathbf{p}_\Theta = [\Theta; 1-\Theta].$$

Since CLT is applicable only in large-λ regime, to draw conclusions for finite λ, we numerically evaluated the expected Fisher information. Fig.4 shows the plot of both true M_λ (without Gaussian approximation) and $\tilde{M}_\lambda$ (Gaussian approximation) for different values of coherence indices. We find that the true M_λ and $\tilde{M}_\lambda$ from the Gaussian approximation have similar shapes and share the same trend with respect to λ, γ and SNR. For larger values of γ, the Gaussian approximation does not appear to be good and needs large values of λ to converge. Fig.5 shows the accuracy of the Gaussian approximation in determining the optimal $\lambda_*(\gamma)$ for different values of γ. We find the Gaussian estimate to be quite close, especially at low values of γ, which are the practical cases of interest.

VI. Conclusions

We focus on the effect of the fading channel on the Bayesian estimation performance in a wireless sensor network. Although, we consider a scalar parameter, the results can be easily extended to vector parameter estimation. Given an energy budget, we provide an optimal spatio-temporal allocation

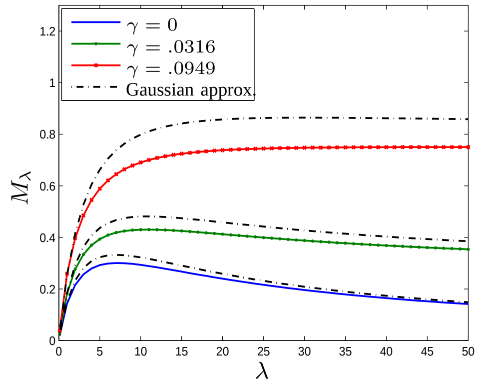

Fig. 4. Performance metric vs. transmission rate. (SNR$= -5$dB, $\sigma_H^2 = 1$, $\Theta \sim \Delta(0.2, 0.8)$)

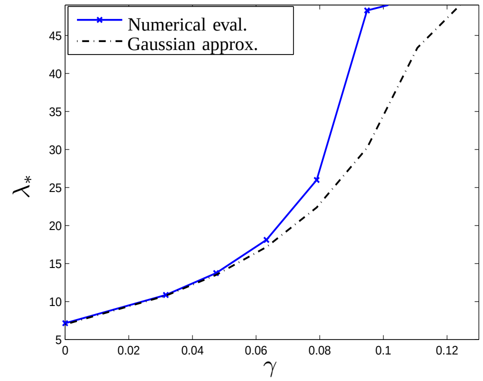

Fig. 5. Gaussian-approximated performance metric vs. transmission rate (SNR$=$ 0dB, $\sigma_H^2 = 1$, $\Theta \sim \Delta(0.2, 0.8)$)

that maximizes the normalized expected Fisher information. The nature of the optimal transmission rate is crucial for network design and is determined by the coherence index. For values of coherence index below a critical index, the optimal transmission rate is bounded whereas for values above it, it is unbounded. This critical index determines whether a sleeping strategy is needed to limit interference between the sensor transmissions. From a practical standpoint, this critical index is given by the product of sensor SNR and channel variance.

We have left several important problems open. We have not dealt with the design of the local-quantization rule. A "cross-layer" optimization of the local quantization, communications and global inference should be of interest. Another possibility is the extension of the problem to a sequential setup with optimal-stopping strategies.

APPENDIX

A. Proof of theorem 1

Let $\mathfrak{o}(\lambda)$ represent a function such that

$$\frac{\mathfrak{o}(\lambda)}{\lambda} \to 0, \quad \text{as} \quad \lambda \to 0.$$

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

$$\mathbb{P}(N_i = 1) = \lambda + \mathfrak{o}(\lambda) \quad \text{and} \quad \mathbb{P}(N_i > 1) = \mathfrak{o}(\lambda).$$

Define the conditional PDF of the sufficient statistic $\mathbf{U}$ given $N = 0, 1$ and θ as

$$\begin{aligned} f_U(\mathbf{u}|N = 0, \Theta = \theta; \lambda) &\triangleq w(\mathbf{u}), \\ f_U(\mathbf{u}|N = 1, \Theta = \theta; \lambda) &\triangleq h_\theta(\mathbf{u}), \\ f_U(\mathbf{u}|N > 1, \Theta = \theta) &\triangleq c_\theta(\mathbf{u}), \end{aligned}$$

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

$$f_U(\mathbf{u}|\theta; \lambda) = (1-\lambda-\mathfrak{o}(\lambda))w(\mathbf{y})+(\lambda+\mathfrak{o}(\lambda))h_\theta(\mathbf{y})+\mathfrak{o}(\lambda)c_\theta(\mathbf{y}).$$

Differentiating with respect to θ,

$$\frac{\partial}{\partial\theta} f_U(\mathbf{u}|\theta; \lambda) = (\lambda + \mathfrak{o}(\lambda))\frac{\partial}{\partial\theta} h_\theta(\mathbf{u}) + \mathfrak{o}(\lambda)\frac{\partial}{\partial\theta} c_\theta(\mathbf{u}).$$

From the definition of Fisher information

$$\frac{I_\lambda(\theta)}{\lambda} = \frac{1}{\lambda}\int_{\mathbf{u}} \left(\frac{\partial}{\partial\theta}\log f_U(\mathbf{u}|\theta;\lambda)\right)^2 f_U(\mathbf{u}|\theta;\lambda)\mathbf{dy}.$$

Since $f_U(\mathbf{u}|\theta; \lambda)$ is a differentiable function of λ and $\mathbf{u}$, M_λ is continuous in λ [11]. Substituting for $f_U(\mathbf{u}|\theta; \lambda)$ and taking the limit,

$$\lim_{\lambda\to 0} M_\lambda = \lim_{\lambda \to 0} \mathbb{E}\frac{I_\lambda(\Theta)}{\lambda} = 0.$$

1) Asymptotic Convergence: For the case when $\lambda \to \infty$, a limiting conditional distribution exists, by theorem 2. Let $\mathbf{Z} \triangleq \frac{\mathbf{Y}}{\sqrt{\lambda}}$. Therefore, the sufficient statistic is $\mathbf{U} = \frac{\mathbf{Y}}{\sqrt{\lambda^3}} = \frac{\mathbf{Z}}{\lambda}$. Let $f_Z(\mathbf{z}|\theta; \lambda)$ be the conditional PDF of $\mathbf{Z}$ and Θ respectively.

$$\begin{aligned} M_\lambda &= \frac{\mathbb{E}[I_\lambda^U(\Theta)]}{\lambda}, \\ &= \frac{\mathbb{E}[I_\lambda^Z(\Theta)]}{\lambda^2}, \end{aligned}$$

where I_λ^Z is the Fisher information of $\mathbf{Z}$ at a given λ. Let $\mathbf{G} \sim \mathcal{N}_c(0, \mathbf{\Sigma}_\theta)$, where $\mathbf{\Sigma}_\theta \triangleq \sigma_H^2 \text{Diag}(\mathbf{p}_\theta)$. The Gaussian Fisher information is given by

$$\begin{aligned} I^G(\theta) &= tr[\mathbf{\Sigma}_\theta^{-1}\frac{\partial \mathbf{\Sigma}_\theta}{\partial\theta}\mathbf{\Sigma}_\theta^{-1}\frac{\partial \mathbf{\Sigma}_\theta}{\partial\theta}], \\ &= \sum_{i=1}^{K} \frac{p_\theta'^2(i)}{p_\theta^2(i)}. \end{aligned}$$

We define $\mathbf{V}$ as

$$\mathbf{V} \triangleq \mathbf{Z} - \boldsymbol{\mu}_\theta, \quad \boldsymbol{\mu}_\theta \triangleq \lambda\mu_H \mathbf{p}_\theta.$$

and let $f_V(\mathbf{v}|\theta;\lambda)$ be the PDF of $\mathbf{V}$. From the local limit theorem for the densities [12], with the assumption that $\mathbb{E}[\mathbf{V}^k] < \infty$, for some $k \geq 3$, we have

$$\lim_{\lambda\to\infty} f_V(\mathbf{z}|\theta;\lambda) = f_G(\mathbf{z}|\theta).$$

Under the assumption of double differentiabilty of f_Z with respect of λ, θ and $\mathbf{z}$, the partial derivatives are also continuous.

$$\begin{aligned}\lim_{\lambda\to\infty} \frac{\partial}{\partial\theta} f_V(\mathbf{z}|\theta;\lambda) &= \lim_{\lambda\to\infty}\lim_{h\to 0} \frac{f_V(\mathbf{v}|\theta+h;\lambda) - f_V(\mathbf{v}|\theta;\lambda)}{h} \\ &= \frac{\partial}{\partial\theta} f_G(\mathbf{z}|\theta),\end{aligned}$$

where the limits can be interchanged, since f is assumed to be continuous in both λ and θ. Since the functions are continuous with respect to $\lambda \in \Re$, the limits and the expectations can also be interchanged. Therefore,

$$\lim_{\lambda\to\infty} I_\lambda^V(\theta) \to I^G(\theta).$$

Similarly, the expectation with respect to θ is also continuous. Now, in order to relate the Fisher information of $\mathbf{V}$ and $\mathbf{Z}$, we have

$$\begin{aligned}\frac{\partial}{\partial\theta}\log f_Z(\mathbf{z}|\theta;\lambda) &= \frac{\partial}{\partial\theta}\log f_V(\mathbf{z}-\boldsymbol{\mu}_\theta|\theta;\lambda) \\ &= \frac{\partial}{\partial\theta}\log f_V(\mathbf{v}|\theta;\lambda)\Big|_{\mathbf{v}=\mathbf{z}-\mu_\theta} \\ &\quad -\frac{\partial}{\partial\mathbf{v}}\log f_V(\mathbf{v}|\theta;\lambda)\Big|_{\mathbf{v}=\mathbf{z}-\mu_\theta}\frac{\partial\boldsymbol{\mu}_\theta^T}{\partial\theta}.\end{aligned}$$

Therefore,

$$\begin{aligned}I_\lambda^Z(\theta) &= \mathbb{E}[\frac{\partial}{\partial\theta}\log f_Z(\mathbf{Z}|\theta;\lambda)]^2 \\ &= I_\lambda^V(\theta) + \mathbb{E}[\frac{\partial}{\partial\mathbf{v}}\log f_V(\mathbf{V}|\theta;\lambda)\Big(\frac{\partial\boldsymbol{\mu}_\theta}{\partial\theta}\Big)^T]^2 \\ &\quad -2\frac{\partial\boldsymbol{\mu}_\theta}{\partial\theta}\mathbb{E}[\frac{\partial}{\partial\mathbf{v}}\log f_V(\mathbf{V}|\theta;\lambda)\frac{\partial}{\partial\theta}\log f_V(\mathbf{V}|\theta;\lambda)].\end{aligned}$$

The last term, under regularity conditions, is

$$\begin{aligned}&\mathbb{E}[\frac{\partial}{\partial\mathbf{v}}\log f_V(\mathbf{V}|\theta;\lambda)\frac{\partial}{\partial\theta}\log f_V(\mathbf{V}|\theta;\lambda)] \\ &= \int_{\mathbf{v}}\frac{\partial}{\partial\mathbf{v}}\log f_V(\mathbf{V}|\theta;\lambda)\frac{\partial}{\partial\theta}f_V(\mathbf{V}|\theta;\lambda)\mathbf{dv} \\ &= \frac{\partial}{\partial\theta}\int_{\mathbf{v}}[\frac{\partial}{\partial\mathbf{v}}\log f_V(\mathbf{V}|\theta;\lambda)]f_V(\mathbf{V}|\theta;\lambda)\mathbf{dv} \\ &= \frac{\partial}{\partial\theta}\int_{\mathbf{v}}\frac{\partial}{\partial\mathbf{v}}f_V(\mathbf{V}|\theta;\lambda)\mathbf{dv} \\ &= f_V(\infty) - f_V(-\infty) = 0,\end{aligned}$$

where we assume that the density is zero at infinity. For the second term, we have

$$\begin{aligned}&\lim_{\lambda\to\infty}\mathbb{E}[\frac{\partial}{\partial\mathbf{v}}\log f_V(\mathbf{V}|\theta;\lambda)\Big(\frac{\partial\boldsymbol{\mu}_\theta}{\partial\theta}\Big)^T]^2 \\ &= \mathbb{E}[\frac{\partial}{\partial\mathbf{v}}\log f_G(\mathbf{V}|\theta)\Big(\frac{\partial\boldsymbol{\mu}_\theta}{\partial\theta}\Big)^T]^2, \\ &= 2\sum_{i=1}^k\mathbb{E}[\frac{V_i^2}{\Sigma_\theta(i)^2}(\frac{\partial\mu_\theta(i)}{\partial\theta})^2], \\ &= 2\sum_{i=1}^k\frac{1}{\Sigma_\theta(i)}(\frac{\partial\mu_\theta(i)}{\partial\theta})^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\lim_{\lambda\to\infty} M_\lambda &= \lim_{\lambda\to\infty}\frac{\mathbb{E}[I_\lambda^Z(\Theta)]}{\lambda^2}, \\ &= \lim_{\lambda\to\infty}\frac{\mathbb{E}[I^G(\Theta)]}{\lambda^2} + 2\gamma\sum_{i=1}^K\mathbb{E}[\frac{p'_\Theta(i)^2}{p_\Theta(i)}], \\ &= 2\gamma\sum_{i=1}^K\mathbb{E}[\frac{p'_\Theta(i)^2}{p_\Theta(i)}].\end{aligned}$$

□

B. Proof of theorem 2

Recall the CLT with random number of summands [10, p. 369]. Let $X_1, X_2, \ldots,$ be IID random variables with mean 0 and variance σ^2, and $S_n = \sum_{i=1}^n X_i$. For each positive t, let ν_t be a random variable assuming positive integers as values; not necessarily independent of X_n. Suppose, there exist positive constants a_t and η such that $a_t \to \infty$, $\frac{\nu_t}{a_t} \xrightarrow{d} \eta$ as $t \to \infty$. Then

$$\frac{S_{\nu_t}}{\sigma\sqrt{\nu_t}} \xrightarrow{d} \mathcal{N}(0,1), \quad \frac{S_{\nu_t}}{\sigma\sqrt{\eta a_t}} \xrightarrow{d} \mathcal{N}(0,1). \tag{27}$$

In our case, parameter a_t corresponds to λ, ν_t to N. We have $\frac{N}{\lambda} \xrightarrow{d} \eta > 0$ and $\frac{1}{\sqrt{\eta\lambda}}\mathbf{W} \xrightarrow{d} \mathbf{0}$ as $\lambda \to \infty$. When N is Poisson, let $N^{(k)}$ be the number of sensors transmitting data level k. Since $N^{(k)}$ is a thinning Poisson process [13, p. 317], $N^{(k)}$ is independent for different data levels for a given θ and

$$N^{(k)}|\theta \sim \text{Poiss}(\lambda p_\theta(k)).$$

Therefore, the vector $\mathbf{Y}$ has independent entries, given θ. Applying the above mentioned central limit theorem for random summands, to each entry of the vector and extending to complex domain, we obtained the needed result. □

C. Proof of theorem 6

The sign of the derivative is crucial in determining the bounded nature of optimal $\tilde{\lambda}$. Differentiating (23) we obtain

$$\frac{\partial \tilde{M}}{\partial \lambda} = 2\mathbb{E}\sum_{k=1}^{K}\Big(\frac{\sigma_H^2 \mathrm{SNR} p'_\Theta(k)}{(\lambda\sigma_H^2 \mathrm{SNR} p_\Theta(k)+1)}\Big)^2 \Big(\frac{\gamma}{\sigma_H^2 \mathrm{SNR}} - \frac{\lambda\sigma_H^2 \mathrm{SNR} p_\Theta(k)-1}{\lambda\sigma_H^2 \mathrm{SNR} p_\Theta(k)+1}\Big).$$

Therefore, the sign of the function inside the expectation is determined by

$$\frac{\gamma}{\sigma_H^2 \mathrm{SNR}} - \frac{\lambda\sigma_H^2 \mathrm{SNR} p_\Theta(k)-1}{\lambda\sigma_H^2 \mathrm{SNR} p_\Theta(k)+1}, \quad k = 1, \ldots, K.$$

The term $\frac{\lambda\sigma_H^2 \mathrm{SNR} p_\Theta(k)-1}{\lambda\sigma_H^2 \mathrm{SNR} p_\Theta(k)+1}$ is an increasing function of λ for $\lambda > 0$ and attains maximum of 1 as $\lambda \to \infty$. The value of γ at which the sign reverses is therefore given by (26). This also implies the unimodality for $\gamma < \gamma_*$. □

REFERENCES

[1] A. Anandkumar and L. Tong, "Distributed statistical inference using Type Based Random Access over multi-access fading channels," in *Proc. of CISS '06*, Princeton, NJ, March 2006.

[2] ——, "Type-Based Random Access for Distributed Detection over Multiaccess Fading Channels," *IEEE Trans. Signal Proc.*, vol. 55, no. 10, pp. 5032–5043, Oct. 2007.

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

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

[6] K. 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.

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

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

[9] H. V. Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, vol. 1.

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

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

[12] V. Petrov, "On Local Limit Theorems for Sums of Independent Random Variables," *Theory of Probability and its Applications*, vol. 9, p. 312, 1964.

[13] S. Resnick, *Adventures in Stochastic Processes*. New York: Springer Verlag, 2002.