BANDWIDTH SCALING FOR EFFICIENT INFERENCE OVER A POWER-LIMITED MAC

Stefano Marano[(a)], *Vincenzo Matta*[(a)], *Lang Tong*[(b)*], *Peter Willett*[(c)]

[(a)]DIIIE, University of Salerno, via Ponte don Melillo, 84084 Fisciano (SA), Italy
[(b)]ECE Department, Cornell University, Ithaca, NY 14853, USA
[(c)] ECE Department, U-2157, University of Connecticut, Storrs, CT 06269, USA
marano@unisa.it vmatta@unisa.it ltong@ece.cornell.edu willett@engr.uconn.edu

ABSTRACT

We introduce a Likelihood Based Multiple Access (LBMA) communication/estimation scheme for nonrandom parameter estimation in wireless sensor networks with additive multiple access channels. Constraining the system in terms of energy and allowing the available number of degrees of freedom to scale as n^{α}, $0.5 < \alpha < 1$, we prove that LBMA is asymptotically efficient. Thus, the new scheme is appropriate for large networks. LBMA is, in addition, simple to implement and relies upon an intuitive approach.

***Index Terms*—** Wireless Sensor Network, Multiple access, Multiterminal inference, TBMA

1. INTRODUCTION

In decentralized inference problems in WSNs one is interested in making inference about a parameter embedded in the nodes' observations. It is here assumed that the nodes of the network deliver messages through a common Multiple Access Channel (MAC) to a single Fusion Center (FC), and a central concern is the design of the messages to be sent over the MAC subject to power and bandwidth constraints. The reference scenario is depicted in Fig. 1, where $s(x_i;t)$ is the waveform that the i^{th} node delivers over the MAC to convey information about the observation x_i. Individual waveforms are added by the MAC that also introduces the noise term $w(t)$. At the receiver (FC) $r(\boldsymbol{x};t)$ is used for computing the estimate of θ. Relying upon the additive structure of the MAC we design suitable waveforms $s(x_i;t)$ as well as the form of the final estimator.

We propose in this paper two multiaccess schemes that lead to the asymptotically efficient parameter estimation at the fusion center. A key feature of the proposed schemes is that the asymptotic efficiency is achieved with respect to the *un-quantized* sensor observations whereas conventional distributed estimation schemes usually assume local quantization before transmission.

*The work of Lang Tong was supported in part by 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

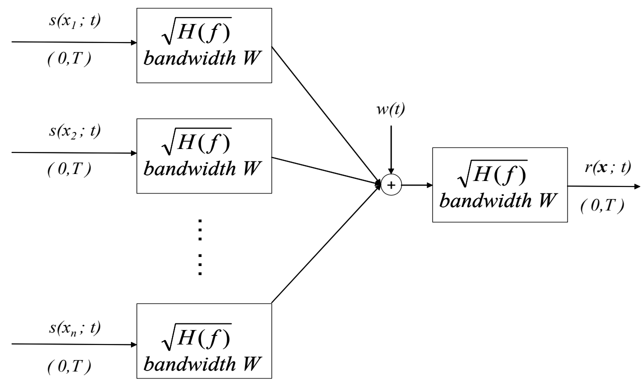

Fig. 1. Block scheme of the additive MAC. The notation $(0,T)$ denotes time windowing of the corresponding signal. $H(f)$ is an ideal $(-W,W)$ band-limiting filter, and $w(t)$ is the noise term.

The idea is to set $s(x_i;t)$ proportional to the local score $\partial \ln p(x_i;t)/\partial t$ (derivative of the log-likelihood), with t spanning the allowable values of the unknown θ. Thus, as the MAC *adds* these signals, the channel output contains a term proportional to the global score $\partial \ln p(\boldsymbol{x};t)/\partial t$. In the absence of noise and of other system constraints, from such a global score, the asymptotically efficient Maximum Likelihood (ML) estimate of θ can be simply obtained.

We show that the simple LBMA scheme is asymptotically efficient provided that transmission bandwidth scales with the number of sensors n in the network as n^{α} with $0.5 < \alpha < 1$. This implies that efficiency can be achieved with bandwidth per sensor node approaching to zero. We also argue informally that an optimal design of the processing at the sensor level, combined with an optimal exploitation of the channel structure, cannot achieve the performances of the LBMA, if these two steps are dealt with separately. In other words, separate (source/channel) schemes are inappropriate for our multiterminal inference task. This comes with no surprise because Shannon's separation theorems usually do not hold true in multiterminal scenarios.

The organization of the paper is as follows. Sections 2 and 3 deal with two different implementations on the LBMA idea, while Sect. 4 explores the performances obtainable with separate approaches. Sect. 5 summarizes the main issues. The main results of this paper are contained in two theorems whose proofs are only sketched, for space reasons. All the details can be found in [1].

2. LIKELIHOOD BASED MULTIPLE ACCESS

We assume that $\theta \in \Theta$ (Θ is an open finite interval) and that the observations x_i's at the nodes are independent and identically distributed (iid) samples drawn from the pdf $p(x;\theta)$. Also, we neglect fading effects as well as any synchronization concern so that when the nodes deliver suitable function of their observations to the FC through the MAC, the output of this latter is simply a bandlimited version of the sum of the inputs, with Gaussian noise added.

Specifically, denoting by $s(x_i;t)$ the waveform sent by the i^{th} sensor and by $s(\boldsymbol{x};t) = \sum_{i=1}^{n} s(x_i;t)$ the sum of that across the n sensors, the signal received at the FC is $r(\boldsymbol{x};t) = \tilde{s}(\boldsymbol{x};t) + w_{BL}(t)$, $0 < t < T$, where $\tilde{s}(\boldsymbol{x};t)$ represents the signal $s(\boldsymbol{x};t)$ processed by the ideal channel low-pass filter, $w_{BL}(t)$ is the bandlimited noise process added by the MAC, and $(0,T)$ is the transmission interval. The set Θ can be always mapped onto the transmission interval and, for simplicity, we henceforth identify Θ with $(0,T)$.

We assume that two constraints must be verified. First, the *average energy* $\mathcal{E}$ spent by each node for transmitting $s(x_i;t)$ must be bounded

$$E\left[\int s^2(x_i;t)dt\right] \leq \mathcal{E}. \tag{1}$$

The second constraint is on the number $2WT$ of degrees of freedom of the waveform set, where W is the available bandwidth. In the following we fix T and allow W to vary.

Assume now that $s(x_i;t) = A\partial \ln p(x_i;t)/\partial t$, $0 < t < T$, so that sensors deliver the derivative of the log-likelihood computed at the observed x_i, *i.e.*, they send over the MAC the *score function*. In the above A is a constant to be chosen in order to fulfill constraint (1). This yields $r(\boldsymbol{x};t) = s(\boldsymbol{x};t) + e(\boldsymbol{x};t) + w(t)$, where $e(\boldsymbol{x};t) = \tilde{s}(\boldsymbol{x};t) - s(\boldsymbol{x};t)$ and

$$s(\boldsymbol{x};t) = A\sum_{i=1}^{n}\frac{\partial \ln p(x_i;t)}{\partial t} = A\frac{\partial \ln p(\boldsymbol{x};t)}{\partial t}. \tag{2}$$

In the absence of noise and filtering, the received signal would be nothing but the score function (that based on the whole observation vector $\boldsymbol{x}$), and it would be a simple matter to compute the ML by integration and maximization. This motivate our definition of the (analog) LBMA estimator

$$\widehat{\theta} = \arg\max_{t\in[0,T]} l(\boldsymbol{x};t) \tag{3}$$

$$\text{with } l(\boldsymbol{x};t) = \int_0^t r(\boldsymbol{x};\xi)d\xi = \ln p(\boldsymbol{x};t) - \ln p(\boldsymbol{x};0) + \int_0^t e(\boldsymbol{x};\xi)d\xi + \int_0^t w(\xi)d\xi. \tag{4}$$

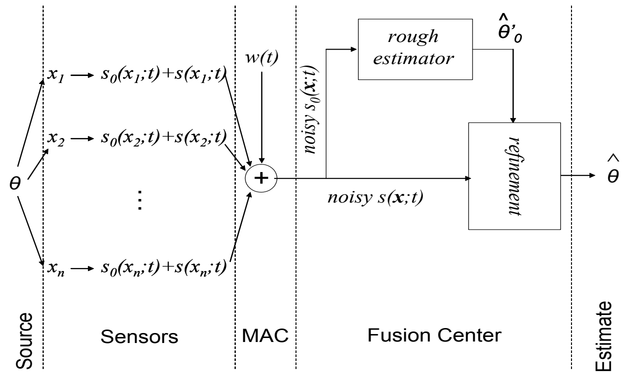

Fig. 2. Conceptual block diagram for discrete LBMA. Sensors observe the x_i's and build two waveforms $s_0(x_i;t)$ and $s(x_i;t)$ carrying information about the rough estimator $\widehat{\theta}_0'$ and about the refinement, respectively. These contributions are recovered at the receiver which outputs the final estimation $\widehat{\theta}$ according to the Fisher scoring approach.

The following theorem states the main properties of the LBMA estimator.

Theorem 1: Let $\widehat{\theta}$ be the LBMA estimator. If the bandwidth scales with the network size as $W \sim n^{\alpha}$, $0.5 < \alpha < 1$, then $\widehat{\theta}$ is asymptotically efficient, i.e., $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1/I(\theta))$*, and complies with the stated constraints on the waveform energy $\mathcal{E}$ and on the number of degrees of freedom.* $\triangle$

In the above $\mathcal{N}(a,b)$ denotes a Gaussian distribution with mean a and variance b, $I(\theta)$ is the Fisher information per sample for the parameter θ, and $\xrightarrow{d}$ means convergence in distribution.

It is worth noting the role of α. As turns out in proving the theorem, a coefficient $\alpha > 0.5$ ensures that the error due to filtering can be asymptotically ignored, while the condition $\alpha < 1$ serves to avoid that too much noise affects the system. In practice, of course, when $\alpha > 1$ we are simply saying that the excess of bandwidth cannot be exploited for improving the system performance.

Note also that the stated convergence holds true regardless of the value of $\mathcal{E}$, this latter only ruling the rate of convergence. It could be also of interest to investigate the scaling behavior with respect to the energy, see also [5].

Sketch of the Proof. The proof amounts to showing that the filtering error can be controlled with the available scaling law of the bandwidth with n, and that the noise term can be neglected when n grows to infinity. For space reasons here we offer only an outline of the proof that can be found in its rigorous and complete form in [1].

In [1] we show that (i), on the average, the filtering error $e(\boldsymbol{x};t)$ is uniformly upper bounded with respect to both the time t and the true parameter θ; and (ii) that the supremum of the noise term goes to zero in probability, namely

$\sup_{t\in(0,T)} \frac{|w_{BL}(t)|}{\sqrt{n}} \xrightarrow{p} 0$. On the other hand, from eqs. (4) and (3) we see that $\widehat{\theta}$ basically differs from the ML estimator for the presence of the filtering error and of the noise and, exploiting the two above results, these can be shown to become negligible in the asymptote of $n \to \infty$. The consequence is that the same properties of the ML estimator can be asymptotically reached, that is the statement of Theorem 1. •

3. DISCRETE LBMA

A *discrete* version of the LBMA estimator can be conceived by a refining process. We assume that the FC computes first a rough estimator, say $\widehat{\theta}_0$, by exploiting suitable waveforms $s_0(x_i;t)$ sent by individual nodes, and then the performance of this rough estimator is improved to ensure asymptotic efficiency. The basic requirements for the rough estimator $\widehat{\theta}_0$ are: (i) it must be $\sqrt{n}$-consistent [2], and (ii) the bandwidth of the waveforms $s_0(x_i;t)$ must not grow with the number of sensors n. The latter property ensures that, supporting the FC with the set $\{s_0(x_i;t)\}_{i=1}^n$, has negligible impact on the degrees-of-freedom constraint, in the asymptote of $n \to \infty$.

Many estimators possess such properties. One choice is based on the so-called Type Based Multiple Access (TBMA) scheme [3, 4]. Thus, assume for the time being that $\widehat{\theta}_0$ is the TBMA estimator computed by arbitrary quantization of the sensors' observations x_i's (the $\sqrt{n}$-consistency easily follows from the known properties of the TBMA estimator, see [3, 4]). The waveforms $s_0(x_i;t)$ provided by the nodes in order to build the TBMA at the FC are just the observed outcomes of the quantized observations, encoded over a signal constellation such as an orthonormal set of κ functions. As κ scales with the number of quantization levels but does not scale with the number of sensors [3], the communication load is asymptotically negligible. Also, the waveforms' energy constraint is easily managed by setting the amplitudes of the signals [3]. Thus, providing the FC with the rough estimator poses no problems.

Consider now the refinement step. Let us introduce $K = 2WT$ orthonormal waveforms (also orthogonal to the previously introduces κ signals), say $\{\psi_k(t)\}_{k=1}^K$, and assume that each sensor conveys K uniformly spaced *samples*[1] of its local score, as the coefficients in a linear combination of the basis function $\{\psi_k(t)\}_{k=1}^K$. The aggregate signal received at the FC is[2]

$$s(\boldsymbol{x};t) = A\sum_{k=1}^{K}\left(\frac{\partial \ln p(\boldsymbol{x};t)}{\partial t}\right)_{t_k}\psi_k(t), \qquad (5)$$

where $t_k = (k-1/2)\,T/K$, $k = 1, 2, \ldots, K$.

The definition of the discrete LBMA is as follows.

- Let us define $\widehat{\theta}_0' = \arg\min_{t_k} |\widehat{\theta}_0 - t_k|$, and let m be the index k attaining the minimum, *i.e.*, $\widehat{\theta}_0' = t_m$. Such a $\widehat{\theta}_0'$ is used as rough estimator in place of $\widehat{\theta}_0$.
- Let $\phi(t)$ be the mth waveform of the basis functions, *i.e.*, $\phi(t) = \psi_m(t)$.
- Let r the projection of the MAC output signal (filtered and noisy version of $s(\boldsymbol{x};t)$ in eq. (5)) onto $\phi(t)$.

The discrete LBMA estimator is

$$\widehat{\theta} = \widehat{\theta}_0' + \frac{r}{nI(\widehat{\theta}_0')}\sqrt{\frac{M_e K}{\mathcal{E}}}. \qquad (6)$$

Exactly as for the analog LBMA case, we have the following result.

Theorem 2: Let $\widehat{\theta}$ be the discrete LBMA estimator. If the bandwidth scales as $W \sim n^\alpha$, $0.5 < \alpha < 1$, then $\widehat{\theta}$ is asymptotically efficient, i.e., $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1/I(\theta))$*, and complies with the stated constraints on the waveform energy $\mathcal{E}$ and on the number of degrees of freedom.* △

As for the analog LBMA a tradeoff exists. An exponent $\alpha > 0.5$ ensures that the t_k's sample accurately the interval $(0,T)$, which yields a 'good' (*i.e.*, $\sqrt{n}$-consistent) rough estimator $\widehat{\theta}_0'$. Conversely, an exponent $\alpha < 1$ limits the number of degrees of freedom used by the system and, as a consequence, limits the noise to a level that becomes asymptotically negligible.

Sketch of the Proof. The complete proof is provided in [1]. The basic ideas are now described. First, it can be easily seen that $\widehat{\theta}_0'$ is $\sqrt{n}$-consistent as $\widehat{\theta}_0$ is. Thus we can use the former instead of the latter as a rough estimator.

As to the refining step, this is based on the so-called Fisher Scoring (FS) method (see, *e.g.*, Theorem 4.19 in [2]): Let $\widehat{\theta}_0$ be a $\sqrt{n}$-consistent estimator of the parameter θ ('starting' estimator); then the refined estimator

$$\widehat{\theta}_{FS} = \widehat{\theta}_0 + \frac{1}{nI(\widehat{\theta}_0)}\left(\frac{\partial \ln p(\boldsymbol{x};t)}{\partial t}\right)_{\widehat{\theta}_0}, \qquad (7)$$

is asymptotically efficient, in the sense that $\sqrt{n}(\widehat{\theta}_{FS} - \theta) \xrightarrow{d} \mathcal{N}(0, 1/I(\theta))$.

It is not difficult to envisage that the second addend on the RHS of eq. (6) provides just the refinement required by the FS method, and the proof of Theorem 2 reduces to show that the effects of filtering and noise can be controlled. In [1] this is done by exploiting the special properties of the prolate-spheroidal wavefunctions (PSWFs) [6]. •

4. SEPARATE SOURCE-CHANNEL CODING

To assess an approximate comparison between the performances of the LBMA and those of a separate (source/channel coding) scheme, let us consider the the so-called quadratic

[1] Note that the score is sampled but these samples are not quantized: The adjective discrete is thus referred to the parameter (*i.e.*, time) axis.

[2] Subindex t_k means that the quantity in parentheses is computed at t_k.

Gaussian CEO problem. There, the observations x_i's, given θ, are normally distributed with mean θ and variance σ^2, and θ itself is Gaussian with zero-mean and variance σ_θ^2. As opposed to our setup, in the CEO problem the parameter to be estimated is random so that the comparison between the two approaches is only indicative (we are comparing two schemes working not under the very same setting).

The distortion rate function for the quadratic Gaussian CEO problem [8, 9] can be bounded as $D(R) \geq \frac{\sigma_\theta^2 \sigma^2}{\sigma_\theta^2 R + \sigma^2}$.

Assume that the source outputs one symbol each T seconds, a bandwidth W is available, and an average power constraint of $\mathcal{E}/T$ is imposed. Following [7] we have that, even in the over-idealized scenario of n co-located (or fully cooperative) sensors, the maximum rate achievable over a Gaussian MAC with noise spectral density $\mathcal{N}_0/2$ is

$$C = WT \log\left[1 + \mathcal{E}n^2/(\mathcal{N}_0 WT)\right].$$

Using now the previous formula for C in the above bound for the distortion rate $D(R)$, and assuming $W \sim n^\alpha$, $0.5 < \alpha < 1$, we get

$$D \geq \frac{\sigma_\theta^2 \sigma^2}{\sigma_\theta^2 C + \sigma^2} \sim \frac{1}{n^\alpha \log n}$$

so that the distortion scales worse than $1/n$. The conclusion is that a separate scheme, although over-idealized, cannot ensure the scaling law of the LBMA.

5. SUMMARY AND DISCUSSION

As in many multiterminal scenarios, the guidelines for designing inference over MAC are not so clear as in a single-terminal setting. This is in part due to the fact that Shannon's separation theorem usually does not apply to multiterminal environments. In addition, making inference about a parameter embedded in an ensemble of observations is a dramatically different problem with respect to that of recovering at the receiver side the original ensemble of samples.

Generally speaking, a convenient approach is to exploit the possible matching between optimum estimation structure and the channel input/output characteristic. In the case of the MAC considered here, the output of the channel is the sum of its inputs (plus noise). Thus, a very simple approach amounts to impose that each sensor delivers over the common MAC the locally measured score function (derivative of the log-likelihood). Indeed, at the FC these are added by the MAC, thus yielding the global score function that is sufficient for estimating θ. In a sense, once that we have carefully chosen the input signals, the optimal estimator is the channel itself.

Exploiting this straightforward idea we propose an estimation/communication strategy that we call LBMA (Likelihood Based Multiple Access) and prove its asymptotic efficiency in the limit of increasingly large number of sensors n. We investigate the system resources that guarantee the LBMA optimality. It turns out that, given an energy constraint for the waveform to be sent over the MAC, an available bandwidth W (or, equivalently, a number of degrees of freedom) scaling as n^α, $0.5 < \alpha < 1$ is sufficient. In fact, an exponent too small, say $\alpha < 0.5$, implies that W grows too slowly and the bandwidth is not sufficient to limit the filtering channel distortion; that is, the requisite log-likelihood function curvature cannot be represented. On the other extreme, if the bandwidth grows too fast, $\alpha > 1$, the noise term that impairs the system performance as too many noise enters the FC.

Two possible implementations of the LBMA idea are explored. Implementation of the analog LBMA is extremely simple in transmission, but requires some complexity at the fusion center where a search over a parameter space is to be carried over. The situation is reversed for the discrete LBMA, which requires a simpler structure at the receiver with a little more complex transmission (sensors') scheme.

In this paper, all the proposed schemes rely upon the assumption of perfect synchronization among sensors. In real bandpass channels, however, phase uncertainty and different link gains are usually expected. These issues can be addressed in strict analogy with [3], where possible remedies are illustrated. At any rate, we have performed numerical simulations (not reported here for the sake of brevity) to investigate the sensitivity of the proposed LBMA schemes. As one would expect, asynchronous transmissions may strongly degrade the system performance, while different gains for different links seem to cause more tolerable performance losses.

6. REFERENCES

[1] S. Marano, V. Matta, L. Tong, and P. Willett, "A likelihood based multiple access for estimation in sensor networks," submitted.

[2] H. Shao, *Mathematical Statistics*, Springer, 2 edition, 2003.

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

[4] Ke Liu and A.M. Sayeed, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," in *Proc. 2004 IEEE ICASSP*, vol. 3, pp. 873–6, 2004.

[5] Ke Liu, H. El Gamal and A.M. Sayeed, "On optimal parametric field estimation in sensor networks," in *Proc. 2005 IEEE SSP*, pp. 1170–5, 2005.

[6] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.

[7] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Info. Theory*, vol. IT-51, no. 3, Mar. 2005.

[8] J. Chen, X. Zhang, T. Berger, and S. Wickler, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," IEEE JSAC, vol. 22, no. 6, pp. 977–987, Aug. 2004.

[9] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," IEEE Trans. Info. Theory, vol. 44, no. 3, May 1988.