A Likelihood-Based Multiple Access for Estimation in Sensor Networks

Stefano Marano, Vincenzo Matta, Lang Tong, *Fellow, IEEE*, and Peter Willett, *Fellow, IEEE*

Abstract—In a wireless sensor network (WSN), the nodes collect independent observations about a nonrandom parameter θ to be estimated, and deliver informations to a fusion center (FC) by transmitting suitable waveforms through a common multiple access channel (MAC). The FC implements some appropriate fusion rule and outputs the final estimate of θ. In this paper, we introduce a new access/estimation scheme, here referred to as likelihood-based multiple access (LBMA), and prove it to be asymptotically efficient in the limit of increasingly large number of sensors n, when the used bandwidth is allowed to scale as $W \sim n^\alpha, 0.5 < \alpha < 1$. The proposed approach is easy to implement, and simply relies upon the very basic property that the log likelihood is additive for independent observations, and upon the fact that the (noiseless) output of the MAC is just the sum of its inputs. Thus, the optimal fusion rule is automatically implemented by the MAC itself.

Index Terms—Decentralized estimation, likelihood-based multiple access (LBMA), multiple access channel (MAC), wireless sensor network (WSN).

I. INTRODUCTION

IN recent years, there has been a renewed interest in designing efficient transmission schemes over multiple access channels (MACs), also driven by the growing interest in effective communications in wireless sensor networks (WSNs). Clearly, earlier schemes aimed at allocating multiple transmissions over the same channel date back many decades and are well known. This is the case, for instance, of time-division multiple access (TDMA), frequency-division multiple access (FDMA), and code-division multiple access (CDMA), where the multiple access is realized by separating different users in time, frequency, and code, respectively [1].

In decentralized inference problems in WSNs, the focus is not to recover with fidelity all the messages from many users, but rather to make inference about a phenomenon commonly observed by different "users," whose role in this context is taken by the sensors (or nodes) of the network. The nodes of the network deliver messages through a common MAC to a fusion center (FC) whose aim is to perform the inference task. Clearly, a central issue is the design of the messages to be sent over the channel in such a way that the eventual inference task can be successful, while complying with the finite resources of the channel, e.g., limitations in input energy, time, and bandwidth. The latter grows with the number of sensors, with an appropriate scaling law. Therefore, the model may be relevant to the dynamic spectrum allocation (see, e.g., [2]) because, as more and more nodes join the network, the overall bandwidth occupancy is judiciously scaled.

Should the system constraints be immaterial, the problem would reduce to the scenario of single-terminal inference and no new issues would arise: If the commonly observed phenomenon is modeled as a nonrandom parameter θ (as assumed hereafter) we are simply faced with a very classical setup in estimation theory [3]. In the presence of limited resources and in the case that the final aim of the system is not data recovery but inference, the appropriate mathematical tools, practical recipes, and design philosophies may drastically change with respect to both the standard estimation theory (see, e.g., [4]) and the standard multiple access approaches quoted earlier.

Accordingly, it comes as no surprise that the decentralized inference over an MAC is a less mature topic with respect to the classical (meaning not inference oriented) multiple access techniques. This notwithstanding, the literature of the last years is rich with valuable contributions about the general decentralized inference problem. A complete survey is out of our scope and we limit ourselves to refer the interested reader to [5]–[7] (see, in particular, [8], [9]) and [10], which also contain many bibliographic references.

A. Summary of Results

In this paper, we propose a new communication/estimation strategy which we call likelihood-based multiple access (LBMA). Our system is constrained in terms of the average energy per sensor $\mathcal{E}$, the time for transmission T, and the available bandwidth W. This latter is allowed to scale with the number of sensors, as not uncommon in multiuser communication systems [11].

We devise two different LBMA schemes, which will be referred to as analog and discrete LBMA, respectively. Implementing the transmission stage of the analog one is straightforward, while the computation of the estimator requires a search inside the parameter space of interest; on the other hand, the discrete one is more complex in terms of transmission, while being easier as to the evaluation of the estimator. The choice seems thus a matter of implementation.

The main result of this paper is that, if the bandwidth scales as $W \sim n^{\alpha}$, where n is the number of sensors, $0.5 < \alpha < 1$, both the analog and discrete LBMA provide an asymptotically efficient estimator $\widehat{\theta}$, that is, as $n \longrightarrow \infty$

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right) \qquad (1)$$

where $\xrightarrow{d}$ denotes convergence in distribution and $I(\theta)$ is the Fisher information per sample of the original (unquantized) observations.

We thus obtain an interesting tradeoff about how fast the bandwidth must scale with n, for the simple protocols considered in this paper to be efficient. An exponent sufficiently large $\alpha > 0.5$ ensures that the effect of bandwidth limitation can be controlled. However, if the bandwidth grows too fast, the performances worsen for the excessive amount of noise that enters the system; thus, in order to control the effect of noise, we need $\alpha < 1$. Clearly, any available bandwidth scaling faster than $\sim n^{1/2}$ guarantees asymptotic efficiency; the upper bound $\alpha < 1$ simply means that one should not exploit the bandwidth in excess because the benefit (more degrees of freedom) is overcome by the drawback (more noise).

It is also relevant that asymptotic efficiency is not paid in the coin of bandwidth requirements: The degrees of freedom per single user converge to zero.

As a by-product of the discrete LBMA scheme, we get a third strategy which is appropriate when a noiseless feedback channel connecting the fusion center to the remote nodes is available. For this scenario, we show that asymptotic efficiency can be obtained even for a number of degrees of freedom which does not increase with the number of sensors.

B. Related Work and Paper Organization

In [12] and [13], the estimation of a random parameter in the presence of a Gaussian MAC is considered, in the context of the so-called Gaussian chief executive officer (CEO) problem. It is shown that an uncoded transmission dramatically outperforms a coded scheme. This is due to the fact that the optimal estimator is additive with respect to the remote nodes' measurements, and the channel is additive, as well. There is a perfect *matching* between the observations' fusion rule to be implemented, and channel input–output characteristic. Thus, the additive nature of the MAC is profitably exploited if the pertinent estimator (or, more in general, a sufficient statistic) depends upon the sum of functions of individual observations, e.g., $\widehat{\theta} = h(\sum_{i=1}^{n} g(x_i))$.

In general, the estimators one is interested in do not meet this requirement. In these cases, an option is first to quantize the continuous observations (note that no optimal recipe is known for this) and then transmit the type (empirical distribution) of the resulting discrete observables. Once the problem is discretized (that is, we are dealing with random variables with a finite number of possible outcomes) the type-based multiple access (TBMA) approach can be exploited. The TBMA is introduced and studied in [14]–[16]; a relevant feature is that TBMA asymptotically reaches the Cramér–Rao bound for quantized observations. In our paper, we will find conditions

that allow achieving the Cramér–Rao pertaining to the original, unquantized data. In all the aforementioned papers, the bandwidth is held fixed, while in our setup, we let it grow with the number of sensors.

Cooperative strategies for designing optimal quantizers for distributed estimation are presented in [17] and [18]. A review on decentralized detection with multiple sensors is offered in [19] and [20]. In [21], distributed detection under a sum-rate constraint is considered, while in [22], quantization for detection in the presence of a MAC is addressed.

This paper is organized as follows. Section II states the decentralized estimation problem. Sections III and IV deal in depth with the analog and discrete LBMA schemes, respectively, including the favorable case that a feedback channel is available. In Section V, some practical issues are discussed, together with the results of numerical simulations. Section VI summarizes the main findings of the paper. Some lemmas exploited in proving the theorems are collected in the Appendix.

II. MODEL

We are faced with the estimation of a nonrandom parameter θ, by a network of n sensors that deliver some function of their observations to an FC through an MAC. The unknown parameter lies in an open finite interval $\Theta : \theta \in \Theta$, and sensors' observations x_i, $i = 1, 2, \ldots, n$, are assumed independent and identically distributed (i.i.d.), according to a probability density function (pdf) parametrized in θ

$$x_i \sim p(x; \theta).$$

It is also assumed that no fading affects the transmission system and that perfect synchronization is maintained among sensors.

The estimation task is accomplished by allowing the ith sensor to send over the MAC a certain waveform, say $s(x_i; t)$, that obviously depends upon the local observation x_i. As said, without posing any restriction on the network resources, the problem would be that of a classical inference task, to be solved with well-known statistical tools. Clearly, a sensor network is usually severely limited in many ways. Here, we work with two main constraints.

The first is on the *average energy* $\mathcal{E}$ spent by each node for transmitting $s(x_i; t)$: We enforce the condition

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

Equation (2) is a *weak* energy constraint: Indeed, in keeping the average energy bounded, we do not rule out those situations where arbitrarily large amounts of energy are requested to a sensor. While this approach turns out to be mathematically clean, in practice, some clever strategy (e.g., a clipping of the waveforms) must be conceived to face the problem.

The second constraint is on the number of degrees of freedom of the waveform set. Such a number can be approximated by $2WT$, where W is the available bandwidth and T represents the time window where $s(x_i; t)$ must lie.

A convenient model to account for these relevant practical aspects is schematically illustrated in Fig. 1 (see, also, [23]). Accordingly, the transmission of the waveforms involves the fol-
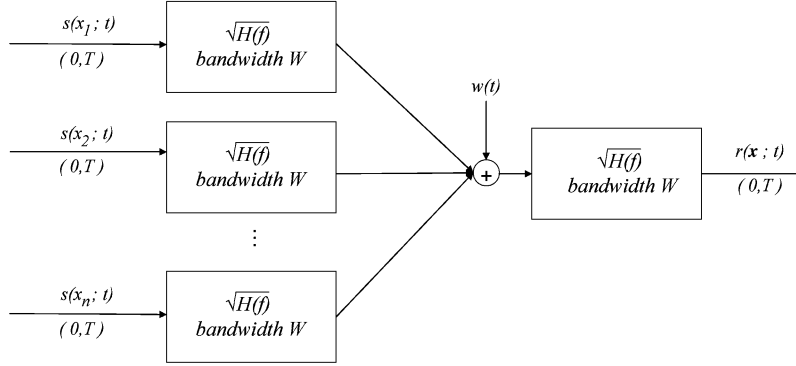
Fig. 1. Block scheme of the communication model adopted in this paper. The notation $(0, T)$ denotes time windowing of the corresponding signal to that time interval. The filter $\sqrt{H(f)}$ is an ideal band-limiting filter aimed to truncate the spectral content to the reference band $(-W, W)$. The waveform $w(t)$ represents a noise term.

lowing basic steps: 1) build up a $(0, T)$ time-limited waveform with (average, in our case) finite energy $\mathcal{E}$, 2) let the aforementioned waveform go through an ideal low-pass filter,[1] chopping the input outside the available bandwidth W, and 3) consider the output of the receiver filter within the original observation interval $(0, T)$.

At the output of the MAC, the white Gaussian noise $w(t)$ is added to the signal, and the resulting waveform is filtered and truncated as in Fig. 1, yielding

$$r(\boldsymbol{x}; t) = \tilde{s}(\boldsymbol{x}; t) + w_{\text{BL}}(t), \qquad 0 < t < T \qquad (3)$$

where $\tilde{s}$ represents the signal s processed by the ideal low-pass filter and w_{BL} is the band-limited noise process. Based on the received signal, the fusion center must produce the best estimate of the parameter θ and the waveforms to be sent by the individual nodes must be accordingly designed. The optimality criterion we elect for judging the goodness of the estimator is asymptotic efficiency [24], as previously detailed in (1).

III. ANALOG LBMA

A good starting point is to look for *analog waveforms* to be sent by individual sensors such that the resulting output of the MAC is suitable for the estimation purposes. As the (noiseless) MAC output is simply the sum of the inputs, we can make an obvious, but key, observation: The log likelihood, as well as any linear function thereof, is additive by its own nature.

To grasp the basic idea, let the parameter space Θ be mapped into a suitable time interval $(0, T)$; specifically, without losing generality, let us restrict[2] Θ to $(0, T)$. Suppose now that sensor i sends over the common channel the waveform

$$s(x_i; t) = A \frac{\partial \ln p(x_i; t)}{\partial t}, \qquad 0 < t < T$$

which is usually referred to (but for the amplitude factor) as the *score function*, and plays a prominent role in the theory of estimation. The amplitude A is suitably tuned to fulfill the energy constraint, as prescribed by (2).

---

[1]In our setup of ideal low-pass filters, the difference between $H(f)$ and $\sqrt{H(f)}$ in Fig. 1 becomes immaterial

[2]Whenever needed, we will assume that all the regularity conditions of our waveforms can be safely extended to the boundaries of the interval $(0, T)$, thus ruling out mathematical pathologies unlikely to occur in practice.

In what follows, we always assume that the pdf of the observations $p(x_i; t)$ meets the (mild) standard regularity conditions required for maximum-likelihood (ML) estimation to be asymptotically efficient (see the Appendix). In addition, we make a further assumption, basically needed for managing the energy constraint

$$\left(\frac{\partial \ln p(x; t)}{\partial t} \right)^2 < G(x) \qquad \forall t \in (0, T) \qquad (4)$$

with

$$\int_{-\infty}^{\infty} G(x) p(x; t) dx \leq M_e < \infty \qquad (5)$$

and with M_e independent of t, $\forall t \in (0, T)$.

Let now $\boldsymbol{x}$ be the vector whose entries are the x_is. In the absence of noise, the output of the MAC is simply the sum of the previous contributions, yielding

$$s(\boldsymbol{x}; t) = A \sum_{i=1}^{n} \frac{\partial \ln p(x_i; t)}{\partial t} = A \frac{\partial \ln p(\boldsymbol{x}; t)}{\partial t}, \qquad 0 < t < T \qquad (6)$$

thus amounting to the score computed on the whole set of data. Instances of the waveforms involved are shown in Fig. 2. This waveform is sufficient to get all the relevant information about the parameter θ, so that the FC may implement an optimal estimator.[3] We would like to emphasize here that the previous scheme essentially relies upon the additive nature of the log-likelihood function, which is ensured by the assumed independence among sensors. Thus, the scheme is expected to work also with different distributions across sensors, with multiple observations per single node (even locally dependent), and when the number of sensors is random or unknown, provided that independence among sensors is preserved, and that all the (energy, bandwidth) local constraints are simultaneously satisfied.

---

[3]In the work by Liu and Sayeed [15], it is shown that, in a decentralized detection problem with MAC, an optimal strategy consists in sending the local log-likelihood values. It is worth noting that the problem is quite different, due to the detection task as opposed to the estimation one we are dealing with. In fact, in the scheme of [15], only values of the log likelihood (i.e., numbers) are to be sent, and this gives no help in an estimation problem. The novelty of the approach we propose, instead, is in transmitting the log-likelihood analog *waveforms* (i.e., functions).
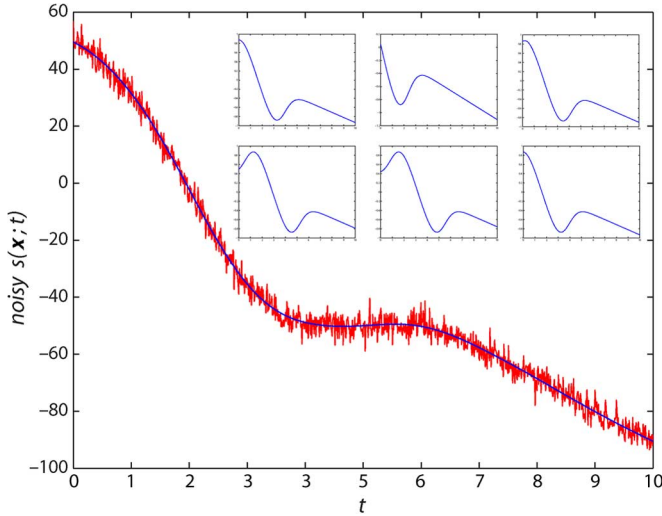
Fig. 2. Big panel depicts a typical waveform at the receiver stage, made of a noisy superposition of local score functions. In the small panels, examples of the constituent (noiseless) score functions are shown as function of time t. These are the waveforms sent over the MAC by individual sensors of the network; note how the specific shape of the scores depend upon the node's observation.

However, all the theorems in this paper are proved under the i.i.d. assumption, with a single observation per sensor, and with know n.

From (3) it is apparent that there are two basic obstacles to be dealt with, namely, the effect of the noise and that of the low-pass filter. Thus, the transmission scheme whose idealization is represented by the signal in (6) would be actually feasible for our estimation purpose provided that the following is true:

1) the noise component can be neglected for large n, due to the increasingly large number of waveforms which are *coherently* added;
2) the relevant (random) waveform is (on average) time limited and band limited (as any practical signal approximately is [25]) so that filtering has negligible effect.

While the first point can be expected to be rigorously proved with arguments similar to those used in [14] for the TBMA, the second issue unavoidably embodies an approximation, whose impact on estimation precision deserves investigation. To this end, let us fix the transmission time T; as the bandwidth W grows, the filtering error vanishes and so does its impact on the asymptotic $(n \longrightarrow \infty)$ performance of the estimator. Thus, an appealing approach is that of precisely defining the asymptotic setting in which the error arising from filtering vanishes, leading to asymptotically efficient estimators. We hence focus on the scaling law ruling the bandwidth growth with the number n of sensors in the network.

In its genuinely analog version, the LBMA scheme prescribes sending over the channel waveform (6). Note that condition (5) guarantees that the amplitude factor can be tuned so as to fulfill the energy constraint. According to (3) and assuming $A = 1$, the received signal can be written as

$$r(\boldsymbol{x};t) = s(\boldsymbol{x};t) + e(\boldsymbol{x};t) + w_{\mathrm{BL}}(t) \qquad (7)$$

where we have defined the error due to the low-pass filtering

$$e(\boldsymbol{x};t) = \tilde{s}(\boldsymbol{x};t) - s(\boldsymbol{x};t). \qquad (8)$$

In the absence of noise and filtering, the received signal would be nothing but the score function and to compute the ML estimator it would suffice to integrate the score function, obtaining the log likelihood to be maximized. Accordingly, it makes sense to define the estimator of θ as follows.

- Compute a noisy and distorted (as a consequence of filtering) log likelihood by integrating the received waveform

$$l(\boldsymbol{x};t) = \int_0^t r(\boldsymbol{x};\xi)d\xi$$

$$= \ln p(\boldsymbol{x};t) - \ln p(\boldsymbol{x};0) + \int_0^t e(\boldsymbol{x};\xi)d\xi + \int_0^t w_{\mathrm{BL}}(\xi)d\xi. \qquad (9)$$

- Define the (analog) LBMA estimator as

$$\widehat{\theta} = \arg \max_{t \in [0,T]} l(\boldsymbol{x};t). \qquad (10)$$

The naively introduced quasi-ML estimator $\widehat{\theta}$ given by (10) is in fact an optimal one, as claimed in the following result showing that the variance of such $\widehat{\theta}$ asymptotically achieves the Cramér–Rao limit of the inverse of Fisher information [26].

Theorem 1 (Asymptotic Efficiency of Analog LBMA): Let $\widehat{\theta}$ be the analog LBMA estimator. If the bandwidth scales as $W \sim n^\alpha$, $0.5 < \alpha < 1$, then $\widehat{\theta}$ is asymptotically efficient

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right). \qquad (11)$$

$\triangle$

Proof: As standard in ML estimation literature, the proof is split in two. First, we prove the consistency of the estimator, and then the asymptotic efficiency is demonstrated.

1) *Consistency.* We show that $\widehat{\theta}$ converges in probability to θ, in symbols $\widehat{\theta} \xrightarrow{p} \theta$. By definition of the ML estimator $\widehat{\theta}_{\mathrm{ML}}$

$$\frac{\ln p(\boldsymbol{x};\widehat{\theta}_{\mathrm{ML}})}{n} \geq \frac{\ln p(\boldsymbol{x};\widehat{\theta})}{n}. \qquad (12)$$

On the other hand, definitions (9) and (10) imply

$$\frac{\ln p(\boldsymbol{x};\widehat{\theta})}{n} \geq \frac{\ln p(\boldsymbol{x};\widehat{\theta}_{\mathrm{ML}})}{n} + \int_{\widehat{\theta}}^{\widehat{\theta}_{\mathrm{ML}}} \frac{e(\boldsymbol{x};\xi)}{n}d\xi + \int_{\widehat{\theta}}^{\widehat{\theta}_{\mathrm{ML}}} \frac{w_{\mathrm{BL}}(\xi)}{n}d\xi. \qquad (13)$$

In view of Lemma 2 (see the Appendix), it is immediate to write

$$E_\theta\left[\left\|\int_{\widehat{\theta}}^{\widehat{\theta}_{\mathrm{ML}}} \frac{e(\boldsymbol{x};\xi)}{n}d\xi\right\|\right] \leq \frac{\beta T}{W}$$

where $\beta > 0$ is an arbitrary constant. The fact that the previous expectation vanishes as W goes to infinity with n, together with a trivial application of Markov inequality, guarantees that the filtering error term vanishes in probability. As to the noise term, Lemma 3 immediately implies that the noise term in (13) vanishes in probability as n goes to infinity. Thus, (12) and (13), together with the asserted convergences, yield

$$\frac{1}{n}\left|\ln p(\boldsymbol{x};\widehat{\theta}) - \ln p(\boldsymbol{x};\widehat{\theta}_{\mathrm{ML}})\right| \xrightarrow{p} 0. \tag{14}$$

Strictly speaking, convergence of $\widehat{\theta}$ toward θ cannot be directly inferred from (14). This problem is commonly encountered when one tries to prove the consistency of the ML estimator, instead of that of a root of the likelihood equation; see, e.g., [27] and [28]. On the other hand, in the standard situation that the parameter is identifiable (i.e., $\theta_1 \neq \theta_2 \Leftrightarrow p(x;\theta_1) \neq p(x;\theta_2)$) and assuming verified some technical conditions (see again [27] and [28]), the ML consistency can be in fact proven. Checking all the technical conditions may be cumbersome [29]. In parallel to these arguments, we assume that (14) is enough for concluding

$$|\widehat{\theta} - \widehat{\theta}_{\mathrm{ML}}| \xrightarrow{p} 0 \tag{15}$$

whence consistency of $\widehat{\theta}$ follows from that of the ML estimator.

2) *Asymptotic efficiency.* By definition, $\widehat{\theta}$ satisfies[4]

$$\left(\frac{\partial l(\boldsymbol{x};t)}{\partial t}\right)_{\widehat{\theta}} = 0 \Leftrightarrow s(\boldsymbol{x};\widehat{\theta}) = -e(\boldsymbol{x};\widehat{\theta}) - w_{\mathrm{BL}}(\widehat{\theta})$$

where the subscript $\widehat{\theta}$ stands for evaluating the function at $t = \widehat{\theta}$. On the other hand, the score can be expanded as

$$s(\boldsymbol{x};\widehat{\theta}) = s(\boldsymbol{x};\theta) + s'(\boldsymbol{x};\theta)(\widehat{\theta}-\theta) + s''(\boldsymbol{x};\xi)\frac{(\widehat{\theta}-\theta)^2}{2}$$

with some ξ lying between $\widehat{\theta}$ and θ. This immediately yields

$$\sqrt{n}(\widehat{\theta}-\theta) = -\frac{s(\boldsymbol{x};\theta)/\sqrt{n}}{s'(\boldsymbol{x};\theta)/n + s''(\boldsymbol{x};\xi)(\widehat{\theta}-\theta)/(2n)}$$
$$-\frac{e(\boldsymbol{x};\widehat{\theta})/\sqrt{n} + w_{\mathrm{BL}}(\widehat{\theta})/\sqrt{n}}{s'(\boldsymbol{x};\theta)/n + s''(\boldsymbol{x};\xi)(\widehat{\theta}-\theta)/(2n)}. \tag{16}$$

The following list of convergences is true:
- $s'(\boldsymbol{x};\theta)/n \xrightarrow{p} -I(\theta)$ by the law of large numbers;
- $s''(\boldsymbol{x};\xi)(\widehat{\theta}-\theta)/(2n) \xrightarrow{p} 0$ by virtue of the proved consistency of $\widehat{\theta}$, and the fact that $s''(\boldsymbol{x};\xi)/n$ is bounded in probability in view of (31) and (32) in the Appendix;
- $s(\boldsymbol{x};\theta)/\sqrt{n} \xrightarrow{d} \mathcal{N}(0,I(\theta))$ thanks to the central limit theorem (CLT) [4];

[4] It is implicitly assumed that $\widehat{\theta}$ does not lie on the boundary of the parameter space.

- in view of Lemma 2

$$E_\theta\left[\frac{\left|e(\boldsymbol{x};\widehat{\theta})\right|}{\sqrt{n}}\right] \leq \sqrt{n}\frac{\beta}{W}$$

which can be made arbitrarily small provided that $W \sim n^\alpha$, for any $\alpha > 0.5$;
- finally, we know by Lemma 3 that the term $|w_{\mathrm{BL}}(\widehat{\theta})|/\sqrt{n}$ converges to zero in probability, provided that $W \sim n^\alpha$ with $\alpha < 1$.

It suffices now to invoke Slutsky's theorem (see, e.g., [24]) in (16) to complete the proof. ●

We would like to stress here the role of the exponent α, as already discussed in the introduction. A coefficient $\alpha > 0.5$ ensures that the error due to filtering can be asymptotically won, while the condition $\alpha < 1$ serves to avoid that too much noise affects the system.

IV. DISCRETE LBMA

In this section, we will devise a *discrete* counterpart to the LBMA proposed in Section III, which reaches asymptotic efficiency the same as analog LBMA does. Key to the development of our method is the following classic estimation procedure.

Fisher-Scoring [24, Th. 4.19]: Let $\widehat{\theta}_0$ be a $\sqrt{n}$-consistent estimator of the parameter θ ("starting" estimator). Then, the estimator

$$\widehat{\theta}_{FS} = \widehat{\theta}_0 + \frac{1}{nI(\widehat{\theta}_0)}\left(\frac{\partial \ln p(\boldsymbol{x};t)}{\partial t}\right)_{\widehat{\theta}_0} \tag{17}$$

is asymptotically efficient

$$\sqrt{n}(\widehat{\theta}_{FS} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right). \tag{18}$$

In order to employ the Fisher-scoring method, we need an initial guess $\widehat{\theta}_0$ which is $\sqrt{n}$-consistent. This is a stronger condition than consistency; see [24]. Obviously, this estimate cannot be computed at the remote sensors, in that each node only observes its own data. On the other hand, in general, there exist many different $\sqrt{n}$-consistent estimators to be plugged into (17), which can be computed by the FC when supported with appropriate data from the sensors. One of these estimators, which is available in the literature, is based upon the already mentioned TBMA.

TBMA assumes that the variables at the sensors are initially quantized. No matter what the quantization scheme is, the TBMA-estimator is asymptotically efficient, *but* with asymptotic variance given by the Fisher information of the quantized data. Notwithstanding this loss with respect to the optimal (no quantization) Fisher information, the efficiency property of TBMA is useful to our purposes in that, as one can easily verify (see, e.g., [24]), this implies its $\sqrt{n}$-consistency. Clearly, using TBMA is not necessary for our scheme to work, and any other possible $\sqrt{n}$-consistent estimator is good as well.

To compute the TBMA at the FC, it is only required that individual nodes transmit the observed outcomes (of the quantized
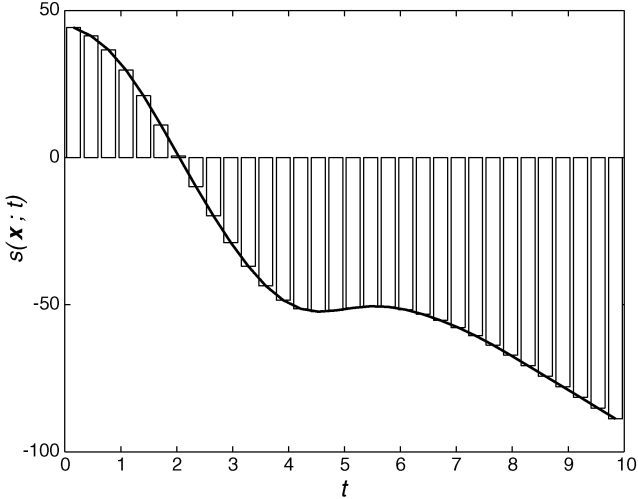
Fig. 3.   Score function, regarded as a time signal, here projected onto rectangular pulses.

observations) by exploiting a signal constellation made of an orthonormal set of functions. The number of these functions scales with the number of quantization levels, but *not* with the number of sensors [14]. Thus, the communication load associated with the TBMA estimator is actually very cheap. As to the energy constraint, it is also easily managed by suitably setting the amplitudes of the signalling waveforms, as shown in [14].

Once the FC is able to compute the initial guess $\widehat{\theta}_0$, we must provide the refinement represented by the second term of the right-hand side of (17). The whole message delivered by each sensor of the network is then made of two parts, with the former conveying information about $\widehat{\theta}_0$ and the latter aimed at recovering the correction term at the FC. In order to compute this refinement, we employ K orthonormal functions $\{\psi_k(t)\}_{k=1}^{K}$ to convey the *samples* of the local score function. Otherwise stated, these samples are simply understood as the coefficients of these K waveforms.[5] The resulting aggregate (sum over all the sensors) signal is[6]

$$ s(\boldsymbol{x};t) = A \sum_{k=1}^{K} \left(\frac{\partial \ln p(\boldsymbol{x};t)}{\partial t} \right)_{t_k} \psi_k(t) \qquad (19) $$

where

$$ t_k = \left(k - \frac{1}{2} \right) \frac{T}{K}, \quad k = 1, 2, \ldots, K \qquad (20) $$

is a set of equally spaced points between 0 and T. For illustration purposes, a typical such waveform is depicted in Fig. 3, in the case that the functions $\psi_k(t)$ are nonoverlapping rectangular pulses.

In addition, let us denote by $s_0(\boldsymbol{x};t)$ the waveform constructed for getting the plug-in estimator at the FC, e.g., the

---

[5]The name "discrete LBMA" can be misleading; we warn the reader that the adjective *discrete* is here used because the score is sampled but these samples are not quantized.

[6]With abuse of notation, we still use the symbol $s(\boldsymbol{x};t)$ but the signal obviously is different from that in (6); similarly, later, we loosely use the symbol $r(\boldsymbol{x};t)$.

kind of waveform employed in [14] for TBMA. Suppose that this signal is obtained as combination of (additional) κ orthonormal functions, i.e., $\psi_k(t), k = K+1, K+2, \ldots, K+\kappa$. The whole waveform travelling over the MAC is the aggregate that includes $s_0(\boldsymbol{x};t)$ and $s(\boldsymbol{x};t)$. A conceptual block scheme of discrete LBMA is depicted in Fig. 4.

Let us consider now the receiver side and suppose for the moment that the transmitted waveform is perfectly recovered. Actually, this is not the case, but we promptly show that the impact of noise and filtering can be made vanishing with increasingly large n.

It is evident that the Fisher-scoring method in (17) could be directly implemented if the plug-in estimator $\widehat{\theta}_0$ would coincide with some of the t_ks in (20). Obviously, this cannot be guaranteed. But the scoring method can still be realized by using as initial guess (plug-in) the t_k that is *closest* to $\widehat{\theta}_0$, in place of this latter. We only need to ensure that the set of the t_ks is sufficiently dense.

To formalize, let us define a $\widehat{\theta}'_0$ as

$$ \widehat{\theta}'_0 = \arg \min_{t_k} |\widehat{\theta}_0 - t_k|. \qquad (21) $$

Then

$$ \sqrt{n} \left| \widehat{\theta}'_0 - \theta \right| \leq \sqrt{n} |\widehat{\theta}_0 - \theta| + \sqrt{n} \left| \widehat{\theta}'_0 - \widehat{\theta}_0 \right| \leq \sqrt{n} |\widehat{\theta}_0 - \theta| + \frac{T}{2} \frac{\sqrt{n}}{K}. $$

Accordingly, the $\sqrt{n}$-consistency of $\widehat{\theta}_0$ implies that of $\widehat{\theta}'_0$, as long as

$$ K \sim n^{\alpha} \qquad (22) $$

for arbitrary $\alpha > 0.5$. Basically, we have replaced $\widehat{\theta}_0$ with an approximation $\widehat{\theta}'_0$ chosen in a sufficiently small neighborhood thereof, as compared to the asymptotic $\sqrt{n}$-decay.

Summarizing, the discrete LBMA consists of using $\widehat{\theta}'_0$ as starting point, and exploiting a slight modification of the Fisher-scoring method. This yields the following definition of the discrete LBMA estimator. Let m be the index in (21) that defines $\widehat{\theta}'_0$. Let r be the projection of the waveform at the output of the MAC over the mth orthonormal function. We define

$$ \widehat{\theta} = \widehat{\theta}'_0 + \frac{r}{nI\left(\widehat{\theta}'_0\right)} \sqrt{\frac{M_e K}{\mathcal{E}}}. \qquad (23) $$

In the following, we demonstrate that the estimation of θ can be implemented by complying with the required constraints (waveform energy $\mathcal{E}$, number of degrees of freedom), and the presence of noise can be circumvented for high n. As a consequence, the optimality (i.e., asymptotic efficiency) of the discrete LBMA is proved. The basic result of this section, which exactly mirrors Theorem 1 in Section III, is now given.

Theorem 2 (Asymptotic Efficiency of Discrete LBMA): 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

$$ \sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)} \right). \qquad (24) $$
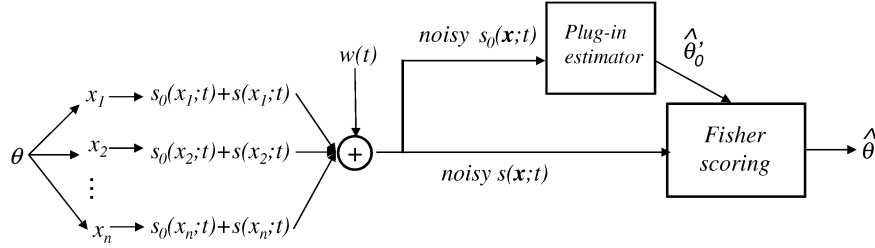
$\triangle$

Fig. 4. Conceptual block diagram for discrete LBMA. Sensors observe the x_is and build two waveforms $s_0(x_i; t)$ and $s(x_i; t)$ carrying information about the plug-in estimator $\widehat{\theta}'_0$ and about the refinement, respectively. These contributions are recovered at the receiver, allowing the FC to make the final estimation $\widehat{\theta}$ according to the Fisher scoring approach.

Proof: Waveform (19) is time limited to the interval $(0, T)$, provided that so are the functions $\psi_k(t)$. We will prove the theorem for the case that $\psi_k(t)$ are the so-called prolate-spheroidal wavefunctions (PSWFs) defined in the interval $(0, T)$, and with baseband scale parameter W [23]. In order to fulfill the energy constraint, the amplitude A must be chosen in such a way that the single-sensor energy

$$E \left\{ A^2 \sum_{k=1}^{K} \left[\left(\frac{\partial \ln p(x_i; t)}{\partial t} \right)_{t_k} \right]^2 \right\}$$

does not exceed $\mathcal{E}$. Taking into account (4) and (5), we immediately get

$$A \leq \sqrt{\frac{\mathcal{E}}{M_e K}}.$$

Using the PSWFs and assuming $A = \sqrt{\mathcal{E}/(M_e K)}$, the (single sensor) output of the band-limiting filter, further truncated again to the interval $(0, T)$, can be written as [23]

$$s(x_i; t) = \sqrt{\frac{\mathcal{E}}{M_e K}} \sum_{k=1}^{K} \left(\frac{\partial \ln p(x_i; t)}{\partial t} \right)_{t_k} \lambda_k \psi_k(t) \qquad (25)$$

where $\lambda_k = \lambda_k(WT)$ represents the fraction of energy of waveform $\psi_k(t)$ lying in the available bandwidth, and is a function of the product WT.

The quantities λ_k have the following relevant properties [25]:

$$1 > \lambda_k > \lambda_{k+1} > 0$$
$$\lim_{WT \to \infty} \lambda_{2WT(1-\delta)} = 1$$
$$\lim_{WT \to \infty} \lambda_{2WT(1+\delta)} = 0 \qquad (26)$$

where $\delta > 0$ is arbitrary. This suggests setting $K = 2WT(1 - \delta)$ as the appropriate number of degrees of freedom for the LBMA scheme. Then, the requirement expressed by (22) translates[7] in a scaling law for the bandwidth $W \sim n^\alpha$.

The output of the MAC (but for the first-stage signal $s_0(\boldsymbol{x}; t)$) is

$$r(\boldsymbol{x}; t) = \sqrt{\frac{\mathcal{E}}{M_e K}} \sum_{k=1}^{K} \left(\frac{\partial \ln p(\boldsymbol{x}; t)}{\partial t} \right)_{t_k} \lambda_k \psi_k(t) + w_{\mathrm{BL}}(t).$$

[7]Actually, the number of orthonormal functions used by the LBMA should take into account the transmission burdens for building the plug-in estimator at the FC. However, as already stated, it is possible (e.g., via TBMA) to derive this latter estimator by using a fixed number of orthonormal functions κ, which is not increasing with n. This would amount to considering $K + \kappa = 2WT(1-\delta)$ waveforms; the constant κ is immaterial for n (hence, WT) large enough.

As previously shown, the estimator in (21) can be retrieved as one out of K admissible values. Suppose that such an estimator corresponds to the mth index. Projecting the received signal onto the mth function, we obtain

$$r = \sqrt{\frac{\mathcal{E}}{M_e K}} \lambda_m \left(\frac{\partial \ln p(\boldsymbol{x}; t)}{\partial t} \right)_{\widehat{\theta}'_0} + w_m. \qquad (27)$$

If the subscript m was not random, w_m would be easily recognized to be Gaussian. On the other hand, m is random, so that we have to elaborate a little more on the noise term.

Given any two projections w_j and w_k, their statistical correlation is given by

$$E[w_j w_k] = \int_0^T \int_0^T E\left[w_{BL}(t) w_{\mathrm{BL}}(\tau) \right] \psi_j(t) \psi_k(\tau) dt d\tau$$

which, recalling that the correlation function of the band-limited noise is $\mathcal{N}_0 W \sin(2\pi W \tau)/(2\pi W \tau)$, can be also written as

$$\frac{\mathcal{N}_0}{2} \int_0^T \psi_j(t) \tilde{\psi}_k(t) dt$$

where $\tilde{\psi}_k(t)$ denotes the waveform $\psi_k(t)$ at the output of the ideal low-pass filter. The properties of the PSWFs [23] allow to conclude that the normalized noise projections $w_k/\sqrt{\lambda_k \mathcal{N}_0/2}$ are all independent, standard Gaussian random variables. In view of this property, the random index m is statistically independent of the first K noise components, because it is obtained (at the first stage) by looking to the subsequent κ components. This implies that, notwithstanding the randomness of m, the normalized projection $w_m/\sqrt{\lambda_m \mathcal{N}_0/2}$ is still a (standard) Gaussian.

We are now in a position to prove the asymptotic efficiency of discrete LBMA. In view of (27), the definition of the discrete LBMA estimator in (23) corresponds to

$$\widehat{\theta} = \widehat{\theta}'_0 + \frac{\lambda_m}{nI(\widehat{\theta}'_0)} \left(\frac{\partial \ln p(\boldsymbol{x}; t)}{\partial t} \right)_{\widehat{\theta}'_0} + \frac{w_m}{nI(\widehat{\theta}'_0)} \sqrt{\frac{M_e K}{\mathcal{E}}}.$$

Using the definition of $\widehat{\theta}_{FS}$ in the previous equation, we get

$$\sqrt{n}(\widehat{\theta} - \theta) = \lambda_m \sqrt{n}(\widehat{\theta}_{FS} - \theta) + (1 - \lambda_m)\sqrt{n} (\theta'_0 - \theta) + \frac{w_m}{I(\widehat{\theta}'_0)} \sqrt{\frac{M_e K}{\mathcal{E}n}}. \qquad (28)$$

We have the following.

- From properties (26), we know that $1 > \lambda_m \geq \lambda_K$ and $\lambda_K \longrightarrow 1$ as W increases in view of the assumed condition $K = 2WT(1 - \delta)$. This implies that λ_m converges to 1 in probability.
- In view of Slutsky's theorem, the previous item and the asymptotic efficiency of $\widehat{\theta}_{FS}$ make the first term on the right-hand side of (28) asymptotically Gaussian with zero-mean and variance $1/I(\theta)$.
- Having the $\sqrt{n}$-consistency of θ'_0, combined with the fact that $\lambda_m \xrightarrow{p} 1$, the second term in the right-hand side of (28) goes to zero in probability as $n \longrightarrow \infty$.
- $I(\widehat{\theta}'_0) \xrightarrow{p} I(\theta)$ by the continuity of Fisher information (usually ensured by the regularity conditions for ML estimation) and the $\sqrt{n}$-consistency of $\widehat{\theta}'_0$. Recalling that $K \sim n^\alpha$, any $0.5 < \alpha < 1$ ensures now that the last term in (28) converges in probability to zero in the limit $n \longrightarrow \infty$.

An immediate application of Slutsky's theorem [24] allows us now to conclude that

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right).$$

We have thus shown that the LBMA estimator $\widehat{\theta}$ defined in (23) is asymptotically efficient, provided that $W \sim n^\alpha$ with $0.5 < \alpha < 1$ ●

As with the analog LBMA, a tradeoff is worth noting. An exponent $\alpha > 0.5$ ensures that the t_ks sample accurately the interval $(0, T)$, which yields a "good" (i.e., $\sqrt{n}$-consistent) plug-in estimator $\widehat{\theta}'_0$; see (20) and (22). Conversely, an exponent $\alpha < 1$ provides that the number of degrees of freedom used by the system is not too large, because otherwise, too much noise impairs the estimator, as quantified by (28).

As a last remark, note that aside from purely theoretical issues such as scaling laws and convergences, there might be practical situations where the first stage estimator is constrained to be relatively rough, and the improvement provided by the second stage becomes particularly desirable.

A. LBMA With Feedback

In this section, we consider the more favorable situation that a feedback channel is available for communication from the FC to the remote sensors. According to an ideal assumption of unlimited resources at the FC, the feedback channel is modeled as perfectly noiseless. We will soon prove that, under these conditions, asymptotic efficiency can be reached by substantially lightening the communication burden from the sensors' perspective.

Let us illustrate how it is possible to capitalize on the Fisher scoring procedure when a feedback channel is available. Once the FC builds a plug-in estimate $\widehat{\theta}_0$ exploiting, for instance, the TBMA approach, the remote nodes can be made aware of this estimate via the feedback channel. Consequently, the sensors have to only transmit the correction (second term on the right-hand side of (17)). This correction is the score function *computed at* $\widehat{\theta}_0$. Thus, after feedback, sensors do not have to send the whole score functions and there is no longer need of a

bandwidth that grows with n. In addition, the presence of noise is asymptotically irrelevant, as can be shown with arguments similar to those used in Section IV. Thanks to the noiseless feedback channel, a finite number of degrees of freedom turns out to be enough for ensuring asymptotic efficiency. We have in fact (informally) proved the following result.

Theorem 3 (Asymptotic Efficiency of LBMA With Feedback): Let $\widehat{\theta}$ be the estimator obtained by LBMA with feedback. Even when the degrees of freedom are finite, $\widehat{\theta}$ is asymptotically efficient

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right). \tag{29}$$

$\triangle$

V. PRACTICAL ISSUES AND NUMERICAL SIMULATIONS

The proposed schemes rely upon the assumption of perfect synchronization among sensors, as it is the case of other multiple access strategies; see, e.g., [12] and [14]. However, wireless bandpass channels usually suffer phase uncertainty due to the timing difference between the transmitters' and receiver's clocks and/or the propagation delays. In addition, in wireless environments, the individual links may have different gains and, in the absence of an accurate power control, the received signals may have different amplitudes.

Investigating these issues in due detail is beyond the scope of this paper, and we limit ourselves to provide some insights and some preliminary numerical results. First, note that the two previous issues can be addressed in strict analogy with [14]. Indeed, for the LBMA scheme to be effective in practice, some kind of transmitter channel side information (CSI) may be needed to compensate for both the channel gains and the phase offsets. Such a CSI can be also obtained in a decentralized fashion, as already proposed in [14], capitalizing on a pilot tone; see, also, [30]. Furthermore, it is expected that the strict requirement of equal channel gains can be somewhat relaxed, as precisely shown in [14] for the TBMA.

Second, in absence of CSI countermeasures, the performance degradation suffered by the proposed multiple access schemes, due to the presence of unequal gains and/or to the lack of synchronization, can be investigated numerically. Accordingly, we have performed extensive simulations via Monte Carlo counting approach and the following general trend emerges: Asynchronous transmissions may strongly degrade the system performance, while different gains for different links may be better tolerable.

Just as an example corroborating the previous claim, let us consider a scenario in which the sensors' observations come from a mixture of two Gaussian random variables with the same mean $\theta = 1$ (unknown) and standard deviations 1 and 3, respectively. Let us assume that $\mathcal{E}/(\mathcal{N}_0 WT) = 0$ dB, and let $n = 250$ be the number of sensors. In this setting, the asymptotic mean square error (MSE) is $[nI(\theta)]^{-1} \approx 1.15 \times 10^{-2}$. The MSEs corresponding to the LBMA schemes in the presence of random gains [uniformly distributed in (0.5,1)] and/or of random time

TABLE I
EXAMPLE OF LBMA'S PERFORMANCE DEGRADATION DUE TO RANDOM
DELAYS AND GAINS. THE NUMBERS ARE THE MSES OBTAINED
WITH 10^3 MONTE CARLO TRIALS

	analog LBMA	discrete LBMA
nominal	1.3×10^{-2}	1.4×10^{-2}
gain	1.5×10^{-2}	1.8×10^{-2}
delay	4.0	1.0
gain & delay	4.0	1.0

delays [modeled by multiplying the synchronous signal by $\cos \phi$ with ϕ uniform in $(0, 2\pi)$] are given in Table I.

VI. CONCLUSION

A WSN is engaged in a nonrandom parameter estimation task. Many nodes observe independent random quantities drawn from a pdf parametrized in the unknown θ and deliver messages to the FC in order to make the final estimate $\hat{\theta}$. The communication medium between nodes and FC is an MAC impaired by additive white Gaussian noise (AWGN), and focus is made on the asymptotic setting of increasingly large number of sensors.

As in many such distributed inference issues, one cannot invoke Shannon's separation theorem as a guideline, in that this usually does not hold true. In fact, we have two peculiar facts to be accounted for. The first is that the network is designed for inference and not for reconstructing at the FC the individual observations made by the remote nodes. The second is the decentralized architecture of the system. As a consequence, the bit fails (or may fail) in being the universal building block to which the designer refers and the landscape becomes less certain.

In this paper, a transmission/estimation strategy that we call LBMA is proposed. Both an analog and a discrete version of the LBMA are considered. In the former case, the nodes simply transmit the score (derivative of the log likelihood) of their local observation; in the latter, a finite number of the samples of the score are sent over the MAC as coefficients of certain orthonormal functions. In both cases, we prove that asymptotic efficiency can be achieved provided that the number K of degrees of freedom assigned to the system (say, e.g., the used bandwidth) scales with the sensor number n faster than $K \sim n^{1/2}$. In addition, we study an architecture, in which the WSN is equipped with an (ideal) feedback channel from the FC to the nodes. In this case, we prove that asymptotic efficiency can be obtained even when the number of degrees of freedom stays constant while n diverges. The proposed methods are easy to implement and exploit a very simple intuition. As the overall log likelihood is the sum of those pertaining to the single observations at the sensors, an optimal estimate can be built by adding these latter at the FC. On the other hand, the output of the MAC is the sum of its inputs. Therefore, the optimum fusion rule is automatically provided at the FC by the output of the MAC.

All the theorems are proved assuming identical distributions across sensors. On the other hand, the log likelihood still is additive for independent but nonidentically distributed observations, so that extensions to this case seem to be possible. For the same reason, the scheme is very appealing for the case that multiple observations are available at each sensor, even if locally dependent. Another assumption has been the Gaussian channel noise,

but we believe that a broader class of stochastic processes will lead to the same asymptotic results, at the price of some additional mathematical burden.

APPENDIX

Here, we collect a number of lemmas exploited in proving the theorems in the main text. Whenever needed, we assume that the regularity conditions usually required for the asymptotic optimality of the ML estimator (see, e.g., [31]) are verified. These include [31]

$$\left| \frac{\partial p(x; t)}{\partial t} \right| < F_1(x) \qquad \left| \frac{\partial^2 p(x; t)}{\partial t^2} \right| < F_2(x) \qquad (30)$$

where $F_1(x)$ and $F_2(x)$ are integrable over $(-\infty, \infty)$ and

$$\left| \frac{\partial^3 \ln p(x; t)}{\partial t^3} \right| < H(x) \qquad (31)$$

with

$$\int_{-\infty}^{\infty} H(x) p(x; t) dx = M < \infty \qquad (32)$$

where M is independent of t. In addition, we require that the Fisher information per sample

$$I(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln p(x; \theta)}{\partial t} \right)_\theta^2 p(x; \theta) dx$$

is bounded inside the observation interval $(0, T)$, i.e.,

$$\sup_{\theta \in (0, T)} I(\theta) < \infty. \qquad (33)$$

Using the previous conditions, it is possible to explore the properties of the score function regarded as a signal in the time domain. We prove the following.

Lemma 1 (Properties of the Score Waveform): The expectations of the score function $s(x; t)$ and its first two derivatives, computed under the true distribution $p(x; \theta)$, are uniformly bounded with respect to both t and θ

$$E_\theta \left[\sup_{t \in (0, T)} \left| \frac{\partial^{(m)} s(x; t)}{\partial t^{(m)}} \right| \right] \leq \zeta < \infty, \qquad m = 0, 1, 2 \quad (34)$$

where ζ is a positive constant. △

Proof: The score function can be expanded as

$$\frac{\partial \ln p(x; t)}{\partial t} = \left(\frac{\partial \ln p(x; t)}{\partial t} \right)_\theta + \left(\frac{\partial^2 \ln p(x; t)}{\partial t^2} \right)_\theta (t - \theta)$$
$$+ \left(\frac{\partial^3 \ln p(x; t)}{\partial t^3} \right)_\xi \frac{(t - \theta)^2}{2} \quad (35)$$

where θ is the true value of the parameter, and ξ is a point lying between t and θ. Computing the derivatives of the logarithms and using (30) and (31), we get

$$\left| \frac{\partial \ln p(x; t)}{\partial t} \right| \leq \frac{F_1(x)}{p(x; \theta)} + \frac{F_2(x)}{p(x; \theta)} T$$
$$+ \left(\frac{\partial \ln p(x; t)}{\partial t} \right)_\theta^2 T + H(x) \frac{T^2}{2}. \quad (36)$$

This implies that the score function can be regarded as a bounded signal in the time domain, once the values x and θ have been fixed. In addition, on average, the score is *uniformly* bounded with respect to t and θ. Indeed, we can write

$$E_\theta\left[\left|\left|\frac{\partial \ln p(x;t)}{\partial t}\right|\right|\right] = \int_{-\infty}^{\infty}\left|\frac{\partial \ln p(x;t)}{\partial t}\right|p(x;\theta)dx. \quad (37)$$

Putting (36) into (37), and further using (32), (33), and the integrability of $F_1(x)$ and $F_2(x)$, we can easily conclude that the score is, on average, uniformly bounded with respect to t and θ. As a by-product of the previous derivation, one can easily recognize that the first and second derivatives of the score are, in turn, bounded functions for every fixed x and θ, and are, in addition, uniformly bounded on average, so that the whole (34) follows.   ●

Lemma 2 (Asymptotic Decay of the Filtering Error): The filtering error $e(\boldsymbol{x};t)$ is on average uniformly upper bounded with respect to both time t and true parameter θ as

$$E_\theta\left[\sup_{t\in(\mu/2,T-\mu/2)}\frac{|e(\boldsymbol{x};t)|}{n}\right] \leq \frac{\beta}{W} \qquad \forall \mu > 0 \quad (38)$$

where β is a suitable positive constant.   △

Proof: Consider the filtering error (8). We have

$$e(\boldsymbol{x};t) = \int_{t-T}^{t} s(\boldsymbol{x};t-\tau)\frac{\sin(2\pi W\tau)}{\pi\tau}d\tau$$

$$-s(\boldsymbol{x};t)\int_{-\infty}^{\infty}\frac{\sin(2\pi W\tau)}{\pi\tau}d\tau$$

which can be further split into the form

$$e(\boldsymbol{x};t) = \int_{t-T}^{t}\frac{s(\boldsymbol{x};t-\tau)-s(\boldsymbol{x};t)}{\pi\tau}\sin(2\pi W\tau)d\tau$$

$$-s(\boldsymbol{x};t)\int_{\tau>t,\tau<t-T}\frac{\sin 2\pi W\tau}{\pi\tau}d\tau. \quad (39)$$

Let us now define the function

$$g(\tau) = \frac{s(\boldsymbol{x};t-\tau)-s(\boldsymbol{x};t)}{\tau}$$

whose first derivative $\forall \tau \neq 0$ can be evaluated as

$$g'(\tau) = \frac{-s'(\boldsymbol{x};t-\tau)\tau+s(\boldsymbol{x};t)-s(\boldsymbol{x};t-\tau)}{\tau^2}.$$

The regularity properties of the score function and its derivatives imply that $g'(\tau)$ is continuous for all $\tau \neq 0$. Continuity can be, more or less obviously, extended to $\tau = 0$ by observing that $\lim_{\tau\to 0} g'(\tau) = s''(\boldsymbol{x};t)/2$. Moreover, using Taylor expansions for the functions s' and s at the numerator of previous equation yields

$$g'(\tau) = \frac{s''(\boldsymbol{x};t-\tau_1)\tau^2 - s''(\boldsymbol{x};t-\tau_2)\tau^2/2}{\tau^2}$$

$$= s''(\boldsymbol{x};t-\tau_1) - \frac{s''(\boldsymbol{x};t-\tau_2)}{2}$$

where τ_1 and τ_2 belong to $[0,\tau]$, implying that a simple upper bound on $g'(\tau)$ can be worked out in terms of the supremum s''_m of $|s''(\boldsymbol{x};t)|$, whose behavior has been already checked in Lemma 1.

The remarked good properties of $g'(\tau)$ allow now invoking the Riemann lemma [32], ensuring that the first error term in (39) is upper bounded by $\beta_1 s''_m/W$, with $\beta_1 > 0$. On the other hand, standard integration by parts allows bounding the second term in (39) as

$$|s(\boldsymbol{x};t)|\left|\int_{\tau>t,\tau<t-T}\frac{\sin 2\pi W\tau}{\pi\tau}d\tau\right| \leq \frac{\beta_2 s_m}{W}\left(\frac{1}{t}+\frac{1}{T-t}\right)$$

with $\beta_2 > 0$ and s_m being the supremum of $|s(\boldsymbol{x};t)|$. If we pose two guard bands of width $\mu/2$ such that the true value of the parameter is known to lie in $(\mu/2, T-\mu/2)$, we can *uniformly* upper bound the latter integral in the region of interest as

$$|s(\boldsymbol{x};t)|\left|\int_{\tau>t,\tau<t-T}\frac{\sin 2\pi W\tau}{\pi\tau}d\tau\right| \leq \frac{\beta_2 s_m}{\mu W}.$$

By application of the triangle inequality and Lemma 1, we conclude that

$$E_\theta\left[\sup_{t\in(\mu/2,T-\mu/2)}|e(\boldsymbol{x};t)|\right] \leq \frac{n\zeta}{W}(\beta_1+\beta_2/\mu)$$

and the proof is now complete.   ●

Lemma 3 (Noise Supremum Convergence): Let $w(t)$ be a zero-mean, stationary, and white Gaussian noise having spectral density $\mathcal{N}_0/2$, and consider its band-limited version $w_{\mathrm{BL}}(t)$, obtained by passing $w(t)$ through an ideal low-pass filter of bandwidth W. Assume the following scaling law for the bandwidth: $W \sim n^\alpha$, $\alpha < 1$. It holds true that

$$\sup_{t\in(0,T)}\frac{|w_{\mathrm{BL}}(t)|}{\sqrt{n}} \xrightarrow{p} 0.$$

△

Proof: The proof is a corollary of the following result due to Rice [33], which holds for every stationary Gaussian process $y(t)$ with almost surely continuous sample functions:

$$\Pr\left[\sup_{t\in(0,T)}y(t)>\gamma\right] \leq \left(\frac{\sigma}{\gamma\sqrt{2\pi}}+\frac{T}{2\pi}\sqrt{\frac{-c''(0)}{\sigma^2}}\right)e^{-\gamma^2/(2\sigma^2)} \quad (40)$$

where σ^2 is the noise variance and $c(\tau)$ is the covariance function of the noise process. In our setup, the relevant noise process is $w_{\mathrm{BL}}(t)/\sqrt{n}$, and it is straightforward to compute $c(\tau) = (\mathcal{N}_0 W/n)\sin(2\pi W\tau)/(2\pi W\tau)$, yielding

$$\sigma^2 = \frac{\mathcal{N}_0 W}{n} \qquad \lim_{\tau\to 0}c''(\tau) = -\frac{4\pi^2\mathcal{N}_0 W^3}{3n}$$

whence (40) becomes

$$\Pr\left[\sup_{t\in(0,T)}\frac{w_{\mathrm{BL}}(t)}{\sqrt{n}}>\gamma\right] \leq \left(\sqrt{\frac{\mathcal{N}_0 W}{2\pi n}}\frac{1}{\gamma}+\frac{WT}{\sqrt{3}}\right)$$

$$\times \exp\left(\frac{-\gamma^2 n}{2\mathcal{N}_0 W}\right). \quad (41)$$

Now, the event $\{\sup_{t\in(0,T)} |w_{\mathrm{BL}}(t)|/\sqrt{n} > \gamma\}$ is equivalent to the following:

$$\left\{ \sup_{t\in(0,T)} \frac{w_{\mathrm{BL}}(t)}{\sqrt{n}} > \gamma \bigcup_{t\in(0,T)} \inf \frac{w_{\mathrm{BL}}(t)}{\sqrt{n}} < -\gamma \right\}.$$

By the union bound and invoking symmetry arguments, we further have

$$\Pr\left[\sup_{t\in(0,T)} \frac{|w_{\mathrm{BL}}(t)|}{\sqrt{n}} > \gamma \right] \le 2\Pr\left[\sup_{t\in(0,T)} \frac{w_{\mathrm{BL}}(t)}{\sqrt{n}} > \gamma \right]$$

which, combined with (41), and further taking into account the scaling law for W, makes the proof complete. ●

REFERENCES

[1] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
[3] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
[4] S. Marano, V. Matta, and P. Willett, "Asymptotic design of quantizers for decentralized MMSE estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 11, Nov. 2007.
[5] B. Chen, W. B. Heinzelman, M. Liu, and A. T. Campbell, "Editorial," *EURASIP Journal on Wireless Communications and Networking*, Special Issue on Wireless Sensor Networks, no. 4, pp. 459–461, 2005.
[6] "Fundamental performance limits of wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, Aug. 2004.
[7] Z.-Q. Luo, M. Gastpar, J. Liu, and A. Swami, "Distributed signal processing in sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 14–15, Jul. 2006.
[8] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, Jul. 2006.
[9] M. Gastpar, M. Vetterli, and P. L. Dragotti, "Sensing reality and communicating bits: A dangerous liaison," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 70–83, Jul. 2006.
[10] B. M. Sadler, "Fundamentals of energy-constrained sensor network systems," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, pt. 2, pp. 17–35, Aug. 2005.
[11] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2059–2078, Sep. 2000.
[12] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.
[13] M. Gastpar, "To code or not to code," Ph.D. dissertation, Faculté Informatique et Communications, Section des Systémes de Communication, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland, 2002.
[14] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 613–626, Feb. 2006.
[15] K. Liu and A. Sayeed, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Montreal, QC, Canada, May 17–21, 2004, vol. 3, pp. iii-873–iii-876.
[16] K. Liu, H. El Gamal, and A. Sayeed, "On optimal parametric field estimation in sensor networks," in *Proc. IEEE/SP 13th Workshop Statist. Signal Process.*, Bordeaux, France, Jul. 17–20, 2005, pp. 1170–1175.
[17] W. M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Trans. Commun.*, vol. 41, no. 11, pp. 1602–1605, Nov. 1993.
[18] J. A. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.
[19] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors I. Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
[20] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors II. Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.
[21] J. F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.
[22] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 2, pp. 469–475, Apr. 1998.
[23] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[24] H. Shao, *Mathematical Statistics*, 2nd ed. New York: Springer-Verlag, 2003.
[25] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, no. 3, pp. 292–300, Mar. 1976.
[26] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part I*. New York: Wiley, 1968.
[27] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York: Springer-Verlag.
[28] A. Stuart, K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics: Classical Inference & the Linear Model*. Oxford, U.K.: Oxford Univ. Press.
[29] A. Wald, "Note on the consistency of the maximum likelihood estimate," *Ann. Math. Statist.*, vol. 20, 1949.
[30] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," presented at the Millitary Commun. Conf. (MILCOM), Boston, MA, Oct. 2003.
[31] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press.
[32] A. Papoulis, *Signal Analysis*, 1st ed. New York: McGraw-Hill, 1977.
[33] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Technical J.*, vol. 24, pp. 46–156, 1945.

Stefano Marano received the Laurea degree in electronic engineering (*cum laude*) and the Ph.D. degree in electronic engineering and computer science from the University of Naples, Italy, in 1993 and 1997, respectively.

Currently, he is a Professor at the University of Salerno, Fisciano, Italy, where he was formerly Assistant Professor. His areas of interest include statistical signal processing with emphasis on distributed inference, sensor networks, and information theory. He published more than 70 papers about these and related topics, including some invited, mainly in top international journals/transactions and proceedings of international conferences. He has also given several invited talks.

Prof. Marano is a Co-Recipient of the 1999 Best Paper Award for the paper published in the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION that included his work on stochastic modeling of electromagnetic propagation in urban areas (jointly with G. Franceschetti and F. Palmieri). He was in the Organizing Committee of the Ninth International Conference on Information Fusion (FUSION 2006). He is in the Organizing Committee of the 2008 IEEE Radar Conference (RADARCON 2008).

Vincenzo Matta received the Laurea degree in electronic engineering and the Ph.D. degree in information engineering from University of Salerno, Fisciano, Italy, in 2001 and 2005, respectively.

Currently, he is an Assistant Professor at the University of Salerno. His main research interests include detection and estimation theory, signal processing, wireless communications, multiterminal inference, and sensor networks.
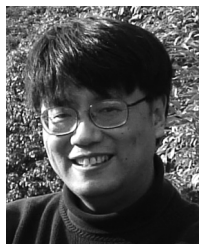
Lang Tong (S'87–M'91–SM'01–F'05) received the B.E. degree from Tsinghua University, Beijing, China, in 1985 and M.S. and Ph.D. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, in 1987 and 1991, respectively.

Currently, he is the Irwin and Joan Jacobs Professor in Engineering at Cornell University, Ithaca, NY. He was on faculty at the West Virginia University and the University of Connecticut. He was also the 2001 Cor Wit Visiting Professor at the Delft University of Technology, Delft, The Netherlands. He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, in 1991. His research is in the general area of statistical signal processing, wireless communications and networking, and information theory.

Dr. Tang received the 1993 Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 best paper award (with M. Dong) from IEEE Signal Processing Society, and the 2004 Leonard G. Abraham Prize Paper Award from the IEEE Communications Society (with P. Venkitasubramaniam and S. Adireddy). He is a coauthor of five student paper awards. He received Young Investigator Award from the U.S. Office of Naval Research. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION THEORY, and the IEEE SIGNAL PROCESSING LETTERS.

Peter Willett (F'03) received the B.A.Sc. degree in engineering science from the University of Toronto, Toronto, ON, Canada, in 1982 and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1986.

He has been a faculty member at the University of Connecticut ever since, and since 1998 has been a Professor. His primary areas of research have been statistical signal processing, detection, machine learning, data fusion, and tracking. He has interests in and has published in the areas of change/abnormality detection, optical pattern recognition, communications, and industrial/security condition monitoring.

Dr. Willett is the Editor-in-Chief of the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, and until recently, was an Associate Editor of the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS (for data fusion and target tracking), the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS. He is also the Associate Editor for the IEEE AEROSPACE AND ELECTRONIC SYSTEMS MAGAZINE, the Editor of the IEEE AEROSPACE AND ELECTRONIC SYSTEMS MAGAZINE's periodic tutorial issues, Associate Editor for the electronic *Journal of Advances in Information Fusion* of the International Society of Information Fusion (ISIF), and is a member of the editorial board of the IEEE SIGNAL PROCESSING MAGAZINE. He has been a member of the IEEE Aerospace and Electronic Systems Society Board of Governors since 2003. He was a General Co-Chair (with S. Coraluppi) for the 2006 ISIF/IEEE Fusion Conference in Florence, Italy, Program Co-Chair (with E. Santos) for the 2003 IEEE Conference on Systems, Man and Cybernetics in Washington, DC, and Program Co-Chair (with P. Varshney) for the 1999 Fusion Conference in Sunnyvale, CA.