

# Decentralized Estimation in an Inhomogeneous Sensing Environment\*

Zhi-Quan Luo<sup>†</sup> and Jin-Jun Xiao<sup>†</sup>

December, 2004 (revised)

## Abstract

We consider decentralized estimation of a noise-corrupted deterministic parameter by a bandwidth constrained sensor network with a fusion center. The sensor noises are assumed to be additive, zero mean, spatially uncorrelated, but otherwise unknown and possibly different across sensors due to varying sensor quality and inhomogeneous sensing environment. The classical Best Linear Unbiased Estimator (BLUE) linearly combines the real-valued sensor observations to minimize the Mean Square Error (MSE). Unfortunately, such a scheme cannot be implemented in a practical bandwidth constrained sensor network due to its requirement to transmit real-valued messages. In this paper, we construct a decentralized estimation scheme (DES) where each sensor compresses its observation to a small number of bits with length proportional to the logarithm of its local Signal to Noise Ratio (SNR). The resulting compressed bits from different sensors are then collected and combined by the fusion center to estimate the unknown parameter. The proposed DES is universal in the sense that each sensor compression scheme requires only the knowledge of local SNR, rather than the noise probability distribution functions (pdf), while the final fusion step is also independent of the local noise pdfs. We show that the MSE of the proposed DES is within a constant factor of 25/8 to that achieved by the classical centralized BLUE estimator.

## 1 Introduction

Recent technological advances have led to the emergence of small, low-power, and possibly mobile devices with limited on-board processing and wireless communication capabilities. When deployed

---

\*This research is supported in part by the Natural Sciences and Engineering Research Council of Canada, Grant No. OPG0090391, by the Canada Research Chair Program and by the National Science Foundation, Grant No. DMS-0312416.

<sup>†</sup>The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455. Emails: {luozq,xiao}@ece.umn.edu.

in large numbers, these devices have the ability to form an intelligent network which can measure aspects and identities of the physical environment in unprecedented scale and precision. In this paper, we focus on a star-like sensor network whereby each sensor in the network collects an observation, computes a local message (either real-valued or discrete) and then sends it to a fusion center, while the latter combines the received sensor messages to produce a final estimate of the environment. We assume that sensor nodes do not communicate with each other, and there is no feedback from fusion center to local sensors. Sensor networks of this type are well-suited for situation awareness applications such as environmental monitoring (air, water, and soil), smart factory instrumentation, military surveillance, precision agriculture, intelligent transportation and space exploration, to name a few.

To reduce the communication requirement from sensors to the fusion center, local data quantization/compression at each sensor site is needed. In fact, a central problem in sensor network research is to design discrete local message functions and the final fusion function in a way that minimizes the total bandwidth requirement while satisfying an overall system performance requirement. Clearly, optimal design of these functions will depend on the underlying sensor noise distributions. Unfortunately, characterizing the exact noise probability distributions for a large number of sensors is impractical, especially for applications in a dynamic sensing environment. What is more realistic in these scenarios is to let each sensor estimate its local signal to noise ratio (SNR) which can be accomplished by simply measuring the received signal power in the presence and absence of the incoming signal. Motivated by these considerations, we are led to the design of bandwidth efficient decentralized estimation schemes (DES) which only depend on the local SNRs, rather than the knowledge of noise pdfs. Throughout this paper, such a DES will be called **universal** with respect to the noise distribution.

The problem of decentralized estimation has been studied first in the context of distributed control [1, 12] and tracking [14], later in data fusion [2, 9], and most recently in wireless sensor networks [10]. Among these studies, the prevailing assumption has been that the joint distribution of the sensor observations is known, with some also making the additional assumption that the communication links can transmit real values and are distortionless. In the case where the communication links can only transmit discrete signals, the work of [4, 10, 11] addressed various design and implementation issues using the knowledge of joint distribution of sensor data. Without the knowledge of noise distribution, the work of [13] proposed to use a training sequence to aid the design of local data compression strategies. Recently, several new universal DES [7, 8] were proposed for distributed parameter estimation in the presence of unknown, additive sensor noises that are bounded and **identically distributed**. These universal decentralized estimation schemes have a low bandwidth requirement: each observation is compressed to exactly **one binary bit** per sensor. In particular,  $\frac{1}{2}$  of the sensors will send to the fusion center the first most significant bit of their observations,  $\frac{1}{4}$  of the sensors will send the second most significant bit of their observations, and so on. When properly combined at the fusion center, these bits can be used to estimate the unknown parameter, resulting

in a mean squared errors (MSE) that is within a constant factor of 4 to the minimum achievable. However, the results of [7, 8] are based on the restrictive assumption that the sensor noises are identically distributed. For applications where the sensors either have varying quality/resolution or are at different distances from the unknown target being monitored, the sensor noise distributions cannot be identical. In fact, the local SNRs will be different across sensors. In such scenarios, a major problem is: How should a sensor with a high SNR compress its data differently from a sensor with a low SNR?

In this paper, we extend the work of [7, 8] to the inhomogeneous sensing environment. When the local message functions are allowed to be real-valued and the sensor noise variances are known by the fusion center, the classical Best Linear Unbiased Estimator (BLUE) can be used to minimize the Mean Square Error (MSE) by linearly combining the (real-valued) sensor observations with weights inversely proportional to sensor noise variances. We show that the same MSE performance (up to a small multiplicative constant) can be achieved by a universal DES with low communication requirement (measured in bits). In particular, we let each sensor first compress its observation to a discrete message with length proportional to the logarithm of its local Signal to Noise Ratio (SNR), and then transmit the resulting message to the fusion center. The final estimate of the unknown parameter is computed at the fusion center by combining the received bits according to a universal fusion rule. It is shown that the MSE of the proposed universal DES is within a constant factor of 25/8 to that achieved by the classical centralized BLUE estimator.

Our paper is organized as follows. In Section 2, we formulate the decentralized estimation problem. Section 3 studies the design of DES for the case where the noise pdf is known, under the assumption that the sensor messages must be one binary bit. The design of DES for unknown noise with bounded range is considered in Section 4 where an universal decentralized linear unbiased estimator is presented. This DES is extended to the case of unbounded sensor noises in Section 5. The final section (Section 6) contains some concluding remarks.

## 2 Problem Formulation

Consider a set of  $K$  distributed sensors, each making observations on an unknown parameter  $\theta \in [-V, V]$ , with  $V > 0$  a given constant. The observations are corrupted by additive noises and are described by

$$x_k = \theta + n_k, \quad k = 1, 2, \dots, K. \quad (2.1)$$

We assume the noises  $\{n_k : k = 1, 2, \dots, K\}$  are zero mean, spatially uncorrelated with a probability density function (or pdf for short)  $f_k(x)$  and variance  $\sigma_k^2$ .

Suppose the sensors and the fusion center wish to jointly estimate  $\theta$  based on the spatially distributed observations  $\{x_k : k = 1, 2, \dots, K\}$ . This can be accomplished as follows (see Figure 1). First, each sensor computes a local message  $m_k(x_k)$  based on its observation  $x_k$  and on the outcome

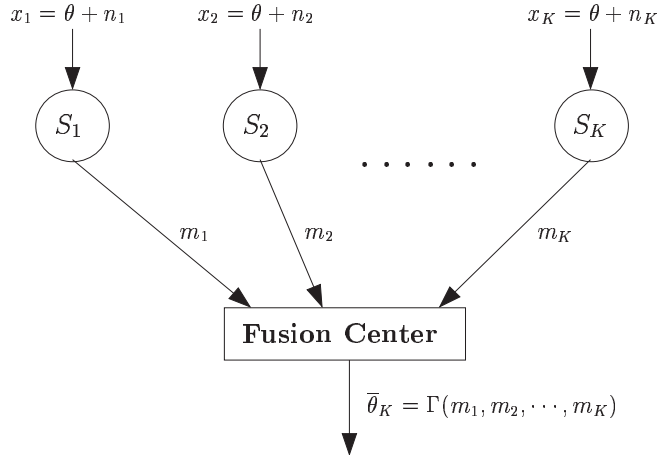


Figure 1: A decentralized estimation scheme.

of a local random variable, where  $m_k$  is a discrete probabilistic message function to be designed. The local random variable plays the role of “coin tossing” at each sensor which effectively randomizes the sensor message. This framework of randomized message functions also includes deterministic message functions as a special case in which the local random variables are chosen as constants. Second, all the sensor messages are transmitted to the fusion center where they are combined to produce a final estimate of  $\theta$  using a real-valued fusion function  $\Gamma$

$$\bar{\theta}_K = \Gamma(m_1(x_1), m_2(x_2), \dots, m_K(x_K)).$$

We will refer  $\{\Gamma, m_k : k = 1, 2, \dots, K\}$  as a decentralized estimation scheme (DES). The problem of decentralized estimation is then to design the local message functions  $\{m_k : k = 1, 2, \dots, K\}$  and the fusion function  $\Gamma$  so that  $\bar{\theta}_K$  is as close to  $\theta$  as possible in a statistical sense. In this paper, we shall adopt the Mean Squared Error (MSE) criterion to measure the quality of an estimator for  $\theta$ :

$$\text{MSE} = \text{E}(|\bar{\theta}_K - \theta|^2),$$

where the expectation is taken with respect to the sensor noises, as well as the local random variables if  $m_k(x_k)$  are randomized.

If the fusion center has knowledge of the sensor noise variances, and the communication links between the sensors and the fusion have sufficient bandwidth, the sensors can send their accurate observations  $\{x_k : k = 1, 2, \dots, K\}$  to the fusion center. In other words, we can set the message functions as  $m_k(x_k) = x_k$  for all  $k = 1, 2, \dots, K$ . Upon receiving these real-valued messages, the fusion center can simply perform the linear minimum MSE estimation to recover  $\theta$  by combining  $x_k$  with weights inversely proportional to  $\sigma_k^2$ . This leads to the following Best Linear Unbiased Estimator (BLUE estimator) [5]

$$\bar{\theta}_K(x_1, x_2, \dots, x_K) = \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1} \sum_{k=1}^K \frac{x_k}{\sigma_k^2}. \quad (2.2)$$

A simple calculation shows that this estimator has a MSE of

$$\mathbb{E}(\bar{\theta}_K - \theta)^2 = \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1}. \quad (2.3)$$

The above scheme requires infinite bandwidth since real-valued messages have to be transmitted. This makes it impractical for implementation in sensor networks. In the remainder of this paper, we assume the communication channels between sensors and the fusion center are bandwidth-constrained, and we will design a DES which can achieve essentially the same MSE as (2.3) while requiring only a small amount of communications from sensors to the fusion center. We will consider two cases depending on whether the noise pdfs are known or unknown. In the former case, we make use of the noise pdfs in designing the message functions  $\{m_k\}$  and the final fusion function  $f$ , while in the latter case, the DES is universal (i.e., independent of the sensor noise pdfs). The message lengths (measured in binary bits) are chosen to be proportional to the logarithm of local maximal SNRs. Both schemes yield a MSE that is within a constant factor of that achieved by the BLUE estimator (cf. (2.3)), with the constant factor depending on the noise pdfs in the former case, and bounded by 25/8 in the latter case.

### 3 A Decentralized Estimator with Known PDF

In this section, we assume that the noise pdfs  $\{f_k(x) : k = 1, 2, \dots, K\}$  are known. The noise variances are  $\{\sigma_k^2 : k = 1, 2, \dots, K\}$ . For convenience we define the normalized noise pdf:

$$p_k(x) = \sigma_k f_k(\sigma_k x) \quad (3.1)$$

which has variance 1, and signifies the distribution type of  $x_k$ . We let  $F_k(x)$  and  $P_k(x_k)$  denote the noise cumulative density functions of  $f_k(x)$  and  $p_k(x)$  respectively. Then it follows that

$$P_k(x) = \int_{-\infty}^x p_k(y) dy = \int_{-\infty}^x \sigma_k f_k(\sigma_k y) dy = \int_{-\infty}^{\sigma_k x} f_k(z) dz = F_k(\sigma_k x). \quad (3.2)$$

Furthermore, we impose the constraint that the local message functions  $\{m_k : k = 1, 2, \dots, K\}$  must be binary, taking values of either 0 or 1. As a result, a natural local quantization scheme is to choose  $m_k(x)$  to be the sign detector:

$$m_k(x_k) = \begin{cases} 1, & \text{if } x_k \geq 0, \\ 0, & \text{if } x_k < 0. \end{cases} \quad (3.3)$$

It is easy to see that

$$\mathbb{E}(m_k) = \mathbb{P}(x_k > 0) = \int_0^{\infty} f_k(x - \theta) dx = 1 - F_k(-\theta), \quad (3.4)$$

where  $F_k(x)$  is continuous whenever  $f_k(x)$  is Riemann-integrable.

Let us define a function

$$G(x) := \sum_{k=1}^K \frac{1 - F_k(-x)}{\sigma_k}.$$

Then it is easy to see that  $G(-\infty) = 0$ ,  $G(+\infty) = \sum_{k=1}^K \sigma_k^{-1}$ , and  $G(x)$  is continuous and monotonically increasing. It follows from (3.4) that

$$G(\theta) = \sum_{k=1}^K \frac{1 - F_k(-\theta)}{\sigma_k} = \sum_{k=1}^K \frac{E(m_k)}{\sigma_k}. \quad (3.5)$$

Therefore, upon receiving the one-bit messages  $\{m_k : k = 1, 2, \dots, K\}$ , the fusion center can first solve the equation (in  $x$ )

$$G(x) = \sum_{k=1}^K \frac{1 - F_k(-x)}{\sigma_k} = \sum_{k=1}^K \frac{m_k}{\sigma_k}. \quad (3.6)$$

Equation (3.6) has at least one solution since  $m_k \in \{0, 1\}$  and hence  $0 \leq \sum_{k=1}^K \frac{m_k}{\sigma_k} \leq \sum_{k=1}^K \sigma_k^{-1}$ . The solution is unique if we further assume that  $\{f_k(x) : k = 1, 2, \dots, K\}$  do not vanish on a common interval in  $\mathbb{R}$ . If the solution is not unique, we can simply pick one of the solutions, and denote it by  $\theta^*$ . The fusion center generates the final estimator  $\bar{\theta}_K$  by projecting  $\theta^*$  onto the interval  $[-V, V]$ , i.e.

$$\bar{\theta}_K = \max\{\min(\theta^*, V), -V\} = \begin{cases} V, & \text{if } \theta^* \geq V, \\ \theta^*, & \text{if } -V < \theta^* < V, \\ -V, & \text{if } \theta^* \leq -V. \end{cases} \quad (3.7)$$

Finding  $\bar{\theta}_K$  can be easily accomplished numerically through bisection search in the interval  $[-V, V]$ . The following theorem provides an estimate of the MSE performance for this DES.

**Theorem 3.1.** *The DES estimator specified by (3.3), (3.6) and (3.7) achieves an MSE of*

$$E\left(|\bar{\theta}_K - \theta|^2\right) \leq \frac{1}{4\mu^2} \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1},$$

where

$$\mu = \min_{1 \leq k \leq K} \inf_{x \in [-V, V]} p_k(x) \quad (3.8)$$

with the normalized pdf  $p_k$  defined by (3.1).

**Proof.** Notice that  $\bar{\theta}_K$  is the projection of  $\theta^*$  to  $[-V, V]$ , which implies  $G(\bar{\theta}_K)$  is the projection of  $G(\theta^*)$  to  $[G(-V), G(V)]$  because  $G(x)$  is monotonically increasing. Also,  $G(\theta) \in [G(-V), G(V)]$  whenever  $\theta \in [-V, V]$ , which further implies

$$|G(\bar{\theta}_K) - G(\theta)| \leq |G(\theta^*) - G(\theta)|. \quad (3.9)$$

Meanwhile, the fact that  $\theta^*$  is a solution of the equation (3.6) implies

$$G(\theta^*) = \sum_{k=1}^K \frac{1 - F_k(-\theta^*)}{\sigma_k} = \sum_{k=1}^K \frac{m_k}{\sigma_k}. \quad (3.10)$$

Therefore we have

$$\begin{aligned}
\mathbb{E} \left( |G(\bar{\theta}_K) - G(\theta)|^2 \right) &\stackrel{(a)}{\leq} \mathbb{E} \left( |G(\theta^*) - G(\theta)|^2 \right) \stackrel{(b)}{=} \mathbb{E} \left( \left| \sum_{k=1}^K \frac{m_k}{\sigma_k} - \sum_{k=1}^K \frac{\mathbb{E}(m_k)}{\sigma_k} \right|^2 \right) \\
&\stackrel{(c)}{=} \sum_{k=1}^K \frac{\mathbb{E}(m_k - \mathbb{E}(m_k))^2}{\sigma_k^2} \stackrel{(d)}{\leq} \sum_{k=1}^K \frac{1}{4\sigma_k^2}, \tag{3.11}
\end{aligned}$$

where (a) follows from (3.9), (b) follows from (3.5) and (3.10), (c) follows from the fact that for any  $1 \leq i \neq j \leq K$ ,  $m_i$  and  $m_j$  are uncorrelated given  $\theta$ , so

$$\mathbb{E}((m_i - \mathbb{E}(m_i))(m_j - \mathbb{E}(m_j)) | \theta) = \mathbb{E}(m_i m_j | \theta) - \mathbb{E}(m_i | \theta) \mathbb{E}(m_j | \theta) = 0,$$

and  $\mathbb{E}((m_i - \mathbb{E}(m_i))(m_j - \mathbb{E}(m_j))) = 0$ . In (d), we have used the fact that  $m_k$  is a 0-1 binary random variable whose variance is at most 1/4.

On the other hand, we have

$$\begin{aligned}
\mathbb{E} \left( |G(\bar{\theta}_K) - G(\theta)|^2 \right) &= \mathbb{E} \left( \left| \sum_{k=1}^K \frac{1 - F_k(-\bar{\theta}_K)}{\sigma_k} - \sum_{k=1}^K \frac{1 - F_k(-\theta)}{\sigma_k} \right|^2 \right) \\
&= \mathbb{E} \left( \left| \sum_{k=1}^K \frac{F_k(-\theta) - F_k(-\bar{\theta}_K)}{\sigma_k} \right|^2 \right) \\
&\stackrel{(a)}{=} \mathbb{E} \left( \left| \sum_{k=1}^K \frac{P_k(-\frac{\theta}{\sigma_k}) - P_k(-\frac{\bar{\theta}_K}{\sigma_k})}{\sigma_k} \right|^2 \right) \\
&\stackrel{(b)}{\geq} \mathbb{E} \left( \left| \sum_{k=1}^K \frac{(\bar{\theta}_K - \theta)\mu}{\sigma_k^2} \right|^2 \right) \\
&= \left( \sum_{k=1}^K \frac{\mu}{\sigma_k^2} \right)^2 \mathbb{E} \left( (\bar{\theta}_K - \theta)^2 \right), \tag{3.12}
\end{aligned}$$

where (a) follows from (3.2), and the inequality in (b) follows from the definition of  $\mu$  in (3.8). Combining (3.11) and (3.12) completes the proof.  $\square$

Notice that the MSE bound in Theorem 3.1 is within a constant factor to that achieved by the centralized BLUE estimator (cf. (2.3)). This suggests that the finite bandwidth constraint (one bit per sample per node) results in only a constant factor increase in the MSE performance. Also, Theorem 3.1 suggests that the MSE performance can degrade significantly when the factor  $\mu$  is small. This will be confirmed by computer simulations later in Section 5 (see Figure 7).

It is worth noting that the DES (3.3)–(3.7) not only requires the knowledge of noise pdf  $f_k$ , but also lets each sensor send exactly one-bit information regardless of sensor qualities (measured by local SNRs). In Section 4 and 5, we will propose a universal DES which allows higher SNR sensors to send more bits. Such a variable-length quantization scheme achieves a MSE performance that is independent of  $\mu$ .

## 4 A Universal DES for the Bounded Noise Case

Suppose the noises  $\{n_k : k = 1, 2, \dots, K\}$  have a bounded range in the interval  $[-U, U]$ , and  $\theta$  is bounded to the interval  $[-V, V]$ . Let  $W = U + V$ , then sensor observations  $x_k \in [-W, W]$  for  $1 \leq k \leq K$ . We normalize  $x_k$  to the range  $[0, 1]$  by a linear transformation  $y_k = (W + x_k)/2W$ . Consider the binary expansion of  $y_k$ :

$$y_k = \frac{W + x_k}{2W} = \sum_{i=1}^{\infty} b_i 2^{-i}, \quad \text{where } b_i = 0, 1.$$

We call  $b_i$  the  $i$ -th most significant bit (MSB) of  $y_k$ .

In a homogeneous sensing environment where all sensors have the same noise variance  $\sigma_k^2 = \sigma^2$  for all  $1 \leq k \leq K$ , it is natural to let each local sensor generate same amount of information (say one bit) to the fusion center. For applications where the sensors either have varying quality/resolution or are at different distances from the unknown target being monitored, the sensor noise distributions cannot be identical. In fact, the local SNRs will be different across sensors. In such scenarios, a sensor with a high SNR should compress its data differently from a sensor with a low SNR. One strategy might be to let high SNR sensors quantize their observations to the more significant MSBs. Unfortunately, this does not work in general. For example, if  $\theta = 0$  and all the noise pdfs are symmetric with respect to the origin, then  $P(b_1 = 0) = P(b_1 = 1) = 1/2$  for any  $\sigma$ . This suggests that, when  $\theta = 0$ , the quality of the first MSB of each sensor is the same irrespective of its local SNR (see Figure 2). This implies that the strategy of simply letting high SNR sensors quantize their observations to the more significant MSBs cannot work.

To extend the existing results of [7, 8] to the inhomogeneous sensing environment, we will allow each sensor to send a multiple-bit message, the length of which will be made dependent on the local SNRs, with high SNR sensors sending more bits than local SNR sensors. To this end, we will need to use a simple probabilistic scheme to quantize a random variable. This is outlined next.

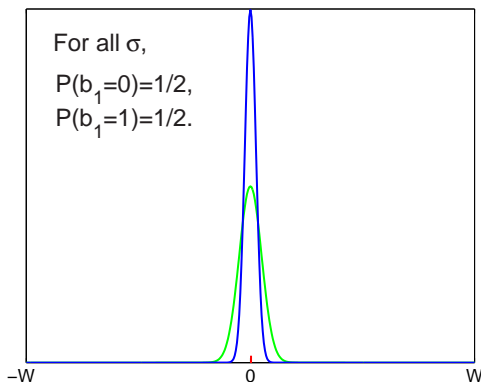


Figure 2: Distribution of  $x_k$  when  $\theta = 0$ .



## 4.1 Probabilistic Quantization of a Bounded Random Variable

Suppose  $x = \theta + n \in [-W, W]$ , where  $W$  is a known constant,  $\theta$  is an unknown parameter to be estimated, and  $n$  is a real-valued noise random variable with mean zero and variance  $\sigma^2$ . Consider the binary expansion of  $y = (x + W)/2W$ :

$$y = \sum_{i=1}^{\infty} b_i 2^{-i}, \quad \text{with } b_i = 0 \text{ or } 1,$$

and define an auxiliary random variable  $a$  as

$$\mathbf{P}(a = i) = 2^{-i}, \quad i = 1, 2, 3, \dots \quad (4.1)$$

We define a randomized  $(M + 1)$ -bit message function  $m$  as follows:

$$m(x, a, M) = \sum_{i=1}^M b_i 2^{-i} + 2^{-M} b_{M+a}. \quad (4.2)$$

Basically,  $m$  consists of the first  $M$  MSBs of  $y$ , plus an extra bit of  $y$  whose position is determined by the auxiliary random variable  $a$ . It is easy to see that  $m$  assumes  $2^M + 1$  discrete values  $\{2^{-M} j : j = 0, 1, 2, \dots, 2^M\}$ . The next lemma shows that this message function is an unbiased estimator of  $\theta$ , with a variance approaching  $\sigma^2$  at an exponential rate as  $M$  increases.

**Lemma 4.1.** *Let  $m$  be defined as above. Then,  $z = W(2m - 1)$  is an unbiased estimator of  $\theta$  and*

$$\begin{cases} E((z - \theta)^2) = W^2 - \theta^2, & \text{if } M = 0, \\ E((z - \theta)^2) \leq \frac{W^2}{2^{2M}} + \sigma^2, & \text{if } M \geq 1. \end{cases} \quad (4.3)$$

**Proof.** Let  $r = 2^M \sum_{i=M+1}^{\infty} b_i 2^{-i}$ , then  $0 \leq r \leq 1$ , and

$$y = \sum_{i=1}^M b_i 2^{-i} + 2^{-M} r. \quad (4.4)$$

From the definition of  $b_{M+a}$  in (4.2), we obtain

$$\mathbf{E}(b_{M+a}|y) = \sum_{i=1}^{\infty} b_{M+i} 2^{-i} = 2^M \sum_{i=M+1}^{\infty} b_i 2^{-i} = r. \quad (4.5)$$

It follows that given  $y$ ,  $b_{M+a}$  is a Bernoulli random variable with

$$\begin{aligned} \mathbf{P}\{b_{M+a} = 1\} &= r, \\ \mathbf{P}\{b_{M+a} = 0\} &= 1 - r. \end{aligned}$$

We also see from (4.2) and (4.4) that

$$m - y = 2^{-M}b_{M+a} - 2^{-M}r = 2^{-M}(b_{M+a} - r). \quad (4.6)$$

Therefore, it immediately follows from (4.5) that  $\mathbb{E}(m|y) = y$ . Further from the definition of  $z$ , we obtain  $\mathbb{E}(z|x) = x$  and  $\mathbb{E}(z) = \mathbb{E}(x) = \theta$ . We have shown that  $z$  is an unbiased estimate of  $\theta$ .

Next, we estimate the MSE of  $z$ :

$$\begin{aligned} \mathbb{E}((z - \theta)^2) &= \mathbb{E}((z - x)^2) + \mathbb{E}((x - \theta)^2) + 2\mathbb{E}((x - \theta)(z - x)) \\ &\stackrel{(a)}{=} 4W^2\mathbb{E}((m - y)^2) + \sigma^2 \\ &\stackrel{(b)}{=} 2^{-2M+2}W^2\mathbb{E}((b_{M+a} - r)^2) + \sigma^2 \\ &\stackrel{(c)}{\leq} 2^{-2M}W^2 + \sigma^2, \quad \text{for all } M \geq 0, \end{aligned} \quad (4.7)$$

where (a) follows from  $\mathbb{E}((x - \theta)(z - x)) = \mathbb{E}((x - \theta)\mathbb{E}(z - x|x)) = 0$ , (b) is due to (4.6), and in (c), we have used the fact that the variance of any 0-1 valued binary random variable is no more than 1/4.

Specifically when  $M = 0$ ,  $m = b_{M+a}$  (c.f. (4.2)) which is a binary message taking values from  $\{0, 1\}$ ,  $\mathbb{E}(m|y) = y$ , and  $\mathbb{E}((m - y)^2|y) = y(1 - y)$ . In this case, we can tighten the bound in (4.7) by calculating

$$\mathbb{E}((m - y)^2) = \mathbb{E}(\mathbb{E}(m - y)^2|y) = \mathbb{E}(y(1 - y)) = \mathbb{E}\left(\frac{W^2 - x^2}{4W^2}\right) = \frac{W^2 - \theta^2 - \sigma^2}{4W^2},$$

and

$$\mathbb{E}((x - \theta)^2) = 4W^2\mathbb{E}((m - y)^2) + \sigma^2 = W^2 - \theta^2, \quad \text{for } M = 0. \quad (4.8)$$

Combining (4.7) with (4.8) yields (4.3), as desired.  $\square$

Lemma 4.1 can be readily used to design a universal DES for the estimation of  $\theta$  in a homogeneous sensing environment. In particular, if sensor noises are i.i.d. with variance  $\sigma^2$ , then we can let each sensor quantize its observation using the same message function (4.2), except with independent copies of  $a$ . Let  $\{m_k : k = 1, 2, \dots, K\}$  denote the set of message functions and  $z_k = W(2m_k - 1)$ , then the fusion center can estimate  $\theta$  by simply averaging  $\{z_k : k = 1, 2, \dots, K\}$ :

$$\bar{\theta}_K = \frac{1}{K} \sum_{k=1}^K z_k = \sum_{k=1}^K W(2m_k - 1).$$

It can be seen from Lemma 4.1 that this DES achieves a MSE of

$$\begin{cases} \mathbb{E}((\bar{\theta}_K - \theta)^2) = \frac{W^2 - \theta^2}{K} \leq \frac{W^2}{K}, & \text{if } M = 0, \\ \mathbb{E}((\bar{\theta}_K - \theta)^2) \leq \frac{1}{K} \left( \frac{W^2}{2^{2M}} + \sigma^2 \right), & \text{if } M \geq 1. \end{cases}$$

When  $M = 0$ , each sensor sends a 1-bit message, and the MSE performance is the same as that obtained in [7, 8]. When  $M \geq 1$ , the MSE of  $\bar{\theta}_K$  converges to the best achievable MSE performance  $\sigma^2/K$  exponentially as  $M \rightarrow \infty$ . Moreover, if we pick a fixed finite message length of  $M = \lceil \log W/\sigma \rceil$ , then the MSE is bounded by  $2\sigma^2/K$  which is a factor of 2 away from the centralized BLUE (c.f. (2.3) when  $\sigma_k^2 = \sigma^2$  for all  $k$ ).

## 4.2 A Universal DES for Bounded Inhomogeneous Noises

When sensor noise levels are different, the centralized BLUE scheme (2.2) differentiates observations by multiplying them with different weights ( $\sigma_k^{-2}$  to be specific). How should we design a DES in which sensors with different noise variances can only send discrete messages to the fusion center? Intuitively, if a sensor has low sensor noise, we have a number of options: (i) let this sensor quantize their observations to the more significant bits; (ii) let this sensor send more bits to the fusion center; (iii) let the fusion center weigh this sensor's message more in the fusion process. As discussed in the beginning of Section 4, simply following any one of these three approaches does not work. It turns out that a good strategy involves a combination of all three approaches. In particular, motivated by Lemma 4.1 and by the discussion at the end of Section 4.1, we will let each sensor send a multiple-bit message with length  $M_k$  approximately proportional to the logarithm of the local SNR:

$$M_k = \left\lceil \log \frac{W}{\sigma_k} \right\rceil. \quad (4.9)$$

In this way, a better quality sensor sends more bits to the fusion center, and the message length  $M_k$  can be used by the fusion center to recover the local noise variance  $\sigma_k^2$  and use it as a weighting factor in the final fusion process. Since the noise range is bounded by  $W$ , it follows that  $\sigma_k < W$  for all  $k$ . This implies that  $M_k \geq 1$ , for all  $k$ .

**Remark 4.2.** *The message length  $M_k$  admits an intuitively appealing interpretation:  $M_k$  can be roughly viewed as the Shannon capacity of a "virtual channel" from Mother nature to the local sensor  $k$ , with  $\theta$  as the channel input and with  $x_k = \theta + n_k$  as the channel output. In particular, the channel capacity, maximized over all possible choices of  $\theta \in [-V, V]$  and all noise distributions with variance bounded by  $\sigma_k^2$ , is*

$$C_k = \frac{1}{2} \log \left( 1 + \frac{V^2}{\sigma_k^2} \right).$$

*A simple calculation yields*

$$C_k \leq M_k < C_k + \frac{1}{2} \log \left( 1 + \frac{V}{U} \right) + 1.$$

*Therefore, the message length of sensor  $k$  is approximately  $C_k$ , especially when  $\sigma_k$  is small. Since  $C_k$  can be viewed as the maximum average number of bits of information carried by the sensor observation  $x_k$ , it is therefore quite natural not to let sensor  $k$  send more than  $C_k$  bits of information to the fusion center.*

Now we are ready to construct the universal DES which consists of the local message functions  $\{m_k : k = 1, 2, \dots, K\}$  and the fusion function  $\Gamma$ . The design is outlined in two steps.

- With the knowledge of  $W$  and the local noise variance  $\sigma_k^2$ , sensor  $k$  determines  $M_k$  according to (4.9). After collecting the observation  $x_k \in [-W, W]$ , sensor  $k$  generates a message  $m_k$  which is a random  $(M_k + 1)$ -bit discrete message

$$m_k(x_k, a_k, M_k) = \sum_{i=1}^{M_k} b_i 2^{-i} + 2^{-M_k} b_{M_k+a_k}, \quad (4.10)$$

where  $\{b_i : i = 1, 2, \dots\}$  are the binary bits associated to the binary expansion of  $y_k = (x_k + W)/2W$ , and  $a_k$  local random variable with distribution defined by (4.1). (Thus, all sensors in the network use i.i.d. local random variables.)

- The final estimator in the fusion center is

$$\bar{\theta}_K = \Gamma(m_1, m_2, \dots, m_K) = \left( \sum_{k=1}^K 2^{2M_k} \right)^{-1} \sum_{k=1}^K 2^{2M_k} W(2m_k - 1). \quad (4.11)$$

Notice that in the above DES, each sensor only needs to know its own noise level (or equivalently local SNR), while the final fusion step (4.11) is completely determined by the received messages  $\{m_k : k = 1, 2, \dots, K\}$ , and thus independent of noise pdfs. Thus, this DES is universal with respect to the noise pdfs. The following theorem shows that the MSE performance of this universal DES is within a small constant factor of 25/8 to that achieved by the centralized BLUE estimator (cf. (2.2)–(2.3)).

**Theorem 4.3.** *The estimator  $\bar{\theta}_K$  in (4.11) is an unbiased estimate of  $\theta$ , i.e.  $E(\bar{\theta}_K) = \theta$ , and  $\bar{\theta}_K$  has an MSE*

$$E(\bar{\theta}_K - \theta)^2 \leq \frac{25}{8} \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1},$$

which is optimal (up to a factor of 25/8) when compared to the BLUE estimator (2.2)–(2.3).

**Proof.** The unbiasedness of  $\bar{\theta}_K$  follows from

$$E(\bar{\theta}_K) = \left( \sum_{k=1}^K 2^{2M_k} \right)^{-1} \sum_{k=1}^K 2^{2M_k} E(W(2m_k - 1)) = \left( \sum_{k=1}^K 2^{2M_k} \right)^{-1} \sum_{k=1}^K 2^{2M_k} \theta = \theta,$$

where second last step is due to Lemma 4.1. To estimate the MSE of the estimator  $\bar{\theta}_K$ , we note from (4.9) that

$$\log \frac{W}{\sigma_k} \leq M_k < \log \frac{W}{\sigma_k} + 1,$$

which implies

$$\frac{\sigma_k^2}{4} < \frac{W^2}{2^{2M_k}} \leq \sigma_k^2, \quad \forall k. \quad (4.12)$$

This, together with Lemma 4.1, further implies

$$\text{Var}(W(2m_k - 1)) \leq \frac{W^2}{2^{2M_k}} + \sigma_k^2 \leq 2\sigma_k^2, \quad \forall k.$$

Therefore, we obtain

$$\begin{aligned} \mathbb{E}((\bar{\theta}_K - \theta)^2) &= \left( \sum_{k=1}^K 2^{2M_k} \right)^{-2} \sum_{k=1}^K 2^{4M_k} \text{Var}(W(2m_k - 1)) \\ &\leq 2 \left( \sum_{k=1}^K 2^{2M_k} \right)^{-2} \sum_{k=1}^K 2^{4M_k} \sigma_k^2 \\ &\leq \frac{25}{8} \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1}, \end{aligned} \quad (4.13)$$

where the proof of last inequality is tedious and is relegated to the appendix.  $\square$

Similar to the BLUE estimator (2.2), the universal DES (4.10)–(4.11) also linearly combines the sensor messages with optimal weights to minimize the MSE. However, unlike the BLUE estimator (2.2) which requires the knowledge of noise variances to determine the weights, the weights used by the DES (4.10)–(4.11) are completely determined by the lengths of the sensor messages. The latter makes the DES (4.10)–(4.11) universal. Moreover, Theorem 4.3 shows that the new DES achieves an MSE performance that is within a factor of 25/8 to that achieved by the BLUE estimator. Thus, the strategy of allowing higher SNR sensors to send more bits not only eliminates the need for the fusion center to know the individual sensors' local SNRs, but also helps to ensure the universal DES (4.10)–(4.11) achieves an (almost) optimal overall estimation accuracy.

Notice that when a sensor's local SNR is too low, the corresponding real-valued observation can be too noisy to be useful. Thus, from a practical implementation point of view, it may be desirable to let each sensor have the option of not sending any message to the fusion center when its local SNR falls below a given threshold. Such an option is also attractive from the power efficiency point of view. Similar to the proof of Theorem 4.3, we can analyze the MSE performance of the DES (4.10)–(4.11) with this option added. Also, notice that the proposed scheme will require more bandwidth for sensors having access to a high SNR observation, leading to possibly unbounded message lengths. This suggests that, in practical implementations, we will need to constrain the maximum message length for sensors with excessively high SNRs so that their message lengths do not grow unbounded. This restriction on the maximum message length will result in a final MSE which is still a constant factor of that of the BLUE, albeit the constant factor will now depend on the maximum allowed message length.

### 4.3 Simulation

We now present some simulations to compare the performances of the universal DES (4.10)–(4.11) and the BLUE estimator. As shown earlier, the MSE of these estimation schemes are of the order  $\left(\sum_{k=1}^K 1/\sigma_k^2\right)^{-1}$ , where  $\{\sigma_k^2 : k = 1, 2, \dots, K\}$  denote sensor noise variances. This motivates us to define, similar to [9], the asymptotic efficiency of a DES as

$$\text{asymptotic efficiency} = \frac{1}{\text{MSE} \cdot \sum_{k=1}^K 1/\sigma_k^2}. \quad (4.14)$$

Thus, the asymptotic efficiency is 1 for the BLUE estimator (cf. (2.2)–(2.3)), and is lower bounded by  $8/25 = 0.32$  for the universal DES (4.10)–(4.11) (cf. Theorem 4.3). Clearly, the larger the asymptotic efficiency, the more efficient is the estimation scheme.

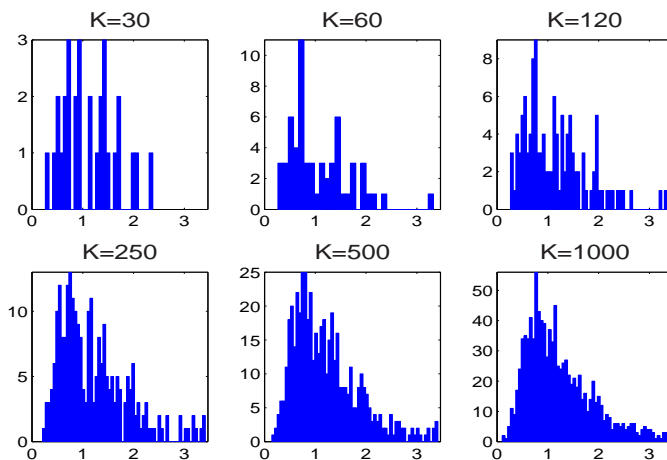


Figure 3: Distribution of sensor noise standard deviations.

In all the simulation runs, we have chosen  $\theta = 1$ ,  $V = 3$ , and  $U = 6$ . Sensor noises are assumed to be uniformly distributed with standard deviations in Figure 3 which are randomly generated according to a central  $F$ -distribution with parameters  $\nu_1 = 10$ ,  $\nu_2 = 10$  (see [6], Ch. 2). The simulations are performed for various sizes of sensor networks:  $K = 30, 60, 120, 250, 500, 1000$ . At each choice of  $K$ , the MSE is obtained by repeating the experiments for 5000 times and averaging the corresponding squared estimation errors.

Figure 4 (left) plots the MSE curves of various schemes. The overall lower bound denotes  $\left(\sum_{k=1}^K 1/\sigma_k^2\right)^{-1}$ , which is the best achievable MSE for linear unbiased estimators, and the universal DES upper bound is  $\frac{25}{8} \left(\sum_{k=1}^K 1/\sigma_k^2\right)^{-1}$  as given by Theorem 4.3. We can see that the universal DES and the BLUE estimator both have MSE converging to 0 as the number of sensors increases. Figure 4 (right) shows that the asymptotic efficiency of our universal DES is about 0.6 for all  $K$ . This suggests that in these simulations, the universal DES has a MSE performance which is about  $1/0.6 \approx 1.67$  of that achieved by the BLUE estimator. This shows that the new universal DES

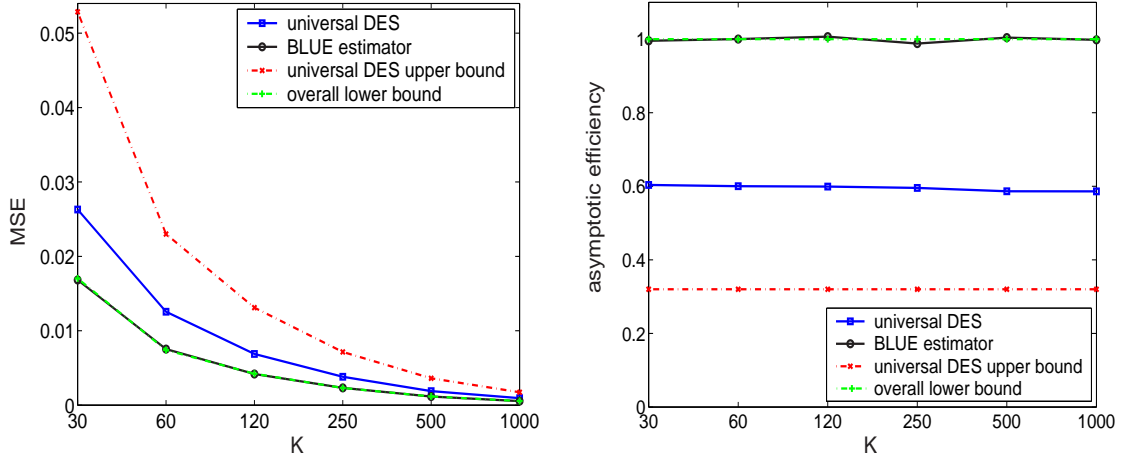


Figure 4: Comparison of MSE and asymptotic efficiencies for bounded noise case.

(4.10)–(4.11), which has a substantially lower communications requirement (see Figure 5), achieves essentially the same MSE performance as the centralized BLUE estimator.

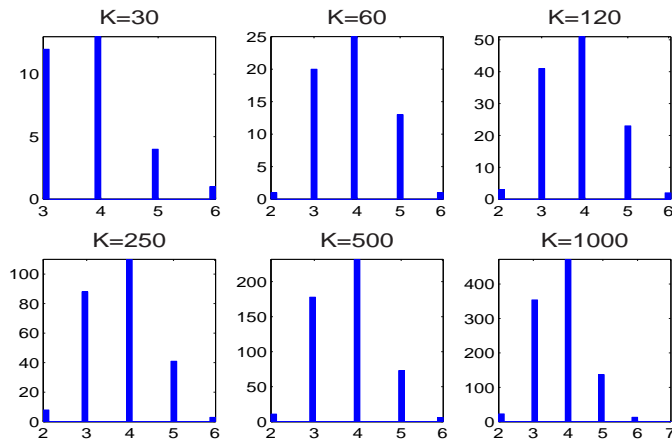


Figure 5: Distribution of  $M_k + 1$  (number of bits transmitted by local sensors).

To examine the communications requirement, we recall that the universal DES lets each sensor compress its observation to a message of length  $M_k + 1 = \lceil \log W/\sigma_k \rceil + 1$  (cf. (4.10)). With  $W = U + V = 9$ , and the noise variances specified in Figure 3, the distributions of the number of bits transmitted by all  $K$  sensors are plotted in Figure 5. The number of bits sent by each sensor is between 2 and 7 for all simulated cases. Moreover, the mean value of  $M_k$  (averaged among all  $K$  sensors) are  $\{3.80, 3.88, 3.83, 3.78, 3.77, 3.77\}$  for  $K = \{30, 60, 120, 250, 500, 1000\}$  respectively. Thus, sensors only need to transmit fewer than 4 bits per sample on average to the fusion center to guarantee the MSE performance specified in Theorem 4.3.

## 5 Design of Universal DES for the Unbounded Noise Case

When the noise range is unbounded (i.e.,  $U = +\infty$ ), the values of observations  $\{x_k : k = 1, 2, \dots, K\}$  range over the entire real line  $\mathbb{R}$ . Our design strategy remains to first construct discrete message functions which are unbiased quantizations of sensor observations, with average message lengths determined by the local SNRs. Then, these message functions are weighted at the fusion center to obtain a final estimator. In what follows, we will consider two separate cases: (i) the fusion center knows all the sensor SNRs; (ii) the fusion center has no information on local SNRs.

### 5.1 A DES for the Known Sensor Variances Case

Assume that fusion center knows sensor noise variances  $\{\sigma_k^2 : k = 1, 2, \dots, K\}$ . In this case, we simply let each sensor send to the fusion center a randomly rounded version of  $x_k/(2\sigma_k)$ , with  $x_k$  denoting the sensor's observation. The fusion center then weighs these random discrete messages optimally with weights inversely proportional to  $\sigma_k^2$ . The whole scheme is outlined as follows.

- Suppose  $x_k \in \mathbb{R}$  denotes the observation of sensor  $k$ . Let  $x'_k = x_k/(2\sigma_k)$  and write  $x'_k = \text{sign}(x'_k)(i_k + r_k)$ , where  $i_k$  and  $r_k$  are the integer and decimal parts of  $|x'_k|$  respectively. Then  $0 \leq r_k < 1$ . We construct sensor  $k$ 's message function as

$$m_k(x_k, \sigma_k) = \text{sign}(x'_k)(i_k + d_k), \quad (5.1)$$

where  $d_k$  is a 0-1 binary random variable with  $P(d_k = 1) = r_k$  and  $P(d_k = 0) = 1 - r_k$ .

- The final estimator of  $\theta$  at the fusion center is

$$\bar{\theta}_K = \Gamma(m_1, m_2, \dots, m_K) = \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1} \sum_{k=1}^K \frac{2m_k}{\sigma_k}. \quad (5.2)$$

**Theorem 5.1.** *For the DES (5.1)–(5.2),  $\bar{\theta}_K$  is an unbiased estimator of  $\theta$ , and the MSE*

$$E\left((\bar{\theta}_K - \theta)^2\right) \leq 2 \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1}.$$

Moreover, for all  $1 \leq k \leq K$ , the average message length of  $m_k$  is no more than  $\log 6 + 0.5 \log(1 + \gamma_k)$  binary bits, where  $\gamma_k = \theta^2/\sigma_k^2$  denotes the local SNR.

**Proof.** It is easy to see that  $E(m_k|x'_k) = x'_k = x_k/(2\sigma_k)$ . So  $E(m_k) = E(x_k)/(2\sigma_k) = \theta/(2\sigma_k)$ . Thus, it follows from (5.2) that  $\bar{\theta}_K$  is an unbiased estimator of  $\theta$ . Also,

$$\begin{aligned} E\left((2\sigma_k m_k - \theta)^2\right) &\stackrel{(a)}{=} E\left((2\sigma_k m_k - x_k)^2\right) + E\left((x_k - \theta)^2\right) \\ &\leq 4\sigma_k^2 E\left((m_k - x'_k)^2\right) + \sigma_k^2 \\ &\stackrel{(b)}{\leq} 2\sigma_k^2, \end{aligned}$$



where in (a), we have applied that  $\mathbb{E}(2\sigma_k m_k | x_k) = x_k$ , so  $\mathbb{E}((2\sigma_k m_k - x_k)(x_k - \theta) | x_k) = 0$ , and  $\mathbb{E}((2\sigma_k m_k - x_k)(x_k - \theta)) = 0$ . The inequality (b) follows from the definition of  $m_k$  in (5.1), and  $\mathbb{E}((m_k - x'_k)^2) = \mathbb{E}((d_k - r_k)^2) = \text{Var}(d_k) \leq 1/4$  since  $d_k$  is a random variable taking values from  $\{0, 1\}$ . Hence, we obtain

$$\begin{aligned}
\mathbb{E}((\bar{\theta}_K - \theta)^2) &= \mathbb{E} \left( \left( \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1} \sum_{k=1}^K \frac{2\sigma_k m_k}{\sigma_k^2} - \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1} \sum_{k=1}^K \frac{\theta}{\sigma_k^2} \right)^2 \right) \\
&= \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-2} \mathbb{E} \left( \left( \sum_{k=1}^K \frac{2\sigma_k m_k - \theta}{\sigma_k^2} \right)^2 \right) \\
&= \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-2} \sum_{k=1}^K \frac{\mathbb{E}((2\sigma_k m_k - \theta)^2)}{\sigma_k^4} \\
&\leq 2 \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1},
\end{aligned}$$

where in the second step, we have applied that  $\mathbb{E}(2\sigma_k m_k) = \theta$ , and for any  $1 \leq i \neq j \leq K$ ,  $m_i$  and  $m_j$  are uncorrelated given  $\theta$ , so  $\mathbb{E}([2\sigma_i m_i - \theta][2\sigma_j m_j - \theta]) = 0$ .

Next, we estimate the average length of the discrete message function  $m_k$ . From (5.1), we know that  $|m_k| \leq \lceil |x'_k| \rceil < \frac{|x_k|}{2\sigma_k} + 1$ . Hence,  $m_k$  has length  $L_k \leq 1 + \lceil \log \left( 1 + \frac{|x_k|}{2\sigma_k} \right) \rceil$ , where the first term accounts for the sign bit of  $m_k$ , and the second term  $\lceil \log \left( 1 + \frac{|x_k|}{2\sigma_k} \right) \rceil$  bounds the binary length of integer  $|m_k|$ . Thus, the average length of  $m_k$  is bounded as follows:

$$\begin{aligned}
\mathbb{E}(L_k) &\leq \mathbb{E} \left( 1 + \left\lceil \log \left( 1 + \frac{|x_k|}{2\sigma_k} \right) \right\rceil \right) \leq 2 + \mathbb{E} \left( \log \left( 1 + \frac{|x_k|}{2\sigma_k} \right) \right) \\
&\stackrel{(a)}{\leq} 2 + \log \left( 1 + \frac{\mathbb{E}(|x_k|)}{2\sigma_k} \right) \stackrel{(b)}{\leq} 2 + \log \left( 1 + \frac{1}{2\sigma_k} \sqrt{\theta^2 + \sigma_k^2} \right) \\
&= 1 + \log \left( 2 + \frac{1}{\sigma_k} \sqrt{\theta^2 + \sigma_k^2} \right) \leq 1 + \log 3 + \log \left( \frac{1}{\sigma_k} \sqrt{\theta^2 + \sigma_k^2} \right) \\
&= \log 6 + \frac{1}{2} \log(1 + \gamma_k),
\end{aligned}$$

where (a) follows from Jensen's inequality, and (b) is due to  $(\mathbb{E}(|x_k|))^2 \leq \mathbb{E}(x_k^2) = \theta^2 + \sigma_k^2$ . The proof is complete.  $\square$

Similar to Remark 4.2, we can think of  $C_k = \frac{1}{2} \log(1 + \gamma_k)$  as the Shannon capacity of a "virtual AWGN channel" from nature to sensor  $k$ , with SNR given by  $\gamma_k$ . By Theorem 5.1, the average length of message  $m_k$  is upper bounded by  $\log 6 + C_k$ . This once again reveals that the message length is decided by the number of "useful" bits contained in  $x_k$  as measured by the channel capacity  $C_k$ .

Theorem 5.1 shows that this choice of message length results in a DES whose MSE performance is within a constant factor of 2 to that of an optimal BLUE estimator.

Notice that Theorem 5.1 does not assume bounded noise range. However, it does require that the noise variances are known at the fusion center. In what follows, we will remove this restrictive assumption.

## 5.2 A Universal DES

We now consider a more practical scenario whereby the fusion center does not have any a priori information on the local SNRs of sensor nodes. Again, the local SNR (or equivalently the local sensor noise variance) is assumed known at each sensor.

Our approach is to first choose a positive number  $W$  (similar to the range parameter used in Section 4), and then decompose  $x_k/2W$  into an integer plus a decimal part, where  $x_k$  denotes the sensor  $k$ 's observation. The integer part will be communicated to the fusion center, whereas the decimal part will be quantized according to the universal DES described in Section 4.2. The overall message length will be a function of  $W$ , though the estimation quality is always upper bounded by 25/8 times that of the BLUE estimator (see Theorem 5.2).

We will consider several different choices of  $W$  later. For the time being, we simply assume  $W$  is chosen, and

$$M_k = \left\lceil \log^+ \frac{W}{\sigma_k} \right\rceil \geq 0, \quad \text{for all } 1 \leq k \leq K, \quad (5.3)$$

where  $\log^+ x = \max\{\log x, 0\}$ . It is easy to see that if  $\sigma_k < W$ , then  $L_k \geq 1$ , and sensor  $k$  will transmit a message. Otherwise, if  $\sigma_k$  is so large that  $M_k$  is negative, then sensor  $k$ 's observation is "too noisy" and can simply be discarded. We now describe the universal DES as follows:

- For sensors with  $\sigma_k \geq W$ , their observations are discarded and no messages are transmitted. For the remaining sensors with  $\sigma_k < W$ , let  $x_k$  denote the observation. Then  $x'_k = x_k/2W$  is decomposed as  $x'_k = \text{sign}(x'_k)(i_k + y'_k)$ , with  $i_k \geq 0$  and  $0 \leq y'_k < 1$  representing the integer and decimal parts of  $|x'_k|$  respectively. Consider the binary expansion of  $y'_k$

$$y'_k = \sum_{i=1}^{\infty} b_i 2^{-i}.$$

The message function  $m_k$  of sensor  $k$  is

$$m_k(x_k, a_k, M_k) = \text{sign}(x'_k) (i_k + m'_k(x'_k, a_k, M_k)), \quad (5.4)$$

where  $a_k$  is an auxiliary random variable with distribution defined by (4.1) and

$$m'_k(x_k, a_k, M_k) = \sum_{i=1}^{M_k} b_i 2^{-i} + 2^{-M_k} b_{M_k+a_k}.$$

- The final estimator of  $\theta$  at the fusion center is

$$\bar{\theta}_K = \Gamma(m_1, m_2, \dots, m_K) = \left( \sum_{k=1}^K 2^{2M_k} \right)^{-1} \sum_{k=1, M_k \geq 1}^K 2^{2M_k} W(2m_k - 1). \quad (5.5)$$

The performance of above universal DES is summarized in the following theorem.

**Theorem 5.2.** *For the universal DES (5.4)–(5.5), the final estimator  $\bar{\theta}_K$  is an unbiased estimate of  $\theta$ , and its MSE is bounded by*

$$E\left((\bar{\theta}_K - \theta)^2\right) \leq \frac{25}{8} \left( \sum_{k=1, \sigma_k < W}^K \frac{1}{\sigma_k^2} \right)^{-1}. \quad (5.6)$$

Moreover, each sensor transmits at most  $M_k = \lceil \log(W/\sigma_k) \rceil$  bits for the decimal part of  $m_k$ , plus on average at most  $N_k = 1 + \log\left(2 + \sqrt{\theta^2 + \sigma_k^2/W}\right)$  extra bits to transmit the integer part of  $m_k$ .

The proof of (5.6) can be modeled after that of Theorem 4.3, and the calculation of  $N_k$ , which is an upper bound of the average number of bits representing the integer part of  $m_k$ , can be established in the same way as the proof of Theorem 5.1. For brevity, we omit the details of the analysis.

There are a number of choices of  $W$  that are worth special consideration.

1. If we want all sensors to send a message, i.e.,  $M_k \geq 1$ , we need to choose  $W = \max_k \{\sigma_k\}$ .
2. If a sensor has a large noise variance, its observation may be too noisy to be useful, and therefore can be discarded to save bandwidth. For example, if we take  $W = V$ , then sensors with  $\sigma_k > V$  should simply discard its observation. In this way,  $M_k \leq 1 + \log \gamma_k < 1 + C_k$ , and  $N_k \leq 1 + \log(2 + \sqrt{2})$ . The corresponding MSE performance is (5.6) with  $W = V$ . Clearly, this strategy is useful when there are enough high SNR sensors so that the performance (5.6) is acceptable.
3. If  $\sigma_{\max} \leq 1$ , then we can simply take  $W = 1/2$ . In this case,  $x'_k = x_k/(2W) = x_k$  and the message  $m_k$  consists of the integer part of  $x_k$ , together with the first  $M_k = \lceil \log(1/2\sigma_k) \rceil$  most significant bits of the decimal part of  $x_k$ , plus an extra random bit.

In practice, the threshold value  $W$  should be chosen according to the final MSE requirement and the observation qualities. Case 1 above gives the largest  $W$  which will allow every sensor to participate in the estimation process. Decreasing  $W$  may “silence” some sensors with “poor” quality observations ( $\sigma_k > W$ ), resulting in energy saving for the corresponding sensors. The issue of adjusting  $W$  to minimize total energy consumption, possibly with the help of appropriate information feedback from fusion center, is a topic of future research.

### 5.3 Simulation

In this section, we present simulation results to illustrate the performances of three estimation schemes: the universal DES (Section 5.2), the DES using known sensor variances (Section 5.1), and the DES using known pdf (Section 3). All these DESs allow sensor noises to be unbounded. In all the simulations, we choose  $\theta = 0.6$  and  $V = 3$ . The noise sources are Gaussian with variances specified in Figure 3. For the universal DES, we choose  $W = V = 3$ . In this case, a small portion of sensors with  $\sigma_k \geq 3$  have their observations discarded for  $K = \{60, 120, 250, 500, 1000\}$ .

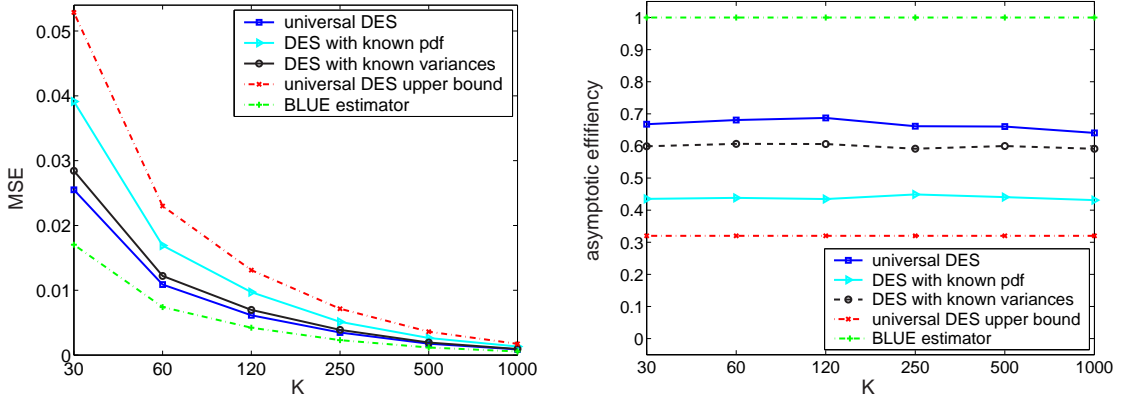


Figure 6: Comparison of MSE performance.

Figure 6 (left) shows the MSE curves of various DESs, while Figure 6 (right) shows the asymptotic efficiency curves. The overall lower bound  $\left(\sum_{k=1}^K 1/\sigma_k^2\right)^{-1}$  is the best achievable MSE by any linear unbiased estimator, and the universal DES upper bound is  $\frac{25}{8} \left(\sum_{k=1, \sigma_k < 3}^K 1/\sigma_k^2\right)^{-1}$  as given in Theorem 5.2. We can see that all three estimators have the same asymptotic MSE performances. Notice that each of the simulated DES has a different bandwidth requirement. Specifically, in the DES using known pdf (Section 3), each sensor transmits exactly 1-bit binary message; the DES using known sensor noise variances (Section 5.1) requires on average  $\{2.13, 2.11, 2.08, 2.09, 2.08, 2.07\}$  binary bits per sensor for  $K = \{30, 60, 120, 250, 500, 1000\}$  respectively in the simulations; while the universal DES (Section 5.2) requires on average  $\{4.73, 4.76, 4.70, 4.67, 4.65, 4.65\}$  bits per sensor for  $K = \{30, 60, 120, 250, 500, 1000\}$  respectively.

We have also plotted in Figure 7 the asymptotic efficiencies of the universal DES and the DES with known pdf when  $K = 250$  and  $\theta \in [-3, 3]$ . Recall that the DES with known pdf has an asymptotic efficiency that is lower bounded by  $4\mu^2$  (Theorem 3.1). Thus, this DES has a good performance when  $\theta$  is small. However, as  $\theta$  moves away from the origin, the vanishing tail of the Gaussian noise pdf leads to small  $\mu$ , resulting in very low asymptotic efficiencies (see Figure 7). In contrast, the universal DES has a stable MSE performance with respect to changing  $\theta$  (and the noise pdf as well), as expected.

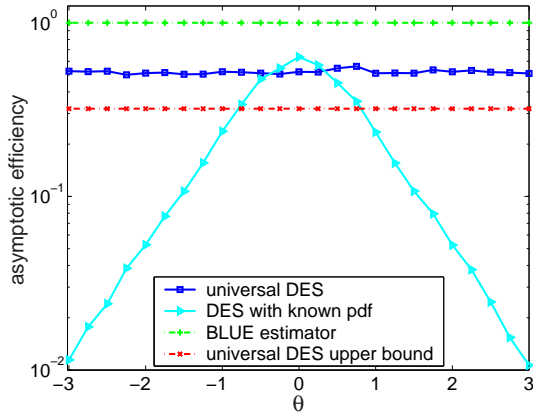


Figure 7: Asymptotic efficiencies v.s.  $\theta$ .

## 6 Conclusion

Motivated by the limited channel capacities between sensors and fusion center and by the difficulty to characterize sensor noise distributions, we have proposed in this paper a decentralized estimation scheme which has low communication requirement and is universal to the sensor noise distributions. The local data compression is probabilistic, with each sensor’s observation compressed to a binary string whose length is proportional to the logarithm of its local SNR. This DES is isotropic in the sense that all sensors in the network operate identically using the same local compression algorithm. This isotropic property greatly simplifies the deployment of a large sensor network. In addition, the proposed DES is robust to sensor failures since all sensors operate independently from each other and the fusion center only needs to linearly combine all of its received messages. In fact, the fusion center is completely oblivious of sensor identities. Furthermore, for the problem of estimating an unknown parameter corrupted by spatially inhomogeneous additive noise, the proposed DES achieves an MSE performance that is within a constant factor of  $25/8$  to being optimal.

Notice that our current work does not impose a fixed bandwidth constraint on each sensor node as was done in [7] for homogeneous sensor networks. Instead, we have adopted an adaptive quantization strategy in our universal DES design where the local message lengths are used as weighting factors at the fusion center as well. We believe this variable bandwidth scheme is more appropriate for the inhomogeneous sensing environment since sensors with more accurate information are allowed to send more bits, resulting in a final estimation performance that is within a constant factor to that of the centralized BLUE. To implement this adaptive quantization scheme in a practical sensor network, the fusion center can use standard carrier-sensing techniques to detect the message lengths  $M_k$ . Alternatively, the sensors can simply add the message length  $M_k$  as part of their information payload. If the sensing environment changes slowly in a quasi-static manner, the values of  $\{M_k\}$  (once acquired) can be re-used multiple times at the fusion center.

The proposed universal DES can be extended to the ad hoc sensor networks in which no central coordinating fusion center is available, and to universal decentralized detection of detecting deterministic signals by remote sensors. Such extensions can be made in a similar fashion as in [8] and [16]. Moreover, our current work can be used to minimize total energy consumption in sensor networks. Indeed, in a follow-up work [15], we have incorporated the proposed universal DES with energy consumption models of [3], and have obtained a power scheduling strategy of universal decentralized estimation to minimize the total transmit power, while ensuring a given mean squared error performance. The power scheduling strategy suggests that sensors with bad channels or poor observation qualities should decrease their quantization resolutions or simply become inactive in order to save power. Finally, as a future work, we plan to apply the proposed DES to universal decentralized tracking where each sensor has a vector observation model  $\mathbf{x}_k = \mathbf{H}_k \boldsymbol{\theta} + \mathbf{n}_k$ . By multiplying the pseudo-inverse  $\mathbf{H}_k^\dagger$  to the observation, we obtain  $\tilde{\mathbf{x}}_k = \boldsymbol{\theta} + \tilde{\mathbf{n}}_k$ . Since each components of the noise  $\tilde{\mathbf{n}}_i = \mathbf{H}_k^\dagger \mathbf{n}_k$  have different variances, the problem is reduced to the one considered in this paper where sensors are allowed to have different noise levels.

## Appendix: Proof of inequality (4.13)

We first show the following lemma.

**Lemma 6.1.** *Suppose  $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$  and  $\{\beta_1, \beta_2, \dots, \beta_K\}$  are positive real numbers and*

$$C_1 \beta_k \leq \alpha_k \leq C_2 \beta_k, \quad \text{for all } 1 \leq k \leq K, \quad (6.1)$$

where  $C_2 \geq C_1 > 0$  are constants. Then

$$\left( \sum_{k=1}^K \alpha_k \right)^{-2} \sum_{k=1}^K \frac{\alpha_k^2}{\beta_k} \leq \frac{(C_1 + C_2)^2}{4C_1 C_2} \left( \sum_{k=1}^K \beta_k \right)^{-1}.$$

This is the usual Cauchy-Schwartz inequality with opposite inequality sign.

**Proof.** Suppose  $\{\beta_1, \beta_2, \dots, \beta_K\}$  are fixed. Let  $f(\alpha_1, \alpha_2, \dots, \alpha_K) = \left( \sum_{k=1}^K \alpha_k \right)^{-2} \sum_{k=1}^K \frac{\alpha_k^2}{\beta_k}$ . We calculate

$$\begin{aligned} \frac{\partial f}{\partial \alpha_j} &= 2 \left( \sum_{k=1}^K \alpha_k \right)^{-3} \left( \frac{\alpha_j}{\beta_j} \sum_{k=1}^K \alpha_k - \sum_{k=1}^K \frac{\alpha_k^2}{\beta_k} \right) \\ &= 2 \left( \sum_{k=1}^K \alpha_k \right)^{-3} \left( \alpha_j \sum_{k=1, k \neq j}^K \frac{\alpha_k}{\beta_j} - \sum_{k=1, k \neq j}^K \frac{\alpha_k^2}{\beta_k} \right). \end{aligned}$$

It is easy to see that  $\frac{\partial f}{\partial \alpha_j} = 0$  has one unique solution

$$\alpha_j' = \left( \sum_{k=1, k \neq j}^K \frac{\alpha_k}{\beta_j} \right)^{-1} \sum_{k=1, k \neq j}^K \frac{\alpha_k^2}{\beta_k}.$$

In addition

$$\begin{cases} \frac{\partial f}{\partial \alpha_j} > 0, & \text{if } \alpha_j > \alpha'_j, \\ \frac{\partial f}{\partial \alpha_j} < 0, & \text{if } \alpha_j < \alpha'_j. \end{cases} \quad (6.2)$$

Hence, if we fix other variables, and only let  $\alpha_j$  vary, then  $g(\alpha_j) = f(\alpha_1, \alpha_2, \dots, \alpha_K)$  reaches the minimum at  $\alpha_j = \alpha'_j$ . From (6.2), we know that the maximum of  $g(\alpha_j)$  for  $\alpha_j \in [C_1\beta_j, C_2\beta_j]$  reaches at the two end points, i.e., either  $\alpha_j = C_1\beta_j$  or  $\alpha_j = C_2\beta_j$ . The same arguments hold for all the other  $\alpha_k$  whenever  $1 \leq k \leq K$ . Thus, when  $f(\alpha_1, \alpha_2, \dots, \alpha_K)$  reaches its maximum at  $(\alpha_1^*, \alpha_2^*, \dots, \alpha_K^*)$ , we can find two index sets  $I_1$  and  $I_2$  with  $I_1 \cup I_2 = \{1, 2, \dots, K\}$  and  $I_1 \cap I_2 = \emptyset$ , such that for  $k \in I_1$ ,  $\alpha_k^* = C_1\beta_k$ , and for  $k \in I_2$ ,  $\alpha_k^* = C_2\beta_k$ . Let

$$S_1 = \sum_{k \in I_1} \beta_k, \quad S_2 = \sum_{k \in I_2} \beta_k.$$

Then, for  $\{\alpha_k, \beta_k : 1 \leq k \leq K\}$  satisfying (6.1), we have that

$$\begin{aligned} \max_{\alpha_1, \alpha_2, \dots, \alpha_K} \left( \left( \sum_{k=1}^K \alpha_k \right)^{-2} \sum_{k=1}^K \frac{\alpha_k^2}{\beta_k} \sum_{k=1}^K \beta_k \right) &= \left( \sum_{k=1}^K \alpha_k^* \right)^{-2} \sum_{k=1}^K \frac{\alpha_k^{*2}}{\beta_k} \sum_{k=1}^K \beta_k \\ &= \frac{(C_1^2 S_1 + C_2^2 S_2)(S_1 + S_2)}{(C_1 S_1 + C_2 S_2)^2} = 1 + \frac{(C_1 - C_2)^2 S_1 S_2}{(C_1 S_1 + C_2 S_2)^2} \\ &\leq 1 + \frac{(C_1 - C_2)^2}{4C_1 C_2} = \frac{(C_1 + C_2)^2}{4C_1 C_2}. \end{aligned}$$

This concludes the proof of the lemma.  $\square$

Now we are ready to establish the inequality (4.13).

**Proof.** Notice that from (4.12), we have

$$\frac{1}{\sigma_k^2} \leq \frac{2^{2M_k}}{4W^2} < \frac{4}{\sigma_k^2}.$$

Invoking Lemma 6.1 with

$$\alpha_k = \frac{2^{2M_k}}{W^2}, \quad \beta_k = \frac{1}{\sigma_k^2}, \quad C_1 = 1, \quad \text{and } C_2 = 4,$$

we obtain

$$\begin{aligned} \left( \sum_{k=1}^K 2^{2M_k} \right)^{-2} \sum_{k=1}^K 2^{4M_k} \sigma_k^2 &= \left( \sum_{k=1}^K \frac{2^{2M_k}}{W^2} \right)^{-2} \sum_{k=1}^K \frac{2^{4M_k}}{W^4} \sigma_k^2 \\ &\leq \frac{(1+4)^2}{4 \cdot 1 \cdot 4} \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1} = \frac{25}{16} \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1}. \end{aligned}$$

This proves the inequality (4.13).  $\square$

## References

- [1] D.A. Castanon and D. Teneketzis, “Distributed Estimation Algorithms for Nonlinear Systems,” *IEEE Transactions on Automatic Control*, Vol. AC-30, pp. 418–425, 1985.
- [2] Z. Chair and P.K. Varshney, “Distributed Bayesian Hypothesis Testing with Distributed Data Fusion,” *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 18, pp. 695–699, 1988.
- [3] S. Cui, A. J. Goldsmith and A. Bahai, “Energy-constrained Modulation Optimizatoin,” to appear in *IEEE Transactions on Wireless Communications*, 2004.
- [4] J.A. Gubner, “Distributed Estimation and Quantization,” *IEEE Transactions on Information Theory*, Vol. 39, pp. 1456–1459, 1993.
- [5] S.M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [6] S.M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall, Englewood Cliffs, New Jersey, 1998.
- [7] Z.-Q. Luo, “Universal Decentralized Estimation in a Bandwidth Constrained Sensor Network,” Submitted to *IEEE Transactions on Information Theory*; revised 2004. Also available at [http://www.ece.umn.edu/users/luozq/recent\\_work.html](http://www.ece.umn.edu/users/luozq/recent_work.html).
- [8] Z.-Q. Luo, “An Isotropic Universal Decentralized Estimation Scheme for a Bandwidth Constrained Ad Hoc Sensor Network,” Accepted for publication in *IEEE Journal on Selected Areas in Communications*, a special issue on Self-Organizing Distributed Collaborative Sensor Networks. Also available at [http://www.ece.umn.edu/users/luozq/recent\\_work.html](http://www.ece.umn.edu/users/luozq/recent_work.html).
- [9] Z.-Q. Luo and J.N. Tsitsiklis, “Data Fusion with Minimal Communication,” *IEEE Transactions on Information Theory*, Vol. 40, pp. 1551–1563, 1994.
- [10] H.C. Papadopoulos, G.W. Wornell, and A.V. Oppenheim, “Sequential Signal Encoding from Noisy Measurements Using Quantizers with Dynamic Bias Control,” *IEEE Transactions on Information Theory*, Vol. 47, pp. 978–1002, 2001.
- [11] W.M. Lam and A.R. Reibman, “Design of Quantizers for Decentralized Systems with Communication Constraints,” *IEEE Transactions on Communication*, Vol. 41, pp. 1602–1605, 1993.
- [12] J.L. Speyer, “Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control Problem,” *IEEE Transactions on Automatic Control*, Vol. AC-24, pp. 266–269, 1979.



- [13] V. Megalooikonomou and Y. Yesha, “Quantizer Design for Distributed Estimation with Communication Constraints and Unknown Observation Statistics,” *IEEE Transactions on Communications*, Vol. 48, pp. 181–184, 2000.
- [14] A.S. Willsky, M. Bello, D.A. Castanon, B.C. Levy, and G. Verghese, “Combining and Updating of Local Estimates and Regional Maps Along Sets of One-dimensional Tracks,” *IEEE Transactions on Automatic Control*, Vol. AC-27, pp. 799–813, 1982.
- [15] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, “Joint Estimation in Sensor Networks under Energy Constraint,” Presented in the *IEEE first conference on Sensor and Ad Hoc Communications and Networks*, Santa Clara, CA, October 2004. Also available at [http://www.ece.umn.edu/users/luozq/recent\\_work.html](http://www.ece.umn.edu/users/luozq/recent_work.html).
- [16] J.-J. Xiao and Z.-Q. Luo, “Universal Decentralized Detection in a Bandwidth Constrained Sensor Network,” Accepted for publication in *IEEE Transactions on Signal Processing*.