

# Minimum Energy Decentralized Estimation in a Wireless Sensor Network with Correlated Sensor Noises

**Alexey Krasnopeev**

*Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, USA  
Email: kras0053@umn.edu*

**Jin-Jun Xiao**

*Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, USA  
Email: xiao0029@umn.edu*

**Zhi-Quan Luo**

*Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, USA  
Email: luozq@umn.edu*

*Received 25 November 2004; Revised 20 May 2005*

Consider the problem of estimating an unknown parameter by a sensor network with a fusion center (FC). Sensor observations are corrupted by additive noises with an arbitrary spatial correlation. Due to bandwidth and energy limitation, each sensor is only able to transmit a finite number of bits to the FC, while the latter must combine the received bits to estimate the unknown parameter. We require the decentralized estimator to have a mean-squared error (MSE) that is within a constant factor to that of the best linear unbiased estimator (BLUE). We minimize the total sensor transmitted energy by selecting sensor quantization levels using the knowledge of noise covariance matrix while meeting the target MSE requirement. Computer simulations show that our designs can achieve energy savings up to 70% when compared to the uniform quantization strategy whereby each sensor generates the same number of bits, irrespective of the quality of its observation and the condition of its channel to the FC.

**Keywords and phrases:** wireless sensor networks, decentralized estimation, power control, energy efficiency.

## 1. INTRODUCTION

Wireless sensor networks (WSNs) are ideal for environmental monitoring applications because of their low implementation cost, agility, and robustness to sensor failures. A popular WSN architecture consists of a fusion center (FC) and a large number of spatially distributed sensors. The FC can be either a standard base station or a mobile access point such as an unmanned aerial vehicle hovering over the sensor field. Each sensor in a WSN is responsible for local data collection as well as occasional transmission of a summary of its observations to the FC via a wireless link. In a practical WSN, each sensor has only limited computation and communication capabilities due to various design considerations such as small size battery, bandwidth, and cost. As a result, it is difficult for sensors to send their entire real-valued observations

to the FC. Instead, a more practical decentralized estimation scheme is to let each sensor quantize its real-valued local measurement to an appropriate length and send the resulting discrete message (typically short) to the FC, while the latter combines all the received messages to produce a final estimate of the unknown parameter. Naturally, the message lengths are dictated by the power and bandwidth limitations, sensor noise characteristics as well as the desired final estimation accuracy.

Recently, several decentralized estimation schemes (DES) [1, 2, 3, 4] have been proposed for parameter estimation in the presence of additive sensor noise. These DESs require each sensor to send only a few bits to the fusion center, with the message length determined by the sensor's local SNR. Performance of the resulting estimator is shown to be within a constant factor of the best linear unbiased estimator (BLUE) performance. While the designs suggested by [1, 2, 3, 4] give a guaranteed estimation performance with low bandwidth requirement, the effect of wireless channel distortion and the important issue of total sensor energy minimization were not directly modelled.

---

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In a practical WSN, the wireless links from sensors to the FC may have different qualities, depending on the sensor locations relative to the FC. Intuitively, local message length should depend not only on the quality of sensor's observation (i.e., local SNR), but also on the quality of its wireless link to the FC. In particular, even if a sensor has a high-quality observation, it should not perform any local quantization or transmission when its wireless link to the FC is weak, in order to conserve sensor energy. In general, minimizing the total sensor energy consumption for a decentralized estimation task is essential to ensure long lifespan of a WSN. Motivated by these considerations, the authors of [5, 6] proposed optimal coded and uncoded transmission strategies for sensor networks which can minimize the required energy per transmitted bit, although no consideration was given to the quantization effect and the accuracy of final estimation. In the recent work of [7, 8], the authors considered the problem of optimal energy scheduling for decentralized estimation where sensor measurements are corrupted by additive noises, while communication links from sensors to the fusion center differ in quality. In particular, [7] used an adaptive modulation scheme with an exponential dependence of energy on the transmitted message size, and then derived optimal sensor power and quantization levels via convex optimization.

The aforementioned results all require an important assumption that sensor observation noises are spatially uncorrelated. Unfortunately, this assumption can be restrictive in a practical WSN, especially when sensors are densely deployed. In this paper, we consider distributed parameter estimation in situations where sensor observations are corrupted by correlated additive noises. Assuming a standard energy model [5, 6], uniform quantization at sensors, and the knowledge of sensor noise correlation matrix, we use convex optimization techniques to derive a nearly-optimal (modulo a minor relaxation) energy scheduling strategy with a mean-squared error performance guaranteed to be within a constant factor to that of the centralized BLUE estimator. Computer simulations show that our designs can achieve energy savings up to 70% when compared to the uniform bit allocation strategy whereby each sensor generates the same number of bits.

Our sensor energy scheduling strategy is suitable for direct application when the sensor noise correlation matrix is available at the FC. In practice, the sensor noise correlation matrix may have to be determined in the sensor network calibration phase, possibly with the help of training signals. In the absence of this knowledge, our scheme is also useful as it provides an upper bound on the performance of all other energy scheduling schemes, both centralized and distributed. In fact, our scheme gives an estimate of the amount of energy "wasted" due to the lack of sensor noise correlation knowledge. The power schedules generated by our design also give insight into the design of distributed energy scheduling algorithms.

Our paper is organized as follows. In Section 2 we describe the DES and formulate total energy minimization problem. In Section 3 we present a convex relaxation of the energy minimization problem and give a nearly-optimal

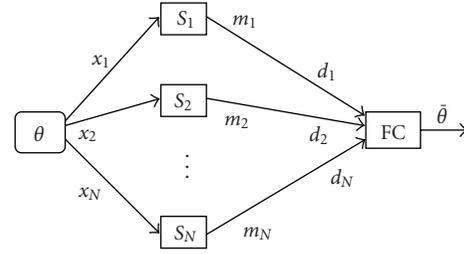


FIGURE 1: Decentralized estimation scheme.

solution in closed form. The performance of our energy-efficient design is analyzed in Section 4 by numerical simulation. Section 5 contains an extension of the work where we formulate an alternative problem of minimizing maximal individual sensor energy and present an analytic solution. Final remarks are given in Section 6.

Throughout, we use the following notations. Matrices and vectors are denoted by boldface letters, capital and small correspondingly, whereas same regular letters with indices denote their elements. Diagonal matrix with nonzero elements  $a_1, \dots, a_N$  is denoted by  $\text{diag}(a_1, \dots, a_N)$ . Logarithms denoted by  $\log(\cdot)$  are taken to the base 2; for natural logarithms notation  $\ln(\cdot)$  is used. For any real number  $x \in \mathbb{R}$ , we use  $\lceil x \rceil$  to denote the smallest integer greater or equal to  $x$ . For any random variable  $R$ , we use  $E_x R$  to denote the expected value of  $R$  taken with respect to random variable  $x$ , while  $E_{x|y} R$  denotes the expected value of  $R$  with respect to  $x$  given  $y$ . Finally,  $\text{var} R$  denotes the variance of random variable  $R$ .

## 2. PROBLEM FORMULATION

Consider the problem of estimating an unknown parameter  $\theta$  by a sensor network consisting of  $N$  sensors. Measurement of each sensor  $x_i$  is corrupted by additive noise  $n_i$  so that

$$x_i = \theta + n_i, \quad i = 1, \dots, N. \quad (1)$$

We assume that both  $\theta$  and  $n_i$  have finite range, so that all  $x_i$  belong to a common finite interval  $[-U, U]$ , with  $U > 0$  a known constant. The noises  $n_i$  are assumed to be zero mean and correlated across sensors with covariance matrix  $\mathbf{C}$ , but otherwise unknown. We assume  $\mathbf{C}$  is known at the FC. Measurements  $x_i$  are quantized to produce messages  $m_i$  to be passed on to the fusion center; the latter then combines received messages in order to estimate  $\theta$ , see Figure 1. The exact form of  $m_i$  will be detailed later.

We assume that each sensor sends messages to FC using a separate channel. This can be achieved by using a multiple access technique such as TDMA or FDMA. Each channel is corrupted by additive white Gaussian noise (AWGN) with power spectral density  $N_0/2$ :

$$\hat{m}_i = d_i^{-\kappa/2} m_i + v_i, \quad (2)$$

where  $\hat{m}_i$  is the received message at FC and  $v_i$  is the AWGN. The signal power received at the FC is assumed to be inversely proportional to  $d_i^\kappa$  where  $d_i$  is the distance between sensor  $i$  and the FC, and  $\kappa$  is the path loss exponent. Suppose that message  $m_i$  has length  $b_i$  bits. We will assume that energy  $W_i$  required for transmission of  $m_i$  is proportional to the number of bits in the message. This is the case, for example, if sensors use  $M$ -QAM or  $M$ -PSK modulation to transmit messages. For example, if  $M$ -QAM is used,  $W_i$  can be found as follows [5, 6]:

$$W_i = \frac{2}{3} N_f N_0 G_0 d_i^\kappa (2^s - 1) \ln \left( \frac{4(1 - 2^{-s})}{s P_b} \right) \frac{b_i}{s} \equiv w_i b_i, \quad (3)$$

where  $s = \log M$  is the number of bits per symbol,  $N_f$  is the receiver noise figure,  $P_b$  is the required bit error probability, and  $G_0$  is the system constant defined as in [5].

### 2.1. Quantization strategy

Suppose that sensor observation  $x_i$  is bounded to a finite interval  $[-U, U]$ . Suppose further that we wish to quantize  $x_i$  in such a way that resulting message  $m_i$  has length  $b_i$  bits, where  $b_i$  is to be determined later. We therefore have  $K_i = 2^{b_i}$  quantization points  $\{a_j^{(i)} \in [-U, U], j = 1, \dots, K_i\}$ . These points are uniformly spaced so that  $a_1^{(i)} = -U < a_2^{(i)} < \dots < a_{K_i}^{(i)} = U$  and  $a_{k+1}^{(i)} - a_k^{(i)} = \Delta_i$  for every  $k$ . Since end points  $\{a_j^{(i)}\}$  divide the observation range into  $K_i - 1$  intervals, it follows that  $\Delta_i = 2U/(K_i - 1)$ . Quantization is done in the following probabilistic manner. Suppose that  $x_i \in [a_k^{(i)}, a_{k+1}^{(i)})$ . Then  $x_i$  is quantized to either  $a_{k+1}^{(i)}$  or  $a_k^{(i)}$  according to

$$\begin{aligned} P(m_i = a_{k+1}^{(i)}) &= \frac{x_i - a_k^{(i)}}{\Delta_i}, \\ P(m_i = a_k^{(i)}) &= \frac{a_{k+1}^{(i)} - x_i}{\Delta_i}. \end{aligned} \quad (4)$$

This probabilistic quantization produces a message  $m_i$  whose expected value equals the observation itself:

$$\begin{aligned} E_{p_i} m_i &= a_{k+1}^{(i)} \Pr(m_i = a_{k+1}^{(i)}) + a_k^{(i)} \Pr(m_i = a_k^{(i)}) \\ &= a_{k+1}^{(i)} \frac{(x_i - a_k^{(i)})}{\Delta_i} + a_k^{(i)} \frac{(a_{k+1}^{(i)} - x_i)}{\Delta_i} \\ &= x_i \frac{a_{k+1}^{(i)} - a_k^{(i)}}{\Delta_i} = x_i, \end{aligned} \quad (5)$$

where the expectation  $E_{p_i}$  is taken with respect to the probabilistic quantization noise model (4).

Next, we consider any fixed observation value of  $x_i$ , and bound the variance  $\text{var } m_i$  (taken with respect to the quantization noise) as follows. Suppose  $x_i$  falls in the interval

$[a_k^{(i)}, a_{k+1}^{(i)})$ . We denote  $r = a_{k+1}^{(i)} - x_i$  and  $p_i = (x_i - a_k^{(i)})/\Delta_i \in [0, 1]$ . Then, we have

$$\begin{aligned} \text{var } m_i &= E_{p_i} (m_i - x_i)^2 \\ &= (\Delta_i - r)^2 (1 - p_i) + r^2 p_i \\ &= \Delta_i^2 \left( (1 - p_i)^2 p_i + p_i^2 (1 - p_i) \right) \\ &= \Delta_i^2 p_i (1 - p_i). \end{aligned} \quad (6)$$

Thus, the maximum variance of  $m_i$  is equal to  $\Delta_i^2/4$  and is achieved when the observation  $x_i$  falls in the middle of quantization interval  $[a_k^{(i)}, a_{k+1}^{(i)})$ .

### 2.2. A linear fusion rule

The classical best linear unbiased estimator (BLUE) for  $\theta$  is given by [9]

$$\hat{\theta} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}, \quad (7)$$

where  $\mathbf{x} = (x_1, \dots, x_N)^T$  and  $\mathbf{1}$  is the vector of all ones. Estimation performance is characterized by the variance of the estimator

$$\text{var } \hat{\theta} = (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^{-1}. \quad (8)$$

To implement BLUE exactly in a WSN setup, we must have  $m_i = x_i$  (i.e., real-valued message) and assume that the channel is distortion-less, both of which are unrealistic in practice. Nonetheless, BLUE estimator serves as a good performance benchmark for the DES to be designed. Motivated by the centralized BLUE, we adopt the following fusion rule: upon receiving sensor messages  $m_i$ , the FC combines them into an estimator  $\bar{\theta}$  given by

$$\bar{\theta} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{m}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}, \quad (9)$$

where  $\mathbf{m} = (m_1, \dots, m_N)^T$ . Equation (5) gives us an important property of  $\bar{\theta}$ : it is an unbiased estimator for  $\theta$ . Indeed, we have

$$E_{p, \mathbf{x}} \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{m}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = E_{\mathbf{x}} \frac{\mathbf{1}^T \mathbf{C}^{-1} E_p \mathbf{m}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = E_{\mathbf{x}} \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = \theta, \quad (10)$$

where  $E_p$  denotes expectation taken with respect to all sensor quantization noises, and the last step is due to  $E_{\mathbf{x}} \mathbf{x} = \theta \mathbf{1}$ . The mean-squared error (MSE) of  $\bar{\theta}$  can be expanded as follows:

$$\begin{aligned} \text{MSE}(\bar{\theta}) &= E(\bar{\theta} - \theta)^2 = E(\bar{\theta} - \hat{\theta} + \hat{\theta} - \theta)^2 \\ &= E(\bar{\theta} - \hat{\theta})^2 + E(\hat{\theta} - \theta)^2 + 2 E(\bar{\theta} - \hat{\theta})(\hat{\theta} - \theta). \end{aligned} \quad (11)$$

Consider the third term in the last expression. We have

$$\begin{aligned} E_{\mathbf{m},\mathbf{x}}(\bar{\theta} - \hat{\theta})(\hat{\theta} - \theta) &= E_{\mathbf{x}}(E_{\mathbf{m}|\mathbf{x}}[(\bar{\theta} - \hat{\theta})(\hat{\theta} - \theta)]) \\ &= E_{\mathbf{x}}[(\hat{\theta} - \theta)E_{\mathbf{m}|\mathbf{x}}(\bar{\theta} - \hat{\theta})] = 0, \end{aligned} \quad (12)$$

where the second step is due to the fact that  $\hat{\theta}$  is independent of  $\mathbf{m}$  for any fixed  $\mathbf{x}$ , and the last step follows from (10). Thus, we can write

$$\begin{aligned} \text{MSE}(\bar{\theta}) &= E(\bar{\theta} - \hat{\theta})^2 + E(\hat{\theta} - \theta)^2 \\ &= E\left(\frac{\mathbf{1}^T \mathbf{C}^{-1}(\mathbf{m} - \mathbf{x})}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}\right)^2 + \text{var } \hat{\theta} \\ &= \left(\frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}\right)^2 \mathbf{1}^T \mathbf{C}^{-1} E(\mathbf{m} - \mathbf{x})(\mathbf{m} - \mathbf{x})^T \mathbf{C}^{-1} \mathbf{1} \\ &\quad + \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \\ &= \left(\frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{Q} \mathbf{C}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} + 1\right) \text{var } \hat{\theta}, \end{aligned} \quad (13)$$

where

$$\mathbf{Q} = E(\mathbf{m} - \mathbf{x})(\mathbf{m} - \mathbf{x})^T \quad (14)$$

is the quantization noise correlation matrix.

In our formulation, we seek an energy-efficient DES which can deliver an MSE performance that is comparable to that of the centralized BLUE estimator. Specifically, we will minimize the transmission energy while maintaining the  $\text{MSE}(\bar{\theta})$  to be within a constant factor of the BLUE performance, that is,  $\text{MSE}(\bar{\theta}) \leq (1 + \alpha) \text{var } \hat{\theta}$  for some constant  $\alpha > 0$ . Therefore, the following condition must hold:

$$\frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{Q} \mathbf{C}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \leq \alpha. \quad (15)$$

The total sensor transmission energy is equal to

$$W = \sum_{i=1}^N W_i = \sum_{i=1}^N w_i b_i, \quad (16)$$

where  $w_i$  is the energy required for transmission of a single bit from sensor  $i$  to the FC; see (3). Therefore, the minimum energy DES design problem becomes

$$\begin{aligned} \text{minimize } W &= \sum_{i=1}^N w_i b_i \\ \text{subject to } &\frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{Q} \mathbf{C}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \leq \alpha, \quad b_i \in \mathbb{N}, \end{aligned} \quad (17)$$

where  $\mathbb{N}$  denotes the set of nonnegative integers.

To complete the formulation, we need to make explicit the dependence of  $\mathbf{Q}$  on  $b_i$ . The unbiasedness of our quantization strategy leads to the following important property on the quantization noise correlation matrix  $\mathbf{Q}$ .

**Lemma 1.** *The quantization noise matrix  $\mathbf{Q}$  is diagonal.*

*Proof.* Consider any  $(i, j)$ th element of the matrix  $\mathbf{Q}$ , with  $i \neq j$ . We have

$$\begin{aligned} Q_{ij} &= E(m_i - x_i)(m_j - x_j) \\ &= E_{x_i, x_j} \left( E_{p_i, p_j | x_i, x_j} (m_i - x_i)(m_j - x_j) | x_i, x_j \right) \\ &= E_{x_i, x_j} \left( E_{p_i | x_i} (m_i - x_i) E_{p_j | x_j} (m_j - x_j) | x_i, x_j \right) = 0. \end{aligned} \quad (18)$$

Here we use the fact that random variables  $m_i$  and  $m_j$  are conditionally independent given corresponding observations  $x_i$  and  $x_j$ , which together with (5) gives the desired result.  $\square$

Lemma 1 states that all the off-diagonal entries of  $\mathbf{Q}$  must be zero. Let  $Q_{ii}$  be the  $i$ th diagonal element of  $\mathbf{Q}$ . Recalling (6), we obtain the following important bound on the diagonal entries of  $\mathbf{Q}$ :

$$Q_{ii} = \text{var } m_i \leq \frac{U^2}{(2^{b_i} - 1)^2}, \quad (19)$$

where  $b_i$  is the number of bits in  $m_i$ . This bound will be useful in our final formulation of the energy minimization problem.

### 2.3. Total energy minimization

We introduce the notation  $\mathbf{c} = \mathbf{C}^{-1} \mathbf{1}$  and  $\beta = \alpha / \text{var } \hat{\theta}$ . Since  $\text{var } \hat{\theta} = 1 / \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}$ , we can rewrite the MSE condition (15) as

$$\mathbf{c}^T \mathbf{Q} \mathbf{c} \leq \beta. \quad (20)$$

This constraint ensures that the MSE performance of the DES is within a factor of  $\alpha$  to the BLUE performance. Since the distribution of  $\mathbf{x}$  is unknown in general, we enforce a stronger condition, namely

$$\max_{\mathbf{x}, p} \mathbf{c}^T \mathbf{Q} \mathbf{c} \leq \beta. \quad (21)$$

Recalling that  $\mathbf{Q}$  is diagonal (cf. Lemma 1), we can use the bound (19) to rewrite the above condition as

$$\max_{\mathbf{x}, p} \mathbf{c}^T \mathbf{Q} \mathbf{c} = \max_{\mathbf{x}, p} \sum_{i=1}^N Q_{ii} c_i^2 = \sum_{i=1}^N \frac{U^2 c_i^2}{(2^{b_i} - 1)^2} \leq \beta. \quad (22)$$

Now we can reformulate the original energy minimization problem (17) explicitly as follows:

$$\begin{aligned} & \text{minimize} \sum_{i=1}^N w_i b_i \\ & \text{subject to} \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} \leq \frac{\beta}{U^2}, \quad b_i \in \mathbb{N}, i = 1, \dots, N. \end{aligned} \quad (23)$$

To relate this formulation to physical parameters, we note that the wireless channel conditions, the choice of modulations/BER, and so forth will determine the values of weighting factors  $w_i$ , as shown in (3). The values of  $c_i$  are determined by the noise correlation matrix  $\mathbf{C}$ . Without loss of generality we assume  $c_i \neq 0$  for all  $i$ . In case  $c_i = 0$  for some sensors, we can exclude corresponding  $m_i$  from fusion consideration, as it does not contribute to the fusion estimate  $\hat{\theta}$ .

### 3. CONVEX RELAXATION WITH A CLOSED-FORM NEARLY-OPTIMAL SOLUTION

Since  $b_i$  can only take integer values, problem (23) is actually a nonlinear integer program whose computational complexity is typically NP-hard. To make this problem computationally tractable, we relax the integer constraints on  $b_i$  to allow them to take real nonnegative values:

$$\begin{aligned} & \text{minimize} \sum_{i=1}^N w_i b_i \\ & \text{subject to} \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} \leq \frac{\beta}{U^2}, \quad b_i \geq 0, i = 1, \dots, N. \end{aligned} \quad (24)$$

The relaxed problem (24) has a linear objective function and convex inequality constraints. Therefore, solution to problem (24) can be efficiently found by the fusion center using convex optimization techniques such as the interior point methods [10]. Once the optimal  $b_i$ 's are found, the fusion center can round this solution to the nearest greater integer and broadcast it to the sensors for power adjustment.

In what follows, we will present an approximately-optimal solution to the problem (24) in closed form. Such a closed-form solution not only simplifies the energy scheduling process, but also provides valuable insight into the optimal power-scheduling scheme. To begin, we first note that, by a simple monotonicity argument, the main MSE constraint will be active (i.e., holds with equality) at any optimum point,<sup>1</sup> while the remaining nonnegativity constraints on  $b_i$  will be inactive since  $b_i = 0$  for some  $i$  would violate the main MSE constraint. Therefore, we can ignore the

<sup>1</sup>Indeed, the left-hand side of MSE constraint is monotonically decreasing in terms of  $b_i$  function. Therefore, if at the optimum the inequality is strict, we could change  $b_k$  in the optimal solution to  $\tilde{b}_k < b_k$  for some  $k$  to decrease the objective function.

nonnegativity constraints (since the Lagrangian multipliers associated with these constraints will be zero). Associating a multiplier  $\lambda$  with the MSE constraint, we can write the Lagrangian for the problem (24) as follows:

$$L(b_i, \lambda) = \sum_{i=1}^N w_i b_i + \lambda \left( \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} - \frac{\beta}{U^2} \right). \quad (25)$$

At the point of optimum we must have  $\partial L / \partial b_i = 0$  for  $i = 1, \dots, N$ , yielding the following set of conditions:

$$\frac{\partial L}{\partial b_i} = w_i - 2\lambda \ln 2 \frac{2^{b_i} c_i^2}{(2^{b_i} - 1)^3} = 0, \quad (26)$$

or alternatively

$$\frac{2^{b_i}}{(2^{b_i} - 1)^3} = \frac{w_i \lambda'}{c_i^2}, \quad (27)$$

where  $\lambda' = 1/2\lambda \ln 2$ . Also, the main MSE constraint holds with equality at optimum point (as noted above), yielding

$$\sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} = \frac{\beta}{U^2}. \quad (28)$$

The optimal solutions  $\{b_i, \lambda'\}$  can be found from the nonlinear equations (27) and (28) which unfortunately cannot be solved in the closed form. To facilitate a closed-form solution, we consider a slightly modified system in variables  $\{b_i^*, \lambda^*\}$ :

$$\sum_{i=1}^N \frac{c_i^2}{(2^{b_i^*} - 1)^2} = \frac{\beta}{U^2}, \quad (29)$$

$$\frac{2^{b_i^*} - 1}{(2^{b_i^*} - 1)^3} = \frac{1}{(2^{b_i^*} - 1)^2} = \frac{w_i \lambda^*}{c_i^2}. \quad (30)$$

The above system is almost identical to the original Karush-Kuhn-Tucker (KKT) system (27) and (28) except for the small change in the numerators of the left-hand sides of (30) and (27). Simple algebraic manipulation shows that (29) and (30) can be solved analytically, yielding

$$\lambda^* = \frac{\beta}{U^2} \left( \sum_{i=1}^N w_i \right)^{-1}. \quad (31)$$

Substituting this  $\lambda^*$  into (30) gives the following feasible solution to the original energy scheduling problem (24):

$$b_i^* = \log \left( 1 + \frac{|c_i|}{\sqrt{\lambda^* w_i}} \right). \quad (32)$$

It remains to quantify the performance of this particular energy scheduling strategy. This is the content of next two lemmas.

Lemma 2. Let  $\{b_i, \lambda'\}$  be the optimal solution to the problem (24) such that  $b_i \geq 1$  for all  $i$ , and let  $\{b_i^*, \lambda^*\}$  be its approximation defined by (29) and (30). Then

$$\lambda^* \leq \lambda' \leq 2\lambda^*. \quad (33)$$

*Proof.* Since

$$\frac{2^{b_i}}{(2^{b_i} - 1)^3} = \frac{1}{(2^{b_i} - 1)^2} + \frac{1}{(2^{b_i} - 1)^3}, \quad (34)$$

an upper bound on  $\lambda'$  can be found using (27) as follows:

$$\lambda' \left( \sum_{i=1}^N w_i \right) = \sum_{i=1}^N \frac{2^{b_i} c_i^2}{(2^{b_i} - 1)^3} \geq \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} = \frac{\beta}{U^2}, \quad (35)$$

and we conclude that  $\lambda^* \leq \lambda'$ . On the other hand, if all  $b_i \geq 1$  we can write

$$\frac{1}{(2^{b_i} - 1)^2} \leq \frac{2^{b_i}}{(2^{b_i} - 1)^3} \leq \frac{2}{(2^{b_i} - 1)^2}, \quad (36)$$

therefore  $\lambda' \leq 2\lambda^*$ , and the result of the lemma follows.  $\square$

We now bound the difference  $|b_i - b_i^*|$ .

Lemma 3. Under the conditions of Lemma 2,

$$b_i^* - \frac{1}{2} < b_i < b_i^* + \frac{1}{2} \quad \forall i = 1, 2, \dots, N. \quad (37)$$

*Proof.* Using left-hand side of (36) and right-hand side of (33) we can write

$$\frac{1}{(2^{b_i} - 1)^2} \leq \frac{2^{b_i}}{(2^{b_i} - 1)^3} \leq \frac{w_i}{c_i^2} 2\lambda^*, \quad (38)$$

which gives the lower bound on  $b_i$ :

$$b_i \geq \log \left( 1 + \frac{|c_i|}{\sqrt{2\lambda^* w_i}} \right) = \log \left( \sqrt{2} + \frac{|c_i|}{\sqrt{\lambda^* w_i}} \right) - \frac{1}{2} > b_i^* - \frac{1}{2}. \quad (39)$$

By analogy, from right-hand side of (36) and left-hand side of (33) we have

$$\frac{2}{(2^{b_i} - 1)^2} \geq \frac{2^{b_i}}{(2^{b_i} - 1)^3} \geq \frac{w_i}{c_i^2} \lambda^*, \quad (40)$$

which further implies

$$b_i \leq \log \left( 1 + \frac{\sqrt{2}|c_i|}{\sqrt{\lambda^* w_i}} \right) = \log \left( \frac{1}{\sqrt{2}} + \frac{|c_i|}{\sqrt{\lambda^* w_i}} \right) + \frac{1}{2} < b_i^* + \frac{1}{2}. \quad (41)$$

This completes the proof.  $\square$

Lemma 3 implies that  $|b_i - b_i^*| < 1$ . Thus, rounded optimal solution  $\lceil b_i \rceil$  is at most one bit away from  $\lceil b_i^* \rceil$ . We can interpret this result as follows: in situation when  $b_i$  are sufficiently large, for example, when high estimation precision is required, the optimal solution behaves approximately as  $\log(1 + |c_i|/\sqrt{\lambda^* w_i})$ . Notice that  $c_i = \mathbf{e}_i^T \mathbf{C}^{-1} \mathbf{1}$  ( $\mathbf{e}_i$  denotes the  $i$ th unit vector), so  $c_i$  signifies the inverse of “noisiness” of signal  $x_i$  in relation to the other sensor observations. Recalling the definition of  $\lambda^*$  we note that product  $\lambda^* w_i$  is proportional to the relative energy per bit  $w_i/\sum w_j$  and the value of  $1/\sqrt{\lambda^* w_i}$  can be interpreted as being proportional to the relative quality of wireless link between sensor  $i$  and the FC. Thus, the local message length  $b_i^*$  can be intuitively interpreted as being proportional to the logarithm of the product of signal quality and channel quality at sensor  $i$ .

We now consider a special case when the use of  $\{b_i^*\}$  is especially appealing. Suppose that covariance matrix  $\mathbf{C}$  has a block-diagonal structure

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_n \end{bmatrix}. \quad (42)$$

This situation may occur when sensors in the network are partitioned into several clusters in such a way that sensors within each group are placed relatively close to each other and far from the rest of the sensors. Thus, sensor observations are uncorrelated unless they are generated from the same cluster. In this case matrix  $\mathbf{C}^{-1}$  is also block-diagonal:

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_n^{-1} \end{bmatrix}. \quad (43)$$

We assume further that sensors within each group can cooperate to learn the corresponding covariance submatrix  $\mathbf{C}_j$ . Value of  $\lambda^*$  can be computed by the fusion center and broadcasted back to the sensors. Thus, each sensor can easily compute  $c_i = [\mathbf{C}_j^{-1} \mathbf{1}]_i$  and independently find its own quantization level  $b_i^*$ . The advantage of this method is that the fusion center needs to broadcast only one universal message for all sensors.

To conclude this section we observe that our strategy can be applied even if sensor noises have infinite range. Indeed, with an appropriate choice of  $U$ , that is, if tails of the noise pdf are negligible, the pdf can be approximated by a finite support function. However, the estimator (9) will no longer be unbiased and cross terms  $E(\bar{\theta} - \hat{\theta})(\hat{\theta} - \theta)$  in the MSE expression will no longer be zero. Thus, inequality (15) only defines a lower bound on estimation performance for some  $\alpha$ , and the gap between left-hand side of (15) and actual MSE is determined by the noise pdf. Therefore, the full pdf knowledge will be required in order to specify constants  $U$  and  $\alpha$  and quantify the estimation bias.

#### 4. NUMERICAL SIMULATIONS

In this section, we present numerical simulations to compare the transmission energy requirement for two energy scheduling strategies: (i) quantization using the closed-form approximate solution (32); (ii) uniform bit allocation when all sensors quantize their observations to the same number of bits to achieve the same MSE. We denote by  $b$  the number of bits used in case of uniform bit allocation. We can find the minimum of  $b$  from the MSE constraint

$$\sum_{i=1}^N \frac{c_i^2}{(2^b - 1)^2} \leq \frac{\beta}{U^2}, \quad (44)$$

which gives

$$b \geq \log \left( 1 + \sqrt{\frac{U^2}{\beta} \sum_{i=1}^N c_i^2} \right). \quad (45)$$

The number of bits can only take integer values, so the total minimal energy is given by

$$W_{\text{uniform}} = \left\lceil \log \left( 1 + \sqrt{\frac{U^2}{\beta} \sum_{i=1}^N c_i^2} \right) \right\rceil \sum_{i=1}^N w_i. \quad (46)$$

Recall that we have relaxed  $b_i$  to take real values to make the problem convex. Therefore, the optimal energy obtained by allowing  $b_i$  to take on real values is a lower bound on the actual optimal energy. If we round  $b_i$  up to the closest integer  $\lceil b_i \rceil$ , we can obtain an upper bound (denoted by  $\overline{W}_{\text{opt}}$ ) on the actual energy. Even though we use  $\lceil b_i^* \rceil$  to approximate the actual optimal solution, significant energy can be saved when compared with the uniform bit allocation strategy in order to achieve the same target distortion. The percentage of saving is defined as

$$\frac{W_{\text{uniform}} - \overline{W}_{\text{opt}}}{W_{\text{uniform}}} \times 100. \quad (47)$$

For a positive random variable  $R$  we define

$$\text{normalized deviation of } R = \frac{\sqrt{\text{var } R}}{ER}, \quad (48)$$

which will be used as a measure of the absolute heterogeneity of  $R$ . The sensor noise variances  $\{\sigma_i^2\}$  are taken to be  $\sigma_i^2 = 1 + a^2 Z_i$ , where  $Z_i$  are i.i.d. random variables with  $Z_i \sim \chi_1^2(z)$ . As can be easily verified,  $\{\sigma_i^2\}$  are also i.i.d. with  $\sigma_i \sim \chi_1^2((x-1)/a^2)$ . We control heterogeneity of sensor noise variances by varying the parameter  $a$ . In Figure 2a, we suppose that sensor noises have tri-diagonal correlation matrix

$$\mathbf{C} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N) \begin{bmatrix} 1 & \rho & \dots & 0 & 0 \\ \rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \rho \\ 0 & 0 & \dots & \rho & 1 \end{bmatrix} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N), \quad (49)$$

where  $\rho = 0.2$ . In Figure 2b, we suppose that sensor noises have correlation matrix

$$\mathbf{C} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N) [(1-\rho)\mathbf{I} + \rho \mathbf{1}\mathbf{1}^T] \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N). \quad (50)$$

In all simulations, the total number of sensors  $N = 200$ . Since all coefficients  $w_i$  are scaled by a common factor, in our simulation,  $\{w_i\}$  are taken to be channel path losses

$$w_i = d_i^\kappa. \quad (51)$$

Assume that the target estimation performance is fixed. From Figure 2 we can see that the amount of energy saving becomes significant when the local noise variances become more and more heterogeneous, assuming that all sensors have identical  $w_i$ . In Figure 3, we plot the percentage of energy savings versus the heterogeneity of channel gains,

supposing that sensors have same observation noise variances with tri-diagonal structure as in (49) where  $\sigma_i^2 = 1$  for all  $i$ , and  $\rho = 0.2$ . Here we suppose that all sensors are uniformly distributed inside a unitary disk whose center is at the FC. It is easy to show that in this case normalized deviation of  $w_i$  depends only on  $\kappa$  (cf. (51)). In our simulation, we choose  $1 \leq \kappa \leq 8$ . We observe that percentage of saving depends more on the heterogeneity of sensor noise variances than that of channel gains. This can be understood regarding expression (32) for  $b_i^*$ , where in the logarithm, the quantity depends on the distribution of  $c_i$ , but only on the distribution of  $1/\sqrt{w_i}$ .

#### 5. AN EXTENSION: MINIMAX FORMULATION

Minimizing total transmission energy results in sensors having different lifetimes. This may induce frequent changes in the network topology. An alternative approach is to minimize

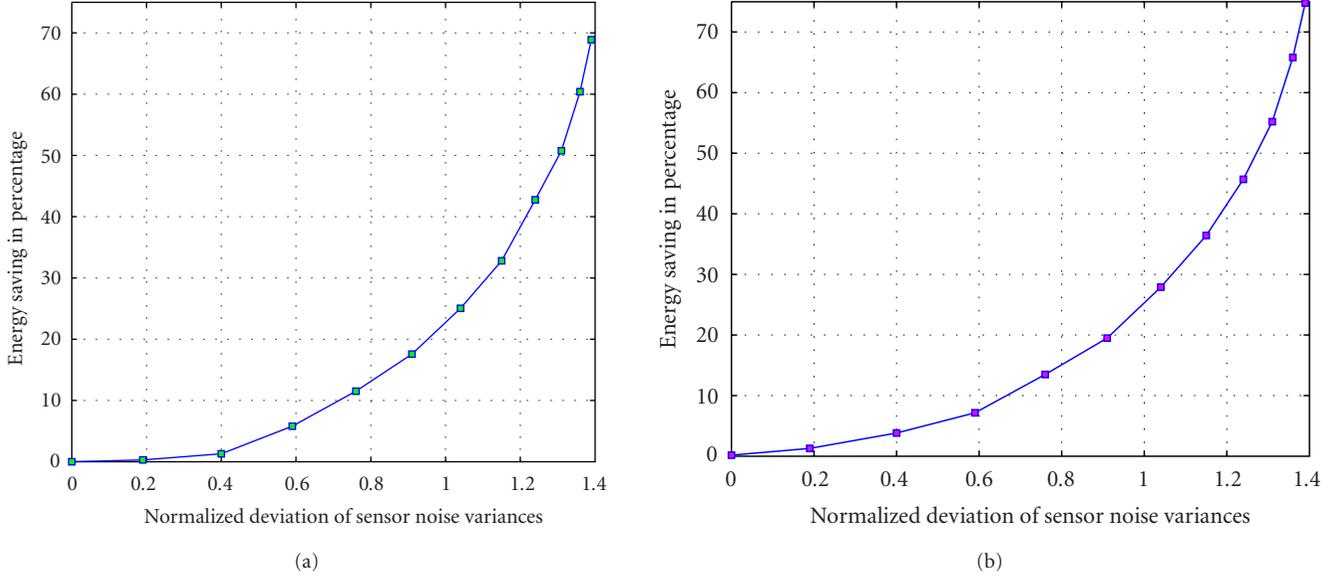


FIGURE 2: Percentage of energy saving increases when sensor noise variances become more heterogeneous.

maximal energy  $W_i$  which leads to maximum network lifetime. Relaxing  $\{b_i\}$  as in (24), we can state the problem as follows:

$$\begin{aligned} & \text{minimize } \max_i w_i b_i \\ & \text{subject to } \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} \leq \frac{\beta}{U^2}, \quad b_i \geq 0, \quad i = 1, \dots, N, \end{aligned} \quad (52)$$

or alternatively

$$\begin{aligned} & \text{minimize } \max t \\ & \text{subject to } w_i b_i \leq t \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} \leq \frac{\beta}{U^2}, \quad b_i \geq 0, \quad i = 1, \dots, N. \end{aligned} \quad (53)$$

As in Section 3, we assume that  $c_i \neq 0$  for all  $i$  and ignore the nonnegativity constraints  $b_i \geq 0$  (which must be inactive at optimum). The Lagrangian for problem (53) is found to be

$$L(t, b_i, \mu_i, \lambda) = t + \sum_{i=1}^N \mu_i (w_i b_i - t) + \lambda \left( \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} - \frac{\beta}{U^2} \right). \quad (54)$$

Differentiating  $L$  with respect to primal variables we obtain the following conditions:

$$\begin{aligned} \frac{\partial L}{\partial t} &= 1 - \sum_{i=1}^N \mu_i = 0, \\ \frac{\partial L}{\partial b_i} &= -2\lambda \ln 2 \frac{2^{b_i} c_i^2}{(2^{b_i} - 1)^3} + \mu_i w_i = 0, \end{aligned} \quad (55)$$

which give

$$\sum_{i=1}^N \mu_i = 1, \quad (56)$$

$$\lambda' \mu_i = \frac{2^{b_i} c_i^2}{(2^{b_i} - 1)^3 w_i}, \quad (57)$$

where as before  $\lambda' = 1/2\lambda \ln 2$ . Taking sum of (57) over all  $i$  we obtain

$$\lambda' = \sum_{i=1}^N \frac{c_i^2}{w_i} \frac{2^{b_i}}{(2^{b_i} - 1)^3}. \quad (58)$$

Since each term in the right-hand side sum in (58) is positive, we conclude that  $\lambda > 0$ , therefore  $\mu_i > 0$ , and complementary slackness condition gives

$$\begin{aligned} \sum_{i=1}^N \frac{c_i^2}{(2^{b_i} - 1)^2} &= \frac{\beta}{U^2}, \\ w_i b_i &= t. \end{aligned} \quad (59)$$

Thus, the optimal value  $t_{\text{opt}}$  can be found as a solution to the following equation:

$$\sum_{i=1}^N \frac{c_i^2}{(2^{t/w_i} - 1)^2} = \frac{\beta}{U^2}. \quad (60)$$

The solution  $t_{\text{opt}}$  is unique due to the monotonicity of the left-hand side function in (60). The FC can solve (60) and broadcast  $t_{\text{opt}}$  to the sensors, which in turn can determine their quantization levels locally. In this case sensor lifetime is not affected by transmitted power.

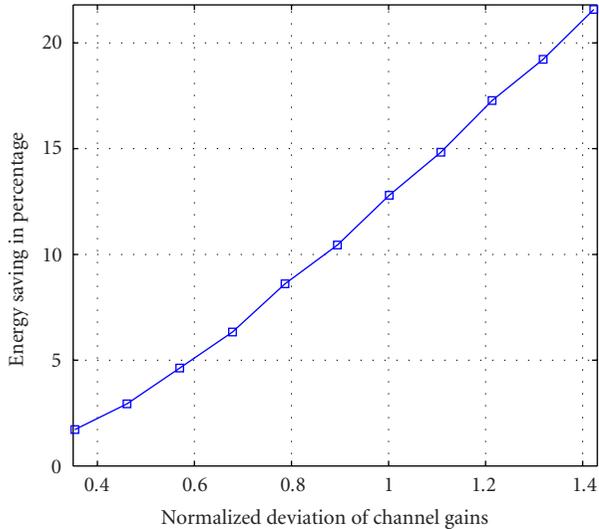


FIGURE 3: Percentage of energy saving increases when channel gains become more heterogeneous.

## 6. CONCLUSION

In this paper we have shown that total energy consumption required for transmission in a sensor network can be minimized if number of quantization levels for each sensor is determined jointly by the fusion center using information about correlation of sensor observations. We have also presented a nearly-optimal solution in closed form to the energy minimization problem which can achieve the same target estimation performance as the optimal solution. It is shown by numerical simulations that to attain the same MSE performance our energy-efficient quantization scheme can achieve energy saving up to 70% when compared to simple uniform bit allocation scheme. We plan to consider various extensions of this work in our future work. These include joint estimation of a common vector signal by a WSN, and distributed least squares and target tracking for dynamic targets.

## ACKNOWLEDGMENTS

Authors would like to thank the anonymous reviewers for their valuable comments that helped improve the quality of this paper. 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.

## REFERENCES

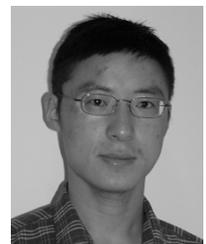
[1] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," to appear in *IEEE Trans. Inform. Theory*.

- [2] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 735–744, 2005.
- [3] Z.-Q. Luo and J.-J. Xiao, "Universal decentralized estimation in an inhomogeneous environment," to appear in *IEEE Trans. Inform. Theory*.
- [4] Z.-Q. Luo and J.-J. Xiao, "Decentralized estimation in an inhomogeneous environment," in *Proc. IEEE International Symposium on Information Theory (ISIT '04)*, pp. 517–517, Chicago, Ill, USA, June–July 2004.
- [5] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," to appear in *IEEE Transaction on Wireless Communications*.
- [6] S. Cui, A. J. Goldsmith, and A. Bahai, "Joint modulation and multiple access optimization under energy constraints," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '04)*, pp. 151–155, Dallas, Tex, USA, November–December 2004.
- [7] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," to appear in *IEEE Trans. Signal Processing*.
- [8] X. Luo and G. B. Giannakis, "Energy-constrained optimal quantization for wireless sensor networks," in *Proc. 1st IEEE Annual Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON '04)*, pp. 272–278, Santa Clara, Calif, USA, October 2004.
- [9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.
- [10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2003.

**Alexey Krasnopeev** received the B.S. degree in applied mathematics and physics in 1999, and the M.S. degree in applied mathematics and physics in 2001, both from the Moscow Institute of Physics and Technology, Moscow, Russia. He is currently pursuing his M.S. degree in electrical engineering at the University of Minnesota. His research interests include wireless sensor networks, information theory, and algebraic coding theory.



**Jin-Jun Xiao** received the B.S. degree in applied mathematics from Jilin University, China, in 1997, and the M.S. degree in mathematics from the University of Minnesota, in 2003. He is currently pursuing the Ph.D. degree in electrical engineering at the University of Minnesota. His research interests are in wireless sensor networks, information theory, and optimization.



**Zhi-Quan Luo** received the B.S. degree in mathematics from Peking University, China, in 1984. During the academic year of 1984/1985, he was with Nankai Institute of Mathematics, Tianjin, China. From 1985 to 1989, he studied at the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, where he received a Ph.D. degree



in operations research. In 1989, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, where he became a Professor in 1998 and held the Canada Research Chair in information processing since 2001. Starting April 2003, he has been a Professor in the Department of Electrical and Computer Engineering at the University of Minnesota, and holds an ADC Chair in digital technology. His research interests lie in the union of large-scale optimization, signal processing, data communications, and information theory. He is a Member of SIAM and MPS. He is presently serving as an Associate Editor for several international journals including SIAM Journal on Optimization, Mathematical Programming, Mathematics of Computation, and Mathematics of Operations Research.