

# Joint Source Coding, Routing and Resource Allocation for Wireless Sensor Networks

Wei Yu and Jun Yuan

Electrical and Computer Engineering Department

University of Toronto, Canada

{weiyu, steveyuan}@comm.utoronto.ca

**Abstract**—This paper presents an optimization framework for a wireless sensor network in which each sensor plays a dual role of sensing the environment and relaying the sensor information. The design of such a network involves two distinct aspects. First, as the observations of the underlying environment are often correlated, distributive source coding methods have the potential to greatly improve the efficiency of the sensor operation. Thus, information theoretical source coding methods are useful in the application layer. Second, as each sensor must send information individually to a central processor, routing and power allocation in the network and physical layers are also important issues. The main focus of this paper is an optimization framework that jointly solves the source coding, routing and power allocation problems in such a network. The main insight is the following: the joint optimization problem for a sensor network, when solved in the dual domain, provides a natural separation between the application layer, the network layer and the physical layer. The interface between the layers are precisely the dual optimization variables. The crucial observation that makes this possible is that the underlying source coding problem in the application layer and the channel coding problem in the physical layer can always be made convex via time-division or frequency-division multiplexing. Convexification in time or frequency enables dual algorithms to reach the global optimum of the overall network optimization problem efficiently.

## I. INTRODUCTION

Sensor networks have been widely regarded as a promising emerging application for future wireless networks. Wireless sensor networks are envisioned to consist of a large number of low-cost low-power sensors deployed over a large area. Each sensor is capable of not only making local observations of the underlying physical environment (e.g. temperature), but also capable of transmitting the observed information to and relaying the information for its neighbours using wireless links. Information from all the sensors is eventually collected at a central processing unit, which produces a global picture of the physical environment.

The design of wireless sensor networks involves two distinct aspects. First, as each sensor observes some local information of the physical environment, such observations are often correlated. Thus, independent quantization and transmission of the observation is not an optimal strategy. Higher quantization efficiency can be obtained using *distributive source coding* techniques from multiuser information theory. Distributive

source coding is usually regarded as an *application layer* issue in network design.

Second, as each sensor is also responsible for information transmission and relaying, the sensors form an ad-hoc network in which design issues such as routing and power allocation are of paramount importance. Routing and power allocation are often considered as *network layer* and *physical layer* issues.

Although the application layer, the network layer and the physical layer are traditionally designed separately, separation is in fact not theoretically justified in a network environment. However, in this paper, we argue that while layering may not necessarily lead to an optimal design in an information theoretical sense, there are in fact clear benefits for layering from a network optimization point of view. Toward this end, we propose an optimization framework for the sensor network that provides a solution to the joint source coding, routing and resource allocation problem. Our main result illustrates that the overall sensor network optimization problem naturally decomposes into distinct parts, corresponding to the traditional layers, with the Lagrangian dual variables playing the roles of coordinating crosslayer interfaces. The crucial ingredient that has made this optimization-theoretical layering approach possible is the following observation. While both the distributive source coding problem in the application layer and the power allocation problem in the physical layer are often non-convex optimization problems, they can be turned into convex problems via time-division or frequency-division multiplexing. This seemingly obvious fact is the key observation that allows the overall sensor network optimization problem to be solved efficiently in the Lagrangian dual domain.

The optimization theoretical layering approach advocated in this paper is inspired by the recent work of Chiang [1], which showed that in a wireless network, the physical layer resource allocation problem and the transport layer TCP congestion control problem can be seen as sub-problems of an overall problem. However, the layered approach contained in [1] is applicable only if the underlying physical layer optimization problem is convex, which is often not the case, especially when nodes in the wireless network interfere with each other. The main contributions of this paper are, first to illustrate that this layered approach is applicable not only between the physical and network layers, but also between network and application layers, and second to show that while convexity is indeed crucial, convexity can always be assured if time-sharing

<sup>1</sup>This work was supported by Natural Sciences and Engineering Council (NSERC) of Canada, and by the Canada Research Chairs program.

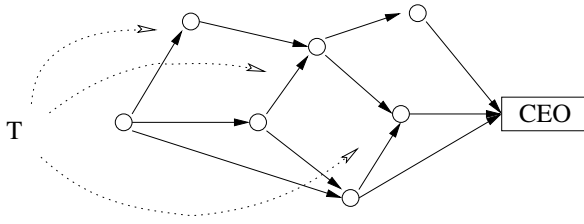


Fig. 1. A sensor network

or frequency-sharing is allowed in both the physical and application layers. This second observation is related to recent advances in optimal spectrum management for physical-layer multiuser transmission systems [2], for which it is shown that dual optimization techniques are capable of solving potentially non-convex capacity maximization problems whenever time-sharing is allowed. This paper shows that a similar approach works also for the distributive source coding problem in the application layer. The optimization approach used in this paper is also inspired by the work of Johansson, Xiao and Boyd [3] where the design of joint routing and power allocation schemes in wireless ad-hoc networks is considered. The network optimization problem considered in this paper is an extension of [3] in combining routing and resource allocation with distributive source coding in an overall framework. The architectural issue and crosslayer design for the wireless sensor networks are also explored in the recent work of Scaglione and Servetto [4], where the interaction of distributive source coding problem and routing is considered. This paper is a generalization of [4] in the sense that physical layer information is also taken into account in the overall network design in addition to routing and source coding.

## II. SENSOR NETWORKS

The wireless sensor network considered in this paper is depicted in Fig. 1 where a large number of sensors make local observations about some underlying physical quantity. The observations are then transmitted and relayed by the same set of nodes to a central processor, called the “CEO” in the Fig. 1. Thus, each sensor plays the roles of both sensing and transmission.

This paper considered a separate source and channel coding approach where the distributive source coding and the information transmission problems are considered in two different layers. Such a separation approach, although not necessarily optimal, is however architecturally appealing. Recent work such as [5] [6] has considered uncoded transmission with joint source and channel coding strategies. Although joint source and channel coding may lead to a truly optimal overall performance in an information theoretical sense, the layered approach advocated in this paper is much easier to design in practice. For this reason, this paper considers only networks with separate application and physical layers and focuses instead on the joint optimization of the two.

Even with separate source and channel coding, fundamen-

tally different network architectures are still possible depending on the roles of the sensors and the relays in the chain of information processing. The main part of this paper focuses on a sensing-and-routing architecture where each sensor distributively quantizes the local observation and routes the information to the CEO. Alternatively, a tree-based architecture [7], where each sensor may combine local observation with quantized observation from neighbouring sensors, is also possible. Such an architecture can be accommodated in the proposed framework as well. A detailed discussion is included at the end of the paper.

### A. Source Coding, Channel Coding and Routing

Source coding is the art of describing the underlying physical quantity as accurately as possible using as few bits as possible. In a distributive source coding environment, multiple sensors observe correlated information about an underlying object, quantize the correlated observations independently, and transmit independent bits to a central processor, which is responsible for reconstructing the physical environment with the smallest distortion possible. The fundamental concept in distributive source coding is the notion of *rate-distortion region*. Consider a sensor network with  $n$  nodes. For a fixed distortion  $D$ , the rate-distortion region is the set of rates  $(R_1, \dots, R_n)$  that are capable of achieving  $D$ . Rate-distortion region is an application layer concept.

Channel coding is the art of transmitting information with an arbitrarily small probability of error. In a network environment, multiple nodes may interfere, cooperate or relay information across the network. The fundamental concept in channel coding is the notion of *capacity region* under a power constraint. In a network of  $l$  links under a fixed power constraint at each transmitter, the capacity region is the set of rates  $(C_1, \dots, C_l)$  that can be achieved with a negligible probability of error. Capacity region is a physical layer concept.

This paper will not investigate the detailed structures of either the rate-distortion region or the channel capacity region. Rather, we assume that these regions are available and instead focus on the interaction between the two. Throughout this paper, distortion is denoted as  $D$ , power constraints are denoted as  $P$ , the source rates are denoted as  $s$ , the link capacities are denoted as  $x$ , the rate-distortion region is denoted as  $\mathcal{R}(D)$  and the channel capacity region is denoted as  $\mathcal{C}(P)$ . Distortion  $D$  is a vector as the underlying physical quantity is usually a vector random variable. Likewise, power  $P$  is also a vector as multiple transmitter power constraints are often involved. To achieve a certain distortion  $D$ , the source rates  $\mathbf{s} = (s_1, \dots, s_n)$  have to satisfy:  $\mathbf{s} \in \mathcal{R}(D)$ . Likewise, under a fixed power constraint, the link capacities  $\mathbf{x} = (x_1, \dots, x_l)$  have to satisfy  $\mathbf{x} \in \mathcal{C}(P)$ .

The rate-distortion region is purely a *node-based* concept. The capacity region is purely a *link-based* concept. To achieve certain distortion under a fixed power constraint, the sensor network has to ensure that the distortion-achievable rates at the sensor nodes can be supported by the links. Routing plays such a role. Mathematically, the routing problem can be formulated

using a multi-commodity flow model [8] via a so-called node-link incidence matrix [3]. Let  $A$  be an  $n \times l$  matrix in a network with  $n$  nodes and  $l$  links. Define

$$a_{nl} = \begin{cases} 1 & \text{if } n \text{ is the start node for link } l \\ -1 & \text{if } n \text{ is the end node for link } l \\ 0 & \text{else} \end{cases} \quad (1)$$

Then, the routing problem is equivalent to the problem of ensuring that a network flow conservation constraint

$$A\mathbf{x} \geq \mathbf{s} \quad (2)$$

is satisfied. The definition of the node-link incidence matrix  $A$  ensures that at each node the difference between outgoing flow and incoming flow is greater than or equal to the source rate generated at the node.

### B. Sensor Network Optimization Problem

We are now ready to state the network optimization problem for a sensor network. The fundamental objective in a sensor network is to minimize the distortion  $D$  incurred in the estimation process. The fundamental constraints in the network are transmit power constraints  $P$  at each node. The goal of the network optimization problem is to characterize a trade-off between the two. Such a trade-off can be parameterized by weighting vectors  $\alpha$  and  $\beta$  in an overall network optimization problem as follows:

$$\begin{aligned} & \text{minimize} && \alpha^T D + \beta^T P && (3) \\ & \text{subject to} && \mathbf{s} \in \mathcal{R}(D) \\ & && \mathbf{x} \in \mathcal{C}(P) \\ & && A\mathbf{x} \geq \mathbf{s} \end{aligned}$$

where  $\alpha$  and  $\beta$  are both vectors. The values of  $\alpha$  and  $\beta$  represent the relative emphasis on distortion and power. Note that both  $D$  and  $P$  are vectors as well.

## III. OPTIMIZATION FRAMEWORK

### A. Optimization Theoretical Layering

The optimization theoretical layering stems from an interpretation of the Lagrangian dual of the above optimization problem with respect to the constraint  $A\mathbf{x} \geq \mathbf{s}$ :

$$L = \alpha^T D + \beta^T P + \lambda^T (\mathbf{s} - A\mathbf{x}) \quad (4)$$

The Lagrangian multiplier  $\lambda$  can be interpreted as the shadow price for bit transmission at each node.

A careful examination of the minimization of the Lagrangian above reveals the following. The minimization problem consists of two sets of variables: application layer variables  $(D, \mathbf{s})$ , and, network and physical layer variables  $(P, \mathbf{x})$ . More specifically, the Lagrangian optimization problem decouples into two distinct parts. The source coding part in application layer is a pure rate-distortion minimization problem:

$$\begin{aligned} & \text{minimize} && \alpha^T D + \lambda^T \mathbf{s} && (5) \\ & \text{subject to} && \mathbf{s} \in \mathcal{R}(D) \end{aligned}$$

The routing and resource allocation is a pure capacity maximization problem:

$$\begin{aligned} & \text{maximize} && \mu^T \mathbf{x} - \beta^T P && (6) \\ & \text{subject to} && \mathbf{x} \in \mathcal{C}(P) \end{aligned}$$

where  $\mu$  is related to the dual variable  $\lambda$  by the link price consistency equation:

$$\mu^T = \lambda^T A \quad (7)$$

In particular, the  $l$ th link price  $\mu_l$  is precisely the difference between the shadow price associated with the start node of link  $l$  and the shadow price associated with the end node.

Thus, the optimization framework naturally provides a layering approach to the sensor network optimization problem. The global minimization problem decomposes into two parts: the source coding problem in the application layer and the channel coding problem in the physical layer. The channel capacity problem ensures that the maximal capacity is provided in individual network links, while the rate distortion problem ensures that the minimum number of bits is used for quantization.

The decoupling of the network optimization problem also reveals that cross-layer design can be done in a theoretically optimal way. The dual variables  $\lambda$  and  $\mu$  play a key role in coordinating the application layer ‘demand’ and physical layer ‘supply’. In particular,  $\lambda_j$ , the  $j$ th component of  $\lambda$  can be interpreted as the cost of quantized bits in node  $j$ . A higher value of  $\lambda_j$  signals to the network that more resources should be devoted to transporting  $s_j$ . Likewise,  $\mu_l$ , the  $l$ th component of  $\mu$ , coordinates between the physical layer and the network layer.  $\mu_l$  can be interpreted as the cost of transmitted bits in link  $l$ . A higher value of  $\mu_l$  signals to the physical layer that capacity in link  $l$  is in high demand, while at the same time, it signals to the network layer that traffic should be diverted from link  $l$  to reduce congestion.

### B. Convexity and Primal-Dual Algorithm

The key requirement that allows the decoupling of the network optimization problem into source coding, routing and resource allocation is the underlying convexity structure of the problem. In this section, we first provide a justification for convexity, then propose a primal-dual algorithm that can be used to solve the joint network optimization problem efficiently.

We start by making the following observation. From an information theoretical point of view, both  $\mathcal{R}(D)$  and  $\mathcal{C}(P)$  must be convex regions, and both  $\mathcal{R}(D)$  and  $\mathcal{C}(P)$  must be convex functions of  $D$  and  $P$ , respectively. This is so because both the rate-distortion and the capacity regions can be convexified via time-sharing if necessary. For the capacity region  $\mathcal{C}(P)$ , if two sets of rates are both achievable under the same power constraint, then their linear combination must also be achievable by simply dividing the frequency (or time) into two subchannels and transmitting using the two different strategies in the two subchannels. Likewise, for the rate-distortion region  $\mathcal{R}(D)$ , if two sets of quantization schemes

achieve the same distortion with two rate tuples, then using one scheme for a fixed proportion of the time and the other scheme for remaining amount of time achieves all the time-sharing points.

A direct implication of convexity is that the rate-distortion sub-problem (5) and the channel capacity sub-problem (6) can be solved efficiently. In particular, these two sub-problems can themselves be dualized. This idea was first explored in [2] in the context of channel capacity problems for multiuser OFDM systems. The dual algorithm in [2] is equally applicable to the source coding problem. More precisely, as shown in [2], if time-sharing or frequency-sharing is allowed, then the capacity maximization problem (and the rate-distortion problem) can be solved in the dual domain. The duality gap is always zero. Therefore, both sub-problems (5) (6) have efficient numerical solutions, in principle.

We now propose the following primal-dual algorithm that solves the entire network optimization problem (3):

- 1) Initialize  $\lambda^{(0)}$
- 2) In primal domain, given  $\lambda^{(t)}$ , solve the following sub-problems:

$$\min_{D, \mathbf{s}} \left\{ \alpha^T D + \lambda^{(t)T} \mathbf{s} \mid \mathbf{s} \in \mathcal{R}(D) \right\} \quad (8)$$

$$\max_{P, \mathbf{x}} \left\{ \lambda^{(t)T} A \mathbf{x} - \beta^T P \mid \mathbf{x} \in \mathcal{C}(P) \right\} \quad (9)$$

- 3) In dual domain, update  $\lambda$  using the following rule:

$$\lambda^{(t+1)} = \lambda^{(t)} + \nu^{(t)}(\mathbf{s} - A \mathbf{x}) \quad (10)$$

- 4) Return to step 2 until convergence.

*Theorem 1:* The above algorithm always converges and it converges to the global optimum of the overall network optimization problem (3) provided that the step sizes  $\nu^{(t)}$  is chosen sufficiently small.

*Proof:* We outline the proof here. The crucial ingredient that makes the algorithm work is convexity. The convexity of  $\mathcal{R}(D)$  and  $\mathcal{C}(P)$  ensures that the network optimization problem (3) is convex. Define the dual objective function

$$g(\lambda) = \min_{D, \mathbf{s}, P, \mathbf{x}} L(D, \mathbf{s}, P, \mathbf{x}, \lambda). \quad (11)$$

The minimization above decomposes into two sub-problems (5) (6). The solutions of the two sub-problems allow  $g(\lambda)$  to be evaluated.

Now, by strong duality, the overall network optimization problem (3) is solved by the following dual maximization problem:

$$\max_{\lambda} g(\lambda) \quad (12)$$

It remains to show that the update steps (10) solve the dual maximization problem. This is due to the fact that the update steps are subgradient updates for  $\lambda$ . It is not difficult to show that  $(\mathbf{s} - A \mathbf{x})$  is a subgradient for  $\lambda$ . Thus, as long as step sizes  $\nu^{(t)}$  are chosen sufficiently small, the subgradient update necessarily converges and it converges to the global optimum of the overall network optimization problem. ■

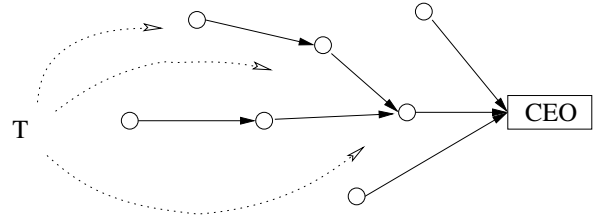


Fig. 2. A tree-based sensor network

### C. Tree-Based Architecture

The sensing-and-routing approach taken in previous sections is not the only sensor network architecture possible. Alternatively, each intermediate node may combine quantized information coming from other sensors with the local observation before sending it to the CEO. This approach is first proposed in [7] and it is most easily understood on a tree-based network topology as shown in Fig. 2. In this setting, the overall distortion becomes a function of the *link* rates, and the rate-distortion region becomes link-based (as opposed to node-based). The network optimization problem then becomes:

$$\begin{aligned} & \text{minimize} && \alpha^T D + \beta^T P && (13) \\ & \text{subject to} && \mathbf{y} \in \mathcal{R}_{\text{tree}}(D) \\ & && \mathbf{x} \in \mathcal{C}_{\text{tree}}(P) \\ & && \mathbf{y} \preceq \mathbf{x} \end{aligned}$$

where  $\mathbf{y}$  denotes admissible link rates under a distortion constraint and  $\preceq$  denotes vector component-wise inequality.

This new network optimization problem can be dualized in the exact same way as before. Define the Lagrangian:

$$L = \alpha^T D + \beta^T P + \lambda^T (\mathbf{y} - \mathbf{x}) \quad (14)$$

The source coding sub-problem now becomes:

$$\begin{aligned} & \text{minimize} && \alpha^T D + \lambda^T \mathbf{y} && (15) \\ & \text{subject to} && \mathbf{y} \in \mathcal{R}_{\text{tree}}(D). \end{aligned}$$

The channel coding sub-problem becomes:

$$\begin{aligned} & \text{maximize} && \lambda^T \mathbf{x} - \beta^T P && (16) \\ & \text{subject to} && \mathbf{x} \in \mathcal{C}_{\text{tree}}(P). \end{aligned}$$

The dual variable  $\lambda$  may be updated as:

$$\lambda^{(t+1)} = \lambda^{(t)} + \nu^{(t)}(\mathbf{y} - \mathbf{x}). \quad (17)$$

Thus, the optimization framework again naturally decomposes into distinct layers with the Lagrangian dual variable  $\lambda$  providing the natural coordination between the layers.

## IV. EXAMPLES

The following example is used to illustrate the main concepts introduced in this paper. A two-sensor three-link network as shown in Fig. 3 is simulated. Each sensor observes a noisy version of an underlying scalar Gaussian random variable  $T$ . The rate distortion region in this setting is known as

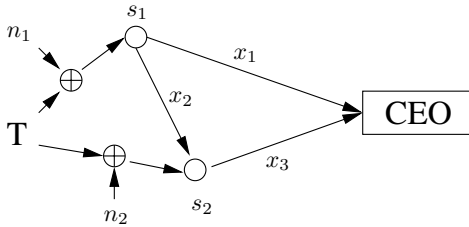


Fig. 3. A two-sensor three-link sensor network

the ‘‘Gaussian CEO’’ problem and the region is completely known [9]. The physical-layer channel model is an interference channel where the three transmitters of the links interfere with each other. A transmission scheme that allows time or frequency sharing is used. This guarantees the convexity of the underlying capacity region and ensures that dual algorithms may be used to efficiently compute the capacity region [2]. The routing problem is simple in this case. Only three routes are possible. The main routing issue is the allocation of bits to different routes.

The subgradient algorithm proposed earlier in the paper is used to find the optimal joint source coding, routing and resource allocation problem. Fig. 4 illustrates the distortion minimization process. The achievable rate-distortion region and a capacity region are plotted. The capacity region is a plot of  $x_1 + x_2$  vs  $x_3 - x_2$ , which are the rates that can be jointly supported by routing and physical layer. As can be seen in the figure, the primal-dual optimization process in this simple two-node example is in fact a matching process between the rate-distortion and capacity regions. The matching process is an illustration of the separation of the physical, routing and application layers. The inter-layer interface is controlled by the dual variables. The convergence behaviors of the dual variables are plotted in Fig. 5. As the figures show, the convergence is very fast. At the convergence point, the process reaches the global optimum for the joint source coding, routing and resource allocation problem in the sensor network.

## V. SUMMARY AND CONCLUSIONS

This paper presents an optimization framework for the joint source coding, routing and resource allocation problem in a sensor network. The main points are the following. First, the traditional layering in networking tremendously simplifies the sensor network optimization problem in the sense that a complex problem is decomposed into well-structured sub-problems corresponding to the layers, with the Lagrangian dual variables serve the role of interfacing between the layers. Second, dual algorithm can be used to solve the joint network optimization problem efficiently. Duality techniques are applicable for networks because the underlying information theoretical rate distortion and capacity regions can always be made convex.

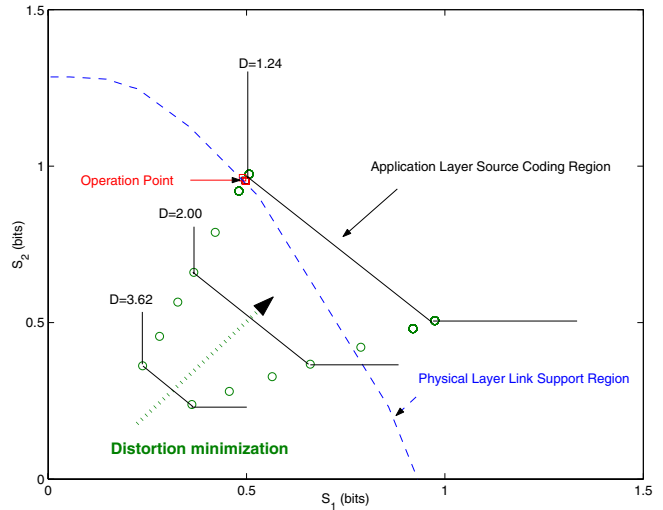


Fig. 4. Convergence of source and channel rates

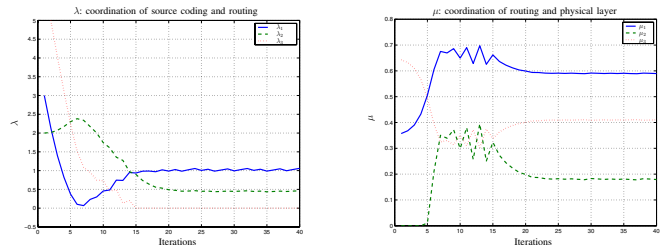


Fig. 5. Convergence of crosslayer dual variables

## REFERENCES

- [1] M. Chiang, ‘‘Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control,’’ *IEEE J. Sel. Areas Comm.*, vol. 23, no. 1, Jan. 2005.
- [2] W. Yu, R. Lui, and R. Cendrillon, ‘‘Dual optimization methods for multiuser orthogonal frequency division multiplex systems,’’ in *IEEE Globecom*, Dallas, TX, USA, 2004.
- [3] M. Johansson, L. Xiao, and S. Boyd, ‘‘Simultaneous routing and power allocation in CDMA wireless data networks,’’ in *IEEE. Inter. Conf. Comm. (ICC)*, Anchorage, Alaska, USA, May 2003.
- [4] A. Scaglione and S. D. Servetto, ‘‘On the interdependence of routing and data compression in multi-hop sensor networks,’’ in *8th ACM Inter. Conf. Mobile Computing and Networking (Mobicom)*, 2002.
- [5] M. Gastpar and M. Vetterli, ‘‘Source-channel communication in sensor networks,’’ in *2nd Inter. Workshop on Information Processing in Sensor Networks (IPSN’03)*, Palo Alto, CA, USA, 2003.
- [6] M. Gastpar and M. Vetterli, ‘‘Power-bandwidth-distortion scaling laws for sensor networks,’’ in *Proc. Third Inter. Symp. Info Proc. in Sensor Networks*, 2004.
- [7] S. C. Draper and G. W. Wornell, ‘‘Side information aware coding strategies for sensor networks,’’ *IEEE J. Select. Areas Commun.*, June 2004.
- [8] D. Bertsekas and R. G. Gallager, *Data Networks*, Prentice Hall, 1991.
- [9] V. Prabhakaran, D. Tse, and K. Ramchandran, ‘‘Rate region of the quadratic Gaussian CEO problem,’’ in *Proc. IEEE. Inter. Symp. Info. Theory*, 2004.