Energy Planning for Progressive Estimation in Multihop Sensor Networks

Yi Huang and Yingbo Hua, *Fellow, IEEE*

***Abstract*—Multihop sensor networks where transmissions are conducted between neighboring sensors can be more efficient in energy and spectrum than single-hop sensor networks where transmissions are conducted directly between each sensor and a fusion center. With the knowledge of a routing tree from all sensors to a destination node, we present a digital transmission energy planning algorithm as well as an analog transmission energy planning algorithm for progressive estimation in multihop sensor networks. Unlike many iterative consensus-type algorithms, the proposed progressive estimation algorithms along with their transmission energy planning further reduce the network transmission energy while guaranteeing any pre-specified estimation performance at the destination node within a finite time. We also show that digital transmission is more efficient in transmission energy than analog transmission if the available transmission time-bandwidth product for each link and each observation sample is not too limited.**

***Index Terms*—Decentralized estimation, distributed estimation, energy scheduling and planning, incremental estimation, multi-hop sensor networks, power scheduling and planning, progressive estimation, wireless sensor networks.**

I. Introduction

With the technological advances of microelectronics and wireless communications, wireless sensor networks are being developed for a wide range of applications such as target tracking in battle fields, environmental monitoring, security surveillance and space exploration. One of the common objectives at the application layer of wireless sensor networks is to estimate parameters of interest.

Because of limited spectrum and limited energy available for communications between sensors, distributed estimation is an important signal processing problem for wireless sensor networks. One of the leading recent efforts on distributed estimation was reported in [1]. In this work, each sensor is allowed to send a single bit to a fusion center where a desired scalar parameter is estimated. This work is followed by a number of extensions. In [2], the noise variances at different sensors are allowed to be different. In [6], the number of bits for uniform quantization at each sensor is designed to further reduce energy consumption. In [3] and [4], the distributed single-bit quantization problem is further explored in terms of the maximum likelihood estimation of the desired parameter. In [5], an optimal scalar quantizer is developed. For sensor networks where the unknown is a vector parameter, a distributed linear compression method is developed in [7]. This method relates to several earlier works on optimal reduced-rank estimation as shown in [8]. The idea of distributed linear compression is further explored in [9] and [10]. In addition to the above signal processing efforts, there has been information-theoretic research on the so called CEO problem [29] which addresses the rate-distortion limit without any constraints on delay and signal processing complexity. The information-theoretic research is outside the scope of this paper.

However, all of the above referenced works are restricted to single-hop networks where each sensor has a direct link to the fusion center. Single-hop networks are not energy efficient unless the number of sensors is small and they are all near the fusion center. For many applications, large-scale sensor networks are needed to collect data from a wide area. For large-scale sensor networks, providing a direct link from each sensor to the fusion center is inefficient in terms of energy and spectrum. An efficient way to transport information within a large-scale sensor network is to transport information only between adjacent sensors. Networks with multihop transmissions will be referred to as multihop networks.

Distributed estimation for multihop networks has also recently attracted strong interests. The main idea behind distributed estimation for multihop networks is that each sensor performs a local estimation by combining the estimation from its neighboring sensors and its local observation, and the estimate from each sensor is transmitted to its neighboring nodes for further processing. This in-network processing continues until a termination criterion is met. There are a number of studies on iterative algorithms for estimating common parameters in the network. In [11], a gradient-based algorithm is proposed to solve an optimization problem where the global cost function is a sum of the local cost functions. In [12], a nonlinearly coupled consensus-type algorithm is investigated. An incremental scheme is proposed in [13], and a diffusion scheme is shown in [14]. In [15], a distributed maximum likelihood method is developed for estimating deterministic parameters. In [16], a distributed maximum *a posteriori* algorithm is developed for estimating random parameters. Another perspective of distributed estimation is shown in [19]. It should be noted that

Manuscript received June 21, 2008; accepted May 03, 2009. First published May 27, 2009; current version published September 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhi Tian. This work was supported in part by the U.S. National Science Foundation under Grant TF-0514736 and the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program. This work was presented in part at the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, November 4–7, 2007.

The authors are with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: yhuang@ee.ucr.edu; yhua@ee.ucr.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2024023

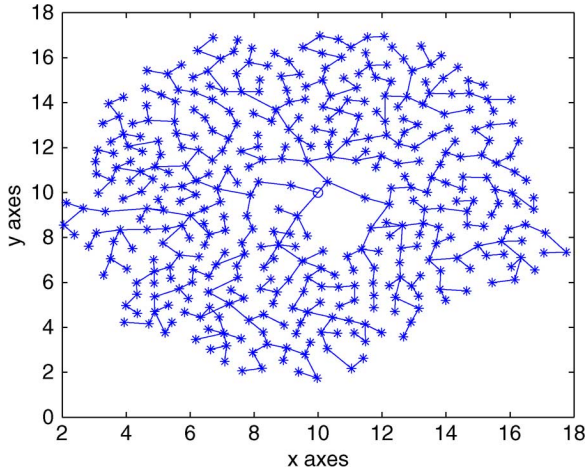


Fig. 1. A 2-D multihop sensor network with a routing tree. The destination node (also referred to as the fusion center) is marked by the circle in the center. Here, there are 400 sensors, each marked by $*$. This network is used for all simulation examples.

there are two types of distributed estimation. One requires a fusion center, and the other does not. The former is also referred to as decentralized estimation. This paper addresses issues of decentralized estimation in the context of multihop networks.

The above mentioned algorithms for multihop networks are all iterative in nature. Their convergence to a desired estimate is asymptotic, and their convergence rate depends on the network topology. Without the knowledge of the network topology, there is no guarantee of performance within any given time window. For applications where there is a strict requirement of guaranteed performance, the network topology must be acquired and exploited by a scheduler.

In this paper, we consider multihop networks where a routing tree between all sensors and their destination node is known to the scheduler. Such a network is illustrated in Fig. 1. The distributed estimation performed by each sensor is in a progressive fashion. In other words, after each sampling instant, each sensor computes an estimate of a desired time-varying space-invariant unknown vector of parameters based on the estimates from its upper stream sensors and its local observation, and the estimate from each sensor is forwarded to its down stream sensor defined by the routing tree. A key function of this scheduler is to determine the amount of transmission energy to be consumed by each sensor per sampling instant such that a pre-specified performance of estimation at the destination node is guaranteed. Unlike the iterative algorithms mentioned previously, the proposed progressive scheme considered in this paper terminates the computational process within a finite time for each sample of the desired unknown vector. In applications, there can be multiple destination nodes (or fusion centers), and a routing tree for each destination node can be assigned. In this case, the theory showed in this paper can be applied separately to each pair of the destination node and its routing tree.

It is important to add that our goal is to estimate and track changing unknown parameters (such as positions of moving targets) by using a wireless sensor network whose routing tree and channel state information can be exploited to minimize the energy consumption while guaranteeing any prespecified estimation quality within a finite time. The existing works on consensus-type algorithms are iterative and do not exploit the network condition to the degree we require but, however, suffer the uncertainty of the estimation quality especially when the unknowns to be estimated vary faster than the convergence rate of these algorithms.

The transmission energy planning problem has also been addressed previously. In [6], the planning problem is limited to a single-hop network. In [17], a planning algorithm is developed for multihop networks and for estimating a scalar unknown parameter. The work in [18] is a preliminary version of this paper.

The original contributions in this paper include 1) a comparison of digital communication and analog communication in terms of transmission energy and transmission errors; 2) an energy planning algorithm for progressive estimation in multihop networks using digital communication; 3) an energy planning algorithm for progressive estimation in multihop networks using analog communication; and 4) the use of L_p norm for both energy planning algorithms (with L_1 norm measuring the sum energy and L_∞ norm measuring the peak energy). We will show that digital communication is more efficient in transmission energy than analog communication unless the available transmission time-bandwidth product (i.e., $\mathcal{TW}$ or its integer part L) for each link and each observation sample is very limited. This fundamental observation provides a useful perspective of the prior works shown in [9] and [10] where analog communication is assumed between sensors and the fusion center. With the time-bandwidth product L for transmission of a message, one can use L parallel subchannels for digital transmission of the same message, the importance of which has been somehow overlooked in the literature on the modeling and analysis of sensor network problems.

Our proposed digital transmission energy planning algorithm is applicable to any L (but no less than one). This is another useful extension from [6] and [17] where the freedom of L is not utilized. The technique we will use for this extension is rooted in the Hölder's inequality. For both the digital and analog cases, we will show that the energy planning algorithms developed in this paper can save much energy in comparison to a uniform energy planning scheme.

The rest of this paper is organized as follows. Section II describes the model of the data observed by each sensor and a probabilistic uniform quantization scheme. Section III describes the transmission energy models for both analog and digital communications. Section IV compares the errors caused by analog and digital communications. Section V presents several progressive estimation algorithms for multihop networks with digital communications. These algorithms are based on the best linear unbiased estimation (BLUE), quasi-BLUE and a simple averaging. The recursive inequality of the covariance matrices based on the averaging algorithm is used for the design of the digital transmission energy planning to be shown in Section VI. In Section VII, progressive estimation algorithms for multihop networks with analog transmissions are proposed, which are based on BLUE and another simple averaging. The analog transmission energy planning is designed in Section VIII. The simulation examples are given in Section IX, which is followed by the conclusion.

II. Data Model and Problem Formulation

In this paper, we consider a wireless sensor network where each sensor can sense, compute, transmit and receive. We assume that a $N \times 1$ data vector $\boldsymbol{x}_k(n)$ is observed by sensor k at (discrete) time n, which is modeled as

$$\boldsymbol{x}_k(n) = \boldsymbol{G}_k\boldsymbol{\theta}(n) + \boldsymbol{\omega}_k(n) \tag{1}$$

where $\boldsymbol{\theta}(n) = [\theta_1(n), \theta_2(n), \cdots, \theta_M(n)]^T$ is a space-invariant vector of parameters to be estimated, $\boldsymbol{G}_k$ is a $N \times M$ observation matrix with $N \geq M$ and full column rank, and $\boldsymbol{\omega}_k(n)$ is the observation noise vector. Here, ^{T} denotes transpose. We assume that $\boldsymbol{\omega}_k(n)$ has zero mean and a time-invariant covariance matrix $\boldsymbol{C}_{\boldsymbol{\omega}_k}$. Since both $\boldsymbol{G}_k$ and $\boldsymbol{C}_{\boldsymbol{\omega}_k}$ are assumed to be known to a scheduler, we can let $\boldsymbol{C}_{\boldsymbol{\omega}_k} = \boldsymbol{I}$ without loss of generality. This is because left-multiplying the square-root of $\boldsymbol{C}_{\boldsymbol{\omega}_k}$ to both sides of (1) makes the noise vector whitened. We assume no knowledge of the statistical distribution of $\boldsymbol{\theta}(n)$ except that each entry of $\boldsymbol{\theta}(n)$ has a known upper bound and a known lower bound.

The estimation of the unknown parameter vector $\boldsymbol{\theta}(n)$ is done in a distributed and multihop fashion. We assume that there is a routing tree from all sensors towards a destination node as illustrated in Fig. 1. There are many ways to establish a routing tree. One example is the technique for finding Steiner tree or the minimum distance tree [23]. But the minimum distance tree may not be optimal for the purpose of distributed estimation. Finding the best tree for distributed estimation remains an important research topic. The destination node in the routing tree may also act as a scheduler which knows the observation matrices of all sensors, plans the transmission energy to be consumed by each sensor, and broadcasts all essential information to all sensors. The purpose of the transmission energy planning is to ensure a desired estimation performance at the destination node while keeping the total energy cost as low as possible. The process of routing tree establishment, transmission energy planning and decision broadcast are all done at a startup of the network or when the network topology changes. With the decision from the scheduler, each sensor in the network performs an estimation of $\boldsymbol{\theta}(n)$ corresponding to each sampling time instant n using both its own observation $\boldsymbol{x}_k(n)$ and the estimates obtained by its upper stream sensors.

We assume that the time interval between $\boldsymbol{x}_k(n)$ and $\boldsymbol{x}_k(n+1)$ is sufficiently long so that during this time interval all communications between sensors and all computations at each sensor can be completed for the estimation of $\boldsymbol{\theta}(n)$ at the destination node. We assume that orthogonal or approximately orthogonal scheduling is made for all links in the network. Since only neighboring nodes are communicating with each other, there can be concurrent and co-channel transmissions without major interferences to each other. We also assume that the network (including network topology and all channel state information) is sufficiently stationary so that each energy planning cycle can span a time window of at least many time instants of $\boldsymbol{\theta}(n)$ and hence the overhead of the energy planning is acceptable.

We will consider two different communication modes between sensors: analog communication and digital communication. When the digital communication is applied, we assume that each sensor quantizes its local estimate using a uniform quantization method. For example, if $\hat{\theta}$ is a local estimate of a scalar parameter θ and is bounded between $[W, -W]$, the quantization of $\hat{\theta}$ is uniform and probabilistic as follows. Let B be the number of bits used to quantize $\hat{\theta}$ within $[W, -W]$, and $\Delta = 2W/\left(2^B - 1\right)$ be the quantization interval. If $-W + i\Delta \leq \hat{\theta} \leq -W + (i+1)\Delta$ and $\hat{\theta} + W - i\Delta = r\Delta$, then $\hat{\theta}$ is quantized to $-W + i\Delta$ with the probability $1 - r$ and to $-W + (i+1)\Delta$ with the probability r. As shown in [2], it is easy to verify that the error of this quantization method has zero mean and the variance σ_q^2 bounded as follows:

$$\sigma_q^2 \leq \max_r r(1-r)\Delta^2 = \frac{\Delta^2}{4} \leq \frac{4W^2}{2^{2B}} \tag{2}$$

where the last inequality holds under $B \geq 1$. Note that if $\hat{\theta}$ is uniformly distributed within $[W, -W]$, then $\sigma_q^2 = \int_0^1 r(1-r)\Delta^2 dr = \Delta^2/6$.

We will not consider more advanced quantization methods such as vector quantization, or quantization based on a statistical knowledge of the random variables. The variance expression of a practical vector quantizer is generally difficult to get. Even for a single Gaussian random variable, the variance expression of the optimal quantization errors is not expressive enough for our purpose, e.g., see [20]. The simple distortion-rate expression in [28] requires an ideal vector quantizer of many observation samples, which is hard to implement. A recent work on source quantization in the context of sensor networks with structured trees is available in [21] where the knowledge of the joint statistical distribution of the source and the side data is required.

III. Communication Channel Model

Both analog communication and digital communication have been considered in the literature on sensor network signal processing. In this paper, we will provide a comparison of energy consumption between the two. In order to do so, we need to first establish the energy model for each communication mode as follows.

We denote by $\mathcal{T}$ and $\mathcal{W}$ respectively the total time and bandwidth available for communication for each link and each observation sample (i.e., for each k and n in $x_k(n)$). We assume that the radio frequency (RF) channels between neighboring sensors have constant channel gains within the bandwidth $\mathcal{W}$ and a whole cycle of transmission energy planning. For a RF channel of single-sided bandwidth $\mathcal{W}$, there is an equivalent complex baseband channel of the double-sided bandwidth $\mathcal{W}$, which corresponds to a pair of in-phase and quadrature channels.

We let h be the complex baseband channel gain from a sensor to its downstream sensor. The channel noise is assumed to be Gaussian with the power spectral density N_0 within the baseband $[-\mathcal{W}/2, \mathcal{W}/2]$.

To transmit a complex symbol s from one sensor to another in the analog mode, we need to use s to modulate (via amplitude, phase, frequency, pulse width, pulse position, etc.) a waveform $c(t)$ which has a time duration effectively no larger than $\mathcal{T}$ and a bandwidth effectively no larger than $\mathcal{W}$. Here, we use the word "effectively" because if zero error is tolerated, then the time duration of any waveform must be infinite

if the bandwidth is finite and vice versa. A fundamental constraint on $\mathcal{T}$ and $\mathcal{W}$ is $\mathcal{TW} \geq 1$. At the output of the baseband channel, a demodulation is performed to obtain an estimate $\hat{s}$ of s. The signal-to-noise ratio (SNR) of the estimated symbol at the channel output is

$$\text{SNR} = \frac{|s|^2}{\mathcal{E}\left\{|\hat{s}-s|^2\right\}} = \mu|h|^2 E/N_0 \tag{3}$$

where $\mathcal{E}$ denotes expectation, E is the total transmitted energy and μ ($\mu > 0$) is a parameter depending on the modulation method. As shown in Appendix I, if the pulse amplitude modulation and the matched filtering demodulation are used, then $\mu = 1$ and $E = |s|^2 E_c$ where E_c is the energy of the waveform $c(t)$.

Note that the combination of the waveform modulator, the continuous-time baseband channel and the waveform demodulation (with sampling) constitute a discrete-time analog channel, which is our model for analog communication of discrete-time symbols. For background knowledge of modulation, demodulation and sampling, there are many textbooks such as [25].

In order to transmit a complex symbol from one sensor to another in the digital mode, we can use the above described discrete-time analog channel preceded by a digital encoder and followed by a digital decoder [25]. Furthermore, if the time-bandwidth product of the RF channel is such that $\mathcal{TW} \geq L \geq 1$ where L is an integer, then we can partition the original channel into L subchannels (in frequency and/or time). For each subchannel, we can apply the above mentioned modulation method to yield an discrete-time analog channel with an output SNR which is proportional to the input energy of the subchannel. In other words, with $\mathcal{TW} \geq L \geq 1$, we can construct L discrete-time analog subchannels and the lth channel has the output SNR: $\text{SNR}_l = \mu|h|^2 E/(LN_0)$, where E denotes the total transmission energy over the L subchannels. Then, the total number of bits that we can transmit digitally over the RF channel within the time $\mathcal{T}$ and bandwidth $\mathcal{W}$ (with negligible errors) is

$$B = \sum_{l=1}^{L} \psi \log_2(1+\phi \text{SNR}_l) = \psi L \log_2\left(1+\phi\mu\frac{|h|^2 E}{LN_0}\right) \tag{4}$$

where ψ ($0 < \psi < 1$) and ϕ ($0 < \phi < 1$) are penalty factors depending on the coding method [24]. This model is appropriate even if the coding is done within a single observation time interval of each link. Alternatively, we can write the transmission energy in terms of the number of bits as follows:

$$E = \frac{LN_0}{\phi\mu|h|^2}\left(2^{\frac{B}{\psi L}} - 1\right). \tag{5}$$

Note that if the subchannels have different channel coefficients, the transmitted energy can be more efficiently distributed according to a water filling algorithm, which, however, will lead to a more complicated expression of the transmission energy in terms of the number of bits. For this reason, we will not consider such a case in this paper.

The model (5) will be used to design the transmission energy planning for digital communication between sensors. The model (3) will be used for the analog case. The energy planning problem will be formulated as minimization of the transmission energy within the entire network subject to that the mean square error (MSE) of the estimate at the destination node is no larger than a pre-specified value.

IV. Analog Versus Digital Transmissions

The communication between adjacent sensors can be digital or analog in principle. Digital communication has many advantages over analog communication, which includes modularity and robustness. These two advantages appear very important for large-scale multihop wireless networks. But from the pure transmission energy point of view, the conclusion is not always in favor of one over the other, which depends on the available time-bandwidth product.

Transmitting a complex symbol s from one sensor to another by the analog mode introduces an additional noise term whose variance is inversely proportional to the transmission energy E according to (3). More specifically, the variance of the additional noise caused by the analog communication is

$$\sigma^2_{\text{analog}} = \frac{|s|^2}{\text{SNR}} = \frac{|s|^2 N_0}{\mu|h|^2 E}. \tag{6}$$

We now consider transmitting the same complex symbol s from one sensor to another by the digital mode using the same amount of total energy E. For digital transmission, there is an additional noise due to quantization. The variance of this quantization error using total B bits for both real part and imaginary part (each bounded between $[-W, W]$) is upper bounded as follows according to (2) and (4):

$$\sigma^2_{\text{digital}} \leq \frac{8W^2}{2^B} = \frac{8W^2}{\left(1+\phi\mu\frac{|h|^2 E}{LN_0}\right)^{\psi L}}. \tag{7}$$

We see that if $\psi L \leq 1$, the exact value of $\sigma^2_{\text{digital}}$ (which depends on the quantization method) can be larger than σ^2_{analog} even when the transmission energy E is large. But if $\psi L > 1$, the upper bound of $\sigma^2_{\text{digital}}$ decreases faster than σ^2_{analog} as E increases. Furthermore, since $\lim_{m\to\infty}(1+1/m)^m = e$, if L is large, we can write

$$\sigma^2_{\text{digital}} \leq 8W^2 \exp\left\{-\psi\phi\mu\frac{|h|^2 E}{N_0}\right\}. \tag{8}$$

Here, we see that the variance of the quantization error decreases exponentially as the transmission energy increases.

We conclude that as long as there is a sufficient transmission time-bandwidth product for each link and each observation sample, the digital transmission is more efficient in transmission energy than the analog transmission. The above analysis also suggests that otherwise the analog transmission can be more efficient than the digital transmission. It should be useful to note that an advantage of analog transmission over digital transmission, as advocated in [26] and [27], was based on a single-hop network, a fixed transmission time-bandwidth and an increasing number of sensors. Their results and ours are complementary to each other. The analog–digital comparison shown in [30] assumes the use of a single channel for digital transmission, and the authors overlooked the importance of subdivision of time-bandwidth product for digital transmission.

V. Progressive Estimation With Digital Transmissions

We will let K be the total number of sensors in the network, I_k be the number of the upper stream sensors of sensor k, and S_k be the set of the indexes of the nodes that are the upper stream sensors of sensor k. Each sensor is assumed to be connected to a destination node via a routing tree; see Fig. 1.

Since the estimation of the unknown vector $\boldsymbol{\theta}(n)$ is done for each sampling time instant n and its procedure is the same for all n, we will remove n for convenience. At each sensor k, the estimation of the unknown vector $\boldsymbol{\theta}$ is based on the local observation $\boldsymbol{x}_k$ and the quantized estimates $\boldsymbol{m}_i$, $i \in S_k$, from the upper stream sensors of sensor k.

A. BLUE

Suppose that $\boldsymbol{\theta}$ is an unknown deterministic vector, and $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{\theta} + \boldsymbol{n}$ is the available data vector, where $\boldsymbol{M}$ is a matrix of full column rank and $\boldsymbol{n}$ is a noise vector of zero mean and the covariance matrix $\boldsymbol{C}$. Then, the best linear unbiased estimate (BLUE) or equivalently the minimum variance unbiased linear estimate of $\boldsymbol{\theta}$ is known as $\hat{\boldsymbol{\theta}} = (\boldsymbol{M}\boldsymbol{C}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}\boldsymbol{C}^{-1}\boldsymbol{y}$, and the covariance matrix of this estimate is $(\boldsymbol{M}\boldsymbol{C}^{-1}\boldsymbol{M})^{-1}$, e.g., see [22].

We now let $\boldsymbol{m}_k$ be an unbiased quantized BLUE estimate at sensor k and $\hat{\boldsymbol{C}}_k$ be the covariance matrix of $\boldsymbol{m}_k$. If $\boldsymbol{m}_i$ and $\hat{\boldsymbol{C}}_i$ for $i \in S_k$ are available at sensor k, the BLUE of $\boldsymbol{\theta}$ at sensor k is given by

$$\hat{\boldsymbol{\theta}}_k = \left(\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{i\in S_k} \hat{\boldsymbol{C}}_i^{-1}\right)^{-1} \left(\boldsymbol{G}_k^H \boldsymbol{x}_k + \sum_{i\in S_k} \hat{\boldsymbol{C}}_i^{-1}\boldsymbol{m}_i\right) \quad (9)$$

and its covariance matrix is $\mathcal{E}\{(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^H\} = (\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{i\in S_k} \hat{\boldsymbol{C}}_i^{-1})^{-1}$. Here, $\mathcal{E}$ denotes expectation and the superscript ^{H} denotes complex conjugate transpose.

Assume that both the real and image part of the mth element of $\hat{\boldsymbol{\theta}}_k$ is bounded between $[-W_m, W_m]$ and quantized using total $B_{k,m}$ bits. (Note that it is meaningful to have a different range for a different element of $\hat{\boldsymbol{\theta}}_k$. This range should be governed by the prior knowledge of the physical nature and resolution requirement of each element of $\boldsymbol{\theta}_k$, which could be temperature, wavelength, distance, etc.) Then the covariance matrix $\hat{\boldsymbol{C}}_k$ of the quantized BLUE estimate $\boldsymbol{m}_k$ at sensor k is bounded as follows:

$$\hat{\boldsymbol{C}}_k \le \left(\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{i\in S_k} \hat{\boldsymbol{C}}_i^{-1}\right)^{-1} + \boldsymbol{C}_{q,k} \quad (10)$$

where the mth diagonal element of $\boldsymbol{C}_{q,k}$ is denoted by $\bar{\sigma}^2_{q,k,m}$ and given by $\bar{\sigma}^2_{q,k,m} = 8W_m^2/2^{B_{k,m}}$. However, except for the upper bound, the exact covariance matrix of the quantized estimate at each sensor is difficult to keep track of. Hence, the exact BLUE is not feasible for our application. Nevertheless, the above discussion is an important part of our systematic treatment.

B. Quasi BLUE

Since the exact $\hat{\boldsymbol{C}}_i$ for $i \in S_k$ are not available at sensor k, we now replace them by their upper bounds $\bar{\boldsymbol{C}}_i$ for $i \in S_k$. Using these upper bounds, the following estimate will be referred to as quasi-BLUE at sensor k:

$$\bar{\boldsymbol{\theta}}_k = \left(\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{i\in S_k} \bar{\boldsymbol{C}}_i^{-1}\right)^{-1} \left(\boldsymbol{G}_k^H \boldsymbol{x}_k + \sum_{i\in S_k} \bar{\boldsymbol{C}}_i^{-1}\boldsymbol{m}_i\right) \quad (11)$$

and the covariance matrix of this estimate is upper bounded by $(\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{i\in S_k} \bar{\boldsymbol{C}}_i^{-1})^{-1}$. Consequently, we can compute the upper bound $\bar{\boldsymbol{C}}_k$ of the covariance matrix of the quantized quasi-BLUE estimate $\boldsymbol{m}_k$ by

$$\bar{\boldsymbol{C}}_k = \left(\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{i\in S_k} \bar{\boldsymbol{C}}_i^{-1}\right)^{-1} + \boldsymbol{C}_{q,k}. \quad (12)$$

Our goal is to determine $B_{k,m}$ for all k and m such that the total transmission energy can be significantly reduced subject to a MSE constraint at the destination node. But according to the recursion of the covariance matrices (12) of the quasi-BLUE, the MSE at the destination is a very complicated function of $B_{k,m}$. Unless the network is very small, finding the optimal $B_{k,m}$ for quasi-BLUE is not feasible.

C. Averaging

There is a simpler method for progressive estimation, i.e., taking a simple average of the quantized estimates from upper stream sensors together with the BLUE estimate solely based on $\boldsymbol{x}_k$. This estimate at sensor k is given by:

$$\tilde{\theta}_k = \frac{1}{I_k+1}\left(\boldsymbol{G}_k^H \boldsymbol{G}_k\right)^{-1} \boldsymbol{G}_k^H \boldsymbol{x}_k + \frac{1}{I_k+1}\sum_{i\in S_k} \boldsymbol{m}_i. \quad (13)$$

If the mth element of $\tilde{\theta}_k$ is quantized with $B_{k,m}$ bits and the upper bounds of the covariance matrices of $\boldsymbol{m}_i$ are denoted by $\widetilde{\boldsymbol{C}}_i$, then the corresponding upper bound $\widetilde{\boldsymbol{C}}_k$ of the covariance matrix of the quantized estimate $\boldsymbol{m}_k$ at sensor k is given by

$$\widetilde{\boldsymbol{C}}_k = \frac{1}{(I_k+1)^2}\left(\boldsymbol{G}_k^H \boldsymbol{G}_k\right)^{-1} + \frac{1}{(I_k+1)^2}\sum_{i\in S_k} \widetilde{\boldsymbol{C}}_i + \boldsymbol{C}_{q,k}. \quad (14)$$

Note that the estimate given by (13) does not need the covariance matrices of the estimates from upper stream sensors. Also note that the recursion of the covariance matrices given by (14) is much simpler than that in (12). It follows from the following lemma that $\widetilde{\boldsymbol{C}}_k$ is simply an upper bound of $\bar{\boldsymbol{C}}_k$, i.e., $\widetilde{\boldsymbol{C}}_k \ge \bar{\boldsymbol{C}}_k$. But as implied by a simulation example shown later, this upper bound is quite loose generally.

Lemma 1: Given n positive definite matrices $\boldsymbol{A}_k$, $k = 1, \cdots, n$, we have

$$\left(\sum_{k=1}^{n} \boldsymbol{A}_k^{-1}\right)^{-1} \le \frac{1}{n^2}\left(\sum_{k=1}^{n} \boldsymbol{A}_k\right). \quad (15)$$

Proof: See Appendix II. ■

VI. Digital Transmission Energy Planning

In this section, we design an algorithm for computing the bit allocations among all sensors. To design such an algorithm, formulating a tractable optimization problem is essential.

The total number of bits used for the quantized estimate at sensor k is $\sum_{m=1}^{M} B_{k,m}$, where $B_{k,m}$ is the number of bits for the mth element of the estimated vector at sensor k. From the energy model (5), the transmission energy to be consumed by sensor k is given by

$$E_k = a_k \left(2^{\frac{1}{\psi L} \sum_{m=1}^{M} B_{k,m}} - 1 \right) \tag{16}$$

where $L = \lfloor \mathcal{TW} \rfloor \geq 1, a_k = N_0 L / \left(\phi \mu |h_k|^2 \right)$, and $|h_k|^2$ is the squared channel gain from sensor k to its down stream sensor. For any n positive real numbers b_i, it is known that $\prod_{i=1}^{n} b_i \leq \sum_{i=1}^{n} b_i^n / n$. It follows that

$$E_k < a_k 2^{\frac{1}{\psi L} \sum_{m=1}^{M} B_{k,m}} \leq \frac{a_k}{M} \sum_{m=1}^{M} 2^{\frac{M}{\psi L} B_{k,m}} = \sum_{m=1}^{M} \bar{E}_{k,m} \tag{17}$$

where $\bar{E}_{k,m} = (a_k/M) 2^{(M/(\psi L)) B_{k,m}}$, which can be viewed as the upper bound of the energy needed to transmit the mth element of the quantized estimation. We now define the following cost function:

$$J_p = \sum_{k=1}^{K-1} \sum_{m=1}^{M} \bar{E}_{k,m}^{p} = \sum_{k=1}^{K-1} \sum_{m=1}^{M} \left(\frac{a_k}{M} \right)^{p} 2^{\frac{pM}{\psi L} B_{k,m}} \tag{18}$$

which is the pth power of the L_p norm of $\bar{E}_{k,m}$ for all k and all m. Note that at the destination node which is labeled as node K, there is no need for quantization. Clearly, when $p = 1$, we have $J_1 = \sum_{k=1}^{K-1} \sum_{m=1}^{M} \bar{E}_{k,m}$, which is an upper bound on the total (sum) transmission energy.

To formulate the MSE constraint at the destination node, we use the linear recursion (14) of the covariance matrices. It follows from (14) that

$$\widetilde{\boldsymbol{C}}_K = \sum_{k=1}^{K} \frac{\alpha_k}{(I_k + 1)^2} \left(\boldsymbol{G}_k^H \boldsymbol{G}_k \right)^{-1} + \sum_{k=1}^{K-1} \alpha_k \boldsymbol{C}_{q,k} \tag{19}$$

where $\alpha_K = 1$, and $\alpha_i = \alpha_k / (I_k + 1)^2$ for $i \in S_k$. Note that in order to compute α_k for all k, one should start with the sensors nearest to the destination node and then proceed outwards recursively. The actual values of α_k depend on the routing tree. Then, the MSE at the destination node is given by

$$\mathrm{MSE}_K = \mathrm{tr}(\widetilde{\boldsymbol{C}}_K) = \xi + \sum_{k=1}^{K-1} \sum_{m=1}^{M} \alpha_k \frac{8 W_m^2}{2^{B_{k,m}}} \tag{20}$$

where $\xi = \sum_{k=1}^{K} (\alpha_k / (I_k + 1)^2) tr((\boldsymbol{G}_k^H \boldsymbol{G}_k)^{-1})$ which is invariant to $B_{k,m}$. If MSE_0 is the desired MSE value at the destination node, it is meaningful to set up the following constraint

$$\sum_{k=1}^{K-1} \sum_{m=1}^{M} \alpha_k \frac{W_m^2}{2^{B_{k,m}}} \leq \eta \tag{21}$$

where $\eta = (1/8)(\mathrm{MSE}_0 - \xi)$.

The problem now is to find $B_{k,m}$ to minimize J_p in (18) subject to (21). To solve this problem, we need the Hölder's inequality: For $x_i > 0, y_i > 0$ where $i = 1, 2, \cdots, n$, if $k > 1$, and $1/k + 1/k' = 1$, then,

$$\left(\sum_{i=1}^{n} x_i^k \right)^{\frac{1}{k}} \left(\sum_{i=1}^{n} y_i^{k'} \right)^{\frac{1}{k'}} \geq \sum_{i=1}^{n} x_i y_i \tag{22}$$

where the equality holds if $x_i^{k-1} = \lambda y_i$ for some constant λ and all i. The equality condition is easy to verify by recognizing $(k-1)k' = k$.

By defining $A_{k,m} = \left((a_k/M)^p 2^{(Mp/(\psi L)) B_{k,m}} \right)^{\psi L/(\psi L + Mp)}$ and $C_{k,m} = \left(\alpha_k W_m^2 / 2^{B_{k,m}} \right)^{Mp/(\psi L + Mp)}$, the Hölder's inequality implies

$$\begin{aligned} &\left\{ \sum_{k=1}^{K-1} \sum_{m=1}^{M} A_{k,m}^{\frac{\psi L + Mp}{\psi L}} \right\}^{\frac{\psi L}{\psi L + Mp}} \left\{ \sum_{k=1}^{K-1} \sum_{m=1}^{M} C_{k,m}^{\frac{\psi L + Mp}{Mp}} \right\}^{\frac{Mp}{\psi L + Mp}} \\ &\quad \geq \sum_{k=1}^{K-1} \sum_{m=1}^{M} A_{k,m} C_{k,m} \\ &\quad = \sum_{k=1}^{K-1} \sum_{m=1}^{M} (a_k/M)^{\frac{\psi L p}{\psi L + Mp}} \left(\alpha_k W_m^2 \right)^{\frac{Mp}{\psi L + Mp}} \\ &\quad \triangleq \gamma \end{aligned} \tag{23}$$

where the equality holds when

$$A_{k,m}^{\frac{Mp}{\psi L}} = \lambda C_{k,m}. \tag{24}$$

It is important to note that γ is independent of $B_{k,m}$. The constraint (21) is equivalent to $\sum_{k=1}^{K-1} \sum_{m=1}^{M} C_{k,m}^{(\psi L + Mp)/(Mp)} \leq \eta$. Then, using (18), (21), and (23), we have

$$J_p = \left\{ \sum_{k=1}^{K-1} \sum_{m=1}^{M} A_{k,m}^{\frac{\psi L + Mp}{\psi L}} \right\} \geq \frac{\gamma^{\frac{\psi L + Mp}{\psi L}}}{\eta^{\frac{Mp}{\psi L}}} \tag{25}$$

The right-hand side of (25) is independent of $B_{k,m}$ and is also the minimum of J_p. This minimum is achieved if (24) holds and the equality in (21) holds. But (24) along with the constraint $B_{k,m} \geq 0$ implies that

$$\begin{aligned} B_{k,m} = \frac{\psi L}{Mp\psi L + M^2 p^2} \Big(&(\psi L + Mp) \log_2 \lambda \\ &+ 2Mp \log_2 (\sqrt{\alpha_k} W_m) - Mp^2 \log_2 \frac{a_k}{M} \Big)^{+} \end{aligned} \tag{26}$$

where $(x)^+ = \max(0, x)$. Using (26) in the equality of (21) yields

$$\lambda = \frac{1}{\eta^{\frac{Mp}{\psi L}}} \left(\sum_{(k,m) \in S^+} (\sqrt{\alpha_k} W_m)^{\frac{2Mp}{\psi L + Mp}} \left(\frac{a_k}{M} \right)^{\frac{p\psi L}{\psi L + Mp}} \right)^{\frac{Mp}{\psi L}} \tag{27}$$

where $S^+ = \{(k,m) | B_{k,m} > 0\}$. The two equations (26) and (27) need to be computed iteratively until convergence. The iteration starts with the computation of λ from (27) with a full set

S^+ where $B_{k,m} > 0$ for all k and m. After convergence, each $B_{k,m}$ is rounded up into an integer.

We now show that the number of iterations between (26) and (27) until convergence is finite. For any given λ, the solution of $B_{k,m}$ from (26) is unique. If none of $B_{k,m}$ is zero after the first iteration, the convergence is achieved. If one or more of $B_{k,m}$ become zero, the set S^+ is reduced which in turn reduces λ via (27). From (26), we see that $B_{k,m}$ does not increase as λ decreases, which means that the previous zero $B_{k,m}$ will remain zero after another iteration. If there is no additional $B_{k,m}$ becoming zero at the end of an iteration, the size of S^+ is not changed and hence the convergence is achieved. Since the size of the initial S^+ is finite, so is the number of iterations required for convergence. It is obvious that the converged solution is invariant to the initial choices of $B_{k,m}$ as long as they are positive for all k and s.

It is also useful to note that (26) and (27) constitute a "water-filling" type algorithm. In other words, (26) can be compressed into the form $B_{k,m} = (l - s_{k,m})^+$ where l resembles a "water level" that is independent of the location parameters k and m in a "water tank," $s_{k,m}$ resembles the heights of the steps at the bottom of the water tank, and $B_{k,m}$ is the depth of the water at the location k and m.

The number L of the subchannels is of great importance to the energy planning. From (5) or (16), one can readily verify that the transmission energy E is a decreasing function of L. However, the effect of increasing L on energy saving diminishes at large L. Taking the limit $L \to \infty$, the energy model (16) becomes

$$E_k = \frac{N_0}{\phi\psi\mu|h_k|^2} \ln 2 \sum_{m=1}^{M} B_{k,m} \tag{28}$$

which is invariant to L. Furthermore, as $L \to \infty$, (26) and (27) can be rewritten as follows:

$$B_{k,m} = \left(\log_2 \lambda' + 2\log_2(\sqrt{\alpha_k} W_m) - p \log_2 \frac{a'_k}{M} \right)^+ \tag{29}$$

$$\lambda' = \frac{1}{\eta} \sum_{(k,m)\in S^+} \left(\frac{a'_k}{M} \right)^p \tag{30}$$

where $a'_k = N_0/\left(\phi\psi\mu|h_k|^2\right)$. Both (29) and (30) are independent of L. Therefore, the energy planning is virtually invariant to L when L is large. The simulation results shown later suggest that $L \geq 4M$ is large enough for the energy planning to become practically invariant to L.

Another important factor is the norm p used to formulate the cost function (18). Selecting $p = 1$ is to minimize the total energy of the network. Using a larger p implies that we want to penalize the larger energy terms $\bar{E}_{k,m}$. In the extreme case where $p = +\infty$, we minimize the maximum value among $\bar{E}_{k,m}$. By taking $p \to \infty$, (26) and (27) can be rewritten as follows:

$$B_{k,m} = \left(\log_2 \lambda'' - \frac{\psi L}{M} \log_2 \frac{a_k}{M} \right)^+ \tag{31}$$

$$\lambda'' = \frac{1}{\eta} \sum_{(k,m)\in S^+} \alpha_k W_m^2 \left(\frac{a_k}{M} \right)^{\frac{\psi L}{M}}. \tag{32}$$

It then follows that unless $B_{k,m} = 0$,

$$\bar{E}_{k,m} = (a_k/M) 2^{(M/(\psi L))B_{k,m}} = (\lambda'')^{M/(\psi L)}$$

which is independent of k and m.

VII. Progressive Estimation With Analog Transmission

We now consider progressive estimation with analog transmissions between sensors. Recall the analog transmission model discussed in Section III and Appendix I. We will assume M identical discrete-time analog subchannels between each sensor and its downstream sensor and the M subchannels are used to transmit in parallel the M elements of $\hat{\boldsymbol{\theta}}_k$ from sensor k to its downstream sensor. Applying (54) and (55), we can formulate an effective channel model for analog transmission between any two sensors as follows: the received symbol at the downstream counterpart of sensor k in the mth subchannel is

$$y_{k,m} = h_{k,m} \sqrt{\frac{E_{c_{k,m}}}{N_0}} \hat{\theta}_{k,m} + \nu_{k,m} \tag{33}$$

where $E_{c_{k,m}}$ is the energy of the waveform used in modulation for the mth element $\hat{\theta}_{k,m}$ of the estimate $\hat{\boldsymbol{\theta}}_k$ at sensor k, and the noise $\nu_{k,m}$ is a complex Gaussian random variable with zero mean and unit variance. The transmission energy for the mth element is $|\hat{\theta}_{k,m}|^2 E_{c_{k,m}}$. Then, by combining all M subchannels, we have the vector channel model

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{A}_k \hat{\boldsymbol{\theta}}_k + \boldsymbol{\nu}_k \tag{34}$$

where $\boldsymbol{A}_k = \mathrm{diag}[\sqrt{E_{c_{k,1}}/N_0}, \sqrt{E_{c_{k,2}}/N_0}, \cdots, \sqrt{E_{c_{k,M}}/N_0}]$ and $\boldsymbol{H}_k = \mathrm{diag}[h_{k,1}, h_{k,2}, \cdots, h_{k,M}]$. For the analog transmission, we allow the channel gains of the subchannels to be possibly different from each other. Denoted by P_k, the total transmission energy for transmitting $\hat{\boldsymbol{\theta}}_k$ from sensor k to its downstream sensor using the analog mode is now $P_k = N_0 Tr\{\boldsymbol{A}_k \hat{\boldsymbol{\theta}}_k \hat{\boldsymbol{\theta}}_k^H \boldsymbol{A}_k\}$. The energy planning for analog transmission is about the design of $\boldsymbol{A}_k$ for all k.

A. BLUE

Based on the local observation $\boldsymbol{x}_k$ at sensor k and the data $\boldsymbol{y}_l$ for $l \in S_k$ received by sensor k from its upstream sensors, the BLUE of $\boldsymbol{\theta}$ at sensor k is

$$\hat{\boldsymbol{\theta}}_k = \left(\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{l\in S_k} \boldsymbol{B}_l^H \boldsymbol{C}_{\boldsymbol{y}_l}^{-1} \boldsymbol{B}_l \right)^{-1} \times \left(\boldsymbol{G}_k^H \boldsymbol{x}_k + \sum_{l\in S_k} \boldsymbol{B}_l^H \boldsymbol{C}_{\boldsymbol{y}_l}^{-1} \boldsymbol{y}_l \right) \tag{35}$$

where $\boldsymbol{B}_l = \boldsymbol{H}_l \boldsymbol{A}_l$, $\boldsymbol{C}_{\boldsymbol{y}_l} = \boldsymbol{B}_l \boldsymbol{C}_l \boldsymbol{B}_l^H + \boldsymbol{I}$, and $\boldsymbol{C}_l$ is the covariance of $\hat{\boldsymbol{\theta}}_l$. The covariance matrix of $\hat{\boldsymbol{\theta}}_k$ is given by

$$\boldsymbol{C}_k = \left(\boldsymbol{G}_k^H \boldsymbol{G}_k + \sum_{l\in S_k} \boldsymbol{B}_l^H \left(\boldsymbol{B}_l \boldsymbol{C}_l \boldsymbol{B}_l^H + \boldsymbol{I} \right)^{-1} \boldsymbol{B}_l \right)^{-1}. \tag{36}$$

Based on (36), the MSE at the destination node is a very complicated function of the amplification matrices $\boldsymbol{A}_k, k = 1, \cdots, K$. It is not feasible to use (36) to design energy planning unless the network is very small.

B. Averaging

Alternatively, at sensor k, we can obtain an estimate of $\boldsymbol{\theta}_k$ by averaging the BLUEs of $\boldsymbol{\theta}$ based on $\boldsymbol{y}_l$ for $l \in S_k$ and $\boldsymbol{x}_k$ individually, i.e.,

$$\tilde{\boldsymbol{\theta}}_k = \frac{1}{I_k+1}\left(\boldsymbol{G}_k^H\boldsymbol{G}_k\right)^{-1}\boldsymbol{G}_k^H\boldsymbol{x}_k + \frac{1}{I_k+1}\sum_{l\in S_k}\boldsymbol{B}_l^{-1}\boldsymbol{y}_l \quad (37)$$

where $\boldsymbol{B}_l$ for $l \in S_k$ are needed at sensor k. Denote the covariance matrix of $\tilde{\boldsymbol{\theta}}_k$ by $\tilde{\boldsymbol{C}}_k$. Then, the covariance matrix of $\boldsymbol{y}_l$ is $\boldsymbol{B}_l\tilde{\boldsymbol{C}}_l\boldsymbol{B}_l^H + \boldsymbol{I}$, and the covariance matrix of $\tilde{\boldsymbol{\theta}}_k$ is given by

$$\tilde{\boldsymbol{C}}_k = \frac{1}{(1+I_k)^2}\left(\left(\boldsymbol{G}_k^H\boldsymbol{G}_k\right)^{-1} + \sum_{l\in S_k}\left(\boldsymbol{B}_l^H\boldsymbol{B}_l\right)^{-1} + \sum_{l\in S_k}\tilde{\boldsymbol{C}}_l\right). \quad (38)$$

The recursion (38) is linear and much simpler than the recursion (36). We will use (38) for energy planning. It follows from Lemma 1 that $\tilde{\boldsymbol{C}}_k \geq \boldsymbol{C}_k$.

VIII. Analog Transmission Energy Planning

Using the recursion (38), the covariance matrix at the destination node, labeled as node K, can be found as

$$\tilde{\boldsymbol{C}}_K = \sum_{k=1}^{K-1}\alpha_k\left(\boldsymbol{B}_k^H\boldsymbol{B}_k\right)^{-1} + \sum_{k=1}^{K}\frac{\alpha_k}{(1+I_k)^2}\left(\boldsymbol{G}_k^H\boldsymbol{G}_k\right)^{-1} \quad (39)$$

where $\alpha_K = 1$ and $\alpha_l = \alpha_k/(1+I_k)^2$ for $l \in S_k$. Then, the MSE at the destination node is

$$\mathrm{MSE}_K = \mathrm{Tr}(\tilde{\boldsymbol{C}}_K) = \sum_{k=1}^{K-1}\sum_{m=1}^{M}\frac{\beta_{k,m}^2}{a_{k,m}^2} + \epsilon \quad (40)$$

where $\epsilon = Tr(\sum_{k=1}^{K}(\alpha_k/(1+I_k)^2)(\boldsymbol{G}_k^H\boldsymbol{G}_k)^{-1})$ and $\beta_{k,m} = \sqrt{\alpha_k}/|h_{k,m}|$. Here, $h_{k,m}$ is the mth diagonal element of $\boldsymbol{H}_k$, and $a_{k,m} = \sqrt{E_{c_{k,m}}/N_0}$ is the mth diagonal element of $\boldsymbol{A}_k$. Then it follows that

$$P_k = N_0\mathrm{Tr}\left(\boldsymbol{A}_k\tilde{\boldsymbol{\theta}}_k\tilde{\boldsymbol{\theta}}_k^H\boldsymbol{A}_k\right) \leq \sum_{m=1}^{M}N_0a_{k,m}^2 2W_m^2 = \sum_{m=1}^{M}\bar{P}_{k,m} \quad (41)$$

where $\bar{P}_{k,m} = N_0a_{k,m}^2 2W_m^2$, which can be viewed as the upper bound of the energy used to transmit the mth element of $\tilde{\boldsymbol{\theta}}_k$. The inequality comes with the assumption that both the real and image part of $\hat{\theta}_{k,m}$ or $\tilde{\theta}_{k,m}$ are bounded within $[-W_m, W_m]$, which is consistent with the digital transmission case.

Similar to the digital transmission energy planning, we define the following cost function for the analog transmission energy planning:

$$F_p = \sum_{k=1}^{K-1}\sum_{m=1}^{M}\bar{P}_{k,m}^p = \sum_{k=1}^{K-1}\sum_{m=1}^{M}a_{k,m}^{2p}(\sqrt{2N_0}W_m)^{2p} \quad (42)$$

which is the pth power of the L_p norm of $\bar{P}_{k,m}$ over all k and m. And the analog transmission energy planning is formulated as follows:

$$\min_{a_{k,m}} F_p \quad (43)$$

subject to the MSE constraint at node K:

$$\sum_{k=1}^{K-1}\sum_{m=1}^{M}\frac{\beta_{k,m}^2}{a_{k,m}^2} \leq \mathrm{MSE}_0 - \epsilon \triangleq \eta \quad (44)$$

We now define $A_{k,m} = (a_{k,m}^{2p}(\sqrt{2N_0}W_m)^{2p})^{1/(p+1)}$ and $C_{k,m} = (\beta_{k,m}^2/a_{k,m}^2)^{p/(p+1)}$. Following the Hölder's inequality, we have

$$\begin{aligned}\left(\sum_{k=1}^{K-1}\sum_{m=1}^{M}A_{k,m}^{(p+1)}\right)^{1/(p+1)}&\left(\sum_{k=1}^{K-1}\sum_{m=1}^{M}C_{k,m}^{(p+1)/p}\right)^{p/(p+1)}\\ &\geq \sum_{k=1}^{K-1}\sum_{m=1}^{M}A_{k,m}C_{k,m}\\ &= (\sqrt{2N_0}W_m)^{2p/(p+1)}\beta_{k,m}^{2p/(p+1)}\\ &\triangleq \gamma \end{aligned} \quad (45)$$

where equality holds when there is a constant λ such that

$$A_{k,m}^p = \lambda C_{k,m}. \quad (46)$$

Combining (44) and (45), we have

$$F_p = \left(\sum_{k=1}^{K-1}\sum_{m=1}^{M}A_{k,m}^{(p+1)}\right) \geq \frac{\gamma^{p+1}}{\eta^p}. \quad (47)$$

The right-hand-side of (47) is independent of $a_{k,m}$ and is the minimum of F_p. This minimum is achieved if the equality in (44) is achieved and the equality (46) holds. From (46), one can easily verify that

$$a_{k,m} = \lambda^{\frac{1}{2p}}\beta_{k,m}^{\frac{1}{p+1}}(\sqrt{2N_0}W_m)^{-\frac{p}{p+1}}. \quad (48)$$

Applying this to the equality in (44) yields

$$\lambda = \left(\frac{\sum_{k=1}^{K-1}\sum_{m=1}^{M}(\sqrt{2N_0}W_m)^{\frac{2p}{p+1}}\beta_{k,m}^{\frac{2p}{p+1}}}{\eta}\right)^p. \quad (49)$$

Combing (48) and (49), we finally get

$$a_{k,m} = \left(\beta_{k,m}^{\frac{1}{p+1}}(\sqrt{2N_0}W_m)^{-\frac{p}{p+1}}\right) \times\sqrt{\frac{\sum_{k=1}^{K-1}\sum_{m=1}^{M}(\sqrt{2N_0}W_m)^{\frac{2p}{p+1}}\beta_{k,m}^{\frac{2p}{p+1}}}{\eta}}. \quad (50)$$

Similar to the digital case, the energy planning here is based on the L_p norm for any p. If $p \to \infty$, (50) can be rewritten as follows:

$$a_{k,m} = (\sqrt{2N_0}W_m)^{-1}\sqrt{\lambda'''} \quad (51)$$

where $\lambda''' = \sum_{k=1}^{K-1}\sum_{m=1}^{M}(\sqrt{2N_0}W_m)^2\beta_{k,m}^2/\eta$, and hence $\bar{P}_{k,m} = N_0 a_{k,m}^2 2W_m^2 = \lambda'''$, which is independent of k and m.

IX. Simulation Results

The network we consider is shown in Fig. 1 where there are $K = 400$ nodes. The destination node is also referred to as fusion center. This network was constructed in such a way that the distance between a sensor and its upper stream sensor is $D\delta$ where δ is uniformly distributed within the range [0.5, 1.5] and D is a normalizing factor.

Recall that for digital transmissions, the energy consumed by sensor k is denoted by E_k in (16) and the p-norm cost function is denoted by J_p in (18). And for analog transmissions, the energy consumed by sensor k is denoted by P_k in (41) and the p-norm cost function is denoted by F_p in (42).

In the simulation, we choose $\mu = 1$, $\phi = 1$, $\psi = 1$, and $N_0 = 1$. We also choose $N = 20$, $M = 10$, $W_m = 1$ for $m = 1, 2, \cdots, M$. We choose $\boldsymbol{G}_k$ for $k = 1, 2, \cdots, K$ randomly , and each $\boldsymbol{G}_k$ is a 20×10 matrix with elements randomly chosen from a Gaussian distribution with zero mean and variance equal to 10. Each entry of $\boldsymbol{\theta}$, both the real and image parts, is chosen randomly from [−1,1]. The squared channel gain from sensor k to its downstream sensor is $|h_k|^2 = d_k^{-\alpha}\rho_k$, where d_k is the distance from sensor k to its down stream sensor, $\alpha = 4$ and ρ_k is randomly chosen from an exponential distribution with mean equal to one.

A. Analog Transmission

The scheme in Section VIII will be referred to as proposed progressive (PP) energy planning scheme for multihop sensor network with analog transmission. For comparison, we also consider a uniform progressive (UP) energy planning scheme for the same network. For the UP scheme, each sensor uses the same transmission matrix $\boldsymbol{A}_k = a\boldsymbol{I}$ where a is chosen to achieve the target MSE at the destination node, i.e., $\text{MSE}_K \leq \text{MSE}_0$. In simulations, once the transmission matrices $\boldsymbol{A}_k$, $k = 1, \cdots, K$, are determined, we use the second term in (41) to calculate the transmission energy at sensor k.

Using the cost F_1, Fig. 2 illustrates the total transmission energy consumed by the entire network (averaged over 100 realizations of $\boldsymbol{\theta}$) versus the target MSE denoted by MSE_0. We see that the PP scheme consumes much less total energy than the UP scheme. We also see that for either the UP scheme or the PP scheme, the BLUE estimation algorithm consumes less energy than the averaging estimation algorithm. To illustrate the advantage of the multi-hop network over the single-hop network in terms of the transmission energy consumption, Fig. 2 also shows a curve of the energy versus the target MSE based on a single-hop tree for the same distribution of sensors shown in Fig. 1. We see an enormously large gap of energy between the single-hop tree and the multi-hop tree. The path loss exponent is assumed to be four.

For $\text{MSE}_0 = 0.05$, Fig. 3 illustrates the transmission energy consumed by each sensor versus the normalized distance (i.e., the distance divided by D) of the sensor from the destination node. A single random realization of $\boldsymbol{\theta}$ is used. For this figure, the PP scheme uses F_1, and the sensors close to the fusion center

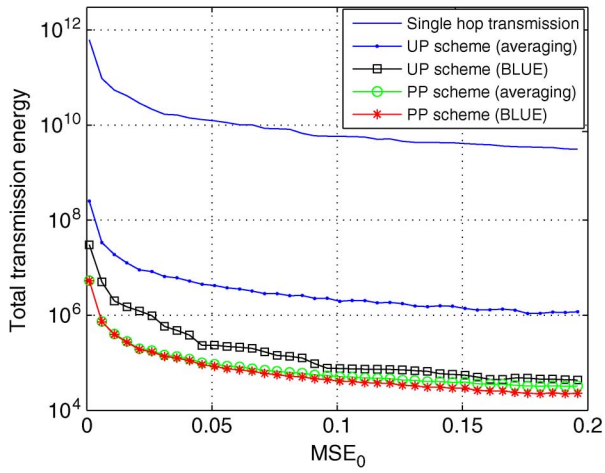


Fig. 2. For the analog PP and UP schemes, the transmission energy consumed by the network versus the target MSE_0. Note that the first curve on the top is based on a single-hop tree for the same distribution of sensors shown in Fig. 1. The path loss exponent is assumed to be four.

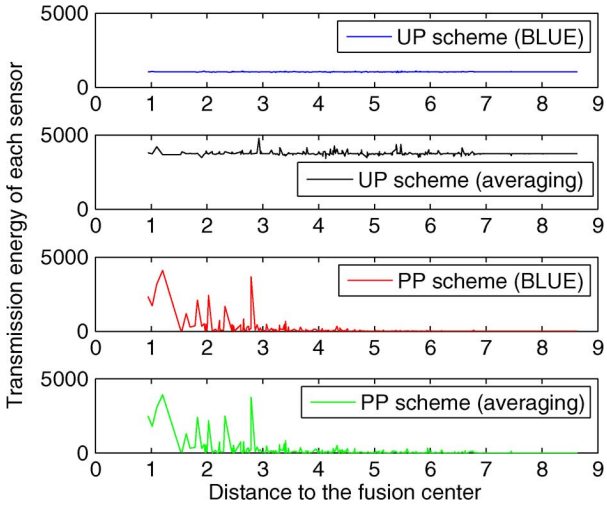


Fig. 3. For the analog PP and UP schemes, the transmission energy consumed by each sensor versus the normalized Euclidean distance between the sensor and the fusion center. $\text{MSE}_0 = 0.05$. The analog PP scheme uses F_1.

needs to transmit much more energy than the sensors far away from the fusion center. But if we choose a large p (say, $p = 50$) for F_p, the PP scheme can yield a virtually constant energy distribution and become virtually the same as the UP scheme.

Fig. 4 shows the effect of p on the analog transmission energy planning. We can see that as p in the cost F_p increases, the total transmission energy (corresponding to F_1) increases. We also see as expected that when $p = \infty$, the PP scheme becomes identical to the UP scheme. See (51).

Fig. 5 illustrates the actual MSE MSE_K (averaged over 100 realizations of $\boldsymbol{\theta}$) at the destination node versus MSE_0. For the PP scheme (using F_1), the averaging algorithm always achieves the target MSE at the destination, but the BLUE algorithm yields a much smaller MSE than the target MSE. This is because the covariance matrix of the estimate by averaging is not a tight upper bound of the covariance matrix of the estimate by the BLUE. In other words, the energy planning algorithm based on

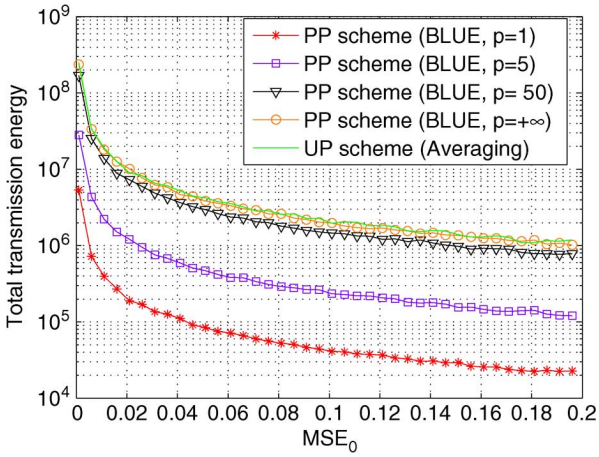


Fig. 4. Illustration of the effect of p in F_p for the analog PP scheme.

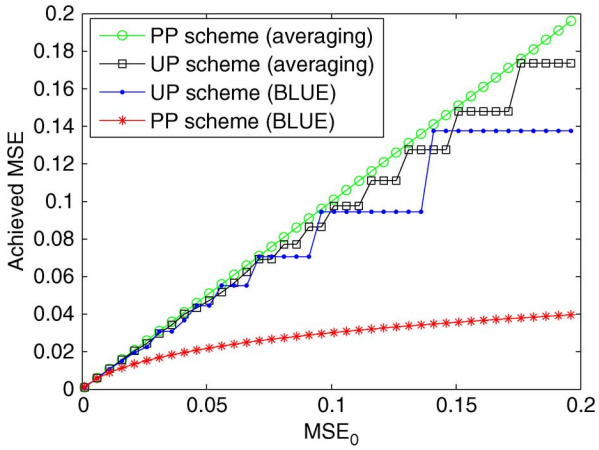


Fig. 5. For the analog PP and UP schemes, the actual MSE at the fusion center versus the target MSE.

the averaging progressive estimation is rather conservative for the BLUE progressive estimation.

B. Digital Transmission

We will compare the proposed progressive (PP) energy planning scheme shown in Section VI and a uniform progressive (UP) energy planning scheme. The UP scheme here is such that the number of bits allocated to each element of the estimate by each sensor is a constant subject to the MSE constraint $\text{MSE}_K \leq \text{MSE}_0$ at the destination node.

For $L = M$, Fig. 6 illustrates the total normalized transmission energy consumed by the network, i.e., $E_{\text{net}} = \sum_{k=1}^{K} E_k/D^{\alpha}$, versus the target MSE. The energy E_k is computed based on (16). We see that the PP scheme with small p consumes much less energy than the UP scheme. But with a large p, the result of the PP scheme becomes similar to, but not exactly the same as, that of the UP scheme.

Under $\text{MSE}_0 = 1.0 \times 10^{-3}$, $L = M$ and J_1, Fig. 7 illustrates the number of bits per element for each sensor, i.e., $\sum_{m=1}^{M} B_{k,m}/M$, versus the Euclidean distance (divided by D) from the sensor to the destination node. We see that under the PP scheme with J_1, the number of bits allocated to each sensor

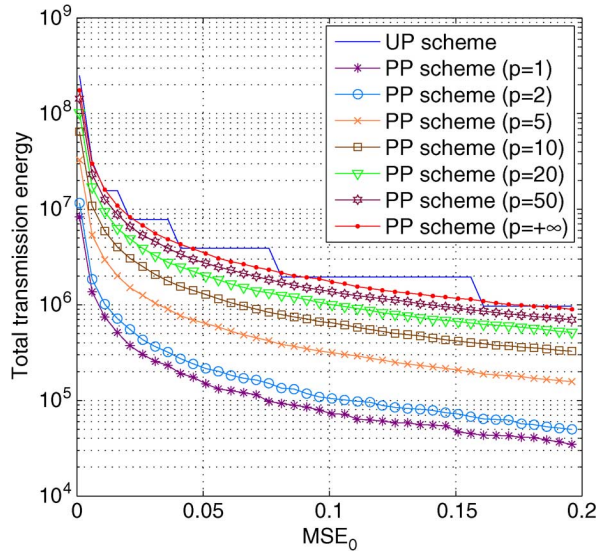


Fig. 6. For the digital PP and UP schemes, the transmission energy consumed by the network versus the target MSE where p is as in J_p for the digital PP scheme.

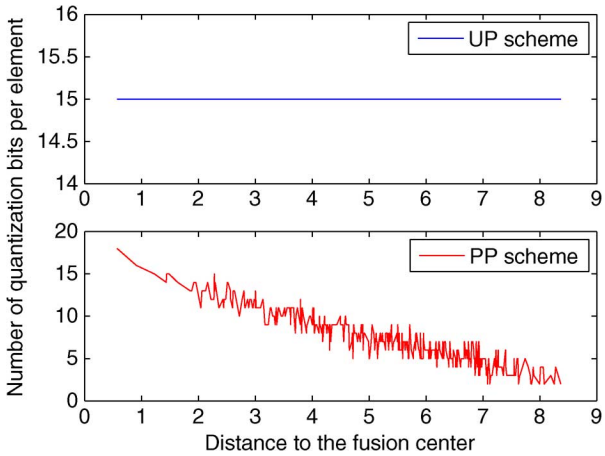


Fig. 7. The number of quantization bits allocated for each sensor per element of the unknown vector versus the normalized Euclidean distance from the sensor to the destination node. The target MSE at the destination node is $\text{MSE}_0 = 1.0 \times 10^{-3}$. The digital PP scheme uses J_1.

generally decreases with the distance from the sensor to the destination node.

Under the same condition as for Fig. 7, Fig. 8 shows the amount of the normalized transmission energy E_k/D^r consumed by each sensor versus the distance (divided by D) from the sensor to the destination node. Note that unlike the analog case, the transmission energy determined for each sensor for the digital case is not affected by the estimation algorithm (quasi BLUE or averaging).

Under $\text{MSE}_0 = 1.0 \times 10^{-3}$ and $L = M$, Fig. 9 shows the effect of a large p on the bit distribution of the digital PP scheme. And Fig. 10 shows the effect of a large p on the energy distribution of the digital PP scheme. We see that the digital PP scheme with a large p is more effective than the digital UP scheme to achieve a constant energy distribution. The digital UP scheme

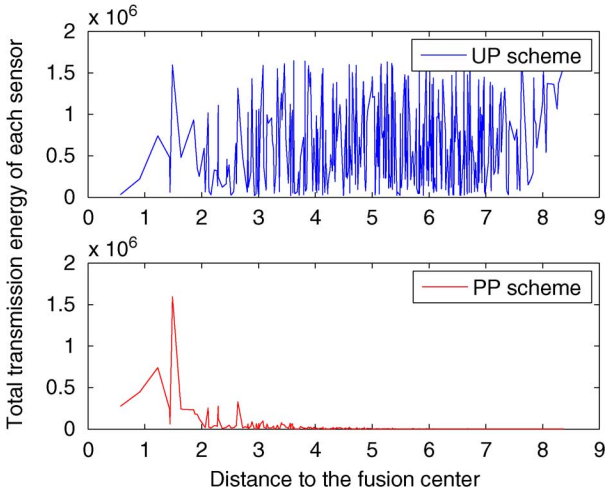


Fig. 8. Normalized digital transmission energy by each sensor versus the normalized Euclidean distance from the sensor to the destination node. $\text{MSE}_0 = 1.0 \times 10^{-3}$. The digital PP scheme uses J_1.

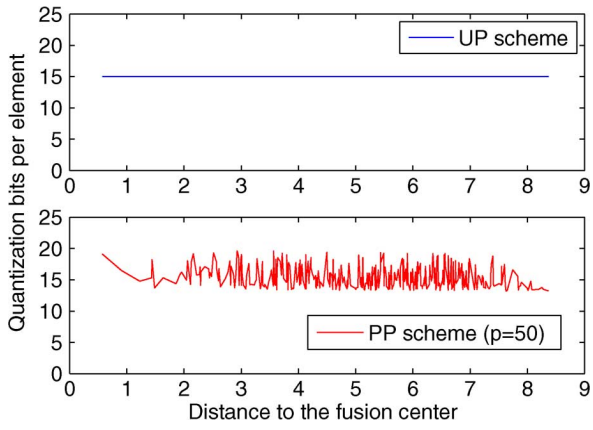


Fig. 9. Same as Fig. 7 except $p = 50$ for the digital PP scheme.

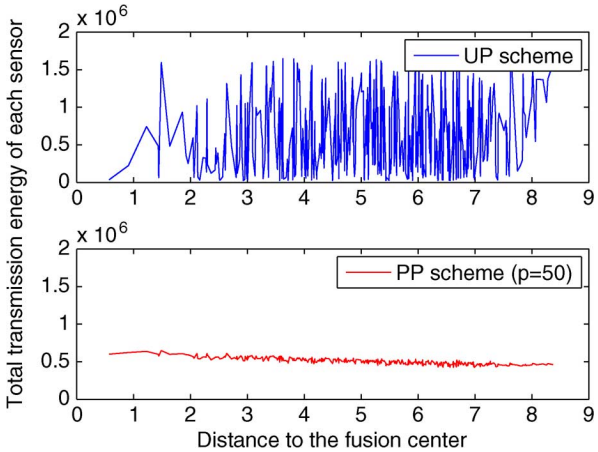


Fig. 10. Same as Fig. 8 except $p = 50$ for the digital PP scheme.

can not achieve a constant energy distribution because of the varying channel gains from sensor to sensor.

We now evaluate the effect of the transmission time-bandwidth product L. Under $\text{MSE}_0 = 1.0 \times 10^{-3}$, Fig. 11 shows the total transmission energy of the network versus the ratio L/M. We see that as L increases, less energy is consumed by any of

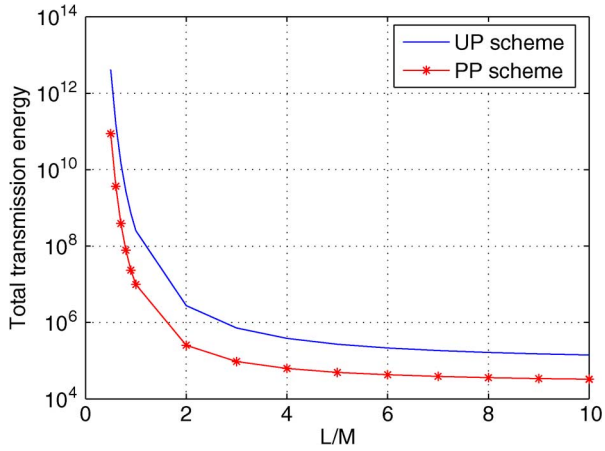


Fig. 11. Total digital transmission energy consumed by the network versus L/M with $M = 10$. $MSE_0 = 1.0 \times 10^{-3}$.

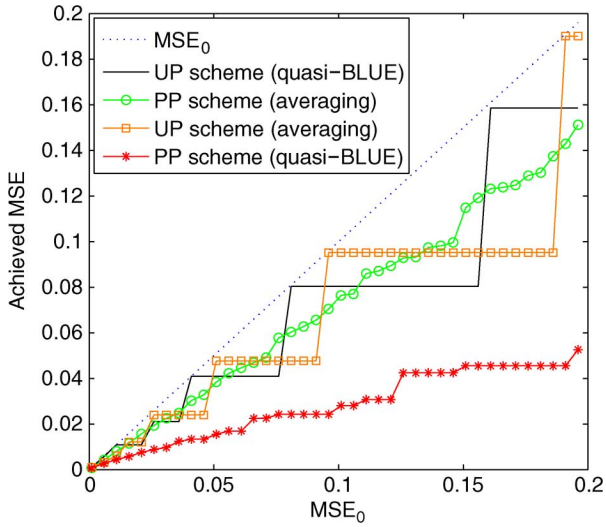


Fig. 12. For the digital PP and UP schemes, the actual MSE value at the destination node versus the target MSE MSE_0.

the two schemes (PP or UP). This figure suggests that if the network has a large bandwidth, we should use a large L to save transmission energy. However, we also see from this figure that the required transmission energy has a nonzero lower bound as L becomes very large, which is proved by (28). For the example shown in Fig. 11, $L = 4M$ is practically large enough to approach the bound.

Fig. 12 shows the actual MSE value at the destination node versus the target MSE value at the destination node. Such a curve depends on both the energy planning algorithm and the estimation algorithm. When the averaging algorithm is used, the actual MSE is quite close to the target MSE, which is expected. However, when the quasi-BLUE algorithm is used, the actual MSE is much smaller than the target MSE. One should recall that the quasi-BLUE algorithm requires each sensor to know the upper bounds of the covariance matrices of the quantized estimates from its upper stream sensors while the averaging algorithm does not have this requirement. Yet, given the large gap of the MSE between the two estimation algorithms as shown in this figure, developing a more efficient bit allocation algorithm

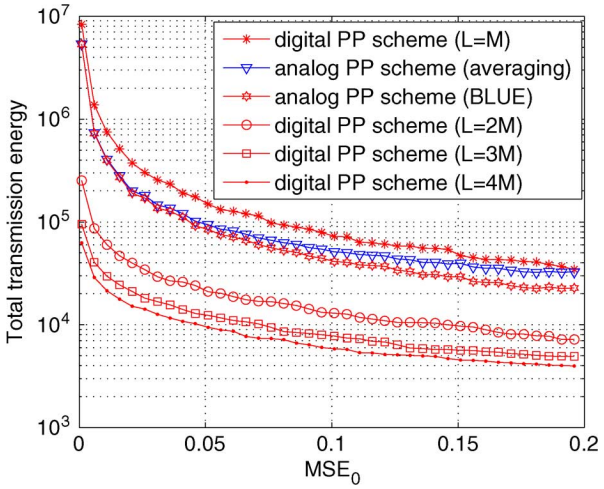


Fig. 13. Comparison between analog transmissions and digital transmissions: total transmission energy versus the target MSE.

for the quasi-BLUE estimation algorithm remains a useful challenge.

C. Analog Versus Digital Transmissions

To compare the analog case with the digital case, we consider the same network topology (Fig. 1), the same observation model (1) and the same RF channel with the same $\mathcal{W}$ the same $\mathcal{T}$. Fig. 13 compares the energy consumptions between the analog PP scheme and the digital PP scheme. We see that when $L = M$, the analog transmissions (either with BLUE or averaging algorithm) consume less energy than the digital transmissions. However, when $L \geq 2M$, the digital transmissions consume less energy than the analog transmissions, which is essentially the same conclusion from Section IV where $M = 1$ was considered. The energy saving by increasing L becomes less significant when L is large, which is consistent with Fig. 11.

X. Conclusion

We have developed a digital transmission energy planning algorithm and an analog transmission energy planning algorithm for progressive estimation in multihop sensor network with routing tree. The routing tree finding and the transmission energy planning are conducted at the startup of the network or once the network condition changes. The network condition (such as topology and channel state information) is assumed to be constant during the time of interest for estimating and tracking spatially invariant parameters. These transmission energy planning algorithms guarantee any pre-specified estimation performance at the destination node. And at the same time, they significantly reduce the required transmission energy for the entire network. Unlike many other consensus-type algorithms, the proposed progressive estimation algorithms along with their energy planning algorithms yield any desired result within a finite time although they require an operational overhead at the startup. The energy planing algorithms shown in this paper provide an optimal energy planning for the proposed progressive estimation algorithms based on averaging. For algorithms based on BLUE or quasi-BLUE, the result of our energy planning algorithms show a rather conservative gap. Whether or not this gap can be narrowed by a more clever energy planning algorithm remains a future research topic.

In practice, any sensor in the network can be a destination node, and there could be multiple destination nodes in the network. The theory and technique shown in this paper are applicable to any given destination node and its associated routing tree. We have seen that most of the energy should be distributed relatively near the destination node (unless a large p is used in the L_p norm). This fact should be taken into account when a routing tree is searched for in a large network. Finally, we note that the proposed algorithms are readily applicable to any single-hop network (as a special case) where each sensor transmits data directly to the destination node. In this special case, no routing tree is needed.

Appendix I
Pulse Modulation

Assume a complex baseband channel between two sensors where the input $x(t)$ and the output $y(t)$ are related as follows:

$$y(t) = hx(t) + v(t) \tag{52}$$

where h is a complex channel coefficient, $v(t)$ is complex Gaussian noise with the energy spectral density function

$$S_v(f) = \begin{cases} N_0, & |f| \leq \mathcal{W}/2 \\ 0, & \text{otherwise.} \end{cases} \tag{53}$$

To transmit a complex symbol s over this channel, we can use the pulse amplitude modulation $x(t) = sc(t)$ where the waveform $c(t)$ has the duration $[0, \mathcal{T}]$ and the double-sided bandwidth $\mathcal{W}$. Then, the output of the channel is $y(t) = hsc(t) + v(t)$. To obtain the maximum likelihood estimate of s based on $y(t)$, we can first compute the sufficient statistics $\bar{y} = \int_0^{\mathcal{T}} y(t)c(t)dt$ which is also known as matched filtering. It follows that

$$\bar{y} = hE_c s + \bar{v} \tag{54}$$

where E_c is the energy of the waveform $c(t)$ and $\bar{v} = \int_0^{\mathcal{T}} v(t)c(t)dt$ is a complex Gaussian random variable with zero mean and the variance

$$\sigma_v^2 = \mathcal{E}\left\{|\bar{v}|^2\right\} = \int_0^{\mathcal{T}} \int_0^{\mathcal{T}} R_v(\tau - t)c(\tau)c(t)d\tau dt \tag{55}$$

where $\mathcal{E}$ denotes expectation, $R_v(\tau)$ is the autocorrelation function of $v(t)$. Since $R_v(\tau) = \int_{-\mathcal{W}/2}^{\mathcal{W}/2} N_0 \exp\{2\pi\tau f\}df$, it follows from (55) that

$$\sigma_v^2 = N_0 \int_{-\mathcal{W}/2}^{\mathcal{W}/2} |C(f)|^2 \, df = N_0 E_c \tag{56}$$

where $C(f)$ is the Fourier transform of $c(t)$. Then, the maximum likelihood estimate of s is $\hat{s} = \bar{y}/(hE_c)$ which has the signal-to-noise ratio

$$\text{SNR} \doteq \frac{|s|^2}{\mathcal{E}\left\{|\hat{s} - s|^2\right\}} = \frac{|s|^2|h|^2E_c^2}{\sigma_v^2} = \frac{|s|^2|h|^2E_c}{N_0} = \frac{|h|^2E}{N_0} \tag{57}$$

where $E = |s|^2 E_c$ is the total transmitted energy, i.e., the energy of $x(t)$ within $[0, \mathcal{T}]$.

APPENDIX II
PROOF OF LEMMA 1

Proof: We first prove that

$$\left(\sum_{k=1}^{2} \boldsymbol{A}_k^{-1}\right)^{-1} \leq \frac{1}{4}\left(\sum_{k=1}^{2} \boldsymbol{A}_k\right). \tag{58}$$

Denote the eigenvalue decomposition $\boldsymbol{A}_1^{H/2}\boldsymbol{A}_2^{-1}\boldsymbol{A}_1^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$. Then, (58) is equivalent to each of the following inequalities:

$$\boldsymbol{A}_1^{-1/2}\left(\boldsymbol{A}_1^{-1}+\boldsymbol{A}_2^{-1}\right)^{-1}\boldsymbol{A}_1^{-H/2} \leq \boldsymbol{A}_1^{-1/2}\left(\frac{\boldsymbol{A}_1+\boldsymbol{A}_2}{4}\right)\boldsymbol{A}_1^{-H/2} \tag{59}$$

$$\left(\boldsymbol{I}+\boldsymbol{A}_1^{H/2}\boldsymbol{A}_2^{-1}\boldsymbol{A}_1^{1/2}\right)^{-1} \leq \frac{1}{4}\left(\boldsymbol{I}+\boldsymbol{A}_1^{-1/2}\boldsymbol{A}_2\boldsymbol{A}_1^{-H/2}\right) \tag{60}$$

$$\boldsymbol{U}(\boldsymbol{I}+\boldsymbol{\Lambda})^{-1}\boldsymbol{U}^H \leq \frac{1}{4}\boldsymbol{U}(\boldsymbol{I}+\boldsymbol{\Lambda}^{-1})\boldsymbol{U}^H \tag{61}$$

$$(\boldsymbol{I}+\boldsymbol{\Lambda})^{-1} \leq \frac{1}{4}(\boldsymbol{I}+\boldsymbol{\Lambda}^{-1}) \tag{62}$$

$$\frac{1}{1+\lambda_i} \leq \frac{1}{4}\left(1+\frac{1}{\lambda_i}\right), \quad \forall i \tag{63}$$

where $\lambda_i > 0$ is the ith element of the diagonal matrix $\boldsymbol{\Lambda}$. It is easy to verify that (63) is equivalent to $(\lambda_i - 1)^2 \geq 0$ which holds always.

We now assume that Lemma 1 holds for $n = m \geq 2$, i.e.,

$$\left(\sum_{k=1}^{m} \boldsymbol{A}_k\right)\left(\sum_{k=1}^{m} \boldsymbol{A}_k^{-1}\right) \geq m^2\boldsymbol{I}. \tag{64}$$

Then, we can write

$$\begin{aligned}
&\left(\sum_{k=1}^{m} \boldsymbol{A}_k + \boldsymbol{A}_{m+1}\right)\left(\sum_{k=1}^{m} \boldsymbol{A}_k^{-1} + \boldsymbol{A}_{m+1}^{-1}\right)\\
&= \left(\sum_{k=1}^{m} \boldsymbol{A}_k\right)\left(\sum_{k=1}^{m} \boldsymbol{A}_k^{-1}\right)\\
&\quad + \left(\sum_{k=1}^{m} \boldsymbol{A}_k\right)\boldsymbol{A}_{m+1}^{-1} + \boldsymbol{A}_{m+1}\left(\sum_{k=1}^{m} \boldsymbol{A}_k^{-1}\right) + \boldsymbol{I}\\
&\geq m^2\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D} + \boldsymbol{I}
\end{aligned} \tag{65}$$

where $\boldsymbol{C} = (\sum_{k=1}^{m} \boldsymbol{A}_k)\boldsymbol{A}_{m+1}^{-1}$ and $\boldsymbol{D} = \boldsymbol{A}_{m+1}(\sum_{k=1}^{m} \boldsymbol{A}_k^{-1})$. Then, it follows that

$$\begin{aligned}
&\boldsymbol{C} + \boldsymbol{D} + 2m\boldsymbol{I}\\
&= \left(\sum_{k=1}^{m}\left(\boldsymbol{A}_k\boldsymbol{A}_{m+1}^{-1} + \boldsymbol{A}_{m+1}\boldsymbol{A}_k^{-1}\right)\right) + 2m\boldsymbol{I}\\
&= \sum_{k=1}^{m}\left(\boldsymbol{A}_k\boldsymbol{A}_{m+1}^{-1} + \boldsymbol{A}_{m+1}\boldsymbol{A}_k^{-1} + 2\boldsymbol{I}\right)\\
&= \sum_{k=1}^{m}\left(\boldsymbol{A}_k\boldsymbol{A}_{m+1}^{-1} + \boldsymbol{A}_{m+1}\boldsymbol{A}_k^{-1} + \boldsymbol{A}_k\boldsymbol{A}_k^{-1} + \boldsymbol{A}_{m+1}\boldsymbol{A}_{m+1}^{-1}\right)\\
&= \sum_{k=1}^{m}(\boldsymbol{A}_k + \boldsymbol{A}_{m+1})\left(\boldsymbol{A}_k^{-1} + \boldsymbol{A}_{m+1}^{-1}\right)\\
&\geq 4m\boldsymbol{I}
\end{aligned} \tag{66}$$

where the last inequality is due to (58). Using (66) in (65) yields

$$\left(\sum_{k=1}^{m} \boldsymbol{A}_k + \boldsymbol{A}_{m+1}\right)\left(\sum_{k=1}^{m} \boldsymbol{A}_k^{-1} + \boldsymbol{A}_{m+1}^{-1}\right) \geq (m+1)^2\boldsymbol{I}. \tag{67}$$

By the above induction, Lemma 1 is proven. ■

ACKNOWLEDGMENT

The authors would like thank Dr. A. Swami for his support and input for this work.

REFERENCES

[1] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2210–2219, Jun. 2005.

[2] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3564–3575, Oct. 2005.

[3] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, Part I: Gaussian PDF," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[4] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained cistributed estimation for wireless sensor networks, Part II: Unknown PDF," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2784–2796, Jul. 2006.

[5] P. Venkitasubramaniam, L. Tong, and A. Swami, "Quantization for maximin ARE in distributed estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3596–3605, Jul. 2007.

[6] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, pp. 413–422, Feb. 2006.

[7] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4284–4299, Aug. 2007.

[8] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced rank estimation and filtering," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 457–469, Mar. 2001.

[9] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.

[10] S. Cui, J.-J. Xiao, A. G. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and power efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[11] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 798–808, Apr. 2005.

[12] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3456–3470, Jul. 2007.

[13] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.

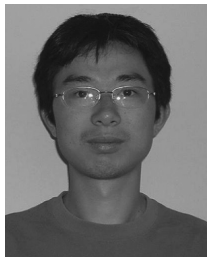
[14] C. G. Lopes and A. H. Sayed, "Diffusion Least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[15] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links-Part I: distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[16] I. D. Schizas, G. B. Giannakis, S. D. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc WSNs with noisy links-Part II: distributed estimation and smoothing of random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1650–1666, Apr. 2008.

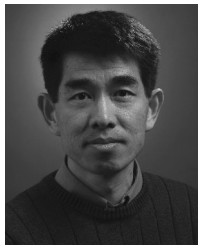
[17] Y. Huang and Y. Hua, "Multihop progressive decentralized estimation in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 1004–1007, Dec. 2007.

[18] Y. Huang and Y. Hua, "Multihop progressive decentralized estimation of a deterministic vector in wireless sensor networks," presented at the Asilomar Conf. Signals, Systems, Computers, Pacific Grove, CA, Nov. 2007.
[19] Y. Hua and Y. Huang, "Progressive estimation and detection," presented at the Workshop on Sensor Signal and Information Processing (SenSIP), Sedona, AZ, May 2008.
[20] S. T. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
[21] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 96–106, Aug. 2004.
[22] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[23] F. C. Harris, Jr., "Steiner minimal trees: An introduction, parallel computation, and future work," in *Handbook of Combinatorial Optimization*. Norwell, MA: Kluwer Academic, 1998.
[24] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.
[25] S. Haykin, *An Introduction to Analog and Digital Communications*. New York: Wiley, 1994.
[26] M. Gastpar and M. Vetterli, "Power, spatial-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 745–754, Apr. 2005.
[27] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gausssian sensor network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.
[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[29] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
[30] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Energy-efficient joint estimation in sensor networks: analog versus digital," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP) Conf.*, Mar. 2005, vol. 4, pp. 18–23.



Yi Huang received the B.E. degree in electronic engineering and information science from the University of Science and Technology of China in 2002 and the M.Phil. degree in information engineering from the Chinese University of Hong Kong in 2004. He is currently working on the Ph.D. degree in electrical engineering at University of California, Riverside.

His current research interests include signal processing in wireless sensor networks and throughput maximization for wireless ad hoc/mesh networks. He is the recipient of UCR Dean's Dissertation Fellowship.



Yingbo Hua (S'86–M'88–SM'92–F'02) received the B.S. degree from Nanjing Institute of Technology, Nanjing, China, in February 1982 and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, in 1983 and 1988, respectively.

From 1988 to 1989, he was a Research Fellow at Syracuse, consulting for the Syracuse Research Company, NY, and Aeritalia Company, Italy. He was a Lecturer from February 1990 to 1992, Senior Lecturer from 1993 to 1995, and Reader and Associate Professor from 1996 to 2001, with the University of Melbourne, Australia. He served as a visiting faculty member with Hong Kong University of Science and Technology from 1999 to 2000, and consulted for the Microsoft Research Company, WA, in summer 2000. Since February 2001, he has been Professor of electrical engineering with the University of California, Riverside. He is an author/coauthor of over 270 articles in journals, conference proceedings, and books, which span the fields of sensor array signal processing, system identification, wireless communications, and sensor networks. He is a co-editor of *Signal Processing Advances in Wireless and Mobile Communications* (Prentice-Hall, 2001), and *High-Resolution and Robust Signal Processing* (Marcel Dekker, 2003).

Dr. Hua served on the Editorial Boards for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, the *IEEE Signal Processing Magazine*, and *Signal Processing* (EURASIP). He also served on the IEEE SPS Technical Committees for Underwater Acoustic Signal Processing, Sensor Array and Multi-channel Signal Processing, and Signal Processing for Communications and Networking, and on numerous international conference organization committees.