DISTRIBUTED ESTIMATION IN RESOURCE-CONSTRAINED WIRELESS SENSOR NETWORKS

A Dissertation
Presented to
The Academic Faculty

by

Junlin Li

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2008

DISTRIBUTED ESTIMATION IN RESOURCE-CONSTRAINED WIRELESS SENSOR NETWORKS

Approved by:

Dr. Ghassan AlRegib, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Monson H. Hayes
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Ying Zhang
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Elliot Moore
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Paul A. Work
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: November 6, 2008

To my family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Ghassan AlRegib, for his persistent support and invaluable advice during my Ph.D. study and research. Without his guidance, I would not be able to successfully complete the dissertation. It has been a great pleasure to have him as my advisor. I would also like to express my gratitude to Dr. Monson H. Hayes, Dr. Ying Zhang, Dr. Ian F. Akyildiz, Dr. Elliot Moore, and Dr. Paul A. Work for their insightful comments and suggestions as part of my supervisory committee.

Besides my thesis research, I have been fortunate to work on several other interesting projects during my Ph.D. study in Georgia Tech. Special thanks go to Dr. Dihong Tian for our extensive collaboration on many projects since I first came to Georgia Tech in 2004. I would also like to thank Dr. Wen H. Chen and Dr. Pi Sheng Chang from Cisco System Inc. for our productive collaboration on video compression research during the past two years. I also would like to thank Dr. Philip A. Chou, Dr. Dinei A. Florencio and Li-wei He from Microsoft Research for giving me the internship opportunity to work in an exceptional industrial research environment during the summer of 2006.

I would like to express my gratitude to the faculty and staff in the School of Electrical and Computer Engineering for providing excellent courses, services, and working environments. I would also like to thank many colleagues and friends in Georgia Tech: Fred Stakem, Mashhour Solh, Wenhui Xu, Jinyu Li, Qiang Fu, Chengyuan Ma, Zhensheng Jia, Wei Zhang, Kun Shi, and Xusheng Sun, to name a few. From outside of Georgia Tech, I also have many friends to thank for all kinds of help and advice from them. Specially, I would like to thank Dr. Yushi Shen. We have a long-time friendship and I enjoyed every conversation we had about Ph.D. study, research, careers, and all other aspects of life.

Last but not least, I would like to express my deepest gratitude to my parents, wife, brothers and sister for their unconditional love and support. It is to them that I dedicate this work.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Wireless sensor networks (WSN) are an emerging technology with a wide range of applications including environment monitoring, security and surveillance, health care, smart homes, etc. Subject to severe resource constraints in wireless sensor networks, in this research, we address the distributed estimation of unknown parameters by studying the correlation among *resource*, *distortion*, and *lifetime*, which are three major concerns for WSN applications.

The objective of the proposed research is to design efficient distributed estimation algorithms for resource-constrained wireless sensor networks, where the major challenge is the integrated design of local signal processing operations and strategies for inter-sensor communication and networking so as to achieve a desirable tradeoff among resource efficiency (bandwidth and energy), system performance (estimation distortion and network lifetime), and implementation simplicity. More specifically, we address the efficient distributed estimation from the following perspectives: (*i*) *rate-distortion* perspective, where the objective is to study the rate-distortion bound for the distributed estimation and to design practical and distributed algorithms suitable for wireless sensor networks to approach the performance bound by optimally allocating the bit rate for each sensor, (*ii*) *energy-distortion* perspective, where the objective is to study the energy-distortion bound for the distributed estimation and to design practical and distributed algorithms suitable for wireless sensor networks to approach the performance bound by optimally allocating the bit rate and transmission energy for each sensor, and (*iii*) *lifetime-distortion* perspective, where the objective is to maximize the network lifetime while meeting estimation distortion requirements by

jointly optimizing the source coding, source throughput and multi-hop routing. Also, energy-efficient *cluster-based distributed estimation* is studied, where the objective is to minimize the overall energy cost by appropriately dividing the sensor field into multiple clusters with data aggregation at cluster heads.

CHAPTER I

INTRODUCTION

Recent advances in low-power micro-electro-mechanical system (MEMS) technology and wireless communications have led to the emergence of a key technology, wireless sensor networks (WSNs) [3]. A wireless sensor network consists of a large number of sensors that can communicate with each other to achieve a specific task. Though each sensor is characterized by low power constraint and limited computation and communication capabilities, when suitably deployed in large scale, potentially powerful networks can be constructed to accomplish various high-level tasks with sensor collaboration [47], thus making wireless sensor networks a promising technology for a wide range of applications.

Current and potential applications of wireless sensor networks include environment monitoring, military sensing, traffic surveillance, health care, and smart homes [3, 22, 25, 28, 29]. For example, distributed sensor networks can be deployed to monitor such features of the natural environment as temperature, water flow, and the condition of glaciers. They also can be used to detect, locate, and identify type and concentration of polluting chemicals in water and air [67, 88, 100]. In the battlefield, networked video, acoustic, or other types of sensors can be used to detect and track suspected targets and coordinate the activities among several unmanned vehicles according to the collected data from the field [35, 71]. In traffic surveillance, image sensors and other types of sensors have been used at roadway intersections to monitor traffic conditions or identify vehicles [36]. In the future, more intelligent sensors are expected to be placed on vehicles to help reduce or avoid accidents. In health care and smart home industries, by combining body-area sensors with environmental sensors embedded in home surroundings, advanced multi-parametric health monitoring may be achieved without compromising the convenience of the patients [34, 45, 84].

A common goal in most WSN applications is to reconstruct the underlying physical phenomenon based on sensor measurements. The distributed estimation of unknown parameters by a set of distributed sensor nodes and a fusion center has become an important topic in signal processing research for wireless sensor networks [93, 96]. In distributed parameter estimation, sensor nodes collect real-valued data, locally process the data, and send the resulting messages to the fusion center (FC), which combines all the received messages to produce a final estimation of the unknown parameter.

Estimation using a WSN requires not only local information processing but also inter-sensor communication because sensors spread over a large geographical area. This particular feature adds a wireless communication and networking component to the problem that is absent from the traditional centralized estimation framework. In fact, a major challenge in WSN research is the integrated design of local signal processing operations and strategies for inter-sensor communication and networking so as to achieve a desirable trade-off among resource efficiency, system performance, and implementation simplicity. Designing distributed signal processing algorithms differs from that of the traditional centralized framework in several important aspects:

- Obtaining the complete signal models for a large number of sensors may be impractical, particularly in dynamic sensing environments. This prevents direct application of optimum estimation algorithms and motivates the development of distributed estimation strategies based on only partially known or unknown data/noise models.

- Constraints on sensor cost, bandwidth, and energy budget dictate that low-quality sensor observations may have to be aggressively quantized (e.g., down to a few bits per observation per sensor). Furthermore, local compression at a sensor node depends not only on the quality of sensor observation, but also on the quality of the wireless communication channels between sensor nodes. Thus, estimators must be developed based on severely quantized versions of very noisy observations.

- Multi-hop transmission of locally processed data from sensor nodes to the fusion center and data aggregation at intermediate nodes on the transmission path are essential to save transmission energy and prolong the network lifetime.

In this dissertation, distributed estimation in resource-constrained wireless sensor networks is addressed, where the main design goals are resource (bandwidth and energy) efficiency, system performance (estimation distortion and network lifetime), and implementation simplicity. More specifically, the distributed estimation problem in wireless sensor networks is addressed from the following aspects:

- *Rate-Distortion Perspective*: The distributed estimation is addressed from the rate-distortion point of view, where the objective is to minimize the estimation distortion under a given bit rate constraint. Also the theoretical rate-distortion bound of the distributed estimation is analyzed.

- *Energy-Distortion Perspective*: The distributed estimation is addressed from the energy-distortion point of view, where the objective is to minimize the estimation distortion under a given energy constraint. Furthermore, the theoretical energy-distortion bound of the distributed estimation is analyzed.

- *Lifetime-Distortion Perspective*: Network lifetime is defined and optimized for the distributed estimation with a given distortion requirement, which involves the joint optimization of source coding, source throughput, and multi-hop routing.

- *Cluster-Based Distributed Estimation*: The whole sensor field is divided into multiple clusters and data aggregation is applied at cluster heads. How to optimally cluster the sensor networks is addressed to minimize the overall energy consumption.

Specifically, this dissertation is organized as follows.

Chapter 2 first presents the general background of wireless sensor networks and the distributed and collaborative signal processing framework in wireless sensor networks; then presents the specific background and state-of-the-art on centralized and decentralized estimation problems and algorithms.

Chapter 3 and Chapter 4 address the resource-constrained distributed estimation subject to severe resource constraints (bandwidth and energy) in wireless sensor networks. In particular, Chapter 3 focuses on the rate-constrained distributed estimation. First, a concept of the equivalent 1-bit MSE (mean square error) function is introduced. Then, based on minimizing the equivalent 1-bit MSE function, a quasi-optimal distributed estimation algorithm is developed and a theoretical rate-distortion bound is presented. Chapter 4 focuses on the energy-constrained distributed estimation. Generally speaking, the energy-constrained distributed estimation is a generalization of the rate-constrained distributed estimation since the energy consumption is generally a function of the transmission bit rate. In this chapter, we first introduce a concept of the equivalent unit-energy MSE function; then, a quasi-optimal distributed estimation algorithm and a theoretical energy-distortion bound are developed by minimizing the equivalent unit-energy MSE function. In this chapter, several different transmission and energy models and different network topologies such as single-hop or multi-hop WSN are considered.

Chapter 5 addresses the lifetime-optimized distributed estimation. In this chapter, we first define a concept of function-based network lifetime, which focuses on whether the network as a whole can perform a given task rather than whether any individual sensor in the network is dead. Then, to maximize the function-based network lifetime under a given estimation distortion requirement, a nonlinear programming (NLP) problem is formulated and solved, which involves a joint optimization of source coding, source throughput, and multi-hop routing.

Chapter 6 studies the cluster-based distributed estimation, where the sensor network is divided into several clusters, each cluster with a cluster head. To reduce the energy consumption, data aggregation is introduced at the cluster heads. Then the major objective is to study how to optimally divide the sensor networks into clusters such that the overall energy consumption is minimized. In this chapter, two algorithms are developed for fixed cluster head case and cluster head rotation case, respectively.

Finally, the dissertation is concluded in Chapter 7 with a summary of contributions and future research directions.

CHAPTER II

BACKGROUND

2.1 Wireless Sensor Networks

A wireless sensor network consists of a large number of sensor nodes, each equipped with three basic functional components: a sensing unit, a processing unit, and a transceiver unit. The sensing unit collects information from the surrounding environment; the processing unit performs some local information processing, such as quantization and compression; and the transceiver unit transmits the locally processed data to a fusion center where the information from different sensor nodes is aggregated and fused to generate the final inference. Figure 1 shows an example of a wireless sensor network for environment monitoring, where the sensors are scattered in a sensor field to observe one or more environmental parameters such as temperature, light level, and soil moisture. Each sensor can route its data (locally processed observations) via single-hop or multi-hop wireless channels to the fusion center / sink node, which makes the final inference based on the received data. This sink node communicates with the end user via public networks over satellite, wireless, or wired links. Another example application of wireless sensor networks is wireless body area networks (WBANs) [40] for health monitoring. A WBAN consists of multiple sensor nodes, each capable of sampling, processing, and communicating one or more vital signs (heart rate, blood pressure, oxygen saturation, activity) or environmental parameters (location, temperature, humidity, light). Typically, these sensors are placed strategically on the human body as tiny patches or hidden in users clothes allowing ubiquitous health monitoring in their native environment for extended periods of time. A typical body sensor is tiny (about $20mm$) and ultra-low power (about $10\mu A$

in active mode and even less than $1\mu A$ in standby mode) [74]. These wireless sensor networks are essentially different from the existing traditional sensor networks, where expensive heavy sensors are laid in a field (such as ocean and desert).

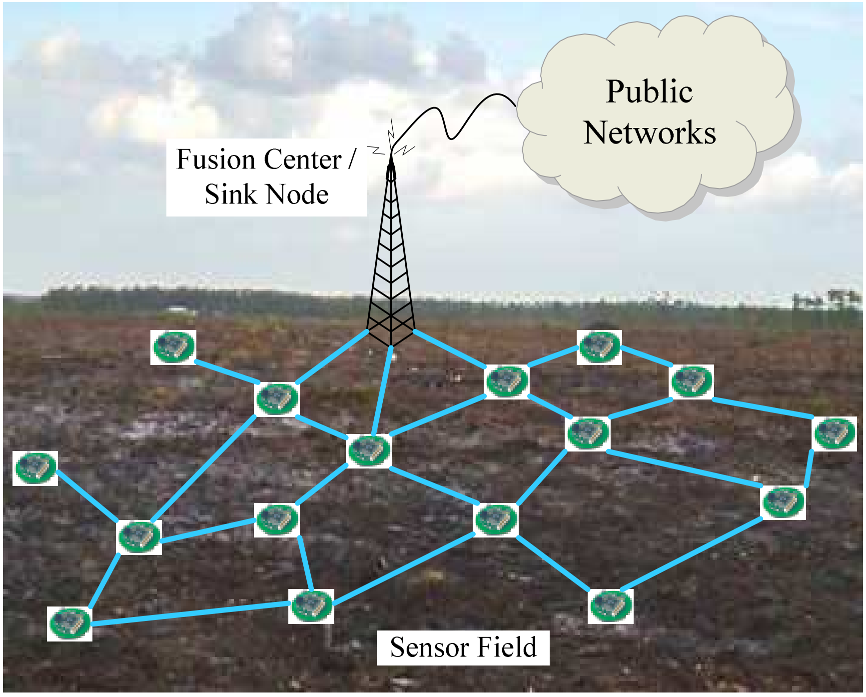


Figure 1: An example of a wireless sensor network for environment monitoring.

While wireless sensor networks share many common features with existing wireless ad hoc network concepts, there are a number of specific characteristics that make wireless sensor networks different. These features can be summarized as follows:

- *Large Network Size*: A wireless sensor network has a large number (hundreds or even thousands) of tiny, low-cost, and low-power nodes densely deployed in a certain geographic area. This not only leads to the advantages of observation redundancy, which may increase the monitoring precision and network robustness, but also creates challenges to sensor collaboration and networking, which requires scalable solutions.

- *Self-Configurability*: The large number of sensor nodes in a wireless sensor network have little or no pre-established infrastructure, and their network topology can change dynamically because of nodes' sleep, nodes' failure, or nodes' move. Thus, similar to other ad hoc networks, wireless sensor networks are required to have the capability of self-configuration. This includes, for example, sensor self-localization, inter-node coordination, and adapting to node failure in a harsh wireless environment.

- *Stringent Energy Constraints*: Sensor nodes are powered by small batteries that are usually not rechargeable or irreplaceable, and thus energy in the network is scarce and energy consumption is a primary design metric to be considered. To prolong the lifetime of the network, the design of all sensor network operations needs to be energy efficient. Energy limitations are, in fact, one of the major differences between wireless sensor networks and other wireless networks, such as wireless local area networks, where energy efficiency is of less concern.

- *Application Specific*: Because of the large number of conceivable combinations of sensing, computing, and communication technology, many different sensor network application scenarios become possible. It is unlikely that there will be a "one-size-fits-all" solution for all types of sensor networks. Usually the low cost and low energy supply require sensor networks to be optimally designed according to a specific application scenario.

- *Data Centric*: The low cost and low energy supply will require, in many application scenarios, redundant deployment of wireless sensor nodes. As a consequence, the importance of any one particular node is considerably reduced compared to traditional networks (such as wireless LAN or cellular phone systems). Instead, the data that is observed by these nodes is the focal point. This results in a shift from node-centric architectures in classical networks toward data-centric architectures in sensor networks.

- *Simple and Easily Implementable*: Because sensor nodes are small and battery-powered with limited onboard processing and communication capabilities, sensor network operations need to be simple. In some applications, sensor networks also need to be deployed in real time. This requires simple and easily implemented designs for sensor networks at all levels, including signal processing algorithms, networking structure, and communication protocols.

To successfully deploy wireless sensor networks, novel and intelligent processing and communication concepts need to be developed that fit the nature of these networks. Given the power constraints in sensors and the massive number of sensors, one of the major objectives of sensor network research is to design energy-efficient devices, protocols, and algorithms, i.e., low-cost sensors to collect information, efficient networking protocols to transmit information among sensor nodes or from sensors to a processing center, and distributed algorithms to process and abstract the core information from the raw data collected by sensors. These three technology areas are not isolated from each other because of the interdisciplinary nature of sensor network design. This dissertation focuses on distributed signal processing in wireless sensor networks, which involves a combined treatment of sensor data processing and communication networks.

2.2 Distributed and Collaborative Signal Processing

Wireless sensor networks present a significant trade-off between power consumed by processing versus communication. Compared with sensing and computation, communication is the most energy-consuming operation in wireless sensor networks; therefore, distributed signal processing at local sensors to reduce the data transmitted among sensors or from sensors to a processing center is essential in saving energy [22]. On the other hand, information sharing and collaborative processing among sensors are also essential to achieve enough precision because the data from a single low-cost sensor is too coarse to derive a reliable inference. Thus, a new paradigm - *distributed and collaborative signal processing* - is necessary to efficiently coordinate the large number of sensors in a network to achieve high-level information processing tasks [2,22,47,103,104].

Figure 2 shows a layered structure for distributed and collaborative signal processing. The bottom layer is sensor scheduling and sensing layer, which is to decide the sensor status and make the desired observations when the sensor is active. The second lowest layers is local signal processing layer, which performs local data processing, such as quantization. Multi-modality fusion also could be performed if there are multiple sensor modules on the same sensor board. These two layers are performed at local sensors in a distributed manner. Then, each sensor communicates its locally processed data to other sensors or a processing center. The data communicated among sensors can be analog waveforms, digitized observations/decisions, or locally derived data models. The upper two layers are collaborative signal processing, including multi-sensor data fusion layer, which is to fuse the data received from multiple sensors and extract the useful information, and application layer, which is to make the final information inference for user applications.

The distributed and collaborative signal processing paradigm provides the foundation for large and highly scalable sensor networks. It is widely adopted to achieve different information processing tasks, such as query dissemination [26], distributed data compression [76,99], distributed estimation [5–9,32,39,46,48,50–58,63–65,69,70,73,75,77,79,80,82,83,85,90–92,94,96–98], distributed detection [4,10,11,14–16,19,20,27,30,37,59,60,62,72,86,89,95], localization [21,43,78], and tracking [33,49,61,103,104].

Distributed and collaborative signal processing and information fusion over a network is an active area of research. Important technical issues include the degree of information sharing between nodes and how nodes fuse the information received from

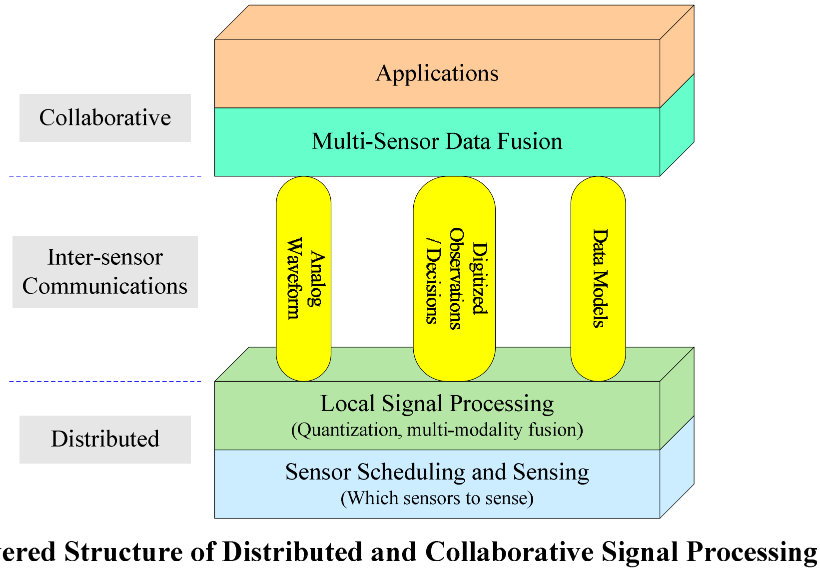


Figure 2: Layered architecture for distributed and collaborative signal processing.

other nodes. Processing data from several sensors generally results in performance improvement but also requires more communication resources (bandwidth and energy). Similarly, when communicating information at a lower level (e.g., raw data), less information is lost, but more bandwidth and energy are required. Therefore, one needs to consider the multiple trade-offs between system performance and resource utilization. Other issues include how to meet latency and reliability requirements and how to maximize sensor network operational lifetime.

In this dissertation, we will apply the distributed and collaborative signal processing paradigm to study the distributed estimation problem. In the context of distributed estimation, all the aforementioned technical issues, including system performance and resource utilization trade-offs and network work lifetime, will be addressed. In the next section, we first review the background and state-of-the-art of the centralized/decentralized estimation.

2.3 Centralized versus Decentralized Estimation

Consider a dense sensor network that includes N distributed sensor nodes and a fusion center to observe and estimate an unknown parameter θ, as shown in Figure 3. Each sensor makes an observation, which is corrupted by additive noise and is described by

$$x_k = \theta + n_k, \quad k = 1, \cdots, N. \tag{1}$$

We assume that the noises n_k ($k = 1, \cdots, N$) are zero mean, spatially uncorrelated with probability density function (pdf) $f_k(x)$ and variance σ_k^2. Then each sensor k may perform a local processing on its observation x_k and transmit its processed message $m_k(x_k)$ to the fusion center over a wireless channel, where the final estimation is made based on all received messages from all sensors.

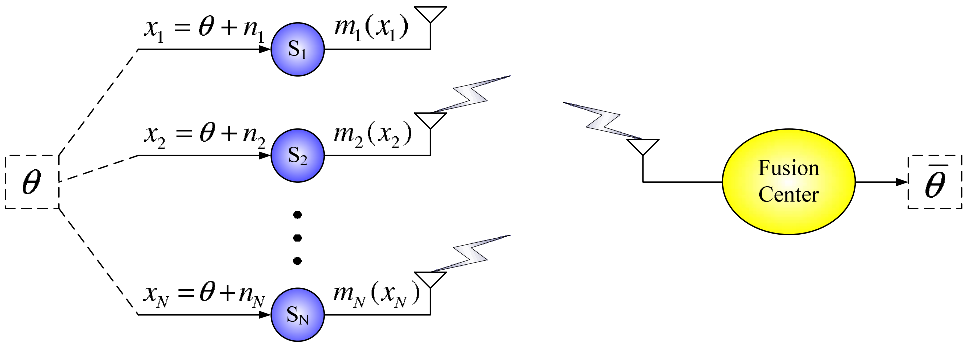


Figure 3: System diagram of centralized/decentralized estimation.

Note that the data model in Equation (1) bears a number of variations in different applications. For example, by a suitable linear scaling, the above data model is equivalent to the one where sensors observe θ with different attenuations, namely, $x_k = h_k\theta + n_k$. Indeed, if we let $x'_k = x_k/h_k$ and $n'_k = n_k/h_k$, then $x'_k = \theta + n'_k$, which is identical to that in Equation (1).

2.3.1 Centralized Estimation

If the fusion center has the knowledge of the sensor noises (pdf f_k and/or variance σ_k^2) and the sensors can perfectly send their observations x_k ($k = 1, \cdots, N$) to the fusion center, the fusion center can estimate the parameter θ based on the observations $\mathbf{x} := (x_1, x_2, \cdots, x_N)$ using two centralized estimators: maximum likelihood estimator (MLE) and best linear unbiased estimator (BLUE) [44]. Their performance serves as a performance benchmark for the decentralized estimation.

- MLE: If the noise distributions f_k $(k = 1, \cdots, N)$ are available as prior information, the centralized MLE of θ based on observations $\mathbf{x}$ is

$$\bar{\theta}_{ML} = \arg\max_{\theta} f(\mathbf{x}|\theta) = \arg\max_{\theta} \sum_{k=1}^{N} \log f_k(x_k - \theta). \tag{2}$$

  MLE, if unique, is asymptotically unbiased and attains the Cramer-Rao Lower Bound (CRLB): $\left(\sum_{k=1}^{N} J(\theta; x_k)\right)^{-1}$, where $J(\theta; x_k)$ is the Fisher information of θ in x_k:

$$J(\theta; x_k) = \int \left(\frac{\partial f_k(x-\theta)}{\partial \theta}\right)^2 \frac{1}{f_k(x-\theta)} dx = \int \frac{f_k'(x)^2}{f_k(x)} dx. \tag{3}$$

- BLUE: If the fusion center only has knowledge of the sensor noise variances σ_k^2 $(k = 1, \cdots, N)$, a best linear unbiased estimator (BLUE) can be performed to recover θ by combining x_k with weights inversely proportional to σ_k^2. This leads to the following BLUE estimation of θ

$$\bar{\theta}_{BLUE} = \left(\sum_{k=1}^{N} \frac{1}{\sigma_k^2}\right)^{-1} \sum_{k=1}^{N} \frac{x_k}{\sigma_k^2}, \tag{4}$$

  and the estimation mean square error (MSE) of the BLUE estimator is

$$E\left(\left|\bar{\theta}_{BLUE} - \theta\right|^2\right) = \left(\sum_{k=1}^{N} \frac{1}{\sigma_k^2}\right)^{-1}. \tag{5}$$

It is noted that the BLUE estimator does not depend on the noise pdf but on the noise variance only, while the MLE estimator needs to know the underlying sensor noise distributions. Unfortunately, characterizing the exact noise probability distribution for a large number of sensors is impractical, especially for applications in a dynamic sensing environment. What is more realistic in these scenarios is to let each sensor estimate its local signal-to-noise ratio (SNR), which can be accomplished by simply measuring the received signal power in the presence and absence of the incoming signal. Motivated by these considerations, we will focus on the BLUE estimator and, therefore, take the MSE in Equation (5) as the major performance benchmark for the decentralized estimation in wireless sensor networks.

2.3.2 Decentralized Estimation

Centralized estimation schemes (MLE and BLUE) require the sensors to send their original real-valued observations x_k ($k = 1, \cdots, N$) to the fusion center perfectly. Therefore, the communication links between sensors and the fusion center need to have sufficient bandwidth. And the transmission energy cost is high. This makes it impractical for implementation in wireless sensor networks, where resources (bandwidth and energy) are severely constrained and the wireless channels between sensors and the fusion center are noisy.

Instead of sending the original real-valued observations to the fusion center, local processing (quantization and compression) at sensors is essential to reduce the communication cost (bandwidth and energy). This is referred to as *decentralized estimation*. It can be accomplished as follows. First, each sensor performs a local quantization $m_k = Q_k(x_k)$, where $Q_k(x_k)$ is a quantization function, and the quantization message m_k is then transmitted to the fusion center. Second, all the quantization messages are combined at the fusion center to produce a final estimation of θ using a real-valued fusion function f: $\bar{\theta} = f(m_1, m_2, \cdots, m_N)$. The quality of an estimation for θ is measured by the MSE criterion: $E(|\bar{\theta} - \theta|^2)$. The distributed estimation of unknown deterministic parameters by a set of distributed sensor nodes and a fusion center has become an important topic in signal processing research for wireless sensor networks [96].

The major issue of decentralized estimation in resource-constrained wireless sensor networks is to co-design the local quantization function, the fusion function, and the resource (bandwidth and energy) allocation strategy to achieve the optimal trade-off between the resource efficiency and overall system performance.

In the context of distributed estimation, subject to the resource (bandwidth and energy) limitation nature of wireless sensor networks, several bandwidth-constrained distributed estimation algorithms [6, 7, 32, 48, 64, 65, 69, 70, 75, 79, 80] have been investigated recently. The work of [32, 48, 75] addressed various design and implementation issues to digitize the transmitted signal into one or several binary bits using the joint distribution of sensors' data. In [79] and [80], a class of maximum likelihood estimators (MLE) was proposed to attain a variance that is close to the clairvoyant estimator when the observations are quantized to one bit. The work of [7] and [6] addressed the maximum likelihood estimation over noisy channel for bandwidth-constrained sensor networks with or without knowing the sensing and channel noise parameters at the fusion center. Without the knowledge of noise distribution, the work of [69] and [70] proposed using a training sequence to aid the design of local data quantization strategies, and the work of [64] and [65] proposed several universal (pdf-unaware) decentralized estimation systems based on best linear unbiased estimation (BLUE) rule for distributed parameter estimation in the presence of unknown, additive sensor noise.

To explicitly address the energy constraint in wireless sensor networks, the minimal-energy distributed estimation problem has also been recently considered in [5, 46, 91, 92, 94]. In [94] and [46], the total sensor transmission energy is minimized by selecting the optimal quantization levels while meeting the target estimation MSE requirements. On the contrary, the work of [5] minimizes the estimation MSE under the given energy constraints. The work of [91, 92] addressed the energy-constrained distributed estimation problem (under the BLUE fusion rule) by exploiting long-term noise variance statistics.

Although both bandwidth and energy constraints for distributed estimation in wireless sensor networks have been widely investigated [5–7, 32, 46, 48, 64, 65, 69, 70, 75, 79, 80, 91, 92, 94], there is a lack of overall optimality analysis in the sense of resource-distortion performance. In the first part of this dissertation, the resource-distortion optimized distributed estimation under a total available resource constraint is studied, where the available resource should be allocated among all sensors jointly and optimally to optimize the estimation performance, and the optimal resource allocation could be unequal because of the heterogeneous nature of sensor networks. In this research, both the theoretical resource-distortion bound for distributed estimation and the practical algorithms with close-to-optimal performance are developed.

In this dissertation, the distributed estimation is addressed not only from *rate-distortion* perspective or *energy-distortion* perspective, but also from *lifetime-distortion* perspective, which is a critical concern in the design of wireless sensor networks. The network lifetime issue for distributed estimation application in wireless sensor networks has not yet been addressed explicitly in the literature. In this dissertation, we define a new notion of *function-based network lifetime*, which focuses on whether the network as a whole can perform a given task, and then address how to optimize the function-based network lifetime for distributed estimation in single-hop and multi-hop wireless sensor networks. In this research, cluster-based distributed estimation is also studied, where data aggregation at the cluster head is introduced and optimal clustering algorithms are developed to minimize the overall energy cost. In the following several chapters, all the aforementioned research efforts and the resulting findings will be described in detail.

CHAPTER III

RATE-CONSTRAINED DISTRIBUTED ESTIMATION

In wireless sensor networks, compared with sensing and computation processes, communication is the most energy consuming operation. Thus, reducing communications between sensors and the fusion center is essential in saving energy. Also, the communication capacity within wireless sensor networks is limited because the wireless channel is shared across the whole network. Therefore, a total communication rate constraint within a sensor network is necessary to avoid communication collisions that waste energy. In this chapter, we study the rate-constrained distributed estimation by imposing a total bit rate constraint for all sensors in the sensor network.

In the rate-constrained distributed estimation problem, data quantization/compression at local sensors is needed to reduce communication requirements. The quantized or compressed data is transmitted to the fusion center to generate the final estimation. In fact, the design of discrete local message functions and the final estimation function is a major topic in the research of wireless sensor networks. Recently, several rate-constrained distributed estimation algorithms have been investigated in [6,7,32,48,64,65,69,70,75,79,80]. Most of the past work on rate-constrained distributed estimation is usually posed for a given number of sensors (one observation per sensor), but there is a lack of overall optimality analysis in the sense of rate-distortion performance. Here, the fundamental issue is the rate-distortion bound. More importantly, how do we achieve the performance bound in a simple and distributed manner such that it is easy to implement for wireless sensor networks?

In this chapter, rate-constrained distributed estimation is addressed from the rate-distortion perspective [51, 53]. To minimize the estimation distortion, the given number of available bits B has to be allocated jointly and optimally among all sensors. Therefore, there exists an interesting trade-off between the number of active sensors and the quantization precision of each active sensor. We address this optimal trade-off and design the optimal distributed estimation mechanism by (i) selecting a subset of sensors to observe the phenomenon and (ii) selecting the quantizer for each active sensor to quantize the real-valued observations. Furthermore, the theoretical rate-distortion bound for the distributed estimation is analyzed, which shows that the proposed algorithm is close to optimal.

The rest of this chapter is organized as follows. Section 3.1 states the distributed estimation problem under the total bit rate constraint. The quantization rule at local sensors and the fusion rule at the fusion center are also described. Section 3.2 introduces a concept of the equivalent 1-bit MSE function. Based on this concept, an optimal distributed estimation algorithm for homogeneous sensor networks and a quasi-optimal distributed estimation algorithm for heterogeneous sensor networks are developed in Section 3.3 and Section 3.4, respectively. Furthermore, a theoretical lower bound of the estimation MSE under the total bit rate constraint is addressed in Section 3.5. Section 3.6 shows some simulation results to demonstrate the performance of the proposed algorithms. Section 3.7 summarizes this chapter. The proofs of some theorems presented in this chapter are delegated to the appendix in Section 3.8.

3.1 System Model and Problem Statement

Consider a dense sensor network that includes N distributed sensors and each sensor can observe, quantize, and transmit its observation to the fusion center, which will estimate the parameter θ based on the received messages. Because of the total bit rate constraint, there is a trade-off between the number of active sensors and the quantization bit rate at each active sensor, that is, only a subset of sensors will be active at each task period. Without loss of generality, we assume that the first K sensors are active. This can be accomplished as follows (Figure 4).

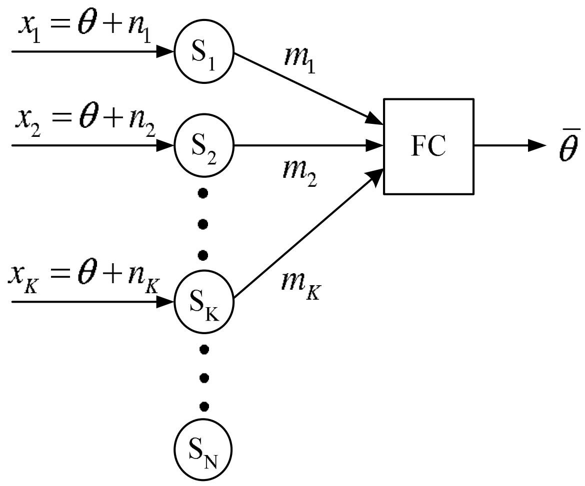


Figure 4: The distributed estimation system under the total bit rate constraint, where K sensors are active and each active sensor performs a local quantization and transmits its quantized message to the fusion center, which will estimate θ based on all the received messages.

First, each sensor makes an observation on the unknown parameter θ. The observations are corrupted by additive noises and are described by

$$x_k = \theta + n_k, \quad k = 1, \cdots, K. \tag{6}$$

We assume that the noises n_k ($k = 1, \cdots, K$) are zero mean, spatially uncorrelated with variances σ_k^2, but otherwise unknown. Second, each active sensor performs a local quantization $m_k = Q_k(x_k)$, where $Q_k(x_k)$ is a quantization function, and the quantized message m_k is transmitted to the fusion center, where all the quantized messages are combined to produce a final estimation of θ using a fusion function f: $\bar{\theta}_K = f(m_1, m_2, \cdots, m_K)$. The quality of an estimation for θ is measured by the mean square error (MSE) criterion. So the primary goal is to perform the following optimization:

$$\begin{aligned} \min \quad & E(\bar{\theta}_K - \theta)^2, \\ s.t. \quad & \sum_{k=1}^{K} b_k \le B, \quad b_k \in \mathbb{Z}^+, \quad k = 1, \cdots, K, \end{aligned} \tag{7}$$

where B is the total bit rate constraint, K is the number of active sensors, and b_k is the quantization bit rate of the active sensor k.

If the fusion center has the knowledge of the sensor noise variances σ_k^2 ($k = 1, \cdots, K$) and the sensors can perfectly send their observations x_k ($k = 1, \cdots, K$) to the fusion center, the best linear unbiased estimator (BLUE) [44] for θ is

$$\bar{\theta}_K = \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right)^{-1} \sum_{k=1}^{K} \frac{x_k}{\sigma_k^2} \tag{8}$$

and the estimation MSE of the BLUE estimator is

$$E(\bar{\theta}_K - \theta)^2 = \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right)^{-1}. \tag{9}$$

But the BLUE scheme is impractical for wireless sensor networks because of the high communication cost. Instead of sending the real-valued observations to the fusion center directly, quantization at local sensors is essential to reduce the communication cost (bandwidth and energy). In this work, we adopt a probabilistic quantization scheme [94] and a quasi-BLUE estimation scheme. Based on that, the optimal trade-off between the number of active sensors and the quantization bit rate of each active sensor is addressed to achieve the optimal rate-distortion performance.

3.1.1 Probabilistic Quantization

Suppose the observed signal of each sensor is bounded to $[-W, W]$, that is, $x = \theta + n \in [-W, W]$, where W is a known parameter decided by the sensor's dynamic range, θ is the unknown signal to be estimated, and n is the zero mean noise with variance σ^2. Regardless of the probability distribution of x, the probabilistic quantization with b bits is summarized as follows: uniformly divide $[-W, W]$ into intervals of length $\Delta = 2W/(2^b - 1)$ and round x to the neighboring endpoints of these small intervals in a probabilistic manner. More specifically, suppose $-W+i\Delta \leq x \leq -W+(i+1)\Delta$, where $0 \leq i \leq 2^b - 2$; then x is quantized to $m(x,b)$ according to

$$\begin{aligned} P\left\{m(x,b) = -W + i\Delta\right\} &= 1 - r, \\ P\left\{m(x,b) = -W + (i+1)\Delta\right\} &= r, \end{aligned} \tag{10}$$

where $r = (x + W - i\Delta)/\Delta \in [1, 1]$.

The following lemma, which is proved in [94], shows that the quantized message $m(x,b)$ is an unbiased estimator of θ with a variance approaching σ^2 at an exponential rate as b increases.

Lemma 3.1 ([94]). *Let $m(x,b)$ be a b-bit quantization of $x \in [-W, W]$ as defined above; then, $m(x,b)$ is an unbiased estimator of θ and*

$$E\left(|m(x,b) - \theta|^2\right) \leq \sigma^2 + \frac{W^2}{\left(2^b - 1\right)^2} = \sigma^2 + \delta^2, \quad \textit{for all } b > 0, \tag{11}$$

where $\delta^2 = W^2/(2^b - 1)^2$ denotes the upper bound of the quantization noise variance.

3.1.2 Quasi-BLUE Estimator

Suppose all the observations of K active sensors x_k $(k = 1, \cdots, K)$ are quantized into b_k-bit discrete messages $m_k(x_k, b_k)$ with the above probabilistic quantization scheme. Based on the quantized messages m_k, the quasi-BLUE estimator at the fusion center has the following form:

$$\bar{\theta}_K = \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2 + \delta_k^2}\right)^{-1} \sum_{k=1}^{K} \frac{m_k}{\sigma_k^2 + \delta_k^2}. \tag{12}$$

Notice that $\bar{\theta}_K$ is an unbiased estimator of θ since every m_k is unbiased. Moreover, the estimation MSE of the quasi-BLUE estimator [94] is

$$E(\bar{\theta}_K - \theta)^2 \leq \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2 + \delta_k^2}\right)^{-1}. \tag{13}$$

3.1.3 Distributed Estimation under Rate Constraints

With the probabilistic quantization scheme and the quasi-BLUE fusion rule, instead of the original problem in Equation (7), we turn to the following modified problem, which minimizes the bound of the estimation MSE under the rate constraint, i.e.,

$$\begin{aligned} \min \quad & \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2 + \frac{W^2}{(2^{b_k}-1)^2}} \right)^{-1}, \\ s.t. \quad & \sum_{k=1}^{K} b_k \leq B, \quad b_k \in \mathbb{Z}^+, \quad k = 1, \cdots, K, \end{aligned} \tag{14}$$

where B is the total bit rate constraint, K is the number of active sensors, and b_k is the quantization bit rate of the active sensor k. Both K and b_k ($k = 1, \cdots, K$) are to be optimized to minimize the estimation MSE. That is, adaptively select the subset of active sensors and the b_k-bit quantizer for each active sensor so that the estimation MSE at the fusion center is minimized.

In the following sections, we address this problem for homogeneous and heterogeneous sensor networks, respectively. To facilitate the solution, we first define an equivalent 1-bit MSE function.

3.2 Equivalent 1-bit MSE Function

As shown in Section 3.1.1, the b-bit quantized message from a sensor with observation noise variance σ^2 is an unbiased estimation of the parameter θ with the estimation MSE $D \leq \sigma^2 + W^2/(2^b - 1)^2$. We denote the estimation MSE bound as

$$f(\sigma^2, b) = \sigma^2 + \frac{W^2}{(2^b - 1)^2}. \tag{15}$$

Definition 3.1 (Equivalent 1-bit MSE function)**.** *For a sensor with b-bit quantization and observation noise variance σ^2, the equivalent 1-bit MSE function is defined as*

$$g(\sigma^2, b) = b \cdot f(\sigma^2, b) = b \left(\sigma^2 + \frac{W^2}{(2^b - 1)^2} \right). \tag{16}$$

With this definition, the estimation MSE of the quasi-BLUE estimator in Equation (13) can be rewritten as

$$
\begin{aligned}
E(\bar{\theta}_K-\theta)^2 &\le \left(\sum_{k=1}^{K}\frac{1}{\sigma_k^2+\delta_k^2}\right)^{-1} = \left(\underbrace{\frac{1}{f(\sigma_1^2,b_1)}+\cdots+\frac{1}{f(\sigma_K^2,b_K)}}_{K}\right)^{-1} \\
&= \left(\underbrace{\underbrace{\frac{1}{g(\sigma_1^2,b_1)}+\cdots+\frac{1}{g(\sigma_1^2,b_1)}}_{b_1}+\cdots+\underbrace{\frac{1}{g(\sigma_K^2,b_K)}+\cdots\frac{1}{g(\sigma_K^2,b_K)}}_{b_K}}_{b_1+\cdots+b_K=B}\right)^{-1}.
\end{aligned}
\tag{17}
$$

From the estimation MSE aspect, a b-bit quantization sensor with the estimation MSE $f(\sigma^2, b)$ can be treated as b equivalent 1-bit quantization sensors, each with the same estimation MSE $g(\sigma^2, b)$ defined as above. This is why the function $g(\sigma^2, b)$ is called the equivalent 1-bit MSE function. Further, the rate-constrained distributed estimation system with K sensors under the total bit rate constraint B can be treated as a distributed estimation system with B equivalent 1-bit quantization sensors, where B is a constant and K is a variable.

Based on the definition of the equivalent 1-bit MSE function $g(\sigma^2, b)$, it is easy to show that it is convex over b, as in Proposition 3.1. Further, we define the optimal equivalent 1-bit MSE function $g^{opt}(\sigma^2)$ and the corresponding optimal quantization bit rate $b^{opt}(\sigma^2)$ for each sensor with observation noise variance σ^2 as

$$
\begin{aligned}
b^{opt}(\sigma^2) &= \arg\min_{b\in\mathbb{Z}^+} g(\sigma^2,b) = \arg\min_{b\in\mathbb{Z}^+}\left[b\left(\sigma^2+\frac{W^2}{(2^b-1)^2}\right)\right], \\
g^{opt}(\sigma^2) &= \min_{b\in\mathbb{Z}^+} g(\sigma^2,b) = g(\sigma^2, b^{opt}(\sigma^2)),
\end{aligned}
\tag{18}
$$

where the minimization in Equation (18) involves just a simple one-dimensional numerical search over $b \in \mathbb{Z}^+$.

Proposition 3.1. *The functions $g(\sigma^2, b)$, $b^{opt}(\sigma^2)$ and $g^{opt}(\sigma^2)$ have the following properties:*

1. *$g(\sigma^2, b)$ increases over σ^2, i.e.,*

$$g(\sigma_i^2, b) < g(\sigma_j^2, b), \quad \text{if } \sigma_i^2 < \sigma_j^2.$$

2. *$g(\sigma^2, b)$ is convex over b ($b > 0$), i.e.,*

$$\begin{cases} g(\sigma^2, b_1) > g(\sigma^2, b_2) \geq g(\sigma^2, b^{opt}(\sigma^2)), & \text{if } b_1 < b_2 \leq b^{opt}(\sigma^2), \\ g(\sigma^2, b^{opt}(\sigma^2)) \leq g(\sigma^2, b_3) < g(\sigma^2, b_4), & \text{if } b^{opt}(\sigma^2) \leq b_3 < b_4. \end{cases}$$

3. *$g^{opt}(\sigma^2)$ increases over σ^2, i.e.,*

$$g^{opt}(\sigma_i^2) < g^{opt}(\sigma_j^2), \quad \text{if } \sigma_i^2 < \sigma_j^2.$$

Proof. Proposition 3.1 is easy to prove as follows:

1. $g(\sigma^2, b)$ increases over σ^2 by definition.

2. $g(\sigma^2, b)$ is convex over b because $\partial^2 g(\sigma^2, b)/\partial b^2 > 0$.

3. For the observation noise variances σ_i^2 and σ_j^2, denote the corresponding optimal quantization bit rates as $b^{opt}(\sigma_i^2)$ and $b^{opt}(\sigma_j^2)$; then

$$\begin{aligned} g^{opt}(\sigma_i^2) &= g(\sigma_i^2, b^{opt}(\sigma_i^2)) \\ &\leq g(\sigma_i^2, b^{opt}(\sigma_j^2)) \\ &< g(\sigma_j^2, b^{opt}(\sigma_j^2)) \\ &= g^{opt}(\sigma_j^2). \end{aligned}$$

□

It is also easy to see that b^{opt} depends on the signal-to-noise ratio (SNR), defined as $SNR = 10 \log_{10}(W^2/\sigma^2)$. Figure 5 shows the optimal quantization bit rate b^{opt} under different SNRs. With the optimal quantization bit rate, the variance of the quantized message is shown in Lemma 3.2 and Figure 6.

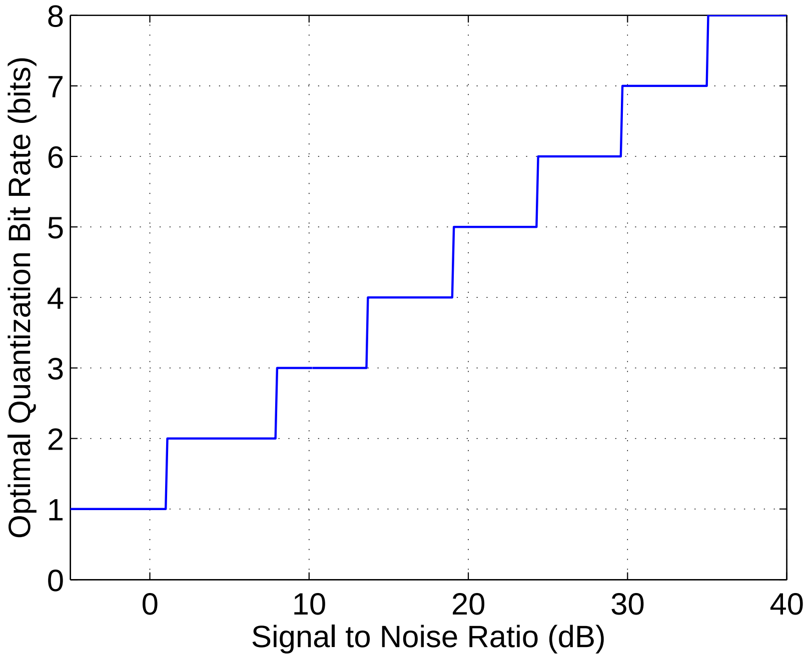


Figure 5: The optimal quantization bit rate versus signal-to-noise ratio (SNR).

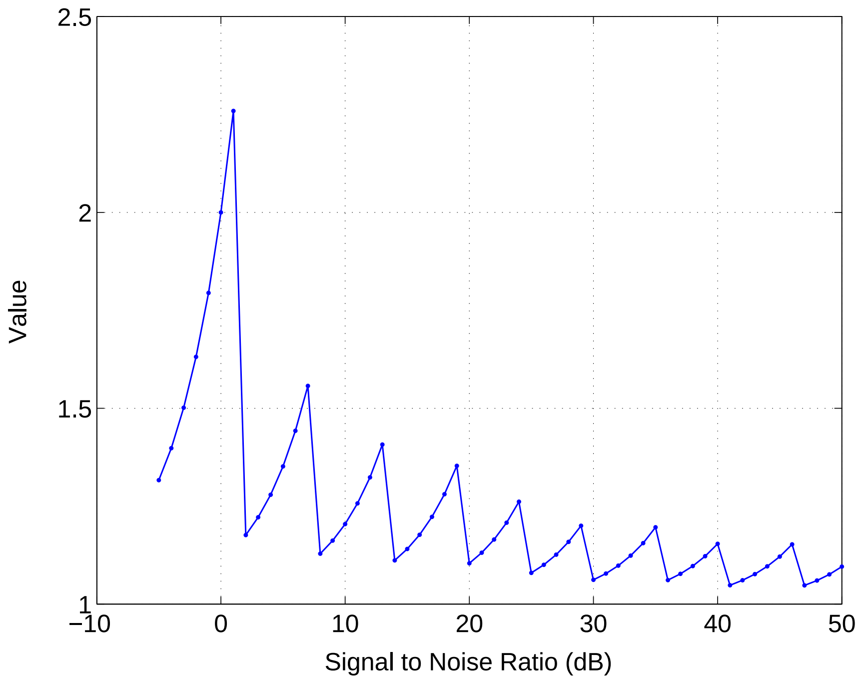


Figure 6: The value of $f(\sigma^2, b^{opt}(\sigma^2))/\sigma^2$ versus signal-to-noise ratio (SNR) defined as $SNR = 10\log_{10}(W^2/\sigma^2)$.

Lemma 3.2. *$f(\sigma^2, b)$ is the estimation MSE bound function defined in Equation (15), and $b^{opt}(\sigma^2)$ is the optimal quantization bit rate defined in Equation (18); then*

$$f(\sigma^2, b^{opt}(\sigma^2)) < 2.2872\sigma^2. \tag{19}$$

Proof. See Appendix 3.8.1. The numerical relationship between $f(\sigma^2, b^{opt}(\sigma^2))$ and σ^2 is shown in Figure 6, from which we can see that $f(\sigma^2, b^{opt}(\sigma^2))/\sigma^2$ is less than 2.2872 and is very close to 1 when the signal-to-noise ratio is high.

□

3.3 Distributed Estimation in Homogeneous Sensor Networks

In this section, we address the distributed estimation under the total bit rate constraint for homogeneous sensor networks, where every sensor has the same observation noise variance, that is, $\sigma_k^2 = \sigma^2$ $(k = 1, \cdots, N)$. In this special case, each active sensor should quantize its observation with the same bit rate $b_k = b$ to minimize the estimation MSE. As a result, the number of active sensors is B/b and the estimation MSE function is simplified to

$$E(\bar{\theta}_K - \theta)^2 \leq \left(\sum_{k=1}^{K} \frac{1}{\sigma^2 + \dfrac{W^2}{(2^{b_k} - 1)^2}} \right)^{-1} = \frac{b\left(\sigma^2 + \dfrac{W^2}{(2^b - 1)^2}\right)}{B}. \tag{20}$$

It is noted that the numerator of the optimization target function in Equation (20) is just the equivalent 1-bit MSE function $g(\sigma^2, b)$ defined in Section 3.2. Hence, for homogeneous sensor networks, the optimal distributed estimation under the total bit rate constraint B can be treated in an alternative way, where there are B identical equivalent 1-bit quantization sensors; thus, minimizing the final estimation MSE bound becomes minimizing the equivalent 1-bit MSE function. The method based on the equivalent 1-bit MSE function is stated as follows:

1. For each sensor, the optimal quantization bit rate is identical and obtained by minimizing the corresponding equivalent 1-bit MSE function, i.e.,

$$b^{opt} = \arg\min_{b \in \mathbb{Z}^+} g(\sigma^2, b) = \arg\min_{b \in \mathbb{Z}^+} \left[b \left(\sigma^2 + \frac{W^2}{(2^b - 1)^2} \right) \right], \tag{21}$$

where the minimization involves just a simple one-dimensional numerical search.

2. The total number of active sensors (K^{opt}) under the total bit rate constraint B is

$$K^{opt} = \left\lfloor \frac{B}{b^{opt}} \right\rfloor. \tag{22}$$

It is obvious that the proposed method based on the equivalent 1-bit MSE function is optimal if B/b^{opt} is an integer, i.e., $K^{opt} = B/b^{opt}$. For the case where B/b^{opt} is not an integer, there are $b_r = B - K^{opt}b^{opt} < b^{opt}$ bits remaining after the two steps above; then, we allocate these remaining b_r bits to one more sensor. Though it is not necessarily optimal, it is quasi-optimal because the estimation MSE bound is between the optimal solutions under the total bit rate constraint $K^{opt}b^{opt}$ and $(K^{opt}+1)b^{opt}$, where $K^{opt}b^{opt} < B < (K^{opt}+1)b^{opt}$.

Remark 3.1. *It is noted that the proposed method based on the equivalent 1-bit MSE function can be implemented in a fully distributed manner. First, the optimal quantization bit rate b^{opt} of each sensor can be obtained locally by minimizing its corresponding equivalent 1-bit MSE function. Second, with the given total bit rate constraint B and the optimal quantization bit rate of each sensor b^{opt}, the number of active sensors at each task period is $K^{opt} = B/b^{opt}$ (we assume B/b^{opt} is integer here); then, each sensor will be in the active mode with a probability of $p = K^{opt}/N$, where N is the total number of sensors. Assume each sensor has a unique index $i \in [0, N-1]$. We design a periodic scheduling for each sensor i as follows: sensor i is active when $t \in [kN+i, kN+i+K^{opt}](k \in \mathbb{Z})$; otherwise, it is in sleep mode. With the given scheduling scheme, there are K^{opt} active sensors at any task period t and each sensor* *will be active for K^{opt} task periods in any consecutive N task duration. Therefore, the energy cost at each sensor node is even, and the network lifetime is maximized, which is defined as the time for the first sensor node in the network to deplete.*

3.4 Distributed Estimation in Heterogeneous Sensor Networks

In this section, we address the general distributed estimation under the total bit rate constraint for heterogeneous sensor networks. Assuming the observation noise variance for every sensor is σ_k^2 $(k = 1, \cdots, N)$, respectively. Without loss of generality, we assume $\sigma_1^2 \leq \cdots \leq \sigma_N^2$; so, if K sensors are needed, we just simply choose the first K sensors, which will minimize the estimation MSE. This scenario leads to the general problem stated in Equation (14).

To find the optimal number of active sensors and the corresponding optimal quantization bit rate for each active sensor that minimizes the estimation MSE bound at the fusion center, we adopt the Lagrange multiplier method to solve the following equivalent problem:

$$\begin{aligned} &\max \quad \sum_{k=1}^{K} \left(\sigma_k^2 + \frac{W^2}{(2^{b_k}-1)^2} \right)^{-1}, \\ &s.t. \quad \sum_{k=1}^{K} b_k \leq B, \quad b_k \in \mathbb{Z}^+, \quad k = 1, \cdots, K. \end{aligned} \tag{23}$$

The Lagrangian G is given as

$$G(b_k, \lambda) = \sum_{k=1}^{K} \left(\sigma_k^2 + \frac{W^2}{(2^{b_k}-1)^2} \right)^{-1} + \lambda \left(\sum_{k=1}^{K} b_k - B \right), \tag{24}$$

which leads to the following two optimization conditions:

$$\begin{aligned} &\frac{\partial \left(\sigma_k^2 + \frac{W^2}{(2^{b_k}-1)^2} \right)^{-1}}{\partial b_k} + \lambda = 0, \ \forall k \in [1, K], \quad \text{and} \\ &\sum_{k=1}^{K} b_k = B. \end{aligned} \tag{25}$$

Unfortunately, the optimal solution b_k $(k = 1, \cdots, K)$ cannot be found in a closed-form from Equation (25). Instead, we develop a quasi-optimal method to solve the problem, which is also based on the equivalent 1-bit MSE function. The procedure is stated as follows:

1. For each sensor, its optimal quantization bit rate b_k^{opt} $(k = 1, \cdots, N)$ is obtained by minimizing its corresponding equivalent 1-bit MSE function, i.e.,

$$b_k^{opt} = \arg\min_{b_k \in \mathbb{Z}^+} g(\sigma_k^2, b_k) = \arg\min_{b_k \in \mathbb{Z}^+} \left[b_k \left(\sigma_k^2 + \frac{W^2}{(2^{b_k} - 1)^2} \right) \right], \tag{26}$$

   where the minimization involves just a simple one-dimensional numerical search.

2. Sort all the sensors by their observation noise variances from the smallest to the largest and denote the total number of active sensors as K^{opt}. Then, k^{opt} is determined by the total bit rate constraint B as follows:

$$K^{opt} = \max K \quad s.t. \quad \sum_{k=1}^{K} b_k^{opt} \leq B. \tag{27}$$

In summary, the whole solution is that the K^{opt} sensors with the smallest observation noise variances are chosen to quantize and transmit their observations with the quantization bit rate b_k^{opt} $(k = 1, \cdots, K^{opt})$. To implement the described algorithm above, each sensor needs to decide (i) whether it should be active or not and (ii) its quantization bit rate if it is active. Both tasks can be achieved in a distributed manner as follows:

- As shown in Equation (27), the subset of active sensors is determined at the fusion center based on the collected network information and the total bit rate constraint B. Denote the maximum observation noise variance of all active sensors as

$$\sigma_{th}^2 = \sigma_{K^{opt}}^2. \tag{28}$$

  Then, the fusion center broadcasts the threshold σ_{th}^2 to all local sensors. Upon receiving the threshold, each sensor compares the threshold with its own observation noise variance σ_k^2. If $\sigma_k^2 \leq \sigma_{th}^2$, sensor k is active; otherwise, it is inactive.

- As shown in Equation (26), the optimal quantization bit rate b_k^{opt} of sensor $k \in [1, N]$ depends only on its own observation noise variance; thus, it can be computed locally at each sensor without requiring information from other sensors.

Next, we will analyze the estimation MSE bound of the proposed method in the following theorem. To simplify the statement, we assume $\sum_{k=1}^{K^{opt}} b_k^{opt} = B$ in the subsequent analysis.

Theorem 3.1. *The estimation MSE of the proposed method based on the equivalent 1-bit MSE function under the total bit rate constraint B is*

$$\left(\sum_{k=1}^{K^{opt}} \frac{1}{\sigma_k^2} \right)^{-1} < E(\bar{\theta}_P - \theta)^2 < 2.2872 \left(\sum_{k=1}^{K^{opt}} \frac{1}{\sigma_k^2} \right)^{-1}, \tag{29}$$

where $\bar{\theta}_P$ denotes the estimation of the parameter θ by the proposed method, and K^{opt} is the optimal number of active sensors, obtained in Equation (27).

Proof. The left part of the theorem is obvious since $\left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2 \right)^{-1}$ is the lower bound of the estimation MSE of the BLUE estimator using K^{opt} active sensors. To prove the right part of the theorem, by Lemma 3.2,

$$\begin{aligned} E(\bar{\theta}_p - \theta)^2 &= \left(\sum_{k=1}^{K^{opt}} \frac{1}{f(\sigma_k^2, b^{opt}(\sigma_k^2))} \right)^{-1} \\ &< \left(\sum_{k=1}^{K^{opt}} \frac{1}{2.2872\sigma_k^2} \right)^{-1} \\ &= 2.2872 \left(\sum_{k=1}^{K^{opt}} \frac{1}{\sigma_k^2} \right)^{-1}. \end{aligned} \tag{30}$$

□

This theorem gives the estimation MSE bound of the proposed method. It is shown that the proposed method is quasi-optimal (up to a factor of 2.2872) compared with the BLUE estimator using the same subset of active sensors.

3.5 Rate-Distortion Bound Analysis

In the previous section, the performance bound of the proposed algorithm is analyzed. Nevertheless, the remaining question is what performance can be achieved if the total B bits are allocated to any number of sensors, say M sensors. More specifically, can a lower bound of the estimation MSE less than $D_0 \equiv \left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$ be achieved using M sensors under the total bit rate constraint B? It is obvious that a lower bound less than D_0 cannot be achieved if $M < K^{opt}$ sensors are used, regardless of the quantization bit rate of each active sensor. It is also obvious that a lower bound $\left(\sum_{k=1}^{M} 1/\sigma_k^2\right)^{-1}$ (less than D_0) can be achieved if $M > K^{opt}$ sensors are used and the quantization bit rate for each sensor is not limited. But under the total bit rate constraint B, whether a lower bound less than D_0 can be achieved if $M > K^{opt}$ sensors are used is a real question. To answer this question, we further analyze the lower bound of the estimation MSE by any quasi-BLUE estimation system with $M > K^{opt}$ active sensors under the total bit rate constraint B in the following theorem.

Theorem 3.2. *For any quasi-BLUE estimation system under the total bit rate constraint B, where there are $M > K^{opt}$ active sensors with the quantization bit rate b_k for sensor k and $\sum_{k=1}^{M} b_k = B$, the lower bound of the estimation MSE is*

$$E(\bar{\theta}_B - \theta)^2 > \left(\sum_{k=1}^{K^{opt}} \frac{1}{\sigma_k^2}\right)^{-1}, \qquad (31)$$

where $\bar{\theta}_B$ denotes the estimation of the parameter θ under the total bit rate constraint B, and K^{opt} is the optimal number of active sensors, obtained by the proposed algorithm, as shown in Equation (27), such that $\sum_{k=1}^{K^{opt}} b_k^{opt} = B$.

Proof. For any given estimation system as stated in the theorem, the basic idea to prove its estimation MSE $D_1 > \left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$ is to construct another corresponding quasi-BLUE estimation system such that its estimation MSE D_2 is smaller than D_1 but larger than $\left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$, i.e., $D_1 > D_2 > \left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$. The proof is based on the concept of the equivalent 1-bit MSE function. Refer to Appendix 3.8.2 for the details.

□

In conclusion, Theorem 3.1 shows that the estimation MSE bound of the proposed method is $\left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1} < E(\bar{\theta}_p - \theta)^2 < 2.2872 \left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$, and Theorem 3.2 shows that $\left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$ is the lower bound of the estimation MSE of any quasi-BLUE estimator under the total bit rate constraint B, regardless of the number of active sensors and the bit allocation among active sensors. Therefore, the proposed algorithm gives a quasi-optimal trade-off between the number of active sensors and the quantization bit rate of each sensor, and its estimation MSE is within a factor 2.2872 of the theoretical non-achievable lower bound.

3.6 Simulation Results

In this section, we present some simulation results for the proposed algorithms in Section 3.3 and 3.4, respectively.

3.6.1 Homogeneous Sensor Networks

In this section, we simulate a homogeneous sensor network with $N = 500$ sensors. Assume the range of the observation signal is $[-1, 1]$, i.e., $W = 1$. Define the signal-to-noise ratio (SNR) as

$$SNR = 10 \log_{10}(W^2/\sigma^2) \tag{32}$$

and generate different SNRs by changing the observation noise variance σ^2. Assuming the total bit rate constraint is $B = 500$ bits, Figure 7 shows the estimation MSE with different quantization bit rates for the active sensors under different SNRs. Notice that a different quantization bit rate for each sensor implies a different number of active sensors to perform the estimation task because of the total bit rate constraint B. For example, in the case of $SNR = 20\ dB$, 125 active sensors with 4-bit quantized message per sensor will produce the minimum estimation MSE under the total bit rate constraint $B = 500$, which is better than all the other possible cases, such as 500 sensors with 1-bit quantized message per sensor, 250 sensors with 2-bit quantized message per sensor, 62 sensors with 8-bit quantized message per sensor and so on. From the results shown in Figure 7, we also can see that for the low SNR case, such as 0 dB, 1-bit quantization per sensor leads to the minimum estimation MSE. On the contrary, for the high SNR case, multiple-bit quantization per sensor significantly decreases the estimation MSE compared with only 1-bit quantization per sensor under the same total bit rate constraint.

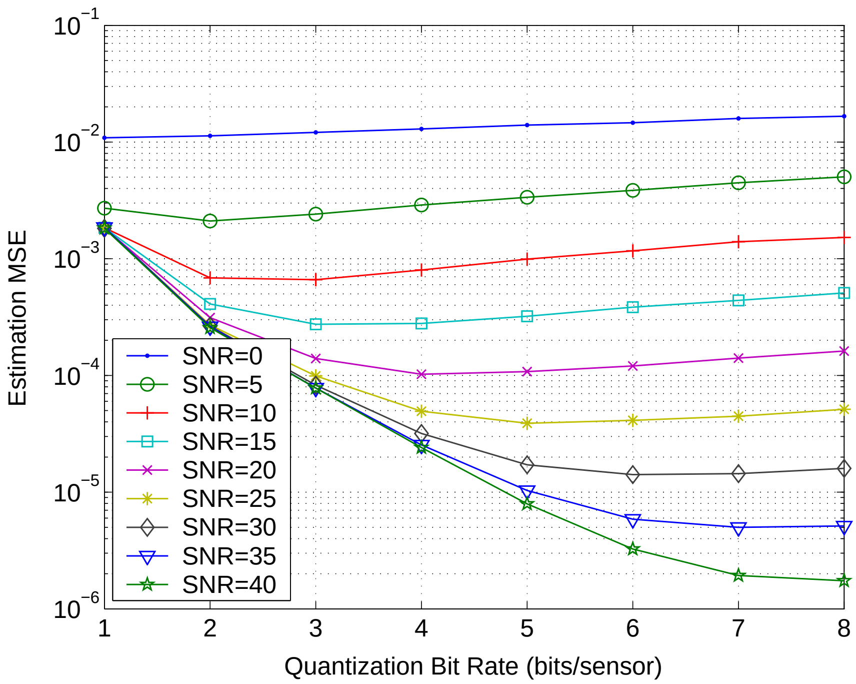


Figure 7: The estimation MSE versus the quantization bit rate per sensor under different signal-to-noise ratios (SNRs) and the total bit rate constraint $B = 500$ bits.

3.6.2 Heterogeneous Sensor Networks

In this section, we simulate a heterogeneous sensor network with $N = 500$ sensors. Assume the range of the observation signal is still $[-1, 1]$. We assume the observation noise variances to be a Chi-squared distribution with one degree of freedom. The distribution of the signal-to-noise ratio of the simulated heterogeneous network is shown in Figure 8(a). The optimal message length for each sensor can be computed by minimizing its corresponding equivalent 1-bit MSE function as shown in Section 3.2. The distribution of the optimal message length of the simulated heterogeneous network is shown in Figure 8(b). In the simulation, for any given total bit rate constraint, the proposed estimation method is implemented to determine the number of active sensors and the quantization bit rate for each active sensor to minimize the estimation MSE. Figure 9 shows the percentage of the active sensors under different total bit rate constraints.

To demonstrate the efficiency of the proposed method, we compare the proposed method with two kinds of uniform schemes:

1. *Uniform-I*: For the given total bit rate constraint, the same subset of active sensors as that used by the proposed method is used, but the quantization bit rate is uniform among all active sensors.

2. *Uniform-II*: All the sensors in the simulated heterogeneous sensor network are used and the quantization bit rate is uniform among all the sensors.

Figure 10 shows the estimation MSE by the proposed method, the *Uniform-I* method and the *Uniform-II* method, and the theoretical lower bound of the estimation MSE presented in Theorem 3.2 under the total bit rate constraint. From Figure 10, we can see that the proposed method outperforms the two uniform schemes. Further, it also can be seen that the estimation MSE of the proposed method is close to the theoretical non-achievable lower bound (about 1.1 times).

Note that both the proposed method and the *Uniform-I* method are based on the same subset of active sensors, and the only difference is that the optimal bit rate allocation is performed in the proposed method, while uniform bit rate allocation is

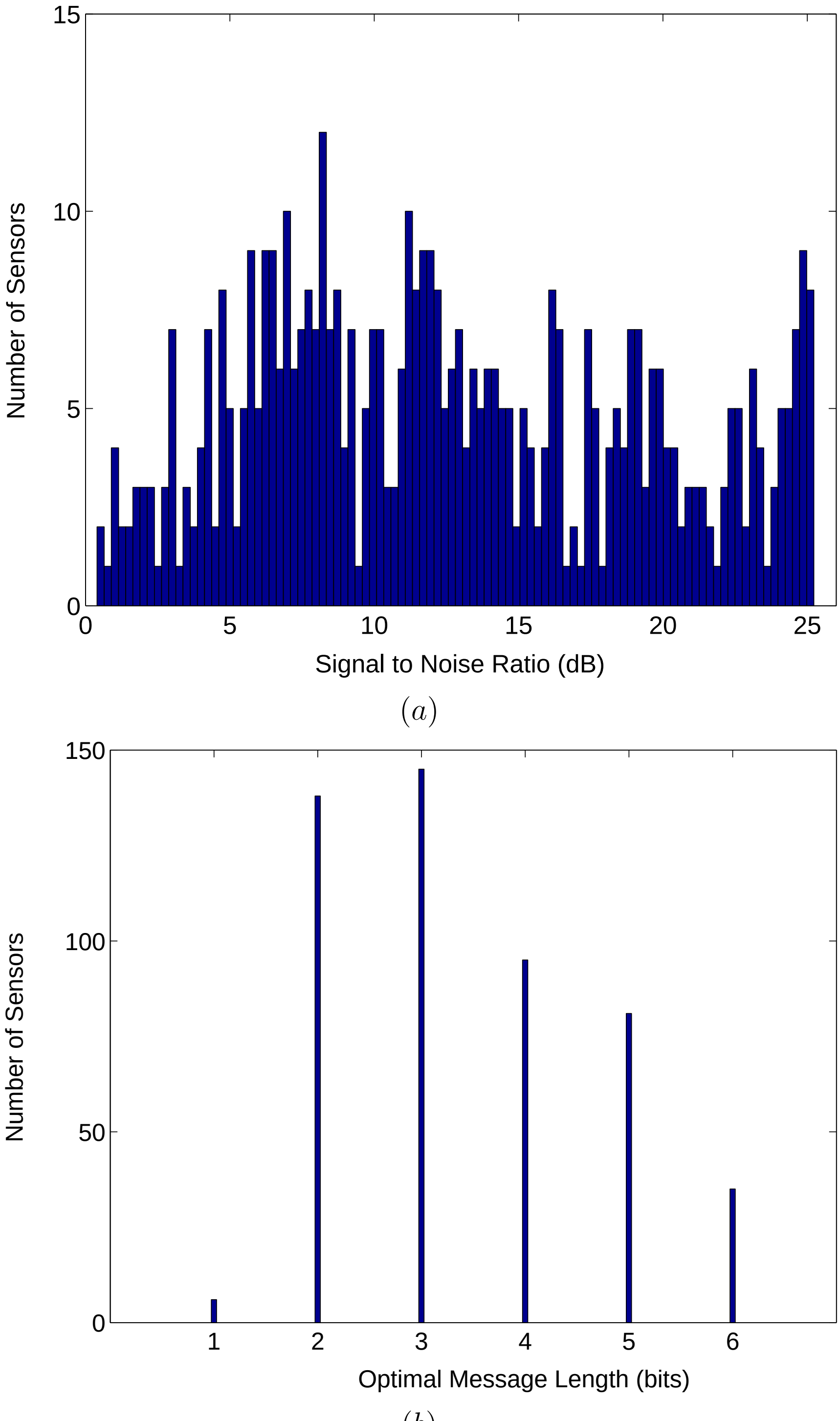


Figure 8: (a) Distribution of the signal-to-noise ratio of the simulated heterogeneous network; (b) Distribution of the optimal message length of the simulated heterogeneous network.

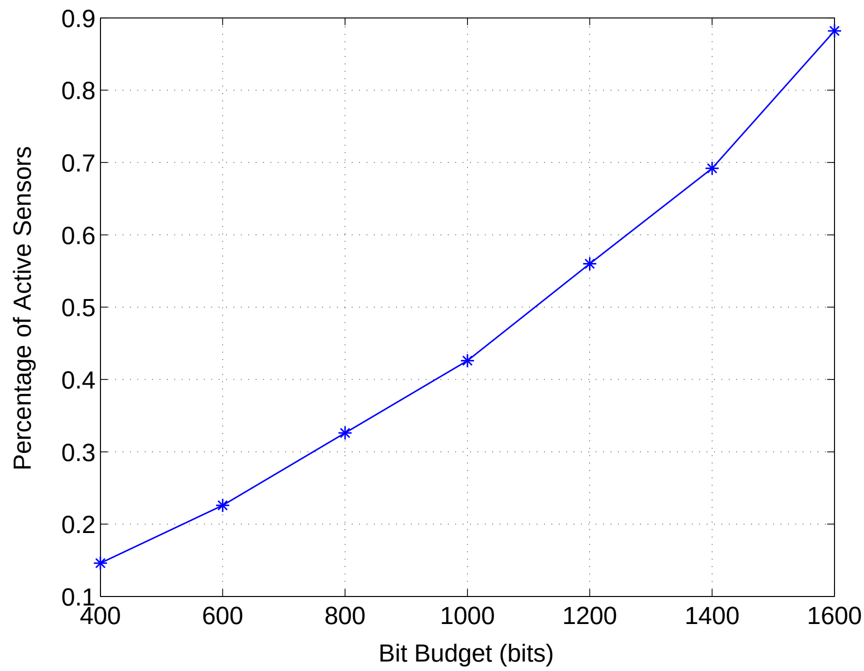


Figure 9: Percentage of the active sensors under different total bit rate constraints.

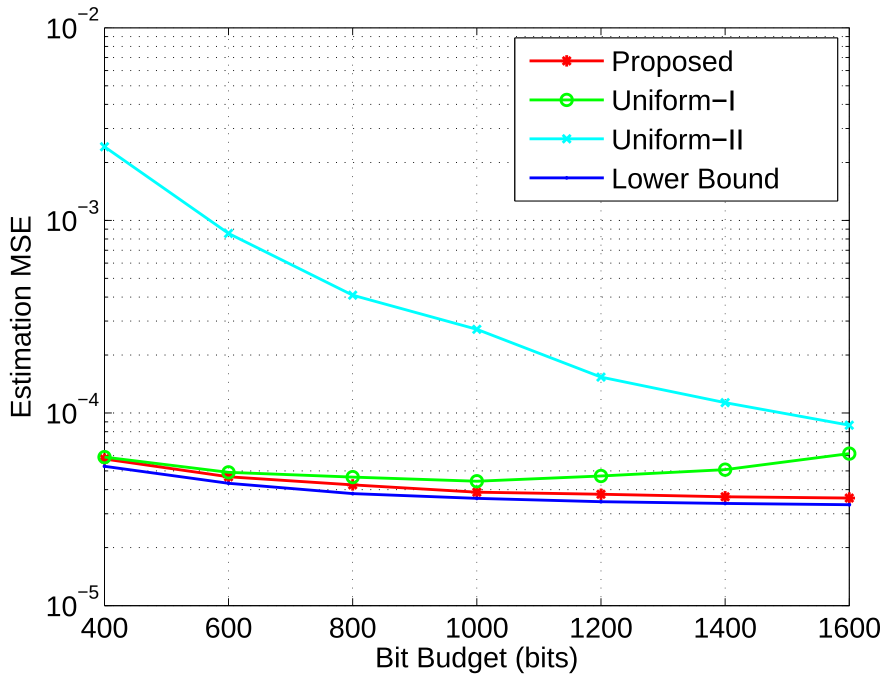


Figure 10: The estimation MSE by the proposed method, *Uniform-I* method, and *Uniform-II* method, and the theoretical non-achievable lower bound of the estimation MSE.

performed in the *Uniform-I* method. Because of the heterogeneity of the network, a better estimation performance is obtained by the proposed method.

Next, we further show how the heterogeneity of the sensor network influences the estimation performance. We define the normalized deviation of sensor noise variances as

$$\alpha = \frac{\sqrt{Var(\sigma^2)}}{E(\sigma^2)}, \tag{33}$$

which will be used as a measure of the heterogeneity of the sensor network. Also we define the reduction in the estimation MSE achieved by the proposed method compared with the *Uniform-I* method as

$$\beta = \frac{D_u - D_p}{D_u}, \tag{34}$$

where D_u denotes the estimation MSE by the *Uniform-I* method, and D_p denotes the estimation MSE by the proposed method. Figure 11 plots the estimation MSE reduction of the proposed method compared with the *Uniform-I* method versus the normalized deviation of sensor noise variances. From Figure 11, we conclude that the amount of estimation MSE reduction of the proposed method becomes more significant when the local sensor noise variances become more heterogeneous.

3.7 Summary

In this chapter, we considered the distributed estimation of a noise-corrupted deterministic parameter under the total bit rate constraint in wireless sensor networks. Because of the total bit rate constraint, a trade-off between the number of active sensors and the quantization bit rate of each active sensor is addressed to minimize the estimation MSE. To determine the optimal quantization bit rate of each sensor, a concept of the equivalent 1-bit MSE function is introduced, based on which an optimal rate-constrained distributed estimation algorithm for homogeneous sensor networks and a quasi-optimal rate-constrained distributed estimation algorithm for heterogeneous sensor networks are developed.

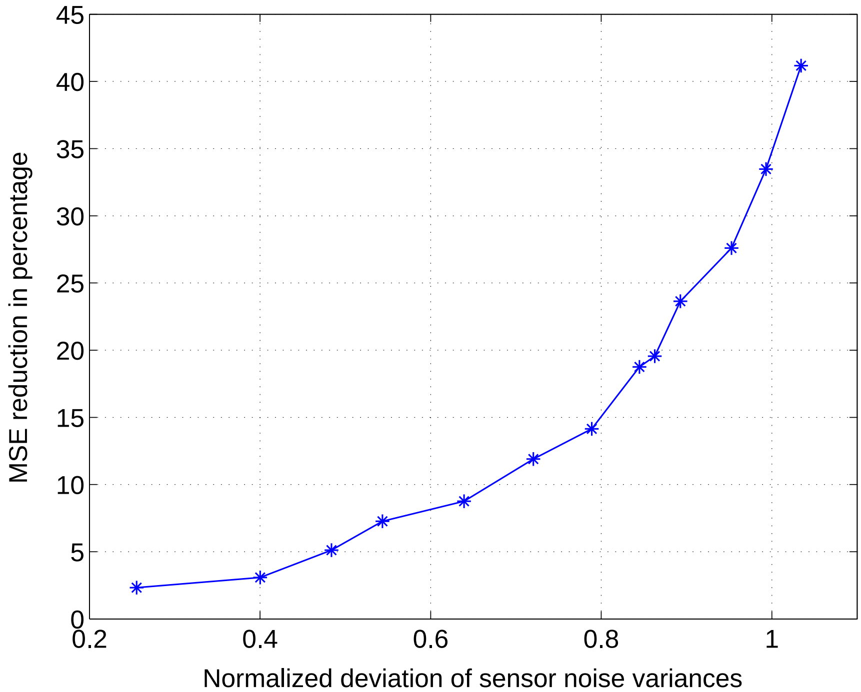


Figure 11: The estimation MSE reduction in percentage of the proposed method compared with the *Uniform-I* method under different normalized deviations of sensor noise variances.

Furthermore, a theoretical analysis on the rate-distortion bound for the distributed estimation is performed and a lower bound of the estimation MSE under any given total bit rate constraint is formulated. It is shown that the proposed algorithm is quasi-optimal within a factor 2.2872 of the theoretical lower bound. Simulation results also show that the proposed algorithm can achieve a significant amount of the estimation MSE reduction compared with several uniform schemes in which each sensor quantizes its observation with the same number of bits.

In this chapter, the distributed estimation problem is addressed from the rate-distortion point of view. By constraining the communication bit rate, the energy consumption is limited; thus, the network lifetime of wireless sensor networks is prolonged. In the next chapter, we will study the distributed estimation from the energy-distortion perspective to explicitly address the energy consumption and to maximize the energy-distortion performance.

3.8 Appendix

3.8.1 Proof of Lemma 3.2

To facilitate the subsequent analysis, we will relax the integer condition $b \in \mathbb{Z}^+$ in Equation (18) to $b \in \mathbb{R}^+$, i.e.,

$$l^{opt}(\sigma^2) = \arg\min_{b \in \mathbb{R}^+} g(\sigma^2, b) = \arg\min_{b \in \mathbb{R}^+} \left[b \left(\sigma^2 + \frac{W^2}{(2^b - 1)^2} \right) \right]. \tag{35}$$

Here, $l^{opt}(\sigma^2) \in \mathbb{R}^+$, while $b^{opt}(\sigma^2) \in \mathbb{Z}^+$ defined in Equation (18). It is obvious that $b^{opt}(\sigma^2) = \lfloor l^{opt}(\sigma^2) \rfloor$ or $\lceil l^{opt}(\sigma^2) \rceil$ since $g(\sigma^2, b)$ is convex over b as stated in Proposition 3.1, where $\lfloor l^{opt}(\sigma^2) \rfloor$ denotes the maximum integer no more than $l^{opt}(\sigma^2)$, and $\lceil l^{opt}(\sigma^2) \rceil$ denotes the minimum integer no less than $l^{opt}(\sigma^2)$.

To solve $l^{opt}(\sigma^2)$ from Equation (35), we need to solve $\partial g(\sigma^2, b)/\partial b = 0$, which leads to the following equation:

$$(2^{l^{opt}(\sigma^2)} - 1)^3 - \frac{W^2}{\sigma^2}[2 \ln 2 \cdot l^{opt}(\sigma^2) \cdot 2^{l^{opt}(\sigma^2)} - 2^{l^{opt}(\sigma^2)} + 1] = 0, \tag{36}$$

so

$$W^2 = \sigma^2 \cdot \frac{(2^{l^{opt}(\sigma^2)} - 1)^3}{2 \ln 2 \cdot l^{opt}(\sigma^2) \cdot 2^{l^{opt}(\sigma^2)} - 2^{l^{opt}(\sigma^2)} + 1}. \tag{37}$$

By solving Equation (18) with Equation (37), we get the following relationship between $b^{opt}(\sigma^2)$ and $l^{opt}(\sigma^2)$:

$$b^{opt}(\sigma^2) = \begin{cases} 1, & \text{if } 0 < l^{opt}(\sigma^2) < 1.41 \\ 2, & \text{if } 1.41 \le l^{opt}(\sigma^2) < 2.44 \\ 3, & \text{if } 2.44 \le l^{opt}(\sigma^2) < 3.45 \\ \text{others.} & \end{cases} \tag{38}$$

Based on the above results, now we turn to $f(\sigma^2, b^{opt}(\sigma^2))$,

$$\begin{aligned} f(\sigma^2, b^{opt}(\sigma^2)) &= \sigma^2 + \frac{W^2}{(2^{b^{opt}(\sigma^2)} - 1)^2} \\ &= \sigma^2 \left[1 + \frac{(2^{l^{opt}(\sigma^2)} - 1)^3}{2 \ln 2 \cdot l^{opt}(\sigma^2) \cdot 2^{l^{opt}(\sigma^2)} - 2^{l^{opt}(\sigma^2)} + 1} \cdot \frac{1}{(2^{b^{opt}(\sigma^2)} - 1)^2} \right] \\ &= \sigma^2 \left[1 + y(l^{opt}(\sigma^2)) \cdot \frac{1}{(2^{b^{opt}(\sigma^2)} - 1)^2} \right], \end{aligned} \tag{39}$$

where

$$y(l^{opt}(\sigma^2)) \equiv \frac{(2^{l^{opt}(\sigma^2)}-1)^3}{2\ln 2 \cdot l^{opt}(\sigma^2) \cdot 2^{l^{opt}(\sigma^2)} - 2^{l^{opt}(\sigma^2)} + 1},$$

and it is easy to verify that $y(l^{opt}(\sigma^2))$ increases over $l^{opt}(\sigma^2) > 0$. Next, we discuss four cases:

1. $0 < l^{opt}(\sigma^2) < 1.41$

   In this case, $b^{opt}(\sigma^2) = 1$ as shown in Equation (38), so

$$\begin{aligned} f(\sigma^2, b^{opt}(\sigma^2)) &= \sigma^2 \left[1 + y(l^{opt}(\sigma^2)) \cdot \frac{1}{(2^{b^{opt}(\sigma^2)}-1)^2}\right] \\ &\le \sigma^2 \left[1 + y(l^{opt}(\sigma^2))\right] \\ &\overset{(*)}{<} 2.2872\sigma^2, \end{aligned} \tag{40}$$

   where the step $(*)$ holds because $y(l^{opt}(\sigma^2))$ increases over $l^{opt}(\sigma^2)$ and $y(1.41) < 1.2872$.

2. $1.41 \le l^{opt}(\sigma^2) < 2.44$

   In this case, $b^{opt}(\sigma^2) = 2$ as shown in Equation (38), so

$$\begin{aligned} f(\sigma^2, b^{opt}(\sigma^2)) &= \sigma^2 \left[1 + y(l^{opt}(\sigma^2)) \cdot \frac{1}{(2^{b^{opt}(\sigma^2)}-1)^2}\right] \\ &\le \sigma^2 \left[1 + y(l^{opt}(\sigma^2)) \cdot \frac{1}{9}\right] \\ &\overset{(*)}{<} 1.6918\sigma^2, \end{aligned} \tag{41}$$

   where the step $(*)$ holds because $y(l^{opt}(\sigma^2))$ increases over $l^{opt}(\sigma^2)$ and $y(2.44) < 6.2265$.

3. $2.44 \le l^{opt}(\sigma^2) < 3.45$

   In this case, $b^{opt}(\sigma^2) = 3$ as shown in Equation (38), so

$$\begin{aligned} f(\sigma^2, b^{opt}(\sigma^2)) &= \sigma^2 \left[1 + y(l^{opt}(\sigma^2)) \cdot \frac{1}{(2^{b^{opt}(\sigma^2)}-1)^2}\right] \\ &\le \sigma^2 \left[1 + y(l^{opt}(\sigma^2)) \cdot \frac{1}{49}\right] \\ &\overset{(*)}{<} 1.4717\sigma^2, \end{aligned} \tag{42}$$

where the step $(*)$ holds because $y(l^{opt}(\sigma^2))$ increases over $l^{opt}(\sigma^2)$ and $y(3.45) < 23.115$.

4. $l^{opt}(\sigma^2) \geq 3.45$

   From the definition of $l^{opt}(\sigma^2)$ and $b^{opt}(\sigma^2)$, it is obvious that $b^{opt}(\sigma^2) > l^{opt}(\sigma^2) - 1$, so

$$\begin{aligned} f(\sigma^2, b^{opt}(\sigma^2)) &= \sigma^2 \left[1 + \frac{(2^{l^{opt}(\sigma^2)} - 1)^3}{2\ln 2 \cdot l^{opt}(\sigma^2) \cdot 2^{l^{opt}(\sigma^2)} - 2^{l^{opt}(\sigma^2)} + 1} \cdot \frac{1}{(2^{b^{opt}(\sigma^2)} - 1)^2} \right] \\ &< \sigma^2 \left[1 + \frac{(2^{l^{opt}(\sigma^2)} - 1)^3}{2\ln 2 \cdot l^{opt}(\sigma^2) \cdot 2^{l^{opt}(\sigma^2)} - 2^{l^{opt}(\sigma^2)} + 1} \cdot \frac{1}{(2^{l^{opt}(\sigma^2)-1} - 1)^2} \right] \\ &\overset{(*)}{\leq} 2.1599\sigma^2, \end{aligned} \tag{43}$$

   where the step $(*)$ holds because

$$z(l^{opt}(\sigma^2)) \equiv \frac{(2^{l^{opt}(\sigma^2)} - 1)^3}{2\ln 2 \cdot l^{opt}(\sigma^2) \cdot 2^{l^{opt}(\sigma^2)} - 2^{l^{opt}(\sigma^2)} + 1} \cdot \frac{1}{(2^{l^{opt}(\sigma^2)-1} - 1)^2}$$

   decreases over $l^{opt}(\sigma^2)$ and $z(3.45) < 1.1599$.

From above, for any given σ^2, we get

$$f(\sigma^2, b^{opt}(\sigma^2)) < 2.2872\sigma^2. \tag{44}$$

3.8.2 Proof of Theorem 3.2

We begin by introducing two definitions and two corresponding lemmas required by the proof of this theorem.

Definition 3.2 (Pairwise Bit Rate Exchange)**.** *Assuming there are two sensors i and j with the observation noise variances $\sigma_i^2 \leq \sigma_j^2$ and the quantization bit rates $b_i < b_j$ in a sensor network with M sensors to estimate an unknown parameter, then exchange the quantization bit rates of the two sensors, i.e., sensor i quantizes its observation using $b_i' = b_j$ bits and sensor j quantizes its observation using $b_j' = b_i$ bits. We call this operation as pairwise bit rate exchange.*

Lemma 3.3. *Let D denote the estimation MSE bound before a pairwise bit rate exchange operation and D_{ex} denote the estimation MSE bound after the exchange operation, then $D_{ex} \leq D$.*

Proof. Let $D = 1/D'$, and $D_{ex} = 1/D'_{ex}$, then

$$\begin{aligned} D' &= \sum_{\substack{k=1 \\ k\neq i,j}}^{M} \frac{1}{\sigma_k^2 + \delta_k^2} + \frac{1}{\sigma_i^2 + \dfrac{W^2}{(2^{b_i}-1)^2}} + \frac{1}{\sigma_j^2 + \dfrac{W^2}{(2^{b_j}-1)^2}}, \\ D'_{ex} &= \sum_{\substack{k=1 \\ k\neq i,j}}^{M} \frac{1}{\sigma_k^2 + \delta_k^2} + \frac{1}{\sigma_i^2 + \dfrac{W^2}{(2^{b_j}-1)^2}} + \frac{1}{\sigma_j^2 + \dfrac{W^2}{(2^{b_i}-1)^2}}. \end{aligned} \tag{45}$$

Let $u = \sigma_i^2$, $v = \sigma_j^2$, $t = W^2/(2^{b_i}-1)^2$ and $s = W^2/(2^{b_j}-1)^2$, then it can be easily verified that $D' - D'_{ex} \leq 0$, thus, $D_{ex} \leq D$, by the following algebra fact: For fixed positive numbers s, t, u, v with $u \leq v$, then

$$\frac{1}{u+s} + \frac{1}{v+t} \geq \frac{1}{u+t} + \frac{1}{v+s}, \quad \text{if } t > s.$$

$\square$

Definition 3.3 (Equivalent 1-bit Quantization Sensor Replacement)**.** *In a sensor network with M sensors to estimate an unknown parameter, if there are two sensors $i \in [1, \cdots, K^{opt}]$ and $j \in [K^{opt}+1, \cdots, N]$ with the observation noise variances $\sigma_i^2 \leq \sigma_j^2$ and the quantization bit rates $b_i \geq b_j$, and $b_i < b_i^{opt}$ (that is, $b_i + 1 \leq b_i^{opt}$), then we replace an equivalent 1-bit quantization sensor corresponding to sensor j by increasing the quantization bit rate of sensor i by 1, that is, sensor i quantizes its observation using $b'_i = b_i + 1$ bits. We call this operation as equivalent 1-bit quantization sensor replacement.*

Lemma 3.4. *Let D denote the estimation MSE bound before an equivalent 1-bit quantization sensor replacement and D_{re} denote the estimation MSE bound after the replacement, then $D_{re} < D$.*

Proof. Let $D = 1/D'$, and $D_{re} = 1/D'_{re}$, then

$$\begin{aligned} D' &= \sum_{\substack{k=1\\k\neq i,j}}^{M} \frac{b_k}{g(\sigma_k^2, b_k)} + \frac{b_i}{g(\sigma_i^2, b_i)} + \frac{b_j}{g(\sigma_j^2, b_j)}, \\ D'_{re} &= \sum_{\substack{k=1\\k\neq i,j}}^{M} \frac{b_k}{g(\sigma_k^2, b_k)} + \frac{b_i+1}{g(\sigma_i^2, b_i+1)} + \frac{b_j-1}{g(\sigma_j^2, b_j)}, \end{aligned} \tag{46}$$

so

$$\begin{aligned} D' - D'_{re} &= \frac{b_i}{g(\sigma_i^2, b_i)} - \frac{b_i+1}{g(\sigma_i^2, b_i+1)} + \frac{1}{g(\sigma_j^2, b_j)} \\ &= \left(\frac{b_i}{g(\sigma_i^2, b_i)} - \frac{b_i}{g(\sigma_i^2, b_i+1)}\right) + \left(\frac{1}{g(\sigma_j^2, b_j)} - \frac{1}{g(\sigma_i^2, b_i+1)}\right) \\ &\overset{(*)}{<} 0 + 0 = 0, \end{aligned} \tag{47}$$

where the step $(*)$ holds because

1. $b_i < b_i + 1 \leq b_i^{opt}$, so from Proposition 3.1, we get

$$g(\sigma_i^2, b_i) > g(\sigma_i^2, b_i + 1). \tag{48}$$

2. $b_j \leq b_i < b_i + 1 \leq b_i^{opt}$, and $\sigma_i^2 \leq \sigma_j^2$, so from Proposition 3.1, we get

$$g(\sigma_j^2, b_j) \geq g(\sigma_i^2, b_j) \geq g(\sigma_i^2, b_i) > g(\sigma_i^2, b_i + 1). \tag{49}$$

So $D' < D'_{re}$, and $D_{re} < D$. □

Now, we begin to prove Theorem 3.2. First, it is noted that $D_0 = \left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$ is just the lower bound of the estimation MSE of BLUE estimator using K^{opt} sensors with observation noise variances $\sigma_1^2, \cdots, \sigma_{K^{opt}}^2$ respectively. Next, we will show that D_0 is also the lower bound of the estimation MSE of any quasi-BLUE estimator using $M > K^{opt}$ sensors under the total bit rate constraint B, i.e., $\sum_{k=1}^{M} b_k = B$.

Assuming M sensors $i_1, \cdots, i_{K^{opt}}, \cdots, i_M$ ($i_1 < \cdots < i_{K^{opt}} < \cdots < i_M$) are used, and the corresponding observation noise variance are $\sigma_{i_1}^2 < \cdots < \sigma_{i_{K^{opt}}}^2 < \cdots < \sigma_{i_M}^2$, respectively, it is obvious that $i_k \geq k$ and $\sigma_{i_k}^2 \geq \sigma_k^2$. The quantization bit rates are

$b_1, \cdots, b_{K^{opt}}, \cdots, b_M$, respectively, and $\sum_{k=1}^{M} b_k = B$. Let D_1 denote the estimation MSE bound under this condition (denoted as C_1).

Step 1: Considering replace the active sensor i_k ($k = 1, \cdots, M$) in the condition C_1 by the sensor k ($k = 1, \cdots, M$), while the quantization bit rate doesn't change. That is to say, the first M sensors are active to observe and quantize their observations using $b_1, \cdots, b_M$ bits, respectively. Let D_2 denote the estimation MSE bound under this condition (denoted as C_2). Obviously, $D_2 \leq D_1$, because

$$D_2 = \left(\sum_{k=1}^{M} \frac{1}{\sigma_k^2 + \delta_k^2} \right)^{-1} \leq \left(\sum_{k=1}^{M} \frac{1}{\sigma_{i_k}^2 + \delta_k^2} \right)^{-1} = D_1. \tag{50}$$

Step 2: Construct another sequence $\{b'_k\}$ ($k = 1, \cdots, M$) by exchanging the order of the sequence $\{b_k\}$ ($k = 1, \cdots, M$) in the condition C_2 to make that $b'_i \geq b'_j$ if $0 < i < j \leq M$, and let the sensor k ($k = 1, \cdots, M$) quantizes its observation with b'_k bits instead of b_k bits. Let D_3 denote the estimation MSE bound under this condition (denoted as C_3). It is obvious that the condition C_3 can be implemented from condition C_2 by a serial of pairwise bit rate exchange operations defined in Definition 3.2. Since each pairwise bit rate exchange operation will not increase the estimation MSE bound as shown in Lemma 3.3, so $D_3 \leq D_2$.

After the two steps above, we constructed a new scenario where the first M sensors k ($k = 1, \cdots, M$) with the smallest observation noise variances $\sigma_1^2 \leq \cdots \leq \sigma_{K^{opt}}^2 \cdots \leq \sigma_M^2$ are used, and the quantization bit rates are $b'_1 \geq \cdots \geq b'_{K^{opt}} \cdots \geq b'_M$. To simplify the notation, in the following we denote the quantization bit rates as b_k ($k = 1, \cdots, M$) and $b_1 \geq \cdots \geq b_{K^{opt}} \cdots \geq b_M$.

Step 3: Expressing the estimation MSE bound D_3 with the concept of the equivalent 1-bit MSE function $g(\sigma^2, b)$ as

$$D_3 = \left(\sum_{k=1}^{M} \frac{1}{\sigma_k^2 + \delta_k^2} \right)^{-1} = \left(\sum_{k=1}^{K^{opt}} \frac{b_k}{g(\sigma_k^2, b_k)} + \sum_{m=K^{opt}+1}^{M} \frac{b_m}{g(\sigma_m^2, b_m)} \right)^{-1}. \tag{51}$$

From the total bit rate constraint, we get

$$\begin{aligned} \sum_{k=1}^{K^{opt}} b_k^{opt} &= B, \quad \text{and} \\ \sum_{k=1}^{M} b_k = \sum_{k=1}^{K^{opt}} b_k &+ \sum_{m=K^{opt}+1}^{M} b_m = B, \end{aligned} \tag{52}$$

so there must exist some $b_k < b_k^{opt}$ $(k = 1, \cdots, K^{opt})$ and

$$\sum_{k=1}^{K^{opt}} I_0(b_k^{opt} - b_k) \geq \sum_{m=K^{opt}+1}^{M} b_m, \tag{53}$$

where $I_0 : R \to R$ is an indicator function defined as follows

$$I_0(u) = \begin{cases} 0, & u \leq 0 \\ u, & u > 0 \end{cases} \tag{54}$$

From Equation (52) and (53), we notice that there are $\sum_{m=K^{opt}+1}^{M} b_m$ equivalent 1-bit quantization sensor corresponding to the sensors m $(m = K^{opt}+1, \cdots, M)$, and they all can be replaced by a serial of the equivalent 1-bit quantization sensor replacement operations defined in Definition 3.3. After finishing the replacement operations, we get a new condition where only sensors k $(k = 1, \cdots, K^{opt})$ are used, and the quantization bit rates are changed to $\bar{b}_k$ ($\bar{b}_k$ is not necessarily equal to b_k^{opt}), and the total bit rate constraint is still satisfied, i.e., $\sum_{k=1}^{K^{opt}} \bar{b}_k = B$. Let D_4 denote the estimation MSE bound of this condition (denoted as C_4). Since every equivalent 1-bit quantization sensor replacement operation will not increase the estimation MSE bound according to Lemma 3.4, so $D_4 < D_3$. On the other hand, in the condition C_4, only sensors k $(k = 1, \cdots, K^{opt})$ are used to quantize their observations with limited bit rates $\bar{b}_k$ $(k = 1, \cdots, K^{opt})$, so it is obvious that $D_4 > D_0 = \left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$.

From all the steps above, we get

$$D_1 \geq D_2 \geq D_3 > D_4 > \left(\sum_{k=1}^{K^{opt}} \frac{1}{\sigma_k^2}\right)^{-1}, \tag{55}$$

which means that the estimation MSE by any quasi-BLUE estimation system with $M > K^{opt}$ sensors under the total bit rate constraint B is larger than $D_0 = \left(\sum_{k=1}^{K^{opt}} 1/\sigma_k^2\right)^{-1}$.

CHAPTER IV

ENERGY-CONSTRAINED DISTRIBUTED ESTIMATION

Subject to the severe bandwidth and energy constraints in wireless sensor networks, in Chapter 3, we addressed the distributed estimation from the rate-distortion perspective and studied the optimal distributed estimation algorithms under a total bit rate constraints. To explicitly address the energy constraints in wireless sensor networks, we further study the energy-constrained distributed estimation in this chapter. Generally, the transmission energy cost is a function of the transmission bit rate, thus the energy-constrained distributed estimation can be treated as a generalization of the rate-constrained distributed estimation. More specifically, the rate-constrained distributed estimation is a special case of the energy-constrained distributed estimation with a special energy cost function that is a constant linear function of the transmission bit rate.

In energy-constrained wireless sensor networks, the energy-constrained distributed estimation algorithms have been recently studied in [5, 46, 91, 92, 94]. In [94] and [46], the total sensor transmission energy is minimized by selecting the optimal quantization levels while meeting the target estimation MSE requirements. On the contrary, the work of [5] is to minimize the estimation MSE under the given energy constraints. The work of [91, 92] addressed the energy-constrained distributed estimation problem (under the BLUE fusion rule) by exploiting long-term noise variance statistics. However, there is a lack of overall optimality analysis in the sense of energy-distortion performance.

In this chapter, we study the distributed estimation from the energy-distortion perspective [52, 54]. Here the fundamental question is: *what is the optimal energy-distortion bound for the distributed estimation and how do we achieve the performance bound in a distributed manner?* More specifically, the problem we address is to minimize the estimation MSE under a given total energy budget by optimally scheduling the quantization bit rate and transmission energy for all sensors. Based on the total energy constraint for all sensors, there exists an interesting trade-off between the number of active sensors and the energy consumed at each active sensor. We solve this optimal trade-off and design the optimal distributed estimation algorithm by (i) selecting a subset of active sensors to observe the phenomenon, and (ii) for each active sensor, determining the quantizer and transmission energy to quantize its real-valued observation and transmit the quantized message to the fusion center to perform the final estimation. Furthermore, we analysis the energy-distortion performance bound for the distributed estimation and show that the proposed algorithm is quasi-optimal within a constant factor of the theoretical lower bound. The proposed algorithm is easy to implement in a distributed manner and it adapts well to the dynamic sensor environments, which both are desirable for wireless sensor network applications. It is also worth noting that the proposed algorithm is applicable to both single-hop and multi-hop wireless sensor networks.

The rest of this chapter is organized as follows. Section 4.1 describes the system model and the distributed estimation problem under the total energy constraint. Section 4.2 introduces a concept of equivalent unit-energy MSE function. Then in Section 4.3 and Section 4.4, we develop an optimal distributed estimation algorithm for homogeneous sensor networks and a quasi-optimal distributed estimation algorithm for heterogeneous sensor networks, respectively. Also the upper bound of the estimation MSE of the proposed algorithm and a theoretical lower bound of the estimation MSE under the total energy constraint are addressed in Section 4.4. Furthermore, the proposed algorithm is extended to the multi-hop wireless sensor networks in Section 4.5. Section 4.6 gives some simulation results that demonstrate the efficiency of the proposed algorithms. Section 4.7 summarizes this chapter. The proofs of some theorems presented in this chapter are delegated to the appendix in Section 4.8.

4.1 System Model and Problem Statement

Consider a dense sensor network that includes N distributed sensors, denoted as $\{1, \cdots, N\}$. Each sensor can observe, quantize and transmit its observation to the fusion center, which will estimate the unknown parameter θ based on the received messages. Since the total energy allowed to be used by all sensors is limited, there exists a trade-off between the number of active sensors and the energy used by each active sensor, that is to say, only a subset of the sensors will be active at each task period. Assume there are K active sensors and denote the subset of active sensors as $S_K = \{i_1, \cdots, i_K\}$ ($i_k \in [1, N]$ for $k = 1, \cdots, K$), the distributed estimation system can be described as follows (Figure 12).

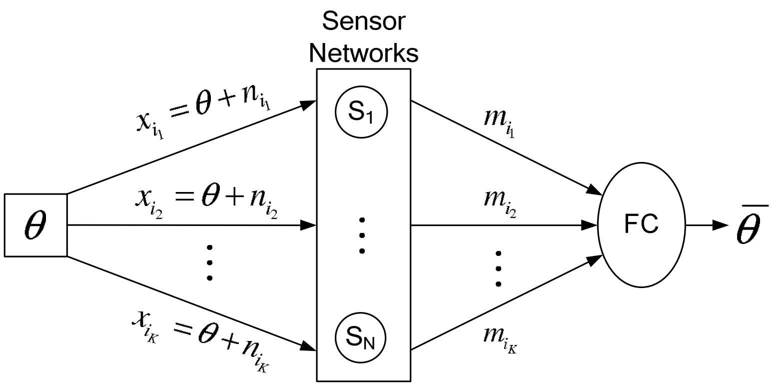


Figure 12: The distributed estimation system under the total energy constraint, where the subset of active sensors are $S_K = \{i_1, \cdots, i_K\}$ and each active sensor $k \in S_K$ performs a local quantization and transmits its quantization message to the fusion center, which will estimate θ based on all the received messages.

First, each active sensor $k \in S_K$ makes an observation on the unknown parameter

θ, which is corrupted by additive noise and is described by

$$x_k = \theta + n_k, \quad k \in S_K. \tag{56}$$

We assume that the observation noises of all sensors n_k ($k = 1, \cdots, N$) are zero mean, spatially uncorrelated with variance σ_k^2, otherwise unknown. Second, each active sensor k performs a local quantization $m_k = Q_k(x_k)$, where $Q_k(x_k)$ is a quantization function, and the quantization message m_k is then transmitted to the fusion center, where all the quantization messages are combined to produce a final estimation of θ using a fusion function. The quality of an estimation for θ is measured by the mean square error (MSE) criterion.

4.1.1 Quantization and Estimation Rules

Similar to that used in Section 3.1, the probabilistic quantization scheme is used at local sensors and the quasi-BLUE estimation scheme is adopted at the fusion center. As shown in Section 3.1.1, let $m(x, b)$ be a b-bit probabilistic quantization of bounded observation signal $x \in [-W, W]$ with noise variance σ^2; then $m(x, b)$ is an unbiased estimation of θ with a variance

$$E\left(|m(x,b) - \theta|^2\right) \le \sigma^2 + \frac{W^2}{(2^b - 1)^2} = \sigma^2 + \delta^2, \tag{57}$$

where $\delta^2 = W^2/(2^b - 1)^2$ for any $b > 0$ denotes the upper bound of the quantization noise variance.

Now suppose all the observations x_k ($k \in S_K$) of the K active sensors are quantized into b_k-bit discrete messages $m_k(x_k, b_k)$ respectively with the probabilistic quantization scheme. Based on the quantized messages m_k, the quasi-BLUE estimator at the fusion center has the following form:

$$\bar{\theta} = \left(\sum_{k \in S_K} \frac{1}{\sigma_k^2 + \delta_k^2}\right)^{-1} \sum_{k \in S_K} \frac{m_k}{\sigma_k^2 + \delta_k^2}, \tag{58}$$

Notice that $\bar{\theta}$ is an unbiased estimator of θ since every m_k is unbiased. Moreover, the estimation MSE of the quasi-BLUE estimator is

$$E(\bar{\theta} - \theta)^2 \leq \left(\sum_{k \in S_K} \frac{1}{\sigma_k^2 + \delta_k^2} \right)^{-1}. \tag{59}$$

4.1.2 Energy Model

To transmit a b-bit message from a sensor to the fusion center, the transmission energy cost P is generally a function of the transmission bit rate b and the transmission distance. Assume that each sensor sends a message to the fusion center using a separate channel, which can be achieved by using a multiple access technique such as TDMA or FDMA; and the channel between the sensor k and the fusion center experiences a path loss proportional to $a_k = d_k^\alpha$, where d_k is the transmission distance and α is the pass loss exponent.

Let's look at several different transmission models: (1) binary transmission model, and (2) uncoded quadrature amplitude modulation (QAM) model, and (3) coded quadrature amplitude modulation model. To reliably transmit b_k-bit message from the sensor k to the fusion center, the transmission energy cost for the binary transmission model, where each bit will be transmitted separately, is

$$P_{BIN}(b_k) = c_1 \cdot a_k \cdot b_k, \tag{60}$$

where c_1 is a system constant. To minimizes the transmission bandwidth and transmission delay, the b_k bits can be transmitted simultaneously using M-ary quadrature amplitude modulation (MQAM) with constellation size 2^{b_k}, then the transmission energy cost [23, 24] is given by

$$P_{QAM}(b_k) = c_2 \cdot a_k \cdot (2^{b_k} - 1), \tag{61}$$

where c_2 is a system constant defined the same as in [23, 24]. Furthermore, with embedded error correction codes, coded MQAM can reduce the transmission energy cost by a constant factor G_c [23, 24], i.e.,

$$P_{CQAM}(b_k) = c_3 \cdot a_k \cdot (2^{b_k} - 1), \tag{62}$$

where $c_3 = c_2/G_c$ is a system constant defined the same as in [23, 24]. Thereafter, we call the system constant c the transceiver characteristic parameter. It is noted that, compared with the binary transmission scheme, the MQAM schemes also minimize the circuit energy consumption since it minimizes the number of transmissions by transmitting the whole b_k-bit message as a single symbol.

4.1.3 Distributed Estimation under Energy Constraint

With the probabilistic quantization scheme and the quasi-BLUE fusion rule, the primary goal of the energy-constrained distributed estimation is to minimize the upper bound of the estimation MSE under the energy constraint, i.e.,

$$\begin{aligned} \min \quad & \left(\sum_{k \in S_K} \frac{1}{\sigma_k^2 + \frac{W^2}{(2^{b_k}-1)^2}} \right)^{-1}, \\ s.t. \quad & \sum_{k \in S_K} P_k \le P_c, \\ & P_k > 0, \quad b_k > 0, \quad k \in S_K, \end{aligned} \tag{63}$$

where S_K is the subset of K active sensors, b_k and P_k are the quantization bit rate and transmission energy of the active sensor $k \in S_K$, and P_c is the total energy allowed to be used by all active sensors.

It is obvious that the solution to the energy-constrained distributed estimation problem stated in Equation (63) depends on the energy model used. For a special energy model, where the energy cost P is assumed to be a constant linear function of the transmission bit rate b, i.e., $P = c \cdot b$, the energy-constrained distributed problems is retrogressed to the rate-constrained distributed estimation problem in Section 3.1.3. In this chapter, we will consider the energy-constrained distributed estimation problem with the QAM-based models. It is worth noting that the similar methodology proposed in this chapter can be extended to solve the energy-constrained distributed estimation with different energy models.

Using the uncoded/coded MQAM models, the original problem in Equation (63) turns to the following problem:

$$
\begin{aligned}
\min \quad & \left(\sum_{k \in S_K} \frac{1}{\sigma_k^2 + \dfrac{W^2}{(2^{b_k}-1)^2}} \right)^{-1}, \\
s.t. \quad & \sum_{k \in S_K} P_k \leq P_c, \\
& P_k = c_k a_k (2^{b_k} - 1), \quad k \in S_K, \\
& b_k > 0, \quad k \in S_K,
\end{aligned}
\tag{64}
$$

where all the variables are defined as before. In practice, the quantization bit rate b_k must be integer, i.e., $b_k \in \mathbb{Z}$. To facilitate the subsequent analysis, we will relax the integer condition $b_k \in \mathbb{Z}$ to $b_k \in \mathbb{R}$. Later, we will discuss how to constraint the quantization bit rate to integer numbers.

It can be verified that the optimal solution for the energy-constrained distributed estimation problem in Equation (64) cannot be found in a closed form. In the following sections, we will address this problem for homogeneous and heterogeneous sensor networks, respectively. To facilitate the solution, we first define an equivalent unit-energy MSE function in the next section.

4.2 Equivalent Unit-Energy MSE Function

As shown in Section 4.1, the b-bit quantization message from a sensor with observation noise variance σ^2 is an unbiased estimation of the parameter θ. We denote the estimation MSE bound as

$$
f(\sigma^2, b) := \sigma^2 + \frac{W^2}{(2^b - 1)^2}. \tag{65}
$$

Definition 4.1 (Equivalent Unit-Energy MSE function). *For a sensor with observation noise variance σ^2, quantization bit rate b, transmission path loss a, transceiver* *parameter c, and transmission energy cost $P(b, a, c)$, the equivalent unit-energy MSE function is defined as*

$$g(\sigma^2, b, a, c) := P(b, a, c) \cdot f(\sigma^2, b). \tag{66}$$

With this definition, the estimation MSE of the quasi-BLUE estimator, shown in Equation (59), can be rewritten as

$$E(\bar{\theta} - \theta)^2 \leq \left(\sum_{k \in S_K} \frac{1}{f(\sigma_k^2, b_k)} \right)^{-1} = \left(\sum_{k \in S_K} \frac{P_k}{g(\sigma_k^2, b_k, a_k, c_k)} \right)^{-1}. \tag{67}$$

From the estimation MSE aspect, a sensor with transmission energy P and estimation MSE $f(\sigma^2, b)$ can be treated as P equivalent unit-energy sensors, each with the same estimation MSE $g(\sigma^2, b, a, c)$ defined as above. That is why the function $g(\sigma^2, b, a, c)$ is called equivalent unit-energy MSE function. Further, the energy-constrained distributed estimation system under the total energy constraint P_c can be treated as another equivalent distributed estimation system with P_c equivalent unit-energy sensors.

With the uncoded/coded MQAM models, the equivalent unit-energy MSE function defined in Equation (66) is

$$g(\sigma^2, b, a, c) = P(b, a, c) \cdot f(\sigma^2, b) = ca(2^b - 1) \left(\sigma^2 + \frac{W^2}{(2^b - 1)^2} \right). \tag{68}$$

As shown in Proposition 4.1, $g(\sigma^2, b, a, c)$ is convex over b. We further define the optimal unit-energy MSE function $g^{opt}(\sigma^2, a, c)$, and the corresponding optimal quantization bit rate $b^{opt}(\sigma^2, a, c)$ and optimal transmission energy $P^{opt}(\sigma^2, a, c)$ for each sensor with observation noise variance σ^2, transmission path loss a, and transceiver parameter c as follows:

$$\begin{aligned} b^{opt}(\sigma^2, a, c) &= \arg\min_{b \in \mathbb{R}^+} g(\sigma^2, b, a, c) = \arg\min_{b \in \mathbb{R}^+} \left[ca(2^b - 1) \left(\sigma^2 + \frac{W^2}{(2^b - 1)^2} \right) \right], \\ g^{opt}(\sigma^2, a, c) &= \min_{b \in \mathbb{R}^+} g(\sigma^2, b, a, c) = g(\sigma^2, b^{opt}(\sigma^2, a, c), a, c), \\ P^{opt}(\sigma^2, a, c) &= ca(2^{b^{opt}(\sigma^2, a, c)} - 1). \end{aligned} \tag{69}$$

Proposition 4.1. *The equivalent unit-energy MSE function $g(\sigma^2, b, a, c)$, the optimal unit-energy MSE function $g^{opt}(\sigma^2, a, c)$, the optimal quantization bit rate function $b^{opt}(\sigma^2, a, c)$, the optimal transmission energy function $P^{opt}(\sigma^2, a, c)$ and the MSE function $f(\sigma^2, b)$ defined before have the following properties:*

1. *$g(\sigma^2, b, a, c)$ is convex over $b \in \mathbb{R}^+$.*

2. *$g^{opt}(\sigma^2, a, c)$ is achieved when the optimal quantization bit rate $b^{opt}(\sigma^2, a, c)$ is used and the optimal transmission energy $P^{opt}(\sigma^2, a, c)$ is allocated, where*

$$\begin{aligned} b^{opt}(\sigma^2, a, c) &= \log_2\left(1 + \frac{W}{\sigma}\right), \\ g^{opt}(\sigma^2, a, c) &= 2ca\sigma W, \\ P^{opt}(\sigma^2, a, c) &= \frac{caW}{\sigma}. \end{aligned} \tag{70}$$

3. *The estimation MSE $f(\sigma^2, b)$ with the optimal quantization bit rate $b^{opt}(\sigma^2, a, c)$ and transmission energy $P^{opt}(\sigma^2, a, c)$ is*

$$f(\sigma^2, b^{opt}(\sigma^2, a, c)) = 2\sigma^2. \tag{71}$$

The proposition 4.1 is easy to prove as follows: the convexity of $g(\sigma^2, b, a, c)$ over b can be proved by checking $\partial^2 g(\sigma^2, b, a, c)/\partial^2 b > 0$ for any $b > 0$; then $b^{opt}(\sigma^2, a, c)$ can be obtained by solving $\partial g(\sigma^2, b, a, c)/\partial b = 0$, and $g^{opt}(\sigma^2, a, c)$, $P^{opt}(\sigma^2, a, c)$ and $f(\sigma^2, b^{opt}(\sigma^2, a, c))$ can be obtained according to the definitions in Equations (65, 68, 69).

It is noted that the optimal transmission energy function $P^{opt}(\sigma^2, a, c)$ depends not only on the signal-to-noise ratio but also on the transmission path loss and transceiver parameter, but the optimal quantization bit rate function $b^{opt}(\sigma^2, a, c)$ depends on only the signal-to-noise ratio, as shown in Equation (70).

4.3 Distributed Estimation in Homogeneous Sensor Networks

In homogeneous sensor networks, the noise variances for all sensors are identical, that is $\sigma_k^2 = \sigma^2$ $(k = 1, \cdots, N)$. We assume equal distances from all sensors to the fusion center; thus the transmission path loss is the same for all sensors too, i.e., $d_k = d$ and $a_k = a$ $(k = 1, \cdots, N)$. Also, assume that the transceiver parameters are the same for all sensors, i.e., $c_k = c$ $(k = 1, \cdots, N)$. Therefore, the equivalent unit-energy MSE function is the same for all sensors. For this homogeneous sensor network model, all active sensor should quantize its observation with the same bit rate $b_k = b$ and transmit its quantized message with same energy $P_k = P$ to minimize the estimation MSE, so the number of active sensors is $K = P_c/P$, and the estimation MSE function shown in Equation (67) is simplified to

$$E(\bar{\theta} - \theta)^2 \leq \left(\sum_{k=1}^{K} \frac{1}{f(\sigma^2, b)} \right)^{-1} = \frac{P \cdot f(\sigma^2, b)}{P_c} = \frac{g(\sigma^2, b, a, c)}{P_c}. \tag{72}$$

It is noted that the numerator of the optimized target function in Equation (72) is just the equivalent unit-energy MSE function. Hence, for homogeneous sensor networks, the optimal distributed estimation under the total energy constraint P_c can be treated in an alternative way, where there are P_c identical equivalent unit-energy sensors, thus minimizing the final estimation MSE becomes minimizing the equivalent unit-energy MSE function. The method based on the unit-energy MSE function is stated as follows:

1. For all sensors, the optimal quantization bit rate b^{opt} and transmission energy P^{opt} are the same and obtained by minimizing the corresponding equivalent unit-energy MSE function, as shown in Proposition 4.1:

$$\begin{aligned} b^{opt} &= \log_2 \left(1 + \frac{W}{\sigma} \right), \\ P^{opt} &= \frac{caW}{\sigma}. \end{aligned} \tag{73}$$

2. The total number of active sensors K^{opt} under the total energy constraint P_c is

$$K^{opt} = \left\lfloor \frac{P_c}{P^{opt}} \right\rfloor . \tag{74}$$

It is obvious that the proposed method based on the equivalent unit-energy MSE function is optimal if P_c/P^{opt} is an integer, otherwise, it is quasi-optimal.

Remark 4.1. *It is noted that the proposed method based on equivalent unit-energy MSE function can be implemented in a fully distributed manner. First, the optimal quantization bit rate b^{opt} and optimal transmission energy P^{opt} of each sensor can be obtained locally by minimizing its corresponding equivalent unit-energy MSE function. Second, the subset of active sensors is chosen in a round-robin manner such that there are $K^{opt} = P_c/P^{opt}$ (we assume P_c/P^{opt} is integer here) active sensors at any task period and each sensor will be active for K^{opt} task periods in any consecutive N task duration. Therefore, the energy cost at each sensor node is even, and the network lifetime is maximized, which is defined as the time for the first sensor node in the network to deplete.*

4.4 Distributed Estimation in Heterogeneous Sensor Networks

In heterogeneous sensor networks, the observation noise variance for sensor k is σ_k^2 ($k = 1, \cdots, N$), respectively. Assume the distance from sensor k to the fusion center is d_k; thus the transmission path loss is $a_k = d_k^{\alpha}$. And assume the transceiver parameter of sensor k is c_k. This scenario leads to the general problem stated in Equation (64). The goal is to find the optimal number of active sensors and the corresponding optimal quantization bit rate and transmission energy allocation for each active sensor to minimize the estimation MSE bound at the fusion center.

Unfortunately, it can be verified that the optimal solution cannot be found in a closed form. Instead, we develop a quasi-optimal method to solve this problem, which is also based on the equivalent unit-energy MSE function. The procedure is stated as follows:

1. For each sensor $k \in [1, N]$, determine its optimal quantization bit rate b_k^{opt}, optimal transmission energy P_k^{opt} and optimal unit-energy MSE function g_k^{opt} as shown in Proposition 4.1:

$$\begin{aligned} b_k^{opt} &= \log_2 \left(1 + \frac{W}{\sigma_k}\right), \\ g_k^{opt} &= 2 c_k a_k \sigma_k W, \\ P_k^{opt} &= \frac{c_k a_k W}{\sigma_k}. \end{aligned} \tag{75}$$

2. Let S_k $(k \in [1, N])$ denote the subset of all sensors consisting of the first k sensors with the minimum optimal unit-energy MSE function, then

$$\begin{cases} S_1 \subset S_2 \subset \cdots \subset S_N = \{1, \cdots, N\}, \\ g_i^{opt} \leq g_j^{opt}, \text{ if } i \in S_k \text{ and } j \in S_k^c, \end{cases} \tag{76}$$

where S_k^c denotes the complemental subset of S_k. Then the optimal number of active sensors K^{opt} under the total energy constraint P_c is determined by

$$K^{opt} = \max k \quad s.t. \quad \sum_{i \in S_k} P_i^{opt} \leq P_c, \tag{77}$$

that is to say, the subset of active sensors is $S_{K^{opt}}$.

In summary, the whole solution is that all sensors in the subset $S_{K^{opt}}$, i.e., the first K^{opt} sensors with the smallest optimal unit-energy MSE function, are active to quantize their observations with quantization bit rate b_k^{opt} and transmit their quantized messages to the fusion center with transmission energy P_k^{opt} $(k \in S_{K^{opt}})$.

To implement the described algorithm above, each sensor needs to decide (i) whether it should be active or not, i.e., whether it belongs to $S_{K^{opt}}$, and (ii) its quantization bit rate and transmission energy if it will be active. Both tasks can be achieved in a distributed manner as follows:

- As shown in Equations (76, 77), the subset of active sensors $S_{K^{opt}}$ is determined at the fusion center based on the collected network information and the total energy constraint P_c. Denote the maximum optimal unit-energy MSE function of all the active sensors in the subset $S_{K^{opt}}$ as

$$g_{th}^{opt} = \arg\max_{k \in S_{K^{opt}}} g_k^{opt}. \tag{78}$$

  Then the fusion center broadcasts the threshold g_{th}^{opt} to all local sensors. Upon receiving the threshold, each sensor compares the threshold with its own optimal unit-energy MSE function g_k^{opt}. If $g_k^{opt} \le g_{th}^{opt}$, then sensor k is active; otherwise, it is inactive.

- As shown in Equation (75), the optimal quantization bit rate b_k^{opt} of sensor $k \in [1, N]$ depends only on its own signal-to-noise ratio, and the optimal transmission energy P_k^{opt} and the optimal unit-energy MSE function g_k^{opt} of sensor k depend only on its own optimal quantization bit rate b_k^{opt}, transmission path loss a_k and transceiver parameter c_k. Therefore, all of b_k^{opt}, P_k^{opt}, and g_k^{opt} can be computed locally at each sensor without requiring information from other sensors.

Remark 4.2. *As shown above, the total energy constraint P_c is to determine the subset of active sensors according to Equations (76, 77). It is interesting to see that if the total energy constraint P_c is changed, we only need to wake up several more sleep sensors (energy constraint increased) or send several active sensors to sleep (energy constraint decreased), but don't need to change the quantization bit rate and transmission energy allocation of each active sensor. So, the proposed method adapts well to the situations when the total energy constraints need to be changed frequently to achieve various estimation MSE performances, which is the case for dynamic sensor environments.*

Next, we will analyze the estimation MSE bound of the proposed method, which is stated in the following theorem. To simplify the statements, we assume $\sum_{k \in S_{K^{opt}}} P_k^{opt} = P_c$ in the subsequent analysis.

Theorem 4.1. *The estimation MSE of the proposed method based on the equivalent unit-energy MSE function under the total energy constraint P_c is*

$$\left(\sum_{k \in S_{K^{opt}}} \frac{1}{\sigma_k^2} \right)^{-1} < E(\bar{\theta}_p - \theta)^2 \leq 2 \left(\sum_{k \in S_{K^{opt}}} \frac{1}{\sigma_k^2} \right)^{-1}, \tag{79}$$

where $\bar{\theta}_p$ denotes the estimation of the parameter θ by the proposed method, and $S_{K^{opt}}$ is the optimal subset of active sensors, obtained in Equations (76, 77).

Proof. The left part of the theorem is obvious since $\left(\sum_{k \in S_{K^{opt}}} 1/\sigma_k^2 \right)^{-1}$ is the lower bound of the estimation MSE of the BLUE estimator using the subset $S_{K^{opt}}$ of sensors without energy constraint. To prove the right part of the theorem, by Proposition 4.1, we have

$$\begin{aligned} E(\bar{\theta}_p - \theta)^2 &\leq \left(\sum_{k \in S_{K^{opt}}} \frac{1}{f(\sigma_k^2, b^{opt}(\sigma_k^2, a_k, c_k))} \right)^{-1} \\ &= \left(\sum_{k \in S_{K^{opt}}} \frac{1}{2\sigma_k^2} \right)^{-1} \\ &= 2 \left(\sum_{k \in S_{K^{opt}}} \frac{1}{\sigma_k^2} \right)^{-1}. \end{aligned} \tag{80}$$

□

This theorem gives the lower and upper bounds of the estimation MSE of the proposed method. It is shown that the proposed method is quasi-optimal (up to a factor of 2) when compared with the BLUE estimator using the same subset of active sensors without energy constraint.

As shown above, the performance bound of the proposed algorithm is analyzed. Nevertheless, the remaining question is that what is the optimal energy-distortion bound for distributed estimation, i.e., what is the minimal estimation MSE that be achieved if the total energy P_c is allocated to any subset of sensors. To answer this question, Theorem 4.2 states the lower bound of the estimation MSE by any quasi-BLUE estimation system with any subset of sensors under the total energy constraint P_c. Surprisingly, under the same total energy constraint P_c, the lower bound of the estimation MSE by any quasi-BLUE estimation system with any subset S of sensors is same as the lower bound of the estimation MSE of the BLUE estimator using the subset $S_{K^{opt}}$ of sensors obtained by the proposed algorithm in Equations (76, 77).

Theorem 4.2. *Assume any subset of sensors $S = \left\{i_1, \cdots, i_k, \cdots, i_{|S|}\right\}$ are used, where $i_k \in [1, N]$ and $|S|$ denotes the cardinality of the set S, i.e., the total number of sensors in the set S. The energy allocated to each active sensor $k \in S$ is P_k, such that $\sum_{k \in S} P_k = P_c$. Then the lower bound of the estimation MSE is*

$$E(\bar{\theta}_c - \theta)^2 > \left(\sum_{k \in S_{K^{opt}}} \frac{1}{\sigma_k^2} \right)^{-1}, \tag{81}$$

where $\bar{\theta}_c$ denotes the estimation of the parameter θ by the subset of active sensors S under the given total energy constraint P_c, and $S_{K^{opt}}$ is the optimal subset of active sensors, obtained by the proposed algorithm as shown in Equations (76, 77) such that $\sum_{k \in S_{K^{opt}}} P_k^{opt} = P_c$.

Proof. For any given estimation system as stated in the Theorem, the basic idea to prove its estimation MSE $D_1 > \left(\sum_{k \in S_{K^{opt}}} (1/\sigma_k^2)\right)^{-1}$ is to construct another corresponding quasi-BLUE estimation system such that its estimation MSE D_2 is smaller than D_1 but larger than $\left(\sum_{k \in S_{K^{opt}}} (1/\sigma_k^2)\right)^{-1}$, i.e., $D_1 > D_2 > \left(\sum_{k \in S_{K^{opt}}} (1/\sigma_k^2)\right)^{-1}$. The proof is based on the concept of equivalent unit-energy MSE function. Refer to Appendix 4.8 for the details. It is worth noting that the similar technique used to prove this theorem also can be used to prove the Theorem 3.2.

□

In conclusion, Theorem 4.1 shows that the bound of estimation MSE of the proposed method is $\left(\sum_{k\in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1} < E(\bar{\theta}_p - \theta)^2 \leq 2\left(\sum_{k\in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1}$, and Theorem 4.2 shows that $\left(\sum_{k\in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1}$ is the lower bound of the estimation MSE of any quasi-BLUE estimator under the total energy constraint P_c, regardless of the subset of active sensors and the energy allocation among the active sensors. Therefore, the proposed algorithm gives a quasi-optimal trade-off between the number of active sensors and the energy allocation at each active sensor, and its estimation MSE is within a factor 2 of the theoretical non-achievable lower bound.

Remark 4.3. *As we mentioned before, in all the prior analysis, we assume the quantization bit rate can be real-valued number. But in practice, the quantization bit rate must be integer. Denote the optimal integer quantization bit rate as* $\bar{b}^{opt}(\sigma^2, a, c) \in \mathbb{Z}^+$, *the corresponding optimal transmission energy as* $\bar{P}^{opt}(\sigma^2, a, c)$ *and optimal equivalent unit-energy MSE function as* $\bar{g}^{opt}(\sigma^2, a, c)$ *for a sensor with observation noise variance* σ^2, *transmission path loss* a, *and transceiver parameter* c, *thus,*

$$\begin{aligned}
&\bar{b}^{opt}(\sigma^2, a, c) = \arg\min_{b\in\mathbb{Z}^+} g(\sigma^2, b, a, c) = \arg\min_{b\in\mathbb{Z}^+} \left[ca(2^b - 1)\left(\sigma^2 + \frac{W^2}{(2^b-1)^2}\right)\right],\\
&\bar{g}^{opt}(\sigma^2, a, c) = g(\sigma^2, \bar{b}^{opt}(\sigma^2, a, c), a, c),\\
&\bar{P}^{opt}(\sigma^2, a, c) = ca(2^{\bar{b}^{opt}(\sigma^2,a,c)} - 1).
\end{aligned} \tag{82}$$

Different from $b^{opt}(\sigma^2, a, c)$, $\bar{b}^{opt}(\sigma^2, a, c)$ *can not be written in a closed form; however, it can be easily solved since the minimization in Equation (82) involves just a simple one-dimensional numerical search. So, in practice, the proposed distributed estimation algorithms above can be easily implemented by using* $\bar{b}^{opt}(\sigma^2, a, c)$, $\bar{P}^{opt}(\sigma^2, a, c)$ *and* $\bar{g}^{opt}(\sigma^2, a, c)$ *instead of* $b^{opt}(\sigma^2, a, c)$, $P^{opt}(\sigma^2, a, c)$ *and* $g^{opt}(\sigma^2, a, c)$.

Since $g(\sigma^2, b, a, c)$ *is convex over* b *as shown in Proposition 4.1,* $\bar{b}^{opt}(\sigma^2, a, c) = \lfloor b^{opt}(\sigma^2, a, c)\rfloor$ *or* $\lceil b^{opt}(\sigma^2, a, c)\rceil$, *where* $\lfloor b^{opt}(\sigma^2, a, c)\rfloor$ *denotes the maximum integer no more than* $b^{opt}(\sigma^2, a, c)$, *and* $\lceil b^{opt}(\sigma^2, a, c)\rceil$ *denotes the minimum integer no less than* $b^{opt}(\sigma^2, a, c)$. *Figure 13 shows the optimal real-valued quantization bit rate* $b^{opt}(\sigma^2, a, c)$

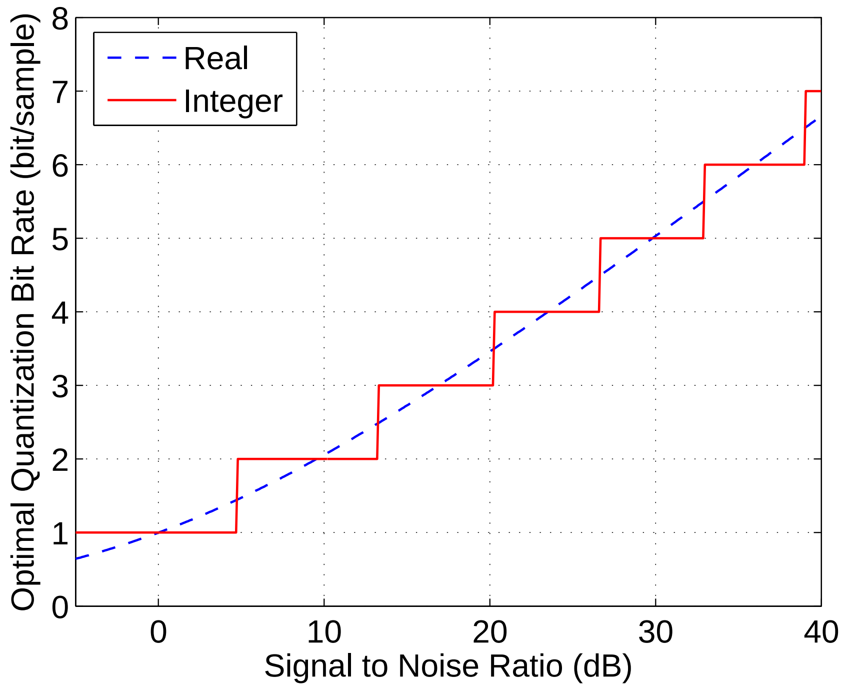


Figure 13: Optimal real-valued and integer quantization bit rates verse SNR.

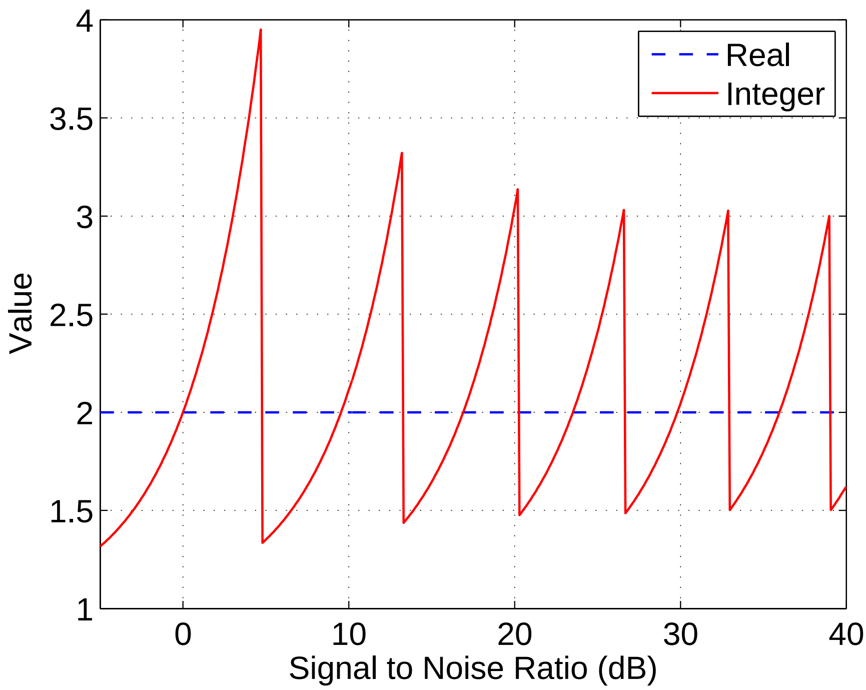


Figure 14: Value of $(f(\sigma^2, b)/\sigma^2)$ using the optimal real-valued and integer quantization bit rates verse SNR.

and the optimal integer quantization bit rate $\bar{b}^{opt}(\sigma^2, a, c)$ verse different signal-to-noise ratios (SNR) defined as $SNR = 10\log_{10}(W^2/\sigma^2)$. Further, Figure 14 shows the ratio of the estimation MSE $f(\sigma^2, b)$ to σ^2 using real-valued quantization bit rate $b = b^{opt}(\sigma^2, a, c)$ with transmission energy $P = P^{opt}(\sigma^2, a, c)$, or integer quantization bit rate $b = \bar{b}^{opt}(\sigma^2, a, c)$ with transmission energy $P = \bar{P}^{opt}(\sigma^2, a, c)$. From Figure 14, we can see that the upper bound of the estimation MSE is twice of the observation noise variance when the optimal real-valued quantization bit rate is used, as it is proved in Proposition 4.1, and the upper bound of the estimation MSE is within a small factor (up to 4) of the observation noise variance when the integer constraint is imposed on the optimal quantization bit rate. It is also worth noting that, when the optimal integer quantization bit rate is used, the theoretical lower bound shown in Theorem 4.2 is still valid and it can be proved in the same way.

4.5 Extension to Multi-Hop Sensor Networks

In the previous section, we focus on the single-hop sensor network case. In this section, we extend the proposed algorithms to the multi-hop sensor network case.

In the single-hop networks, each sensor locally processes its observation and then transmits the processed message to the fusion center directly. As shown in Equations (60, 61, 62), the transmission energy function P_s is as follows:

$$P_s = cd^\alpha p(b), \tag{83}$$

where c is the transceiver parameter of the sensor node, d is the transmission distance, α is the transmission path loss exponent ($2 \leq \alpha \leq 4$), and $p(b)$ is a function of the transmission bit rate b. Denote the single-hop transmission energy factor for the sensor k to the fusion center as $C_s(k) = c_k d_k^\alpha$. Since the energy cost function is proportional to d^α, it could save more energy to transmit the messages through multiple relay paths with short distance for each relay instead of a single path with a long distance.

In the multi-hop networks, each sensor locally processes its observation and then transmits its processed message to the fusion center along a multi-hop routing path. Assume in the multi-hop routing path, the processed message from sensor k_0 is sequentially relayed by sensor $k_1, k_2, \cdots, k_n$ to the fusion center, then the total transmission energy cost is

$$P_m = \sum_{i=0}^{n} (c_i {d_i}^{\alpha}) p(b), \tag{84}$$

where c_i $(i = 0, \cdots, n)$ is the transceiver parameter of the sensor k_i, d_i $(i = 0, \cdots, n-1)$ is the transmission distance from the sensor k_i to the sensor k_{i+1}, and d_n is the transmission distance from the sensor k_n to the fusion center. Denote $C_m(k) = \sum_{i=0}^{n} (c_i {d_i}^{\alpha})$ as the multi-hop transmission energy factor for the sensor k to the fusion center. The multi-hop routing tree can be established using any routing algorithms. To minimize the transmission energy cost from each sensor to the fusion center in Equation (84), the shortest path tree routing is desirable.

With the multi-hop routing, the energy-constrained distributed estimation problem in Equation (63) turns to the following problem:

$$\begin{aligned} \min \quad & \left(\sum_{k \in S_K} \frac{1}{\sigma_k^2 + \dfrac{W^2}{(2^{b_k} - 1)^2}} \right)^{-1}, \\ s.t. \quad & \sum_{k \in S_K} P_m(k) \le P_c, \\ & P_m(k) = C_m(k)(2^{b_k} - 1), \quad k \in S_K, \\ & b_k > 0, \quad k \in S_K, \end{aligned} \tag{85}$$

where, all the variables are defined as before. Compared the problem for the single-hop network case stated in Equation (64) with the problem for the multi-hop network case stated in Equation (85), they are equivalent with the variable replacement from $C_s(k) = c_k d_k^{\alpha}$ in Equation (64) to $C_m(k)$ in Equation (85). Since both $C_s(k)$ and $C_m(k)$ are constants for the given network and the given multi-hop routing tree, the same algorithms proposed in Section 4.2, 4.3, and 4.4 are applicable to the general multi-hop network case.

4.6 Simulation Results

In this section, we present some simulation results for the proposed algorithms in Section 4.3 and 4.4, respectively. In all the simulations, we assume the transceiver parameters are the same for all sensors, i.e., $c_k = c$, and the quantization bit rates to be integer number as discussed in Remark 4.3. All the final results are obtained by repeating the experiments for 10000 times and averaging the corresponding results.

4.6.1 Homogeneous Sensor Networks

In this section, we simulate a homogeneous sensor network with $N = 500$ sensors, where the noise variances of all sensors are the same and the distances from all sensors to the fusion center are also the same. Without loss of generality, we assume the range of the observation signal is $[-1, 1]$, i.e., $W = 1$, and the distance from each sensor to the fusion center is $d = 1$. Define the signal-to-noise ratio (SNR) as $SNR = 10 \log_{10}(W^2/\sigma^2)$ and generate different SNR by changing the observation noise variance σ^2. Define the normalized energy as $P' = P/c = a(2^b - 1)$, where c is the transceiver parameter, a is the transmission path loss, and b is the quantization bit rate.

Assuming the normalized total energy constraint is $P'_c = 500$, Figure 15 shows the estimation MSE with different quantization bit rates for the active sensors under different SNR, where different quantization bit rates, amounting to different energy allocation, imply different total number of active sensors to perform the estimation task because of the total energy constraint. Explicitly, for the given total normalized energy constraint $P'_c = 500$, we can have 500 active sensors with 1-bit quantization message for each sensor, or 167 active sensors with 2-bit quantization message for each sensor, or 71 active sensors with 3-bit quantization message for each sensor, or 33 active sensors with 4-bit quantization massage for each sensor and so on. For example, for the case of $SNR = 20\ dB$, totally 71 active sensors out of all 500 sensors with 3-bit quantization message per sensor will produce the minimum estimation MSE among all the possible energy allocation strategies as shown in Figure 15. From Figure 15, we also can see that there exists an optimal quantization bit rate for any given SNR under the total energy constraint, and that too small or too big quantization bit rate will sacrifice the estimation MSE performance significantly. More specifically, 1-bit quantization per sensor will lead to the minimum estimation MSE for low SNR cases, such as 0 dB, while for high SNR cases, multiple-bit quantization per sensor will significantly decrease the estimation MSE compared to only 1-bit quantization per sensor under the same total energy constraint.

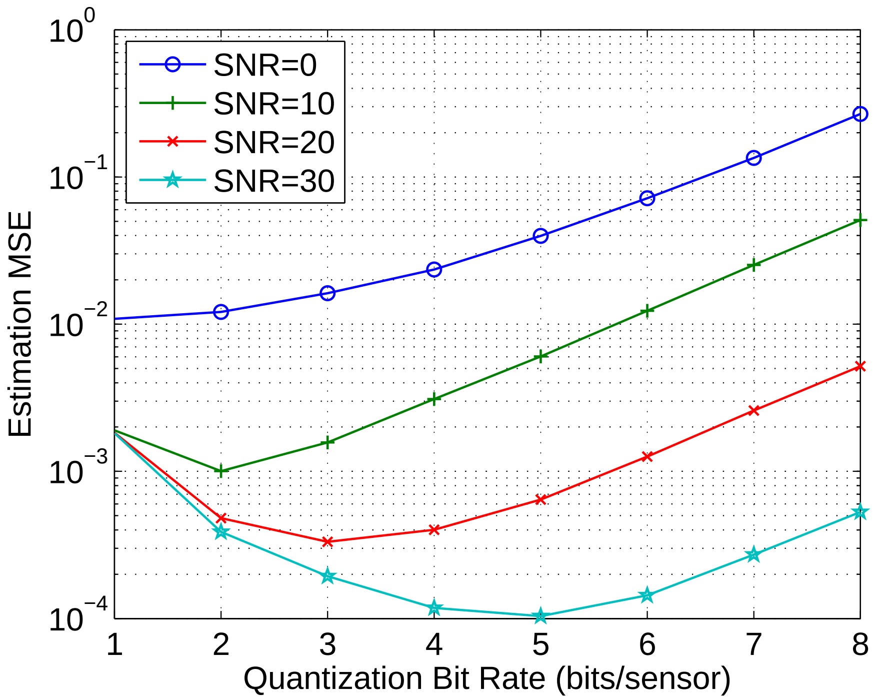


Figure 15: The estimation MSE versus the quantization bit rate per sensor and signal-to-noise ratios (SNR) under the total energy constraint.

4.6.2 Heterogeneous Sensor Networks with Equal Distances

In this section, we simulate a heterogeneous sensor network with $N = 500$ sensors, where the noise variance of each sensor is different, which is assumed to be a Chi-squared distribution with one degree of freedom, while the distance from each sensor to the fusion center is the same. Same as before, we assume the range of the observation signal is $[-1, 1]$ and the distance from each sensor to the fusion center is $d = 1$.

For any given total energy constraint, the proposed estimation method in Section 4.4 is implemented to determine the subset of active sensors and the energy allocation at each active sensor to minimize the estimation MSE. To demonstrate the efficiency of the proposed method, we compare the proposed method with two uniform schemes:

1. *Uniform-I*: For the given total energy constraint, the same subset of active sensors as that used by the proposed method is used, but the energy is uniformly allocated among all the active sensors.

2. *Uniform-II*: all sensors in the simulated heterogeneous sensor network are used and the energy is uniformly allocated among all sensors.

Figure 16 shows the estimation MSE by the proposed method, the *Uniform-I* method, and the *Uniform-II* method, and the theoretical lower bound of the estimation MSE presented in Theorem 4.2 under the total energy constraint. From Figure 16, we can see that the proposed method outperforms the two uniform schemes. Further, it also can be seen that the estimation MSE of the proposed method is within a factor 2 of the theoretical non-achievable lower bound.

Note that both the proposed method and the *Uniform-I* method are based on the same subset of active sensors, and the only difference is that the optimal energy allocation is performed in the proposed method, while uniform energy allocation is

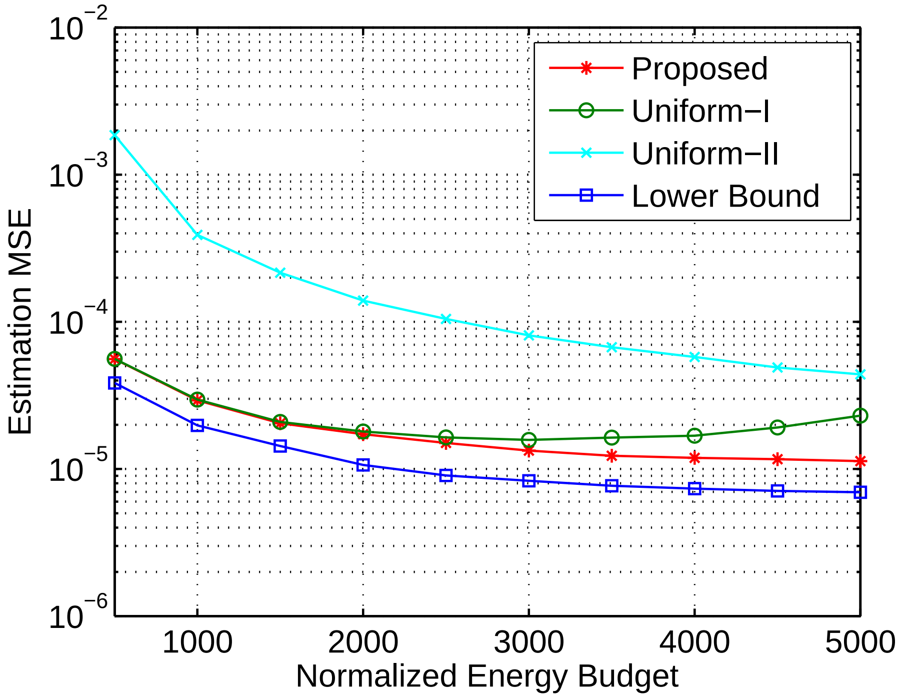


Figure 16: The estimation MSE by the proposed method, *Uniform-I* method, and *Uniform-II* method, and the theoretical non-achievable lower bound of the estimation MSE for heterogeneous sensor networks with equal distances.

performed in the *Uniform-I* method. Because of the heterogeneity of the network, a better estimation performance is obtained using the proposed method. Define the normalized deviation of sensor noise variances as

$$\alpha = \frac{\sqrt{Var(\sigma^2)}}{E(\sigma^2)}, \tag{86}$$

which will be used as a measure of the heterogeneity of sensor networks. And define the reduction in the estimation MSE achieved by the proposed method in comparison with the *Uniform-I* method as

$$\beta = \frac{D_u - D_p}{D_u}, \tag{87}$$

where D_u denotes the estimation MSE by the *Uniform-I* method, and D_p denotes the estimation MSE by the proposed method. Figure 17 plots the estimation MSE reduction of the proposed method compared with the *Uniform-I* method versus the normalized deviations of sensor noise variances. From Figure 17, we conclude that, when compared with the *Uniform-I* method, the amount of estimation MSE reduction of the proposed method becomes more significant when the local sensor noise variances become more heterogeneous.

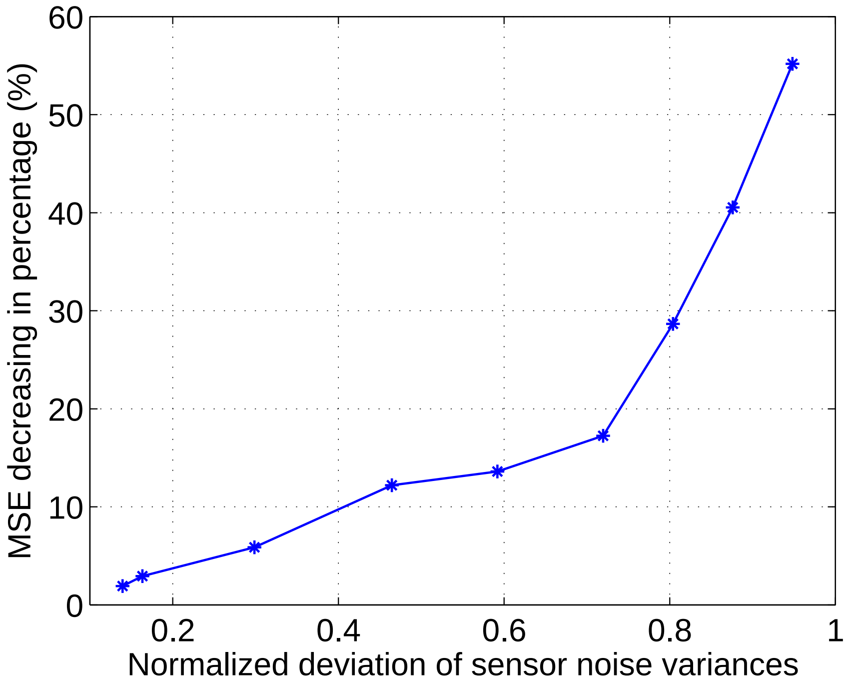


Figure 17: The estimation MSE reduction in percentage of the proposed method compared with the *Uniform-I* method under different normalized deviations of sensor noise variances.

4.6.3 Multi-Hop Heterogeneous Sensor Networks

In this part of the simulation, we relax the assumption in Section 4.6.2 that the distance from each sensor to the fusion center is the same. We assume that all sensors are independently and uniformly distributed in a rectangular region of $[0, 20, 0, 20]$, and the fusion center is located at the central point of the region, i.e., $(0, 0)$. Same with in Section 4.6.2, we simulate a heterogeneous sensor network with $N = 500$ sensors, where the noise variances of all sensors are different and are assumed to be a Chi-squared distribution with one degree of freedom. In this simulation, both the single-hop transmission scheme and the multi-hop transmission scheme are considered. For the multi-hop transmission, the shortest path routing tree is established as shown in Figure 18.

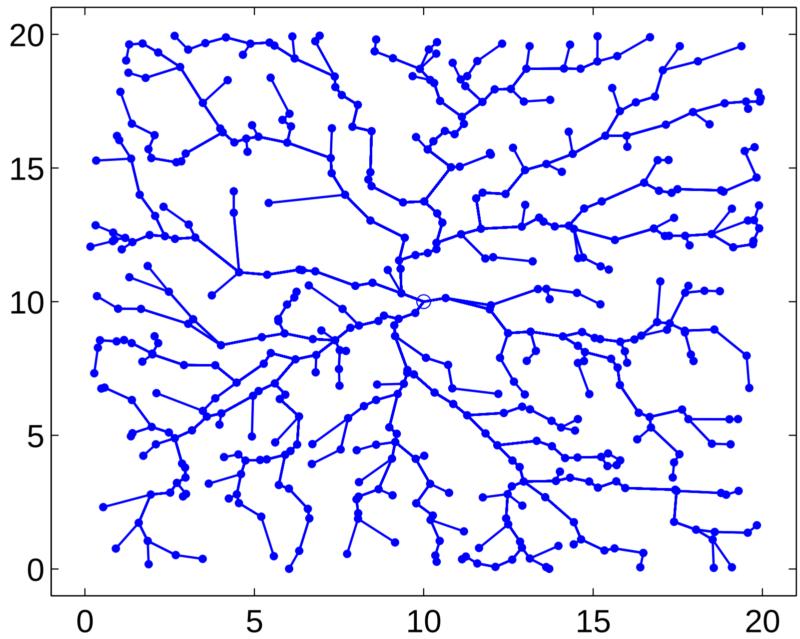


Figure 18: An example of a wireless sensor network with a multi-hop routing tree. The fusion center is denoted by the circle in the center. There are 500 sensors, each denoted by a dot. The shortest path from each sensor to the fusion center is established and shown in solid lines.

For the single-hop transmission case, the proposed method in Section 4.4 is implemented and compared with two uniform schemes: *Uniform-I* and *Uniform-II* method defined as before. Furthermore, the proposed method for the multi-hop transmission case in Section 4.5 is also implemented. Figure 19 shows the estimation MSE by the proposed method, the *Uniform-I* method, and the *Uniform-II* method with the single-hop transmission scheme, and the estimation MSE by the proposed method

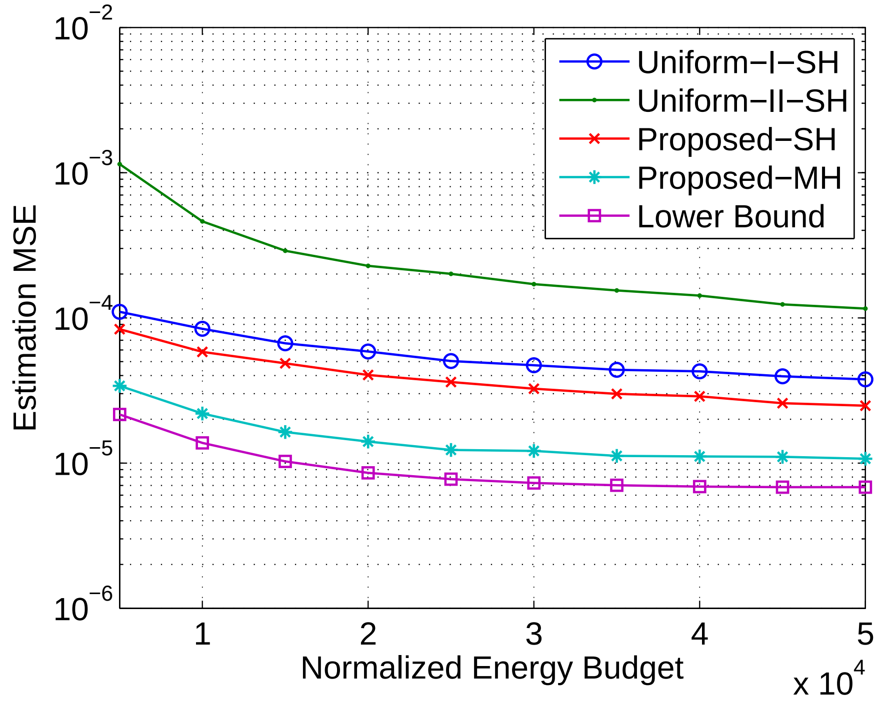


Figure 19: The estimation MSE by the proposed method, *Uniform-I* method, and *Uniform-II* method with the single-hop transmission scheme, and the estimation MSE by the proposed method with the multi-hop transmission scheme, and the theoretical non-achievable lower bound of the estimation MSE with the multi-hop transmission scheme for a randomly deployed heterogeneous sensor network.

with the multi-hop transmission scheme. From Figure 19, we can see that the proposed method outperforms the two uniform schemes and the estimation MSE is significantly reduced by the multi-hop transmission scheme compared with the single-hop transmission scheme under the same total transmission energy. Furthermore, the theoretical lower bound of the estimation MSE with the multi-hop transmission scheme under the total transmission energy is also shown in Figure 19, and it is shown that the estimation MSE of the proposed method is within a factor 2 of the theoretical non-achievable lower bound. Comparing the results in Figure 19 and Figure 16, it can be seen that the proposed method obtains more gain for the heterogeneous networks with random distances than for the heterogeneous networks with equal distances, especially when the total energy constraint is more stringent, since there exists more randomness in the networks.

4.7 Summary

In this chapter, we considered the distributed parameter estimation in energy-constrained wireless sensor networks from the energy-distortion perspective. For a given constraint on the allowable total energy to be used by all sensors at each estimation cycle, we studied the optimal trade-off between the subset of active sensors and the energy used by each active sensor to minimize the estimation MSE. To facilitate the solution, a concept of equivalent unit-energy MSE function was introduced. Then, an optimal distributed estimation algorithm for homogeneous sensor networks and a quasi-optimal distributed estimation algorithm for heterogeneous sensor networks, which are both based on the equivalent unit-energy MSE function, were developed.

Furthermore, the lower and upper bounds of the estimation MSE of the proposed algorithm were discussed and a theoretical energy-distortion bound for the distributed estimation was proved. It is shown that the proposed algorithm is quasi-optimal within a factor 2 of the theoretical lower bound. Simulation results also show that a significant reduction in estimation MSE is achieved by the proposed algorithm compared with other uniform methods.

In Chapter 3 and this chapter, the distributed estimation is addressed from the resource-distortion perspective, where the major goal is to minimize the estimation MSE under a given total resource (bandwidth or energy) constraint for a single estimation cycle. From the application point of view, not only the estimation distortion in each estimation cycle but also the longevity of the whole network needs to be optimized. In the next chapter, we will address the network lifetime optimization for the distributed estimation in the resource-limited wireless sensor networks.

4.8 Appendix: Proof of Theorem 4.2

Assume a subset of sensors $S_\xi = \left\{i_1, \cdots, i_k, \cdots, i_{|S_\xi|}\right\}$ ($i_k \in [1, N]$) are used, and the quantization bit rate of each active sensor $k \in S_\xi$ is b_k^ξ and the corresponding transmission energy allocated is P_k^ξ, such that $\sum_{k \in S_\xi} P_k^\xi = P_c$. Denote this estimation system as C_ξ, the estimation of θ as $\bar{\theta}_\xi$, and its estimation MSE as D_ξ, so the objective is to show that $D_\xi = E(\bar{\theta}_\xi - \theta)^2 > \left(\sum_{k \in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1}$, where $S_{K^{opt}}$ is the optimal subset of active sensors, obtained by the proposed algorithm as shown in Equations (76, 77) such that $\sum_{k \in S_{K^{opt}}} P_k^{opt} = P_c$. The basic idea to prove this statement is to construct another quasi-BLUE estimation system, denoted as C_η, with estimation MSE D_η such that $D_\xi \geq D_\eta > \left(\sum_{k \in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1}$. The estimation system C_η is constructed as follows: only the sensors in $S_{K^{opt}}$ are used, i.e., the subset of active sensors is $S_\eta = S_{K^{opt}}$, the quantization bit rate of each active sensor sensor $k \in S_\eta$ is b_k^η, and the corresponding transmission energy allocated is P_k^η. More specifically,

$$b_k^\eta = \begin{cases} \max(b_k^\xi, b_k^{opt}), & \text{if } k \in S_{K^{opt}} \cap S_\xi, \\ b_k^{opt}, & \text{if } k \in S_{K^{opt}} \backslash S_\xi, \\ 0, & \text{otherwise}, \end{cases} \tag{88}$$

thus

$$P_k^\eta = \begin{cases} \max(P_k^\xi, P_k^{opt}), & \text{if } k \in S_{K^{opt}} \cap S_\xi, \\ P_k^{opt}, & \text{if } k \in S_{K^{opt}} \backslash S_\xi, \\ 0, & \text{otherwise}, \end{cases} \tag{89}$$

where, $k \in S_{K^{opt}} \backslash S_\xi$ means that $k \in S_{K^{opt}}$ but $k \notin S_\xi$. It is noted that $(S_{K^{opt}} \cap S_\xi) \cup (S_{K^{opt}} \backslash S_\xi) = S_{K^{opt}}$.

(1) Show that $D_\eta > \left(\sum_{k \in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1}$.

Since in the constructed estimation system C_η, only the sensors in the subset $S_{K^{opt}}$ are active and limited quantization bit rate b_k^η and limited transmission energy P_k^η are used for each sensor $k \in S_{K^{opt}}$, and $D_0 = \left(\sum_{k \in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1}$ is the lower bound of the estimation MSE of BLUE estimator using the subset of sensors $S_{K^{opt}}$ without quantization bit rate and transmission energy constraints, so $D_\eta > D_0 = \left(\sum_{k\in S_{K^{opt}}} 1/\sigma_k^2\right)^{-1}$.

(2) Show that $D_\eta \le D_\xi$.

Divide S_ξ into three disjoint subset $S_{\xi1}$, $S_{\xi2}$ and $S_{\xi3}$ as follows:

$$\begin{aligned} S_{\xi1} &= \{k : b_k^\xi \ge b_k^{opt}, \text{ and } k \in S_{K^{opt}} \cap S_\xi\}, \\ S_{\xi2} &= \{k : b_k^\xi < b_k^{opt}, \text{ and } k \in S_{K^{opt}} \cap S_\xi\}, \\ S_{\xi3} &= S_\xi \backslash S_{K^{opt}}. \end{aligned} \tag{90}$$

Similarly, divide S_η into three disjoint subset $S_{\eta1}$, $S_{\eta2}$ and $S_{\eta3}$ as follows:

$$\begin{aligned} S_{\eta1} &= \{k : b_k^\xi \ge b_k^{opt}, \text{ and } k \in S_{K^{opt}} \cap S_\xi\}, \\ S_{\eta2} &= \{k : b_k^\xi < b_k^{opt}, \text{ and } k \in S_{K^{opt}} \cap S_\xi\}, \\ S_{\eta3} &= S_{K^{opt}} \backslash S_\xi. \end{aligned} \tag{91}$$

Proposition 4.2. *According to the definitions of $S_{K^{opt}}$, S_ξ, S_η b_k^η, and P_k^η in Equations (77, 88, 89, 90, 91), it is easy to see that:*

1. $S_{\xi1} \cup S_{\xi2} \cup S_{\xi3} = S_\xi$ *and* $S_{\eta1} \cup S_{\eta2} \cup S_{\eta3} = S_\eta = S_{K^{opt}}$,
2. $S_{\eta1} = S_{\xi1}$, $b_k^\eta = b_k^\xi \ge b_k^{opt}$ *and* $P_k^\eta = P_k^\xi \ge P_k^{opt}$ *for any* $k \in S_{\eta1}$,
3. $S_{\eta2} = S_{\xi2}$, $b_k^\eta = b_k^{opt} > b_k^\xi$ *and* $P_k^\eta = P_k^{opt} > P_k^\xi$ *for any* $k \in S_{\eta2}$,
4. $b_k^\eta = b_k^{opt}$ *and* $P_k^\eta = P_k^{opt}$ *for any* $k \in S_{\eta3}$,
5. $S_{\eta2} \subseteq S_{K^{opt}}$, $S_{\eta3} \subseteq S_{K^{opt}}$, *and* $S_{\xi3} \subseteq S_{K^{opt}}^c$, *thus for any* $i \in S_{\eta2} \cup S_{\eta3}$ *and* $j \in S_{\xi3}$, $g(\sigma_i^2, b_i^\eta, a_i, c_i) = g_i^{opt} \le g_j^{opt} \le g(\sigma_j^2, b_j^\xi, a_j, c_j)$ *according to Equation (76). Let* $g_1 = \max\limits_{i\in S_{\eta2}\cup S_{\eta3}} g(\sigma_i^2, b_i^\eta, a_i, c_i)$ *and* $g_2 = \min\limits_{j\in S_{\xi3}} g(\sigma_j^2, b_j^\xi, a_j, c_j)$, *then* $g_1 \le g_2$.

Let $D_\xi = 1/D'_\xi$, and $D_\eta = 1/D'_\eta$. Expressing D'_ξ and D'_η with the concept of the equivalent unit-energy MSE functions as follows:

$$\begin{aligned} D'_\xi &= \sum_{k\in S_{\xi1}\cup S_{\xi2}\cup S_{\xi3}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \\ D'_\eta &= \sum_{k\in S_{\eta1}\cup S_{\eta2}\cup S_{\eta3}} \frac{P_k^\eta}{g(\sigma_k^2, b_k^\eta, a_k, c_k)} \end{aligned} \tag{92}$$

according to Proposition 4.2, then

$$
\begin{aligned}
D'_\eta - D'_\xi = & \left(\sum_{k\in S_{\eta 1}} \frac{P_k^\eta}{g(\sigma_k^2, b_k^\eta, a_k, c_k)} - \sum_{k\in S_{\xi 1}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \right) + \\
& \left(\sum_{k\in S_{\eta 2}} \frac{P_k^\eta}{g(\sigma_k^2, b_k^\eta, a_k, c_k)} - \sum_{k\in S_{\xi 2}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \right) + \\
& \left(\sum_{k\in S_{\eta 3}} \frac{P_k^\eta}{g(\sigma_k^2, b_k^\eta, a_k, c_k)} - \sum_{k\in S_{\xi 3}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \right) \\
= & \left(\sum_{k\in S_{\eta 2}} \frac{P_k^{opt}}{g(\sigma_k^2, b_k^{opt}, a_k, c_k)} - \sum_{k\in S_{\xi 2}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \right) + \\
& \left(\sum_{k\in S_{\eta 3}} \frac{P_k^{opt}}{g(\sigma_k^2, b_k^{opt}, a_k, c_k)} - \sum_{k\in S_{\xi 3}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \right) \\
= & \left(\sum_{k\in S_{\eta 2}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^{opt}, a_k, c_k)} - \sum_{k\in S_{\xi 2}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \right) \\
& + \sum_{k\in S_{\eta 2}} \frac{P_k^{opt} - P_k^\xi}{g(\sigma_k^2, b_k^{opt}, a_k, c_k)} + \sum_{k\in S_{\eta 3}} \frac{P_k^{opt}}{g(\sigma_k^2, b_k^{opt}, a_k, c_k)} \\
& - \sum_{k\in S_{\xi 3}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \\
\geq & \sum_{k\in S_{\eta 2}} \frac{P_k^{opt} - P_k^\xi}{g(\sigma_k^2, b_k^{opt}, a_k, c_k)} + \sum_{k\in S_{\eta 3}} \frac{P_k^{opt}}{g(\sigma_k^2, b_k^{opt}, a_k, c_k)} \\
& - \sum_{k\in S_{\xi 3}} \frac{P_k^\xi}{g(\sigma_k^2, b_k^\xi, a_k, c_k)} \\
\geq & \left(\sum_{k\in S_{\eta 2}} (P_k^{opt} - P_k^\xi) + \sum_{k\in S_{\eta 3}} P_k^{opt} \right) \frac{1}{g_1} - \left(\sum_{k\in S_{\xi 3}} P_k^\xi \right) \frac{1}{g_2}, \\
\geq & \left(\sum_{k\in S_{\eta 2}\cup S_{\eta 3}} P_k^{opt} - \sum_{k\in S_{\xi 2}\cup S_{\xi 3}} P_k^\xi \right) \frac{1}{g_1}.
\end{aligned} \tag{93}
$$

From the total energy constraint, we have

$$\sum_{k\in S_{\xi 1}\cup S_{\xi 2}\cup S_{\xi 3}} P_k^\xi = \sum_{k\in S_{\eta 1}\cup S_{\eta 2}\cup S_{\eta 3}} P_k^{opt} = P_c \tag{94}$$

Since $S_{\eta 1} = S_{\xi 1}$ and $P_k^\xi \geq P_k^{opt}$ for any $k \in S_{\xi 1}$ as shown in Proposition 4.2, then

$$\sum_{k\in S_{\xi 2}\cup S_{\xi 3}} P_k^\xi \leq \sum_{k\in S_{\eta 2}\cup S_{\eta 3}} P_k^{opt}, \tag{95}$$

thus

$$D'_\eta - D'_\xi \geq \left(\sum_{k\in S_{\eta 2}\cup S_{\eta 3}} P_k^{opt} - \sum_{k\in S_{\xi 2}\cup S_{\xi 3}} P_k^\xi \right) \frac{1}{g_1} \geq 0, \tag{96}$$

therefore,

$$D_\xi \geq D_\eta. \tag{97}$$

From (1) and (2) above, we get

$$D_\xi \geq D_\eta > \left(\sum_{k \in S_{Kopt}} \frac{1}{\sigma_k^2} \right)^{-1}, \tag{98}$$

thus the theorem is proved.

CHAPTER V

NETWORK LIFETIME OPTIMIZATION FOR DISTRIBUTED ESTIMATION

Distortion is one of the major system performance measurements for the distributed estimation in wireless sensor networks. To minimize the estimation distortion while meeting the resource (bandwidth and energy) limitation of wireless sensor networks, the rate-constrained distributed estimation algorithms [6, 7, 32, 48, 53, 64, 65, 69, 70, 75, 79,80] and the energy-constrained distributed estimation algorithms [5,46,52,91,92,94] have been widely investigated. However, all these algorithms are not necessarily optimal in the sense of network lifetime.

Network lifetime is another major system performance measurement in wireless sensor networks and it is also widely addressed in the literature [12, 13, 17, 18, 31, 38, 41, 42, 66, 68, 81, 87, 101, 102]. However, the network lifetime issue for the distributed estimation applications in wireless sensor networks has not yet been addressed explicitly. Furthermore, it is desirable to address both the estimation distortion and the network lifetime jointly, thus it involves not only the local information processing but also inter-sensor communication and networking.

In this chapter, we study the *lifetime-distortion* issue for the estimation applications in wireless sensor networks, where the lifetime is defined as *the estimation task cycles successfully accomplished until the network can not perform the task with a given distortion requirement any more*. In resource-limited wireless sensor networks, both local quantization and multi-hop transmission are essential to save transmission energy and thus prolong the network lifetime. To maximize the network lifetime for the estimation application, three factors are needed to be optimized together: *(i)* *source coding, i.e., quantization level of each observation, (ii) source throughput, i.e. total number of observations or total information bits generated by each sensor, and (iii) multi-hop routing path to transmit the observations from all sensors to the fusion center.*

The rest of the chapter is organized as follows. Section 5.1 introduces the system model of the distributed estimation in multi-hop wireless sensor networks. Section 5.2 introduces a new notion of network lifetime, called function-based network lifetime, and formulates its upper bound. Section 5.3 addresses the network lifetime maximization for the single-hop wireless sensor networks, where the optimal source coding based on the equivalent unit-resource MSE function is developed. Section 5.4 formulates the network lifetime bound maximization problem for the multi-hop wireless sensor networks as a nonlinear programming (NLP) problem, and then decouples the original problem into two sub-problems, i.e., (*i*) source coding optimization, and (*ii*) joint source throughput and multi-hop routing optimization, without compromising the optimality. These two problems are addressed in Section 5.4.3 and section 5.5, respectively. Section 5.6 gives some simulation results that demonstrate the efficiency of the proposed algorithms. Finally, conclusions are given in Section 5.7. The proofs of some theorems presented in this chapter are delegated to the appendix in Section 5.8.

5.1 System Model and Preliminaries

Consider a dense sensor network including N distributed sensor nodes and a fusion center, denoted as node $N+1$, to observe and estimate an unknown parameter θ. An example network is shown in Figure 20.

First, each sensor k can make observations on the unknown parameter θ. The observations are corrupted by additive noise and described by

$$x_k = \theta + n_k, \quad k = 1, \cdots, N. \tag{99}$$

We assume that the observation noises of all sensors n_k ($k = 1, \cdots, N$) are zero mean,

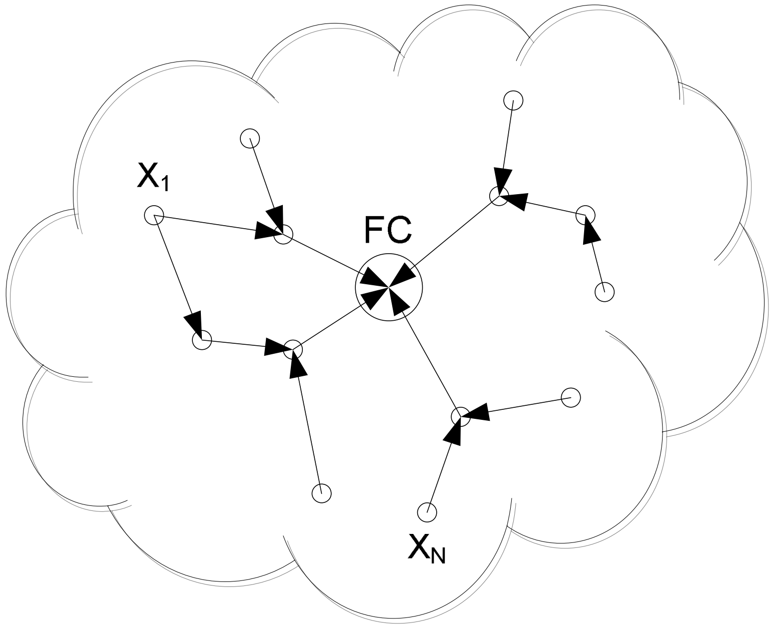


Figure 20: An example of a wireless sensor network with N distributed sensor nodes. Each sensor can observe the phenomenon, quantize and transmit its observation to the fusion center (FC) via multi-hop wireless channel, and the fusion center makes the final estimation based on all the received messages. In directed solid lines, a chosen multi-hop routing path is shown, where the data from a sensor can be relayed by multiple sensors, meanwhile a sensor can relay data for multiple sensors.

spatially uncorrelated with variance σ_k^2, while the noise at each sensor is assumed to be temporally i.i.d distributed, otherwise unknown.

Subject to severe bandwidth and energy limitations, each sensor is prevented from transmitting real-valued (analogy) data to the fusion center, that is, a local quantization $m_k = Q_k(x_k)$ is performed before transmission, where $Q_k(x_k)$ is a quantization function, and only the quantization message m_k is transmitted to the fusion center via multi-hop wireless channel. Similar with that used in Section 3.1 and Section 4.1, the probabilistic quantization scheme is used at local sensors. As shown in Section 3.1.1, let $m(x, b)$ be a b-bit probabilistic quantization of bounded observation signal $x \in [-W, W]$ with noise variance σ^2; then $m(x, b)$ is an unbiased estimation of θ with a variance

$$E\left(|m(x,b)-\theta|^2\right) \leq \sigma^2 + \frac{W^2}{\left(2^b-1\right)^2} := \pi^2(\sigma^2, b), \tag{100}$$

where $W^2/(2^b-1)^2$ for $b>0$ denotes the upper bound of the quantization noise variance.

Assume there are K received observations $(m_1, m_2, \cdots, m_K)$ at the fusion center, then the fusion center produces a final estimation of θ by combining all the available observations using a fusion function $f : \bar{\theta} = f(m_1, m_2, \cdots, m_K)$. Similar with that used in Section 3.1 and Section 4.1, the quasi-BLUE estimation scheme is adopted at the fusion center. Suppose all the observations of the K active sensors $x_k (k = 1, \cdots, K)$ are quantized into b_k-bits discrete messages $m_k(x_k, b_k)$ respectively with the probabilistic quantization scheme. Based on the quantized messages m_k, the quasi-BLUE estimator at the fusion center has the following form:

$$\bar{\theta} = \left(\sum_{k=1}^{K} \frac{1}{\pi_k^2(\sigma_k^2, b_k)} \right)^{-1} \sum_{k=1}^{K} \frac{m_k}{\pi_k^2(\sigma_k^2, b_k)}. \tag{101}$$

Notice that $\bar{\theta}$ is an unbiased estimation of θ because every m_k is unbiased. Moreover, the estimation MSE of the quasi-BLUE estimator is

$$E(\bar{\theta} - \theta)^2 \leq \left(\sum_{k=1}^{K} \frac{1}{\pi_k^2(\sigma_k^2, b_k)} \right)^{-1}. \tag{102}$$

For the given estimation system, the major goal of this chapter is to study the optimal estimation scheme to maximize the network lifetime. In the next section, we first introduce a novel definition of network lifetime, called function-based network lifetime; then we give a upper-bound on the function-based network lifetime for a given sensor networks with limited-energy supply.

5.2 Network Lifetime for Estimation

Network lifetime is a critical concern in the design of wireless sensor networks. In this section, we first define the network lifetime and then formulate the network lifetime maximization problem.

5.2.1 Function-based Network Lifetime

In the literature, many different lifetime definitions are used, such as duration of time until the first sensor failure due to battery depletion [17], fraction of surviving nodes in a network [81, 101], mean expiration time [87], etc. However, these notions of network lifetime mainly focus on the time until the first node or a fraction of nodes deplete even though the remaining network may be still functional from the application perspective. In this research, we introduce a notion of function-based network lifetime, which focuses on whether the network can perform a given task instead of whether any individual sensor is dead.

Definition 5.1 (Function-based Network Lifetime)**.** *For the estimation application, the network is considered functional if it can produce an estimation satisfying a given distortion requirement D_r; otherwise it is nonfunctional. The network lifetime L is defined as the estimation task cycles accomplished before the network becomes non-functional, where each time when the sensor network makes an estimation is denoted as an estimation task cycle.*

5.2.2 Upper-Bound on Function-based Network Lifetime

At different estimation cycles, the parameter θ is assumed to be unrelated, and the estimation at each cycle is performed independently using only the observations made by all sensors in the given estimation cycle. Based on the system model in Section 5.1, assume a sensor network with N sensors, each with observation noise variance σ_k^2 $(k = 1, \cdots, N)$. To satisfy the given estimation distortion requirement D_r at each estimation cycle, a subset of sensors is required to observe the parameter θ and transmit their quantized measurements to the fusion center to make the final estimation.

Proposition 5.1. *Assume sensor k $(k = 1, \cdots, N)$ make a total of M_k measurements and quantize its measurements using probabilistic quantization scheme to $b_{k,i}$ $(i = 1, \cdots, M_k)$ bits, respectively, before it depletes. Then the function-based network* *lifetime L for the estimation application is bounded as follows:*

$$L \leq D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right), \tag{103}$$

where N, σ_k^2, M_k, $b_{k,i}$, and D_r are defined as above, and $\pi_k^2(\sigma_k^2, b_{k,i})$ as defined in Equation (100).

Proof. See Appendix 5.8.1. □

It is noted that the upper bound shown in Proposition 5.1 could be closely approached by appropriately scheduling the subset of active sensors in each estimation cycle such that the actual estimation MSE obtained is equal to or slightly smaller than D_r.

Based on the estimation system model and the definition of function-based network lifetime, in the following sections, we will study how to maximize the upper-bound of the function-based network lifetime in Equation (103) under the energy resource constraint of each sensor. In the next section, we first study a special case – single-hop wireless sensor networks.

5.3 Single-Hop Wireless Sensor Networks

In the single-hop wireless sensor networks, each sensor transmits its observations to the fusion center directly, so all energy of each sensor can be used to transmit its own data instead of relaying other sensors' data. Then the network lifetime optimization problem under the energy resource constraint of each sensor can be cast as follows:

$$\begin{aligned} \max \quad & D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right), \\ s.t. \quad & \sum_{i=1}^{M_k} e_k(b_{k,i}) \leq P_k, \qquad \forall k \in [1, N], \end{aligned} \tag{104}$$

where P_k is the total energy resource of sensor k, $e_k(b_{k,i})$ is the transmission energy cost for sensor k to transmit a $b_{k,i}$-bit quantization message to the fusion center, and $M_k \geq 0$ and $b_{k,i} \geq 0$ defined as before are variables to be optimized.

5.3.1 Equivalent Unit-Resource MSE Function

To facilitate the solution to Equation (104), we first introduce a concept of the equivalent unit-resource MSE function.

Definition 5.2 (Equivalent Unit-Resource MSE Function). *For a quantized message from a sensor with observation noise variance σ^2 and quantization bit rate b, the estimation variance is $\pi^2(\sigma^2, b) := \sigma^2 + W^2/(2^b - 1)^2$ as shown in Section 5.1. Denote the resource cost by this message is $r(b)$. Then, the equivalent unit-resource MSE function is defined as*

$$g(\sigma^2, b) := r(b) \cdot \pi^2(\sigma^2, b) = r(b) \cdot \left(\sigma^2 + \frac{W^2}{(2^b - 1)^2}\right). \tag{105}$$

Based on this definition, from the estimation MSE aspect, a sensor with quantization bit rate b, resource cost $r(b)$ and estimation MSE $\pi^2(\sigma^2, b)$, can be treated as $r(b)$ equivalent unit-resource sensors, each with the same estimation MSE $g(\sigma^2, b)$. That is why $g(\sigma^2, b)$ is called equivalent unit-resource MSE function. It is worth noting that this definition is quite generic, where the resource can be bandwidth, energy, etc. When the bandwidth resource is considered, i.e., $r(b) = b$, the equivalent unit-resource MSE function is the same as the equivalent 1-bit MSE function defined in Section 3.2. When the energy resource is considered, i.e., $r(b) = e(b)$, the equivalent unit-resource MSE function is the same as the equivalent unit-energy MSE function defined in Section 4.2. So the equivalent unit-resource MSE function is a generalization of the equivalent 1-bit MSE function and the equivalent unit-energy MSE function.

Let $r(b)$ in Equation (105) be the transmission energy cost $e(b)$. We consider two different transmission models: (1) the binary transmission model and (2) the quadrature amplitude modulation (QAM) based transmission model. Assume the transmission distance from sensor k to the fusion center is d_k, and the channel power attenuation factor is $a_k = d_k^\alpha$, where α is the path loss exponent. Then, as shown in

Section 4.1.2, the transmission energy cost for the binary transmission model is

$$e_1(b_k) = c_1 \cdot a_k \cdot b_k, \tag{106}$$

where c_1 is a system constant. The transmission energy cost for the QAM-based model [23, 24] is

$$e_2(b_k) = c_2 \cdot a_k \cdot (2^{b_k} - 1), \tag{107}$$

where c_2 is a system constant. For both transmission models above, it can be shown that the corresponding equivalent unit-resource MSE functions $g(\sigma^2, b)$ defined in Equation (105) are convex over b.

Based on the convexity of $g(\sigma^2, b)$, we further define the optimal unit-resource MSE function $g^{opt}(\sigma^2)$, and the corresponding optimal quantization bit rate $b^{opt}(\sigma^2)$ and optimal transmission energy $e^{opt}(\sigma^2)$ as follows:

$$\begin{aligned} b^{opt}(\sigma^2) &= \arg\min_{b \in \mathbb{Z}^+} g(\sigma^2, b), \\ g^{opt}(\sigma^2) &= \min_{b \in \mathbb{Z}^+} g(\sigma^2, b) = g(\sigma^2, b^{opt}(\sigma^2)), \\ e^{opt}(\sigma^2) &= e(b^{opt}(\sigma^2)). \end{aligned} \tag{108}$$

It is noted that the minimization in Equation (108) involves just a simple one-dimensional numerical search over $b \in \mathbb{Z}^+$.

5.3.2 Network Lifetime Maximization for Single-Hop WSNs

Based on the concept of the equivalent unit-energy MSE function, the upper bound of the function-based network lifetime in Proposition 5.1 can be maximized as in the following Theorem.

Theorem 5.1. *The bound of function-based network lifetime for estimation is*

$$L \leq D_r \left(\sum_{k=1}^{N} \frac{P_k}{g_k^{opt}(\sigma_k^2)} \right) = D_r \left(\sum_{k=1}^{N} \frac{P_k}{e_k^{opt}(\sigma_k^2) \cdot \pi_k^2(\sigma_k^2, b_k^{opt}(\sigma_k^2))} \right), \tag{109}$$

where N, σ_k^2, M_k, $b_{k,i}$, and D_r are defined as before, and $g_k^{opt}(\sigma_k^2)$, $b_k^{opt}(\sigma_k^2)$ and $e_k^{opt}(\sigma_k^2)$ are the optimal unit-resource MSE function, optimal quantization bit rate, and optimal transmission energy per observation, of sensor k, respectively.

Proof. Assume sensor k makes M_k measurements, each with quantization bit rate $b_{k,i}$ and transmission energy cost $e_k(b_{k,i})$, before the sensor depletes, i.e., $\sum_{i=1}^{M_k} e_k(b_{k,i}) \le P_k$. Then as shown in Equation (103), the network lifetime bound is

$$L \le D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right). \tag{110}$$

According to the definition of $g(\sigma^2, b)$ and $g^{opt}(\sigma^2)$ and the energy constraints in Equation (104),

$$\begin{aligned} L &\le D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right) \\ &= D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{e_k(b_{k,i})}{g_k(\sigma_k^2, b_{k,i})} \right) \\ &\le D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{e_k(b_{k,i})}{g_k^{opt}(\sigma_k^2)} \right) \\ &\le D_r \left(\sum_{k=1}^{N} \frac{P_k}{g_k^{opt}(\sigma_k^2)} \right) \end{aligned} \tag{111}$$

thus the theorem is proved. □

Note that the equality in Equation (111) is achieved when each sensor node adopts optimal quantization bit rate $b^{opt}(\sigma^2)$ and optimal transmission energy $e^{opt}(\sigma^2)$ to quantize and transmit its observations. As shown before, the optimal quantization bit rate $b^{opt}(\sigma^2)$ and optimal transmission energy $e^{opt}(\sigma^2)$ of each sensor can be easily obtained by minimizing its equivalent unit-resource MSE function, which only depends on its own observation noise variance and transmission system parameters; therefore, this optimization can be done in a completely distributed manner.

5.4 Multi-Hop Wireless Sensor Networks

In the energy-limited wireless sensor networks, multi-hop transmission is essential to save transmission energy and thus prolong the network lifetime. In this section, we will address the network lifetime optimization problem for the general multi-hop wireless sensor networks, where the data is transmitted from each sensor to the fusion center through multi-hop channels. In this section, only the binary transmission model is used.

5.4.1 Nonlinear Programming (NLP) Formulation

Model the wireless sensor network as a directed graph $G(V, E)$, where V is the set consisting of all the N sensor nodes and the fusion center (node $N+1$), i.e., $V = [1, N+1]$, E is the set of directed links in the network. An edge $(i, j) \in E$ iff $d_{i,j} \leq R$, where $d_{i,j}$ is the distance between node i and node j, and R is the maximum transmission range. The link cost to transmit a unit bit information from node i to node j, denoted as $C_{i,j}$, depends on the distance $d_{i,j}$ between them based on the energy model in Equation (106) as follows,

$$C_{i,j} = \begin{cases} cd_{i,j}^{\alpha}, & \text{if } d_{i,j} \leq R \\ +\infty, & \text{otherwise} \end{cases} \tag{112}$$

where c and α are defined as before.

Assume each sensor has a limited energy supply P_k ($k = 1, \cdots, N$). During the lifetime of the network, assume sensor k make a total of M_k measurements and quantize its measurements to $b_{k,i}$ ($i = 1, \cdots, M_k$) bits, respectively. Denote the source throughput of sensor node k, i.e., the total amount of data in bits generated at sensor node k as S_k, and the amount of data in bits transmitted from sensor node i to sensor node j as $f_{i,j}$. According to network lifetime bound shown in Equation (103), the network lifetime maximization problem can be formulated as a nonlinear programming (NLP) problem as follows:

$$\text{maximize} \quad D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right) \tag{113}$$

subject to

$$\sum_{i=1, i \neq k}^{N} f_{i,k} + S_k = \sum_{j=1, j \neq k}^{N+1} f_{k,j}, \quad \forall k \in [1, N] \tag{114}$$

$$\sum_{j=1, j \neq k}^{N+1} f_{k,j} C_{k,j} \leqslant P_k, \quad \forall k \in [1, N] \tag{115}$$

$$S_k = \sum_{i=1}^{M_k} b_{k,i}, \quad \forall k \in [1, N] \tag{116}$$

where

$$\begin{aligned} &S_k \geq 0, \quad \forall k \in [1, N] \\ &M_k \geq 0, \quad \forall k \in [1, N] \\ &b_{k,i} \geq 0, \quad \forall k \in [1, N], i \in [1, M_k] \\ &f_{i,j} \geq 0, \quad \forall i \in [1, N], j \in [1, N+1] \end{aligned} \tag{117}$$

where Equation (114) and Equation (115) represent two constraints of the optimization problem:

1. *flow conservation*: the amount of data transmitted by a sensor node is equal to the sum of the amount of data received by the sensor node and the amount of data generated by the sensor node itself.

2. *energy constraint*: the amount of data transmitted by a sensor node is limited by the energy supply of the sensor node.

It is noted that the problem given above is a nonlinear programming problem since the objective function in Equation (113) nonlinearly depends on the variables $b_{k,i}$.

5.4.2 Separation of Source Coding with Multi-Hop Routing

To maximize the objective function in Equation (113), there are three factors needed to be optimized together: (i) source coding at each sensor, i.e., quantization level $b_{k,i}$ for each observation i of each sensor k, (ii) source throughput of each sensor, i.e., the total number of observations M_k and the total amount of data in bits S_k generated at each sensor k, and (iii) multi-hop routing, i.e., the feasible network flow $\{f_{i,j} : i, j \in [1, N+1]\}$ satisfying both the flow conservation constraint in Equation (114) and energy constraint in Equation (115). Fortunately, the source coding optimization can be decoupled from the source throughput and multi-hop routing optimization as shown in Proposition 5.2.

Proposition 5.2. *For the nonlinear programming model stated in Equations (113, 114, 115, 116), given the source throughput* $\{S_k, k \in [1, N]\}$*, the source coding optimization can be decoupled from multi-hop routing optimization.*

Proof. As shown in Equation (113), the objective function is

$$D_r\left(\sum_{k=1}^{N}\sum_{i=1}^{M_k}\frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})}\right),$$

which only depends on the source throughput $S_k = \sum_{i=1}^{M_k} b_{k,i}$, ($k \in [1, N]$) and source coding scheme, but does not depend on how the source data is transmitted to the fusion center. On the other hand, the flow conservation in Equation (114) and the energy constraint in Equation (115) only depends on the source throughput S_k, but does not depend on the source coding. Thus, given the source throughput S_k of each sensor k, the source coding optimization is independent from the multi-hop routing optimization. □

According to the separation principle of source coding optimization with multi-hop routing optimization, we can solve the original optimization problem stated in Equations (113, 114, 115, 116) in two steps without loss of optimality: (i) optimizing the source coding for given source throughput, and (ii) optimizing the source throughput and multi-hop routing jointly, based on the optimal source coding. In the next two sections, we will address these two sub-problems, respectively.

5.4.3 Source Coding Optimization and Network Lifetime Bound

In this section, we optimize the source coding for a given source throughput S_k of each sensor $k \in [1, N]$, i.e., find the optimal quantization level $b_{k,i}$ for each observation i of each sensor k to maximize the network lifetime bound. Mathematically, the problem is formulated as follows:

$$\begin{aligned}&\max\ D_r\left(\sum_{k=1}^{N}\sum_{i=1}^{M_k}\frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})}\right)\\ &s.t.\ \sum_{i=1}^{M_k} b_{k,i} = S_k, \quad \forall k \in [1, N],\end{aligned} \tag{118}$$

where $M_k \geq 0$ and $b_{k,i} \geq 0$ defined as before are variables to be optimized.

Compared the problem in Equation (118) with that in Equation (104), it is easy to see that the problem in Equation (118) is a special case of the problem in Equation (104) with $e(b) = b$ and $P_k = S_k$. So the source coding method based on the equivalent unit-resource MSE function in Section 5.3 can be used. Here the resource is the transmission bit rate, i.e., $r(b) = b$, and the equivalent unit-resource MSE function is retrogressed to the equivalent 1-bit MSE function.

First, the equivalent 1-bit MSE function is defined as

$$g(\sigma^2, b) := b \cdot \pi^2(\sigma^2, b) = b \cdot \left(\sigma^2 + \frac{W^2}{(2^b - 1)^2}\right). \tag{119}$$

Since $g(\sigma^2, b)$ is convex over $b > 0$, we further define the optimal 1-bit MSE function $g^{opt}(\sigma^2)$ and the corresponding optimal quantization bit rate $b^{opt}(\sigma^2)$ as follows:

$$\begin{aligned} b^{opt}(\sigma^2) &= \arg\min_{b \in \mathbb{Z}^+} g(\sigma^2, b), \\ g^{opt}(\sigma^2) &= \min_{b \in \mathbb{Z}^+} g(\sigma^2, b) = g(\sigma^2, b^{opt}(\sigma^2)), \end{aligned} \tag{120}$$

where the minimization involves just a simple one-dimensional numerical search. As shown in Proposition 3.1, the optimal 1-bit MSE function $g^{opt}(\sigma^2)$ increases over the observation noise variance σ^2.

Based on the definitions above, the network lifetime bound for estimation can be reformulated as a linear function of the source throughput S_k $(k = 1, \cdots, N)$ as shown in Theorem 5.2.

Theorem 5.2. *Given the source throughput S_k of all sensor nodes $k \in [1, N]$ and the estimation distortion requirement D_r, the bound of function-based network lifetime for estimation is*

$$L \leq D_r \left(\sum_{k=1}^{N} \frac{S_k}{g_k^{opt}(\sigma_k^2)} \right), \tag{121}$$

where $g_k^{opt}(\sigma_k^2)$ is the optimal 1-bit MSE function of sensor node k.

Proof. Assume sensor $k \in [1, N]$ makes a total of M_k measurements, each with quantization bit rate $b_{k,i}$, respectively, such that $\sum_{i=1}^{M_k} b_{k,i} \leq S_k$. Then as shown in Equation (103), the network lifetime bound is

$$L \leq D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right). \tag{122}$$

According to the definition of $g(\sigma^2, b)$ and $g^{opt}(\sigma^2)$ in Equation (119, 120) and the source throughput constraints in Equation (118),

$$\begin{aligned} L &\leq D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right) \\ &= D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{b_{k,i}}{g_k(\sigma_k^2, b_{k,i})} \right) \\ &\leq D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{b_{k,i}}{g_k^{opt}(\sigma_k^2)} \right) \\ &\leq D_r \left(\sum_{k=1}^{N} \frac{S_k}{g_k^{opt}(\sigma_k^2)} \right) \end{aligned} \tag{123}$$

thus the theorem is proved. □

Note that the equality in Equation (123) is achieved when each sensor node adopts optimal source coding, i.e., optimal quantization bit rate $b^{opt}(\sigma^2)$ to quantize its observations. As shown before, the optimal quantization bit rate $b^{opt}(\sigma^2)$ of each sensor can be easily obtained by minimizing its equivalent 1-bit MSE function, which only depends on its own observation noise variance, therefore, this optimization can be done in a distributed manner. It is also worth noting that the optimal source coding is independent from the source throughput, while the source throughput at each sensor determines the total number of observations the sensor makes.

5.5 Joint Optimization of Source Throughput and Multi-Hop Routing

As shown in Equation (121) in Theorem 5.2, the network lifetime bound depends on the source throughput S_k for all sensors $k \in [1, N]$, which are unknown variables to be optimized. In multi-hop wireless sensor networks, each sensor not only transmits the data generated by itself, but also relays the data for other sensors. Since the total amount of data each sensor can transmit and relay is limited by the energy supply of the sensor node, the source throughput of each sensor and the multi-hop routing path from each sensor to the sink node need to be optimized together.

5.5.1 Linear Programming (LP) Formulation

As shown in Theorem 5.2, the nonlinear objective function in Equation (113) can be reformulated as a linear function of the source throughput S_k ($k \in [1, N]$) by the optimal source coding, then the original network lifetime bound maximization problem for multi-hop wireless sensor networks shown in Section 5.4.1 can be reformulated as a linear programming (LP) problem as follows:

$$\text{maximize} \quad D_r \left(\sum_{k=1}^{N} \frac{S_k}{g_k^{opt}(\sigma_k^2)} \right) \tag{124}$$

subject to

$$\sum_{i=1, i \neq k}^{N} f_{i,k} + S_k = \sum_{j=1, j \neq k}^{N+1} f_{k,j}, \quad \forall k \in [1, N] \tag{125}$$

$$\sum_{j=1, j \neq k}^{N+1} f_{k,j} C_{k,j} \leqslant P_k, \quad \forall k \in [1, N] \tag{126}$$

where

$$\begin{aligned} &S_k \geq 0, \quad \forall k \in [1, N] \\ &f_{i,j} \geq 0, \quad \forall i \in [1, N], j \in [1, N+1] \end{aligned} \tag{127}$$

and all variables are defined as before.

In summary, the network lifetime bound maximization for estimation can be formulated as a linear programming problem as shown in Equation (124, 125, 126), which can be easily solved using any LP solver, such as [1] used in our simulations.

The linear programming problem in Equation (124, 125, 126) can be understood as a *weighted data gathering* problem since the objective function in Equation (124) is the weighted sum of the amount of data generated at all sensors, where the weight of the data from sensor k ($k = 1, \cdots, N$) is the inverse of its corresponding optimal 1-bit MSE function $g_k^{opt}(\sigma_k^2)$ as shown in Equation (124). As shown in Proposition 3.1, $g^{opt}(\sigma^2)$ increases over σ^2, so the weight decreases over σ^2, that is, it is more desirable to get data from the sensor nodes with small observation noise. It is also noted that if some sensors in the networks can act only as a relay, i.e., no observation capabilities, the linear programming model above still works by simply setting the weights of the data from the relay-only sensors as 0.

5.5.2 Character-based Routing

Though the multi-hop routing path for the weighted data gathering problem can be easily obtained by solving the associated linear programming problem using any LP solver, it is interesting to note that, in the optimal multi-hop routing structure for this problem, a sensor node only relays data generated by sensor nodes with higher importance, i.e., bigger weight, as shown in Theorem 5.3. That is to say, the optimal routing is based on the character (fidelity and importance) of the sensor nodes, thus it is called *character-based routing*. Character-based routing is a new notion for routing and it is different from the traditional distance-based routing, such as shortest path tree, where a sensor node closer to the sink node relays information for sensor nodes farther away from the sink node.

Theorem 5.3. *The optimal routing structure for the weighted data gathering problem shown in Equations (124, 125, 126) is character-based routing, where a sensor node only relays data generated by sensor nodes with higher importance, i.e., bigger weight. More specifically, in the optimal flow and routing solution, let η be a sub flow with data volume S, generated at sensor i_0 and relayed by sensors $i_1, \cdots, i_T$ sequentially to the fusion center, i.e.,*

$$S_{i_0}^{\eta} = f_{i_0,i_1}^{\eta} = f_{i_1,i_2}^{\eta} = \cdots = f_{i_{T-1},i_T}^{\eta} = f_{i_T,N+1}^{\eta} = S, \tag{128}$$

then,

$$\sigma_{i_0}^2 \leq \sigma_{i_t}^2, \quad \forall t \in [1, T]. \tag{129}$$

Proof. See Appendix 5.8.2. □

5.5.3 Special Case: Homogeneous Networks

In homogeneous wireless sensor networks, where each sensor has the same observation noise variance, i.e., $\sigma_k^2 = \sigma^2$ $(k = 1, \cdots, N)$, single-hop routing path, i.e., all sensors transmit their observations to the fusion center directly, can maximize the weighted data gathering as shown in Proposition 5.3. Furthermore, the network lifetime bound for estimation can be easily quantified as shown in Proposition 5.4.

Proposition 5.3. *In a homogeneous network with N sensors and observation noise variance σ^2, single-hop routing can maximize the weighted data gathering as in Equations (124, 125, 126).*

Proof. See Appendix 5.8.3. □

Proposition 5.4. *In a homogeneous network with N sensors and observation noise variance σ^2, denote the energy supply of sensor k $(k = 1, \cdots, N)$ as P_k, then the network lifetime bound for estimation is*

$$L \leq D_r \left(\sum_{k=1}^{N} \frac{P_k}{C_{k,N+1} \cdot g^{opt}(\sigma^2)} \right), \tag{130}$$

where D_r is the estimation distortion requirement, and $C_{k,N+1}$ defined as in Equation (112) denotes the energy cost for sensor k to transmit 1-bit message to the fusion center directly.

Proof. As shown in Proposition 5.3, single-hop routing can maximize weighted data gathering, thus the network lifetime bound for estimation. In single-hop wireless sensor network, each sensor transmits all its measurements to the fusion center directly, and no energy is used to relay other sensors' data, thus the maximum source throughput of each sensor node is easily obtained as $S_k = P_k/C_{k,N+1}$. Therefore, according to Theorem 5.2, the network lifetime bound for estimation in a homogeneous network is $L \leq D_r \left(\sum_{k=1}^{N} \frac{P_k}{C_{k,N+1} \cdot g^{opt}(\sigma^2)} \right)$. $\square$

5.6 Simulation Results

In homogeneous sensor networks, the network lifetime bound for estimation is maximized by single-hop routing and optimal source coding as shown in Proposition 5.4, while in heterogeneous sensor networks, the network lifetime bound for estimation is maximized by optimal source coding and optimal multi-hop routing jointly. To demonstrate the performances of the optimal coding scheme in Section 5.4.3 and the optimal multi-hop routing in Section 5.5, we simulate a homogeneous sensor network and a heterogeneous sensor network, respectively.

5.6.1 Homogeneous Sensor Networks

In this section, we simulate a homogeneous sensor network with $N = 500$ sensors, where the noise variance σ_k^2, the initial energy P_k, and the distances to the fusion center d_k for all sensors are the same. Without loss of generality, we assume $d_k = 1$, the normalized initial energy $P_k/c = 10000$, the range of the observation signal is $[-1, 1]$, i.e., $W = 1$, and path loss exponent $\alpha = 2$ (free space). Define the signal-to-noise ratio (SNR) as $SNR = 10 \log_{10}(W^2/\sigma^2)$ and generate different SNR by changing the observation noise variance σ^2. In order to demonstrate the efficiency of the proposed method, we compare the proposed algorithm with a heuristic method, where each sensor uses the same amount of energy to achieve the distortion requirement at each estimation task period, thus all the sensors will deplete at the same time.

Denote the estimation MSE of the clairvoyant estimator as $D_0 = \left(\sum_{k=1}^{N}(1/\sigma_k^2)\right)^{-1}$ and define the normalized estimation MSE requirement as $D_n = D_r/D_0$. Figure 21(a) and Figure 21(b) show the ratio of network lifetime bound by the proposed algorithm

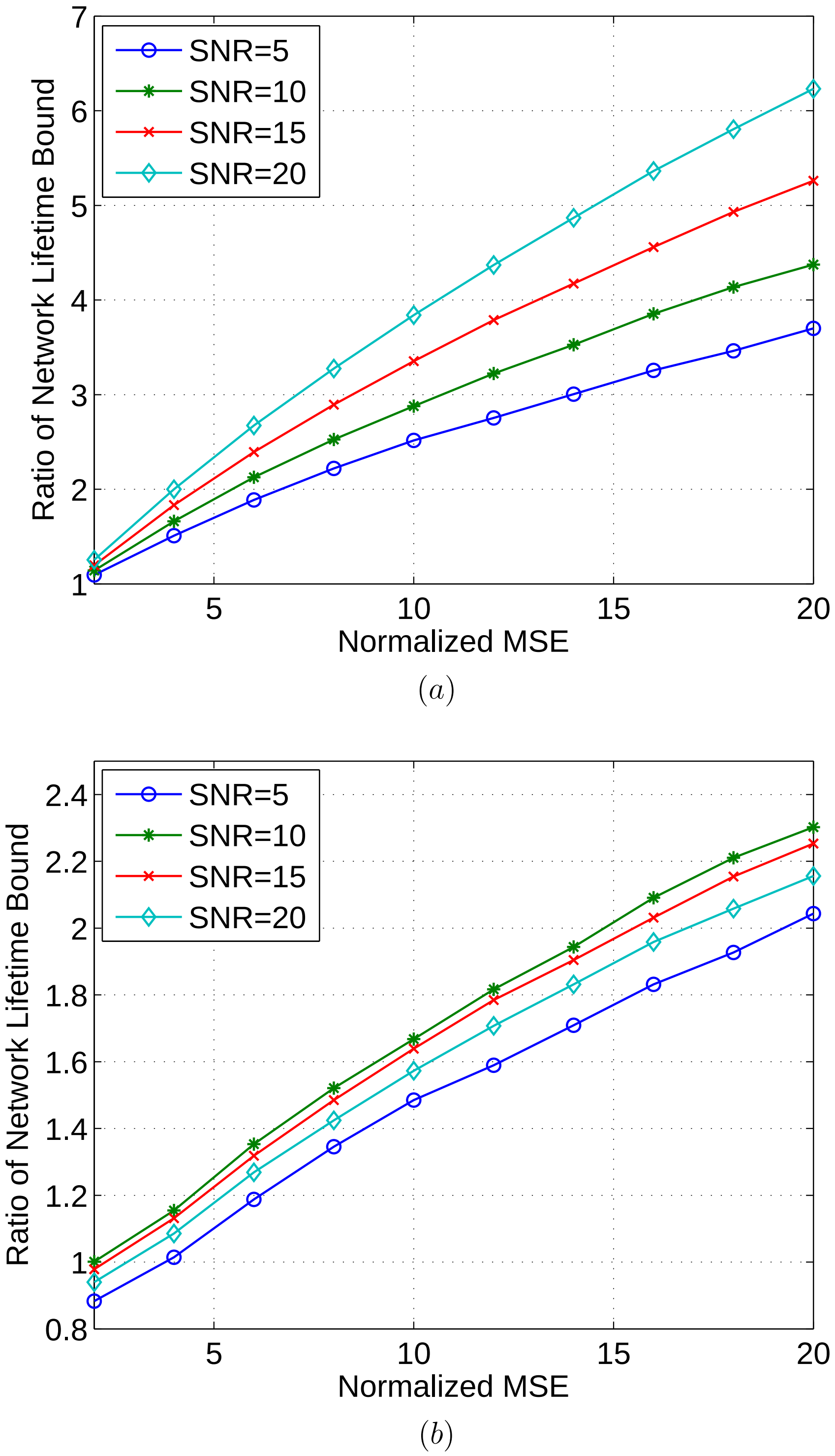


Figure 21: Ratio of network lifetime bound for homogeneous sensor networks under different SNRs and normalized estimation MSE requirements: (a) binary model, (b) QAM model.

to that by the heuristic method under different SNRs and different normalized estimation MSE requirements using the binary and QAM based transmission models, respectively. From Figure 21(a) and Figure 21(b), we can see that a significant gain on network lifetime is achieved by the proposed algorithm compared with heuristic method for both energy models, and the gain for binary model is larger than the gain for QAM model. Also the gain increases with normalized estimation MSE requirement increasing, which is because that the energy is less optimally used by the heuristic method when the normalized estimation MSE requirement increases.

5.6.2 Heterogeneous Sensor Networks

In this section, we simulate a heterogeneous sensor network with N sensors, where the observation noise variance of each sensor is assumed to be

$$\sigma_k^2 = \beta + \gamma z_k, \quad k = 1, \cdots, N, \tag{131}$$

where β models the network-wide noise variance threshold, γ controls the underlying variation from sensor to sensor, and $z_k \sim \chi_1^2$ is a Chi-Square distributed random variable with one degree of freedom. It is noted that the network is homogeneous for the special case of $\gamma = 0$. In the experiments, we assume $\beta = 0.01$ and $\gamma = 0.00, 0.05, 0.10, 0.15,$ or 0.20. Assume all sensors are independently and uniformly distributed in a rectangular region of $[-5, 5, -5, 5]$, and the fusion center is located at the central point of the region, i.e., $(0, 0)$. And the initial energy is still assumed to be the same for all sensors. Assume the estimation MSE requirement is $D_r = 5D_0$ and the binary transmission model is used.

For a given network setting, the optimal source coding and optimal multi-hop routing solutions are determined to maximize the network lifetime bound for estimation. To demonstrate the efficiency of the proposed algorithms, we compare it with two heuristic methods:

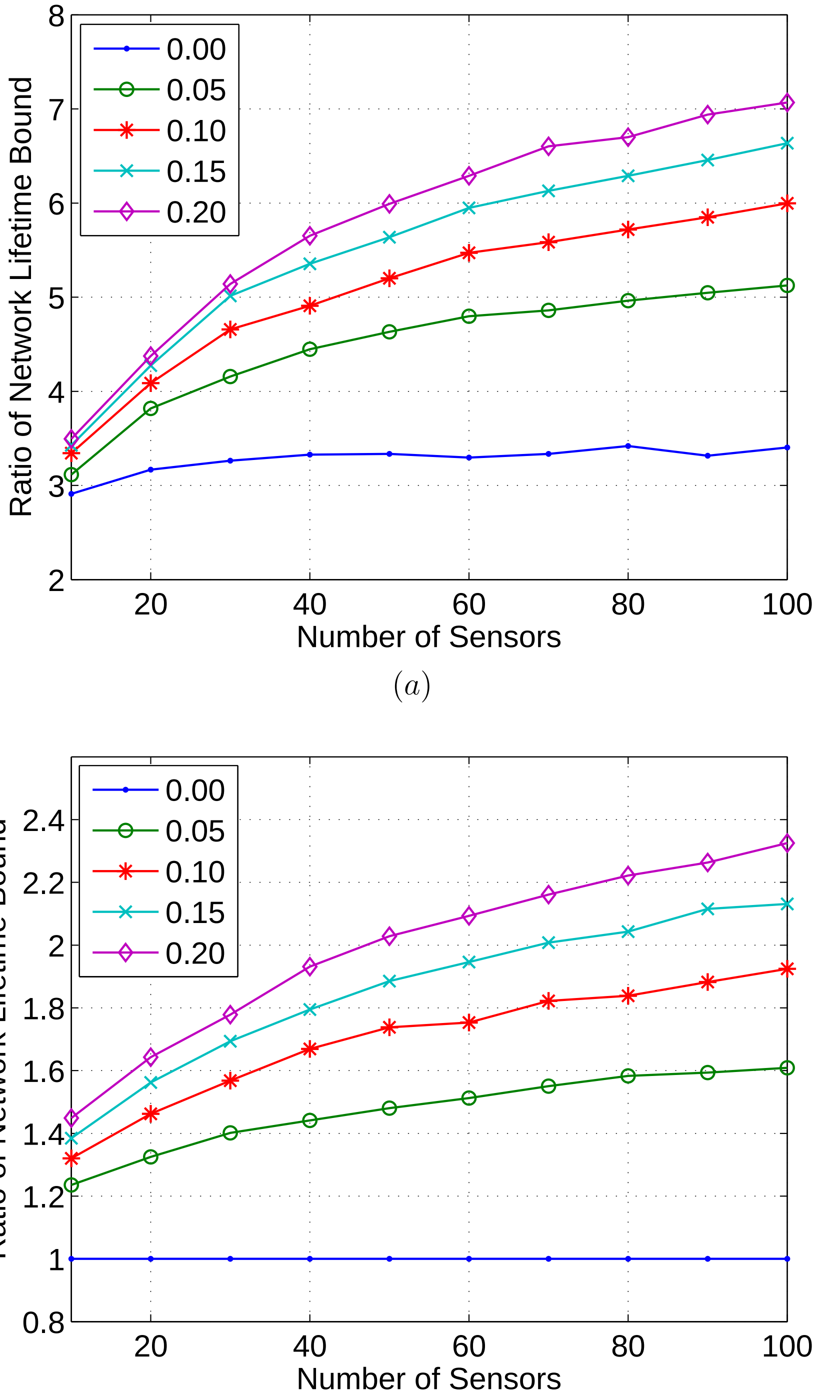
1. *Heuristic-I*: single-hop routing with uniform energy scheduling for each sensor.

2. *Heuristic-II*: single-hop routing with optimal source coding and energy scheduling.

Figure 22(a) and Figure 22(b) show the ratio of network lifetime bound achieved by the proposed algorithm to that by the *Heuristic-I* and *Heuristic-II* methods under different total number of sensors and different sensor noise variation parameters γ, respectively, where all the simulation results are obtained by repeating the experiments for 2000 times and averaging the individual results. From Figure 22(a) and Figure 22(b), we can see that the proposed algorithms improve the network lifetime bound significantly compared with both *Heuristic-I* and *Heuristic-II* methods, and the gain becomes more significant when the sensor network becomes denser or the observation noise variances become more diverse, i.e., γ becomes bigger. It is also worth noting that the similar conclusions can be drawn for different estimation MSE requirements D_r except that the actual value in Figure 22(a) will be even bigger with bigger D_r as we have shown in Figure 21(a).

It is noted that both the optimal method and the *Heuristic-II* method use optimal source coding, and the only difference is that optimal multi-hop routing is used by the optimal solution, while single-hop routing is used by the *Heuristic-II* method. From Figure 22(b), we see that the *Heuristic-II* method is also optimal when $\gamma = 0.00$, which confirms our conclusion in Proposition 5.3 that single-hop routing can maximize the network lifetime bound for homogeneous networks. From Figure 22(b), we also can see that optimal multi-hop routing improves the network lifetime bound significantly compared with single-hop routing for heterogeneous networks. Furthermore, the gain is more significant when the network is denser since there are more opportunities for multi-hop routing. Also the gain is more significant when the observation noise variances are more diverse since the optimal multi-hop routing is character-based as shown in Section 5.5.2.

To further demonstrate the character-based routing, Figure 23(a) and Figure 23(b)



(a)

(b)

Figure 22: Ratio of network lifetime bound by the proposed algorithm to that by two heuristic methods under different total number of sensors and different sensor noise variation parameters ($\gamma = 0.00, 0.05, 0.10, 0.15, 0.20$): (a) compared with *Heuristic-I*, (b) compared with *Heuristic-II*.

show two example heterogeneous sensor networks with $N = 10$ sensor nodes, where each circle denotes a sensor node. There are two numbers in the brackets around each sensor node, where the first one denotes its index and the second one denotes its observation noise variance. In these two networks, the sensor locations are the same, while the observation noise variances are different. From Figure 23(a) and Figure 23(b), we can see that the optimal routing completely changed due to the different observation noise variances, and the sensors only relay information generated at other sensors with smaller observation noise variance, such as in Figure 23(a), sensor 8 relays information from sensor 3, while in Figure 23(b), sensor 3 relays information from sensor 8 even though sensor 3 is farther away from the fusion center than sensor 8. The intuitive explanation is that sensor 8 has a very small observation noise variance, then it is desirable to gather as much data as possible from sensor 8, thus sensor 8 should transmit its data to its nearest neighbor (sensor 3) if possible to save transmission energy and improve source throughput. It is also noted that in Figure 23(b), sensor 7 relays some information from sensor 3 even though sensor 7 has smaller observation noises than sensor 3 because the relayed information is originally generated at sensor 8 other than sensor 3.

5.7 Summary

In this chapter, we consider the distributed estimation in energy-limited wireless sensor networks from the *lifetime-distortion* perspective, which is rarely addressed in the literature. From the application aspect, we are interested in the estimation task cycles the network can accomplish before the network becomes nonfunctional other than whether any individual sensor node is dead, thus we introduce a concept of function-based network lifetime. Based on this concept, it is shown that the network lifetime bound maximization for the distributed estimation can be formulated as a nonlinear programming (NLP) problem, where there are three factors needed to be

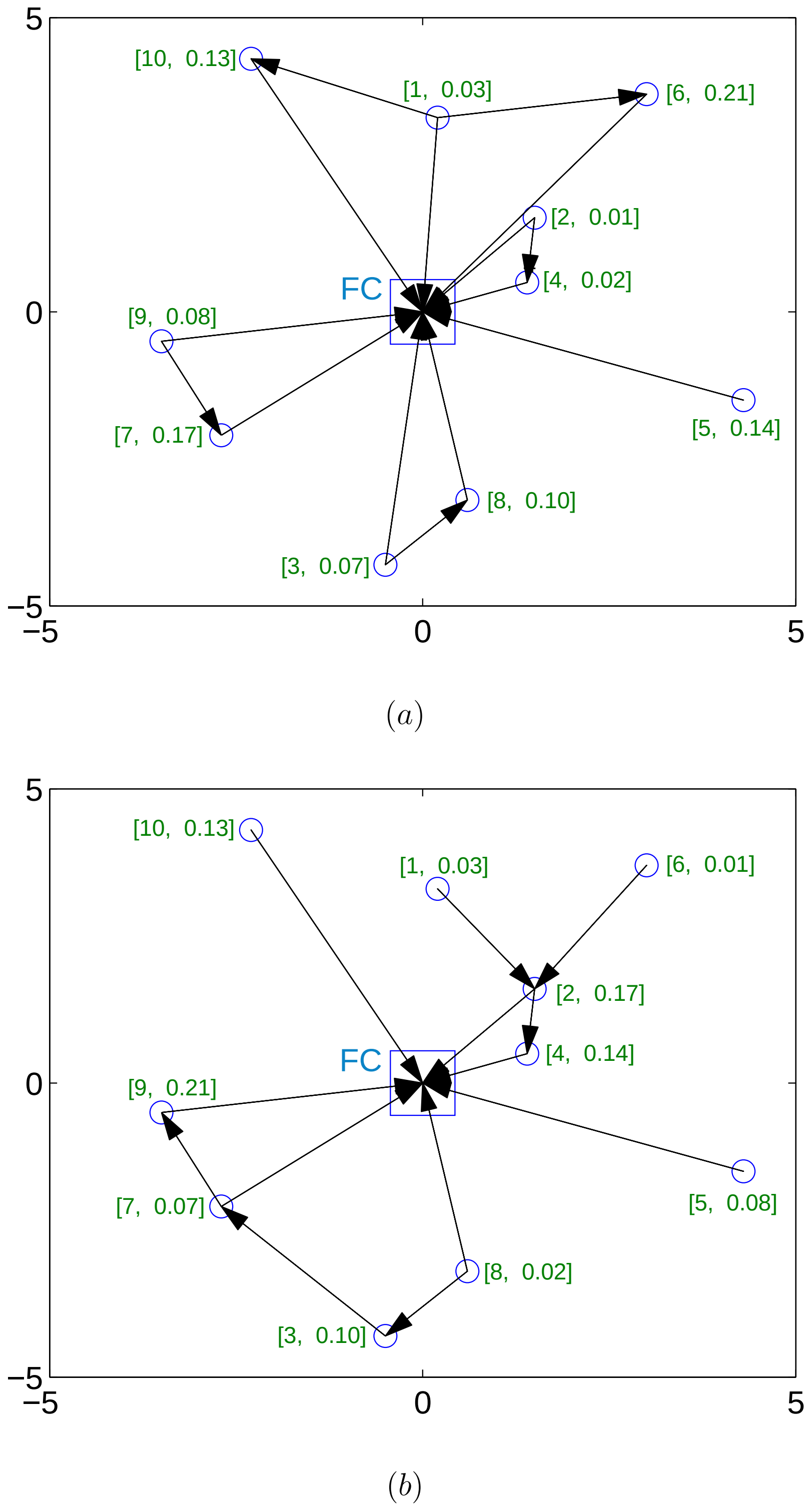


Figure 23: The optimal multi-hop routing path for two example heterogeneous sensor networks with same sensor locations but different observation noise variances.

optimized together: (*i*) source coding at each sensor, (*ii*) source throughput of each sensor, and (*iii*) multi-hop routing. We further show that the source coding can be optimized independently from the source throughput and multi-hop routing, and the optimal source coding is achieved by maximizing the equivalent unit-resource MSE function. Then based on the optimal source coding, the nonlinear programming (NLP) problem of network lifetime bound maximization can be reformulated as a linear programming (LP) problem, which can be easily solved by any LP solver.

On the other hand, the linear programming formulation for the network lifetime bound maximization problem can be understood as a *weighted data gathering* problem, where the objective is to maximize the weighted sum of the amount of data generated at all sensors. The weight of each sensor is inversely proportional to its observation noise variance, which is meaningful since the data from sensors with small noise variance is more useful. Furthermore, we find out that the optimal routing solution is *character-based routing*, where a sensor node only relays data from sensor nodes with smaller observation noise variance. Different from the traditional distance-based routing, where the routing path is selected based on the distance to the destination, *character-based routing* explicitly takes into account the heterogeneous nature of the information in wireless sensor networks.

It is worth to point out that the concepts of function-based network lifetime and character-based routing concepts proposed in this research are promising to be generalized to a wide range of applications other than the distributed estimation in wireless sensor networks.

5.8 Appendix

5.8.1 Proof of Proposition 5.1

Assume a sensor network with N sensors, each with observation noise variance σ_k^2. Assume sensor k ($k = 1, \cdots, N$) can make totally M_k measurements and quantize its measurements using probabilistic quantization scheme to $b_{k,i}$ $(i = 1, \cdots, M_k)$ bits, respectively, before it depletes. To satisfy the given estimation distortion requirement D_r, at each estimation cycle, a subset of sensors are required to observe the parameter θ and transmit their quantized measurements to the fusion center to make the final estimation.

Assume the network lifetime for this network is L. At each estimation cycle $l \in [1, L]$, denote the subset of observations each sensor k makes and sends to the fusion center is $O_{k,l}$. Then for any sensor $k \in [1, N]$, we have

$$\begin{aligned} & O_{k,i} \cap O_{k,j} = \emptyset, \quad \forall i, j \in [1, L], \text{ and } i \neq j, \\ & \bigcup_{l=1}^{L} O_{k,l} \subseteq \{1, \cdots, M_k\}, \end{aligned} \tag{132}$$

and for any estimation cycle $l \in [1, L]$, we have

$$\left(\sum_{k=1}^{N} \sum_{i \in O_{k,l}} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right)^{-1} \leq D_r. \tag{133}$$

So,

$$\sum_{l=1}^{L} \sum_{k=1}^{N} \sum_{i \in O_{k,l}} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \geq \frac{L}{D_r}, \tag{134}$$

i.e.,

$$\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \geq \frac{L}{D_r}, \tag{135}$$

therefore,

$$L \leq D_r \left(\sum_{k=1}^{N} \sum_{i=1}^{M_k} \frac{1}{\pi_k^2(\sigma_k^2, b_{k,i})} \right). \tag{136}$$

5.8.2 Proof of Theorem 5.3

The theorem is proved by contradiction. Assume a sensor i_m $(m \in [1, T])$ on the routing path of the sub flow η has a smaller observation noise variance than the source node i_0, i.e.,

$$\sigma_{i_m}^2 < \sigma_{i_0}^2, \tag{137}$$

then, remove the sub flow η and add a new sub flow ξ with the same data volume, which is generated at sensor node i_m and transmitted to the fusion center through sensor nodes $i_{m+1}, \cdots, i_T$ sequentially, i.e.,

$$\begin{aligned} S_{i_0}^{\eta} &= f_{i_0,i_1}^{\eta} = f_{i_1,i_2}^{\eta} = \cdots = f_{i_{T-1},i_T}^{\eta} = f_{i_T,N+1}^{\eta} = 0, \\ S_{i_m}^{\xi} &= f_{i_m,i_{m+1}}^{\xi} = \cdots = f_{i_{T-1},i_T}^{\xi} = f_{i_T,N+1}^{\xi} = S. \end{aligned} \tag{138}$$

First, we need to show that the new flow is feasible, that is to say, it satisfies the flow conservation and energy constraints as shown in Equation (125, 126). It is easy to show as follows:

1. By removing the sub flow η and adding the new sub flow ξ, the flow conservation at sensor node i_0 is satisfied since both the data generated at this node and the outcoming flow from this node are reduced by the same amount; the flow conservation at sensor node $i_1, \cdots, i_{m-1}$ are satisfied since both the incoming flow and outcoming flow are reduced by the same amount; the flow conservation at sensor node i_m is satisfied since the data generated at this node increases and the incoming flow decreases by the same amount and the outcoming flow is not changed; and the flow conservation at the sensor nodes $i_{m+1}, \cdots, i_K$ are also satisfied since both the incoming flow and outcoming flow are not changed. Also, the flow conservation is satisfied for all other sensor nodes since the data generated, incoming flow and outcoming flow are not changed at all.

2. By removing the sub flow η and adding the new sub flow ξ, the energy cost at the sensor nodes $i_0, \cdots, i_{m-1}$ is reduced since the data transmitted by these sensor nodes are reduced, and the energy cost at the sensor nodes $i_m, \cdots, i_K$ is not changed since the data volume transmitted by these sensor nodes are not changed even though the data content is changed. The energy cost for all other nodes is not changed since the data transmitted by these nodes are not changed at all.

Next, assume the total data volume generated at each sensor k is S_k and denote ϕ_0 and ϕ_1 as the objective function divided by D_r before or after removing the sub flow η and adding the new sub flow ξ, i.e.,

$$\begin{aligned} \phi_0 &= \sum_{k=1}^{N} \frac{S_k}{g_k^{opt}(\sigma_k^2)}, \\ \phi_1 &= \sum_{k=1,k\neq i_0,i_m}^{N} \frac{S_k}{g_k^{opt}(\sigma_k^2)} + \frac{S_{i_0} - S}{g_{i_0}^{opt}(\sigma_{i_0}^2)} + \frac{S_{i_m} + S}{g_{i_m}^{opt}(\sigma_{i_m}^2)}, \end{aligned} \tag{139}$$

then,

$$\phi_1 - \phi_0 = \frac{S}{g_{i_m}^{opt}(\sigma_{i_m}^2)} - \frac{S}{g_{i_0}^{opt}(\sigma_{i_0}^2)} > 0 \tag{140}$$

because $g_{i_m}^{opt}(\sigma_{i_m}^2) < g_{i_0}^{opt}(\sigma_{i_0}^2)$ when $\sigma_{i_m}^2 < \sigma_{i_0}^2$ as shown in Proposition 3.1. It means that the objective function in Equation (124) is increased by removing the sub flow η and adding the new sub flow ξ, which contradicts with the optimality of the original flow and routing solution. So the assumption made in Equation (137) does not hold, therefore, the Theorem is proved.

5.8.3 Proof of Proposition 5.3

This proposition is proved by contradiction. In the optimal flow and routing solution for the weighted data gathering problem in homogeneous wireless sensor networks, assume there is a multi-hop sub flow η with data volume S, generated at sensor i_0 and transmitted to the fusion center through sensors $i_1, \cdots, i_T$ sequentially, i.e.,

$$S_{i_0}^{\eta} = f_{i_0,i_1}^{\eta} = f_{i_1,i_2}^{\eta} = \cdots = f_{i_{T-1},i_T}^{\eta} = f_{i_T,N+1}^{\eta} = S \tag{141}$$

then, remove this multi-hop sub flow η and add a serial of single-hop sub flow $\xi_0, \cdots, \xi_T$ as follows:

$$\begin{aligned} S_{i_t}^{\xi_t} &= f_{i_t,N+1}^{\xi_t} = \frac{C_{i_t,i_{t+1}}}{C_{i_t,N+1}} \cdot S, \quad \forall t \in [0, T-1] \\ S_{i_T}^{\xi_T} &= f_{i_T,N+1}^{\xi_T} = S. \end{aligned} \tag{142}$$

Similar with the proof for Theorem 5.3, it is easy to show that both the flow conservation and energy constraints as shown in Equations (125, 126) are satisfied by removing the sub flow η and adding the new sub flows $\xi_0, \cdots, \xi_T$.

Next, assume the total data volume generated at each sensor k is S_k and denote ϕ_0 and ϕ_1 as the objective function divided by D_r before or after removing the sub flow η and adding the new sub flows $\xi_0, \cdots, \xi_T$, i.e.,

$$\begin{aligned}
\phi_0 &= \frac{1}{g^{opt}(\sigma^2)} \sum_{k=1}^{N} S_k, \\
\phi_1 &= \frac{1}{g^{opt}(\sigma^2)} \left(\sum_{k=1, k \neq i_0, i_T}^{N} S_k + (S_{i_0} - S) + (S_{i_T} + S) \right) + \frac{1}{g^{opt}(\sigma^2)} \sum_{t=0}^{T-1} \frac{C_{i_t, i_{t+1}}}{C_{i_t, N+1}} \cdot S,
\end{aligned} \tag{143}$$

then,

$$\phi_1 - \phi_0 = \frac{1}{g^{opt}(\sigma^2)} \sum_{t=0}^{T-1} \frac{C_{i_t, i_{t+1}}}{C_{i_t, N+1}} \cdot S \geq 0, \tag{144}$$

where, the equality holds only when the fusion center is not in the transmission range of all sensors $i_0, \cdots, i_{T-1}$, i.e., $C_{i_t, N+1} = \infty$ for all $t \in [0, T-1]$, otherwise, $\phi_1 - \phi_0 > 0$. It means that for homogeneous networks with unlimited transmission range for each sensor, single-hop routing can gather greater amount of data than multi-hop routing, while for homogeneous networks with limited transmission range for each sensor, single-hop routing can gather no less amount of data than multi-hop routing.

CHAPTER VI

CLUSTER-BASED DISTRIBUTED ESTIMATION

In wireless sensor networks, one way to reduce energy consumption is to process the data locally, such as data compression at the local sensors. Many rate-constrained distributed estimation algorithms [6, 7, 32, 48, 53, 64, 65, 69, 70, 75, 79, 80] and energy-constrained distributed estimation algorithms [5, 46, 52, 91, 92, 94] have been proposed along this line. In these schemes, denoted as parallel scheme thereafter, each sensor transmits its measurements using few bits to the fusion center directly. Another way to reduce energy consumption is to aggregate data at the intermediate sensor nodes, i.e., the intermediate sensor nodes make a local estimation by combining their own observations and the received messages from other sensor nodes and only send the local estimation to the fusion center (FC). In [39, 90], the progressive estimation schemes are proposed, where a sensor performs estimation based on its own measurement and the intermediate estimation from its immediate upstream sensors and then transmits its estimation to its immediate downstream sensor.

In this chapter, we study the distributed estimation in clustered sensor networks as shown in Figure 24, where the whole sensor field is divided into several clusters, each with a cluster head (CH). All the cluster members send their observations to the cluster head. Then, the cluster head makes a local estimation based on its own observation and the received messages from its cluster members. The cluster head then quantizes its local estimation and sends it to the fusion center. Finally, the fusion center makes the final estimation based on its received quantization messages.

In cluster-based distributed estimation, because of the intermediate data aggregation at the cluster heads, the communication between sensor nodes and the fusion

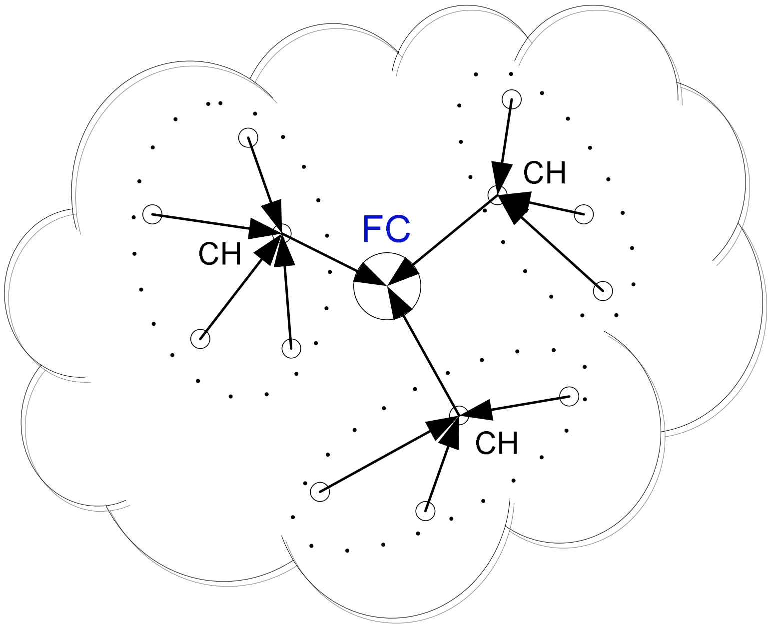


Figure 24: System diagram of cluster-based distributed estimation in wireless sensor networks, where each dashed circle denotes one cluster and each cluster has a cluster head (CH).

center is decreased; thus, the overall energy consumption is reduced. In clustered sensor networks, the energy cost includes two parts: intra-cluster energy cost, i.e., energy cost for communication between the cluster members and the cluster heads, and inter-cluster energy cost, i.e., energy cost for communication between the cluster heads and the fusion center. With a small number of large clusters, the intra-cluster energy cost is high, but the inter-cluster energy cost is low, while with a large number of small clusters, the intra-cluster energy cost is low, but the inter-cluster energy cost is high. So there exists an optimal trade-off between the cluster size and the number of clusters to minimize the overall energy consumption. Optimally clustering the sensor networks for cluster-based distributed estimation is the major objective of this research.

Furthermore, the cluster head inside each cluster can be fixed or can be rotated among all the cluster members. A fixed cluster head, if appropriately chosen, can reduce the overall energy consumption, while cluster head rotation can balance the energy consumption among all sensors. For these two cases, the optimal cluster structures could be different to minimize the overall energy consumption.

The rest of the paper is organized as follows. Section 6.1 describes a cluster-based estimation scheme, and analyzes its potential to save energy and the major challenge to maximize the energy saving. Section 6.2 discusses a special case – ring network, where the optimal clustering can be analytically obtained. Then the optimal clustering for general sensor networks are addressed in Section 6.3, and the simulation results for the proposed algorithms are presented in Section 6.4. Finally, the conclusion is given in Section 6.5.

6.1 Cluster-based Distributed Estimation

In the cluster-based distributed estimation scheme, as shown in Figure 24, the whole sensor field is divided into several clusters, all the cluster members send their original observations (with fine quantization[1]) to the cluster head, then the cluster head makes a local estimation, quantizes its local estimation and sends it to the fusion center, which makes the final estimation. Next, we first introduce a cluster-based estimation scheme and show its potential advantage compared with the parallel estimation scheme. Then, we will highlight the major challenges in designing cluster-based estimation method, i.e., optimally determining the trade-off between the cluster size and the number of clusters to minimize the overall energy cost.

6.1.1 Cluster-based Estimation Scheme

Consider a dense sensor network that includes N sensors, each sensor makes an observation on the unknown parameter θ. The observations are corrupted by additive noises and are described by

$$x_k = \theta + n_k, \quad k = 1, \cdots, N. \tag{145}$$

[1] Here, fine quantization means the quantization bit rate is high enough such that the quantization error is neglectable compared with the observation noise.

We assume that the noises n_k $(k = 1, \cdots, N)$ are zero mean, spatially uncorrelated with variance σ_k^2, but otherwise unknown. There are K clusters, each consists of $N_i (i = 1, \cdots, K)$ sensors such that $\sum_{i=1}^{K} N_i = N$, one of which is the cluster head. The cluster head can be fixed or rotated among all sensors in the same cluster. Then the cluster-based estimation method can be described as follows. First, in each cluster i, all cluster members send their original observations with fine quantization to the cluster head; then, the cluster head produces a local estimation $\bar{\theta}_i$ using the BLUE estimator as follows:

$$\bar{\theta}_i = \left(\sum_{k=1}^{N_i} \frac{1}{\sigma_{i_k}^2} \right)^{-1} \sum_{k=1}^{N_i} \frac{x_{i_k}}{\sigma_{i_k}^2}, \tag{146}$$

where, $\bar{\theta}_i$ denotes the local estimation of the i-th cluster, while x_{i_k} and $\sigma_{i_k}^2$ denote the observation and the observation noise variance of the k-th sensor in the i-th cluster, respectively. And denote the variance of the local estimation $\bar{\theta}_i$ as σ_i^2, then

$$\sigma_i^2 = E(\bar{\theta}_i - \theta)^2 = \left(\sum_{k=1}^{N_i} \frac{1}{\sigma_{i_k}^2} \right)^{-1}. \tag{147}$$

Second, the cluster head of each cluster i quantizes its local estimation $\bar{\theta}_i$ using b_i-bit probabilistic quantization in Section 3.1.1; then, the variance of the quantization message m_i is

$$E(m_i - \theta)^2 \leq \left(\sum_{k=1}^{N_i} \frac{1}{\sigma_{i_k}^2} \right)^{-1} + \frac{W^2}{(2^{b_i} - 1)^2} = \sigma_i^2 + \delta_i^2, \tag{148}$$

where σ_i^2 is the equivalent observation noise variance, while δ_i^2 is the upper bound of the quantization noise variance. Then the cluster head transmits its quantization message to the fusion center, which makes the final estimation using quasi-BLUE estimator in Section 3.1.2 based on all the received quantization messages from all cluster heads as follows:

$$\bar{\theta} = \left(\sum_{i=1}^{K} \frac{1}{\sigma_i^2 + \delta_i^2} \right)^{-1} \sum_{i=1}^{K} \frac{m_i}{\sigma_i^2 + \delta_i^2}, \tag{149}$$

and the final estimation MSE at the fusion center is

$$E(\bar{\theta} - \theta)^2 \leq \left(\sum_{i=1}^{K} \frac{1}{\sigma_i^2 + \delta_i^2} \right)^{-1}. \tag{150}$$

In short, the cluster-based estimation method combines the BLUE estimator in each cluster and the quasi-BLUE estimator at the fusion center. Based on the above procedure, the following Lemma, which is easy to prove, shows that the cluster-based scheme is promising to reduce the communication requirements and to save energy.

Lemma 6.1. *Assume there are n sensors with observation noise variance σ^2 to estimation an unknown deterministic parameter θ. If the parallel estimation scheme is used, each sensor quantizes its observation with b-bit, and the fusion center performs the quasi-BLUE estimation based on the n quantization messages; then, the final estimation variance bound is $f_p(\sigma^2, b) = (1/n)\left(\sigma^2 + W^2/(2^b - 1)^2\right)$. For the cluster-based estimation scheme, assuming that these n sensors construct a cluster, all the cluster members send their original observations to the cluster head; then, the cluster head makes a local estimation using BLUE estimator and the variance of the local estimation is σ^2/n. Then the cluster head quantizes the local estimation with k bits and sends the k-bit quantization message to the fusion center, which makes the final estimation; thus, the final estimation variance bound is $f_c(\sigma^2, k) = \sigma^2/n + W^2/(2^k - 1)^2$. Comparing the cluster-based estimation scheme with the parallel estimation scheme:*

- *If $k = nb$, then for $b \geq 1$ and $n \geq 1$,*

$$f_p(\sigma^2, b) = \frac{1}{n}\left(\sigma^2 + \frac{W^2}{(2^b - 1)^2}\right) \geq \frac{\sigma^2}{n} + \frac{W^2}{(2^{nb} - 1)^2} = f_c(\sigma^2, nb). \tag{151}$$

- *$f_c(\sigma^2, k) \leq f_p(\sigma^2, b)$, if only $k \geq \log_2 n + b$. It is obvious that $k \ll nb$ when n is large. Specially, if $b = 1$, then $k \geq \log_2 (n + 1)$ and $k \ll n$.*

In Lemma 6.1, the first conclusion implies that with the same communication rate between the fusion center and the local sensors, the cluster-based scheme will lead to better estimation MSE performance than the parallel scheme, while the second conclusion means that much less communication rate between the fusion center and the local sensors is required for the cluster-based estimation scheme than the parallel estimation scheme to gain the same estimation MSE performance.

6.1.2 Optimal Trade-off between Cluster Size and Number of Clusters

From Section 6.1.1, we see that for the same estimation MSE performance, the communication rate between the fusion center and the local sensors will be greatly reduced by cluster-based estimation scheme compared with the parallel estimation scheme. Thus, the communication energy will be greatly saved when we do not consider the energy cost for the intra-cluster communication between the cluster head and the cluster members. If energy cost for the intra-cluster communication is also taken into account, the total energy cost for the cluster-based estimation method is

$$C_{total} = C_{intra} + C_{inter}, \tag{152}$$

where C_{intra} denotes the communication energy cost within the clusters, and C_{inter} denotes the communication energy cost from the cluster heads to the fusion center. We assume that the energy consumption to transmit b-bit message is $C = b \cdot \beta d^{\gamma}$, where β is a constant factor, d is the transmission distance, and γ is the path loss exponent ($2 \leq \gamma \leq 4$). Therefore, the fundamental challenge in the cluster-based estimation method is to determine the optimal trade-off between the inter-cluster communication and the intra-cluster communication to minimize the total energy cost. In other words, we need to determine the optimal cluster sizes and the number of clusters. This trade-off will depend on the network topology, network density, the distance between local sensors and the fusion center as well as the local quantization bit rate for each sensor.

Before we address this optimal trade-off problem for the general sensor network case, in the next section, we first discuss this trade-off for a special network topology –

ring network, which not only shows the gain of the cluster-based estimation scheme in saving energy compared with the parallel estimation scheme, but also provides some insight to construct the optimal clusters for general network topology to minimize the total energy cost.

6.2 Case Study: Ring Network

Consider a special network topology, ring network, as shown in Figure 25, where all sensor nodes, say m sensors, are uniformly located on a circle whose center is the fusion center. The circle's radius is r and the angle between any two adjacent nodes is α radians, i.e, $m\alpha = 2\pi$. In this topology, the distance between each sensor node to the fusion center is the same, which will simplify the analysis and formulation of the optimal clustering.

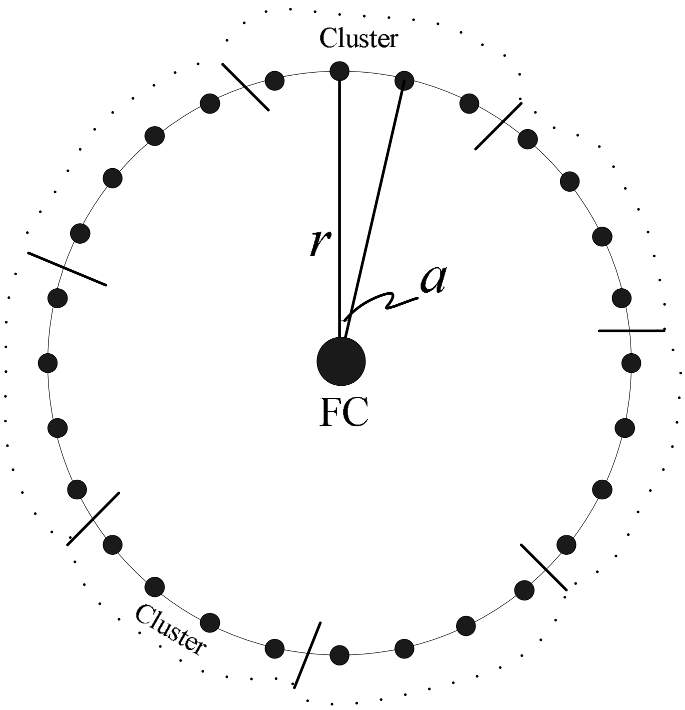


Figure 25: Illustration of a ring network where there are seven clusters with four sensors in each cluster.

It is obvious that for this special network topology each cluster should have the same number of sensors, say n sensors. Further, assume each sensor sends one bit to the fusion center in the parallel estimation scheme, then in the cluster-based scheme, each cluster head should send $\log_2(n+1)$ bits to the fusion center to achieve the same estimation performance according to Lemma 6.1. Assuming the path loss exponent parameter to be $\gamma = 2$, the total energy cost of the parallel estimation method is given as $C_p = m \cdot \beta r^2$. For the cluster-based method, the energy cost of intra-cluster communication and inter-cluster communication can be easily formulated as follows:

$$
\begin{aligned}
C_{intra} &\leq \frac{m}{12}(n^2+2)(r\alpha)^2 \cdot b_0\beta, \\
C_{inter} &= \frac{m}{n}\log_2(n+1) \cdot \beta r^2,
\end{aligned}
\tag{153}
$$

where b_0 denotes the bit rate of each cluster member sending to the cluster head. In our simulation, we assume $b_0 = 16$. Then the total energy cost of the cluster-based scheme is $C_{total} = C_{intra} + C_{inter}$. By appropriately choosing the cluster size n, we can minimize the total energy cost C_c. For a uniform ring network with totally 200 sensor nodes, Figure 26 shows the ratio of the energy cost using the cluster-based estimation scheme to that using the parallel estimation scheme under various cluster sizes. It can be observed that there exists an optimal cluster size to minimize the energy cost. Figure 27 shows the energy cost ratio of the cluster-based estimation scheme with optimal cluster size compared with the parallel estimation scheme. It is noted that with the cluster-based estimation scheme, more energy will be saved when the network gets denser and the saving energy will be up to 80% when there are 1000 sensors in the ring network. Figure 28 shows the optimal cluster radius, which is defined as the largest distance between the cluster members and the cluster head, verse the radius of the ring network. It implies that the optimal radius of the cluster is proportional to the distance between the sensors and the fusion center, that is, the optimal cluster radius is not equal and it is bigger in the region far away from the fusion center.

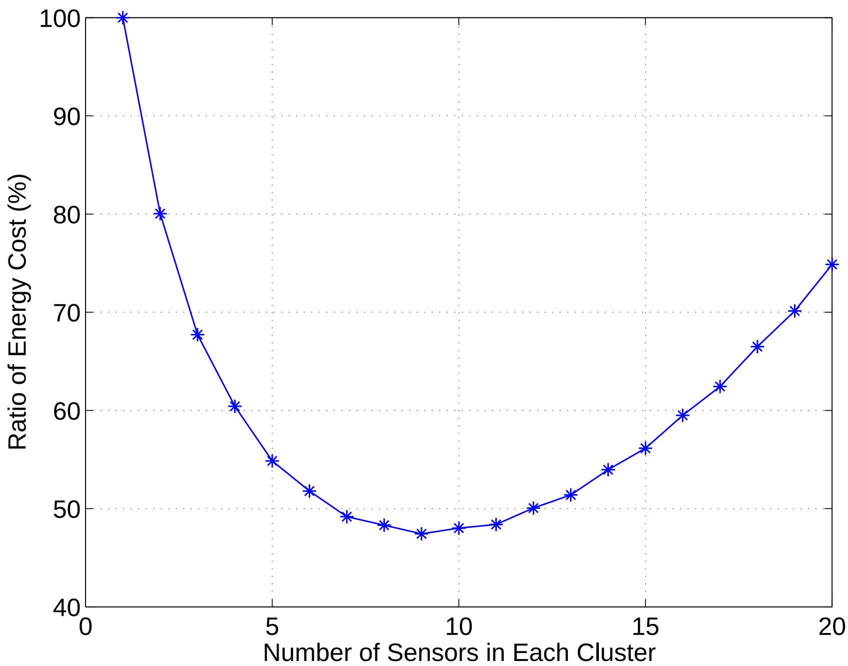


Figure 26: Simulation results of cluster-based estimation scheme for a ring network: Energy cost ratio vs. the number of sensors in each cluster for a ring network with 200 sensor nodes.

6.3 Clustering for General Networks

From the special case study in Section 6.2, we show that the cluster-based estimation scheme will lead to significant energy saving. In this section, we consider how to optimally cluster a sensor network with arbitrary sensor distribution, i.e, all the sensors are arbitrarily distributed in the observation field, to minimize the total energy cost of the cluster-based estimation scheme.

From the discussion before, we can get the following hints to design the optimal clustering scheme:

- From Lemma 6.1, we can see that the inter-communication cost will reduce when the cluster becomes bigger. On the contrary, the intra-communication will increase. So there exists a trade-off between the cluster size and the number of clusters to minimize the total cost.

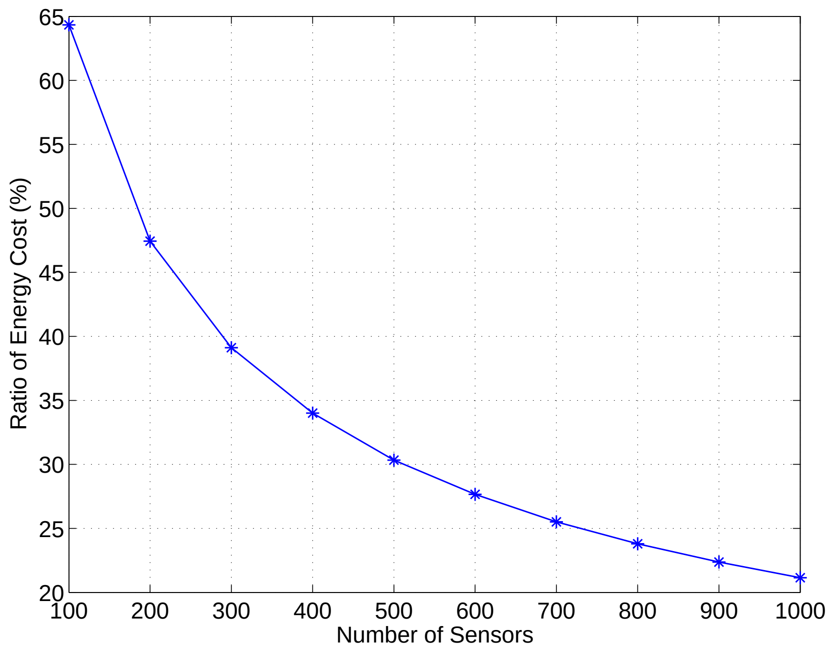


Figure 27: Simulation results of cluster-based estimation scheme for a ring network: Energy cost ratio vs. the total number of sensors.

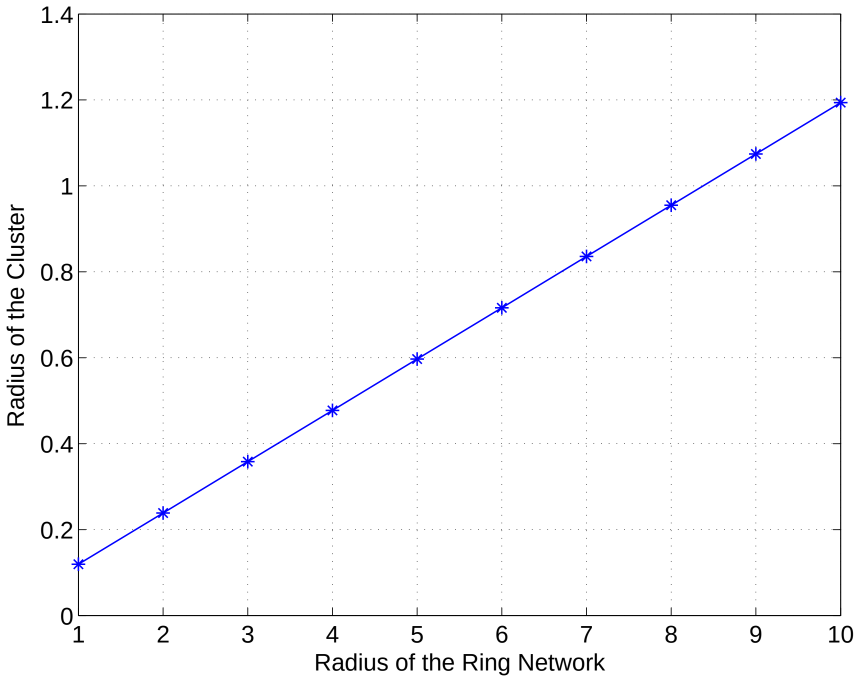


Figure 28: Simulation results of cluster-based estimation scheme for a ring network: Optimal cluster radius vs. the radius of the ring network.

- The optimal cluster size depends on the the distances between the sensors and the fusion center, the network density, and the network topology. From the case study of the ring network, we conclude that the optimal clustering is non-uniform, specifically, the optimal cluster radius is bigger in the region farther away from the fusion center.

6.3.1 Clustering with Fixed Cluster Head

In this section, we consider partitioning a wireless sensor network into clusters, each one with a fixed cluster head (CH) and some ordinary sensors as its members. We model the cluster-based sensor networks with fixed cluster head for each cluster as a *directed graph* G. The corresponding directed graph G (called cluster graph) as shown in Figure 24 is constructed such that each sensor node is a vertex in the graph and an directed edge $u \to v$ is drawn if vertex u is a member of the cluster whose cluster head is vertex v. Furthermore, we model the intra-cluster communication energy cost from the cluster member u to the cluster head v as the *edge weight* of the edge $u \to v$, and the inter-cluster communication energy cost from the cluster head v to the fusion center as the *point weight* of the cluster head v. Thus the total communication cost of each cluster consists of all the edge weights in the cluster and the point weight of the cluster head. Then clustering the sensor network to minimize the total energy cost can be modelled as constructing the corresponding directed graph G with smallest total weight which consists of the edge weights of all edges between the cluster members and its cluster head and the point weights of all the cluster heads.

Intuitively, sensor u will not join a cluster whose cluster head is sensor v such that the intra-cluster communication energy cost from sensor u to sensor v is even larger than the energy cost that sensor u directly communicates with the fusion center. That is, sensor node u can be connected to sensor node v only if

$$\frac{C_{u\to v}}{C_{u\to FC}} = \frac{b_0 \cdot \beta d_{u\to v}^{\gamma}}{b_q \cdot \beta d_{u\to FC}^{\gamma}} < r_{th}, \tag{154}$$

where β and γ are defined as before, r_{th} is a threshold parameter, $C_{u\to v}$ denotes the energy cost of sensor u sends its original observation (with b_0-bit fine quantization) to sensor v and $d_{u\to v}$ denotes the corresponding transmission distance, while $C_{u\to FC}$ denotes the energy cost of sensor u sends its quantized observation (with b_q-bit) to the fusion center directly and $d_{u\to FC}$ denotes the corresponding transmission distance. Then we construct a directed graph G' that includes all the possible edges which satisfy Equation (154). It is noted that an edge $v \to u$ may not exist even though an edge $u \to v$ exists in the graph G', thus the graph G' is a directed graph. It is also noted that in the graph G', each sensor u may be connected to many different sensors, and many different sensors may be connected to sensor u at the same time, that means any sensor u can be as a cluster head, and all the sensors connected to sensor u can be as its cluster members.

Next, based on the graph G', we choose the appropriate sensor nodes as the cluster head and choose their corresponding cluster members with Algorithm 1, which is a greedy algorithm. The basic idea is to repeat identifying the cluster (cluster head and its corresponding cluster members) which leads to the most average energy saving compared with the parallel scheme from the uncovered subgraph of G' until all the sensors are covered.

6.3.2 Clustering with Cluster Head Rotation

In Section 6.3.1, we considered clustering the sensor networks with fixed cluster head for each cluster, as we will show in the Section 6.4, the total energy cost is greatly reduced compared with the parallel scheme. But the cluster head will consume its energy much quickly since the cluster head is fixed and will communicate with the fusion center directly. Therefore, cluster head rotation among all the members in the

Algorithm 1 Clustering with Fixed Cluster Head

1: **Input**: directed graph G' generated from the given network.

2: **Output**: clustering result for the given network.

3: Initialize all vertices in the graph G' as uncovered.

4: **while** (there are vertices uncovered in the graph G') **do**

5: **for all** (uncovered vertices v in the graph G') **do**

6: Assume v as the cluster head, then the set of its cluster members $S(v)$, the number of sensors in the cluster $N(v)$, the communication rate from the cluster head v to the fusion center $b_c(v)$, the intra-cluster communication energy cost between the cluster members and the cluster head $C_{intra}(v)$, the inter-cluster communication energy cost from the cluster head v to the fusion center $C_{inter}(v)$, the total energy cost of this cluster $C_{cluster}(v)$, the total energy cost of all the sensor nodes in this cluster using the parallel estimation scheme $C_{parallel}(v)$, the total and average energy saving of the cluster-based scheme compared with the parallel scheme $C_{save}(v)$ and C_{s_ave} are computed as follows:

$$\begin{aligned}
&S(v) : \text{includes all the uncovered vertices connected to } v \\
&N(v) = \sum_{u \in S(v)} 1 \\
&b_c(v) = \log_2 N(v) + b_q \\
&C_{intra}(v) = \sum_{\substack{u \in S(v) \\ u \neq v}} b_0 \cdot \beta d^{\gamma}_{u \to v} \\
&C_{inter}(v) = b_c(v) \cdot \beta d^{\gamma}_{v \to FC} \\
&C_{cluster}(v) = C_{intra}(v) + C_{inter}(v) \\
&C_{parallel}(v) = \sum_{u \in S(v)} b_q \cdot \beta d^{\gamma}_{u \to FC} \\
&C_{save}(v) = C_{parallel}(v) - C_{cluster}(v) \\
&C_{s_ave}(v) = \frac{C_{save}(v)}{N(v)}
\end{aligned} \tag{155}$$

7: **end for**

8: Choose the sensor v' such that the average energy saving is maximized, i.e.,

$$v' = \max_{v} C_{s\_ave}(v). \tag{156}$$

9: Let v' be the cluster head, and all the uncovered sensors u connected to it as its cluster members. Label all the sensor nodes in this cluster as covered.

10: **end while**

cluster is essential to even the energy consumption within the cluster.

Since Algorithm 1 does not take the cluster head rotation into account, it is not necessarily optimal when the cluster head rotation is actually adopted. In this section, we design another clustering algorithm which explicitly take the cluster head rotation into account and minimize the total energy cost. Similarly, we still model the cluster-based sensor networks with cluster head rotation as a graph, but it is a *undirected graph* here instead of a *directed graph.* The corresponding undirected graph G (called cluster graph) is constructed such that each sensor node is a vertex in the graph and an undirected edge (u, v) is drawn if vertex u and vertex v belong to the same cluster. So all the sensors in the same cluster are connected with each other, that is, a cluster is a fully connected subgraph of G.

To obtain the cluster graph G, we first construct a undirected graph G' that includes all the possible edges. An edge (u, v) to connect sensor node u and sensor node v exists, i.e., sensor u and v can be in the same cluster only if they are close enough such that

$$\frac{C_{u\to v} + C_{v\to u}}{C_{u\to FC} + C_{v\to FC}} = \frac{b_0 \cdot \beta d^{\gamma}_{u\to v} + b_0 \cdot \beta d^{\gamma}_{v\to u}}{b_q \cdot \beta d^{\gamma}_{u\to FC} + b_q \cdot \beta d^{\gamma}_{v\to FC}} < r_{th} \tag{157}$$

where β and γ are defined as before, r_{th} is a threshold parameter, $C_{u\to v}$ denotes the energy cost of sensor u sends its original observation (with b_0-bit fine quantization) to sensor node v and $d_{u\to v}$ denotes the corresponding transmission distance, while $C_{u\to FC}$ denotes the energy cost of sensor u sends its quantized observation (with b_q-bit) to the fusion center directly and $d_{u\to FC}$ denotes the corresponding transmission distance.

Next, we partition the network into clusters based on the graph G', i.e, partition the graph G' into fully connected subgraphs, each subgraph is a cluster, to minimize the total energy cost. Algorithm 2 is developed for this task. This algorithm consists of two major steps, that is, repeat generating the fully connected subgraphs stemming from each uncovered vertex v with a greedy criterion, and identifying the cluster which

Algorithm 2 Clustering with Cluster Head Rotation

Input: undirected graph G' generated from the given network.
Output: clustering result for the given network.
Initialize all vertices in the graph G' as uncovered and denote the set of all the uncovered vertices as S_{uc};
while ($S_{uc} \neq \varnothing$) **do**
  for all (vertices $v \in S_{uc}$) **do**
    Initialize the cluster stemming from v is $S(v) = \{v\}$;
    $\bar{S}(v) = S_{uc} \backslash S(v)$;
    while (there are vertices in $\bar{S}(v)$ connected to all the vertices in $S(v)$) **do**
      for all (vertices u in $\bar{S}(v)$ connected to all the vertices in $S(v)$) **do**
        Compute the intra-communication energy cost between sensor u and all the sensors in $S(v)$ as follows:

$$C_{u \to S(v)} = \sum_{x \in S(v)} b_0 \cdot \beta d_{u \to x}^{\gamma} \tag{158}$$

      end for
      Find the vertex u' which minimizes $C_{u \to S(v)}$, i.e.,

$$u' = \min_{u} C_{u \to S(v)}. \tag{159}$$

      Update the cluster $S(v)$ as follows:

$$\begin{aligned} S(v) &= S(v) \cup u' \\ \bar{S}(v) &= S_{uc} \backslash S(v) \end{aligned} \tag{160}$$

    end while
    Then for the possible cluster $S(v)$ stemming from vertex v, the number of member sensors $N(v)$, the communication rate from the cluster head v to the fusion center $b_c(v)$, the intra-cluster communication energy cost and the communication energy cost from the cluster head to the fusion center $C_{intra}(v)$ and $C_{inter}(v)$ assuming each sensor in the same cluster as the cluster head equally, the total energy cost of this cluster $C_{cluster}(v)$, the total energy cost of all the sensor nodes of this cluster using the parallel estimation scheme $C_{parallel}(v)$, the total and average energy saving of the cluster-based scheme compared with the parallel scheme $C_{save}(v)$ and C_{s_ave} are computed as follows:

$$\begin{aligned} N(v) &= \sum_{u \in S(v)} 1 \\ b_c(v) &= \log_2 N(v) + b_q \\ C_{intra}(v) &= \frac{1}{N(v)} \sum_{u \in S(v)} \sum_{\substack{x \in S(v) \\ x \neq u}} b_0 \cdot d_{x \to u}^{\gamma} \\ C_{inter}(v) &= \frac{1}{N(v)} \sum_{u \in S(v)} b_c(v) \cdot d_{u \to FC}^{\gamma} \\ C_{cluster}(v) &= C_{intra}(v) + C_{inter}(v) \\ C_{parallel}(v) &= \sum_{u \in S(v)} b_q \cdot d_{u \to FC}^{\gamma} \\ C_{save}(v) &= C_{parallel}(v) - C_{cluster}(v) \\ C_{s\_ave}(v) &= \frac{C_{save}(v)}{N(v)} \end{aligned} \tag{161}$$

  end for
  Choose the cluster stemming from the sensor v' to maximize the average energy saving, i.e., $v' = \max_{v} C_{s\_ave}(v)$.
  Label all the sensor nodes in this cluster as covered and let $S_{uc} = S_{uc} \backslash S(v')$.
end while

leads to the most average energy saving compared with the parallel scheme until all the sensors are covered.

6.4 Simulation Results

In this section, we present some simulation results for the proposed algorithms in Section 6.3. We simulate a sensor network, where N sensors are uniformly distributed in a rectangular region with dimension 100×100 and the fusion center is located at the central point, i.e., $(50, 50)$.

In order to demonstrate the efficiency of the proposed clustering algorithms, we compare them with k-mean clustering method. With k-mean clustering method, we cluster the given sensor field into different numbers of clusters to determine the optimal cluster number and the corresponding cluster structure which minimizes the total energy cost of the cluster-based estimation scheme. Then if the cluster head is fixed, we appropriately assign the cluster head for each cluster to minimize the total energy cost, otherwise the cluster head is rotated. Here, we first show the performance of k-mean clustering scheme for the cluster-based estimation problem. Figure 29(a) shows the ratio of energy cost of the cluster-based estimation scheme using k-mean clustering method with different number of clusters to the energy cost of the parallel estimation scheme, where the simulated network consists of a total of 200 sensors. It can be seen that there is an optimal number of clusters to minimize the energy cost. Figure 29(b) shows the energy cost ratio of the cluster-based estimation scheme using optimal number of clusters versus the total number of sensors in the network. In both Figures 29(a) and 29(b), two cases are shown: (*i*) the cluster head (CH) is fixed; and (*ii*) the cluster head is rotated among all the sensors in the cluster.

Next, we compare the proposed algorithms with k-mean clustering method. For the fixed cluster head case, Figure 30 shows the energy cost ratio of the cluster-based

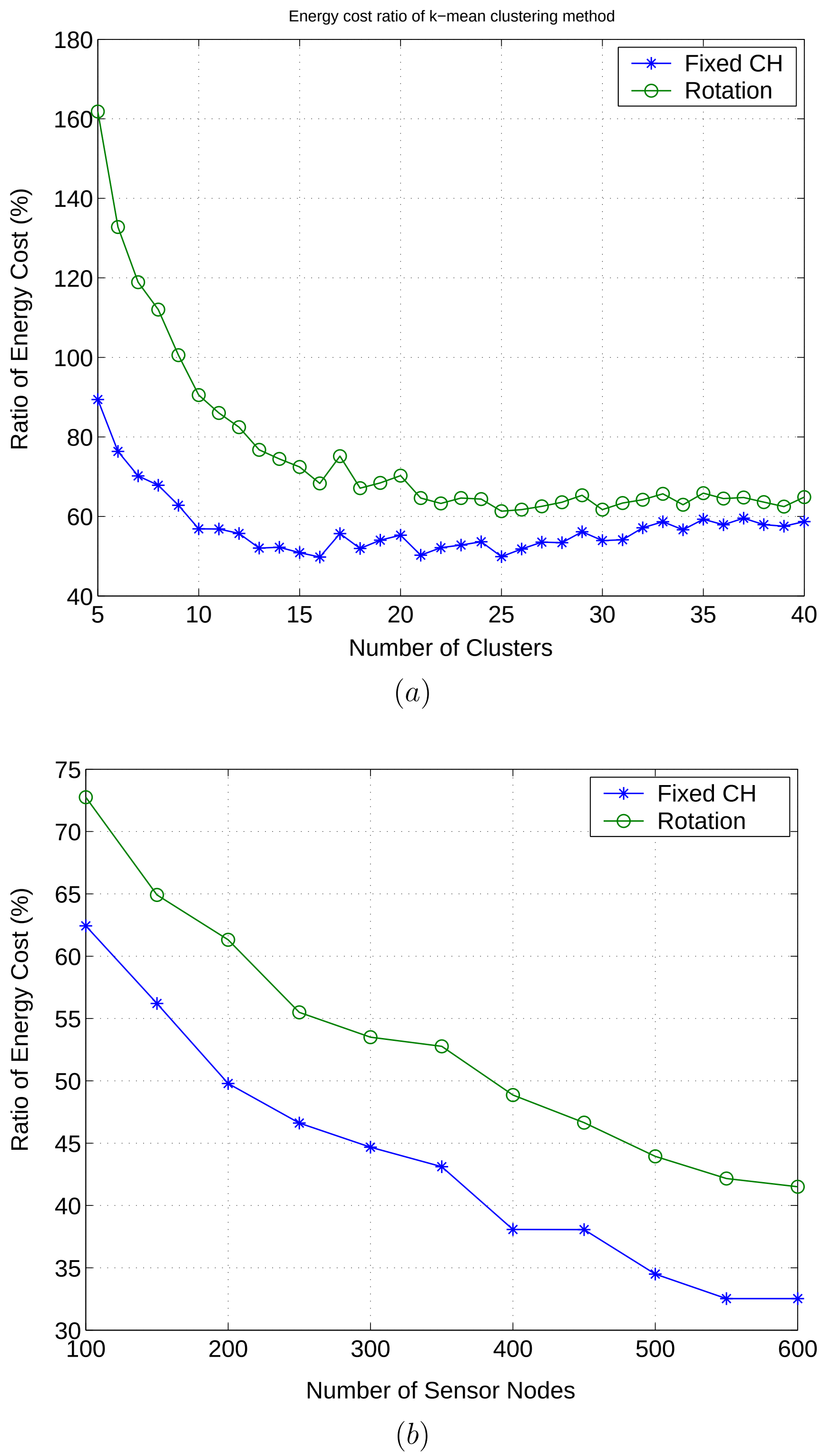


Figure 29: Performance of k-mean clustering method: (a) Energy cost vs. the number of clusters for a sensor field with 200 sensor nodes; (b) Energy cost vs. the total number of sensors.

estimation scheme using k-mean clustering method and the proposed clustering Algorithm 1 to the parallel estimation scheme. It is shown that the energy cost is reduced by cluster-based estimation scheme when compared with the parallel estimation scheme, furthermore more energy is saved with the proposed Algorithm 1 when compared with the k-mean clustering method. It is also shown that more energy is saved when the network gets denser.

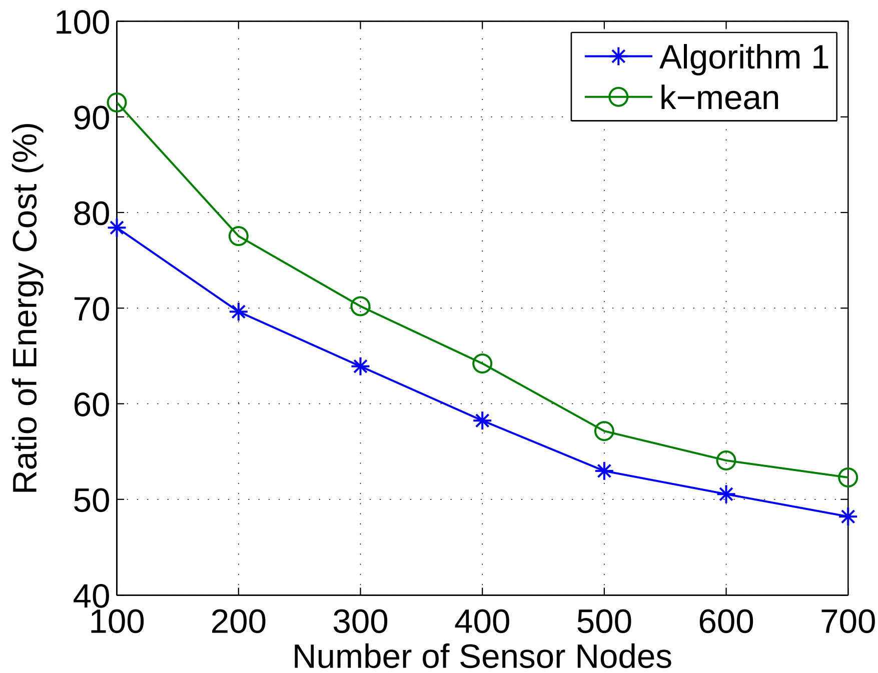


Figure 30: Performances of the proposed Algorithm 1 and k-mean clustering method for the fixed cluster head case.

If the cluster head is rotated among all sensors in the same cluster, we compare the proposed Algorithm 2 with the proposed Algorithm 1 and k-mean clustering method. Figure 31 shows the energy cost ratio of the cluster-based estimation scheme using three different clustering methods to the parallel estimation scheme. It is shown that energy cost is reduced by cluster-based estimation scheme when compared with the parallel estimation scheme. And compared with the k-mean clustering method, the proposed Algorithm 1 and Algorithm 2 will save more energy. Furthermore, Algorithm 2 saves a little more energy than Algorithm 1 as we expected since it takes the cluster head rotation into account explicitly while Algorithm 1 does not.

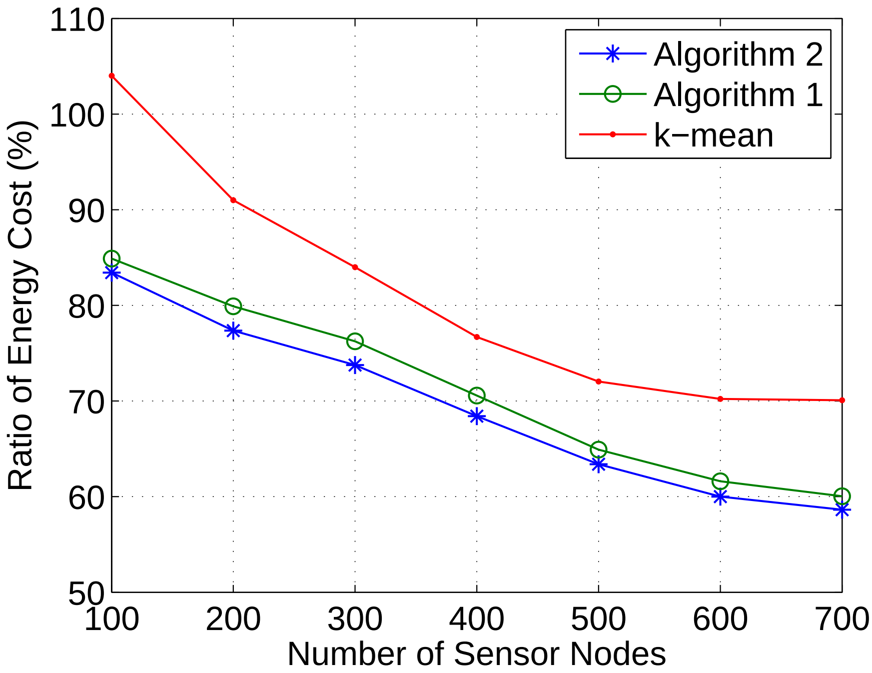


Figure 31: Performances of the proposed Algorithm 2, Algorithm 1 and k-mean clustering method for cluster head rotation case.

In addition, Figure 32 gives an example output of the proposed Algorithm 1 where the total number of sensors is 300 and the fusion center is located at the origin point. From Figure 32, we can see that the radius of the clusters are nonequal, specifically, the clusters far away from the fusion center usually consist of many sensor nodes, while the clusters very close to the fusion center may include only one sensor node, i.e., the sensor node will communicate with the fusion center directly.

6.5 Summary

In this chapter, we consider the cluster-based distributed estimation in wireless sensor networks to save energy. First, a hybrid cluster-based estimator is introduced and its potential to save energy is shown. To maximize the energy saving, optimally clustering the sensor networks is essential. We first study a simple network – ring network, and then discuss how to optimally cluster the general sensor networks to minimize the total

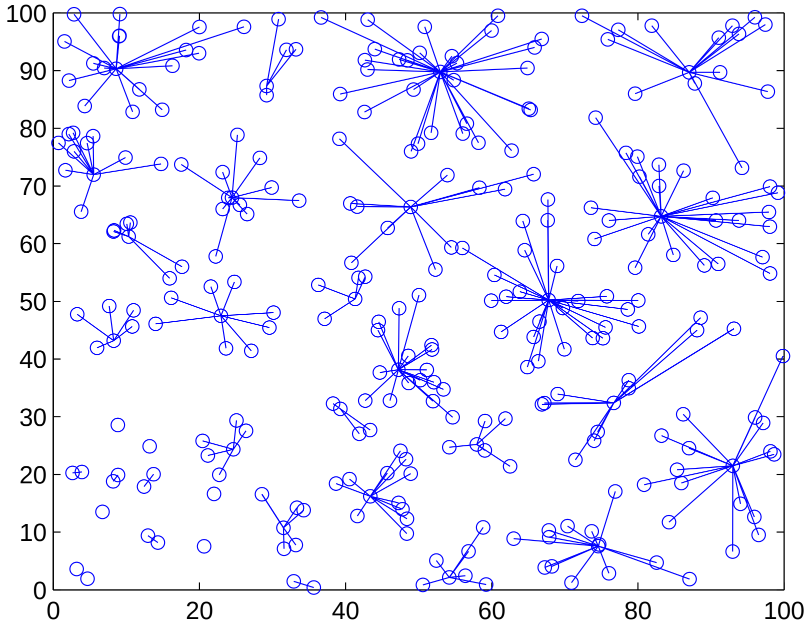


Figure 32: An example clustering result by Algorithm 1.

energy cost. It is shown that the clustering depends on the the network topology, network density and the sensor node distribution and that the optimal cluster size is nonequal, specifically, the clusters farther away from the fusion center have a larger size. Simulation results show that significant energy saving is achieved by the cluster-based estimation scheme and the proposed clustering algorithms compared with the parallel estimation scheme and the k-mean clustering methods.

CHAPTER VII

CONCLUSIONS

In this chapter, the dissertation is concluded with a summary of contributions and future research directions.

7.1 Summary of Contributions

In this dissertation, we have investigated the distributed estimation in resource-constrained wireless sensor networks, where the major challenge is the integrated design of local signal processing operations and strategies for inter-sensor communication and networking so as to achieve a desirable trade-off among resource efficiency (bandwidth and energy), system performance (estimation distortion and network lifetime), and implementation simplicity. More specifically, the efficient distributed estimation has been addressed in this dissertation from the following perspectives: (*i*) *resource-distortion* perspective, where the objective is to minimize the estimation distortion under resource constraints, and (*ii*) *lifetime-distortion* perspectives, where the objective is to maximize the network lifetime while meeting estimation distortion requirements, and (*iii*) *cluster-based distributed estimation*, where the objective is to minimize the overall energy cost by clustering the sensor field appropriately with data aggregation at cluster heads.

7.1.1 Resource-Constrained Distributed Estimation

Subject to severe resource (bandwidth and energy) constraints in wireless sensor networks, we addressed the resource-constrained distributed estimation, which includes rate-constrained distributed estimation and energy-constrained distributed estimation. The available resource should be allocated among all sensors jointly and optimally to optimize the estimation performance. To address the optimal trade-off between the resource efficiency and the estimation distortion, we first introduced a concept of *equivalent unit-resource MSE function*, where the resource could be bandwidth or energy. Then, based on minimizing the equivalent unit-resource MSE function at local sensors, quasi-optimal distributed estimation algorithms were developed, which suggested an optimal trade-off between the number of active sensors and the resource allocation for each active sensor. The proposed algorithms also suggested that each active sensor compresses its observation into a small number of bits determined only by its local signal-to-noise ratio (SNR), so it is easy to implement in a distributed manner. Furthermore, a theoretical lower bound on the estimation distortion under resource constraints was developed and it is shown that the estimation MSE of the proposed algorithms is within a small factor of the theoretical lower bound. It is worthy noting that the developed algorithm in this dissertation provides a generic framework to address different types of resource constraints (bandwidth and energy) under different estimation system models and transmission models.

7.1.2 Network Lifetime Optimization

Network lifetime other than distortion is another major system performance measurement and a critical design concern in wireless sensor networks. In this dissertation, we addressed the network lifetime optimized distributed estimation. First, a new notion of *function-based network lifetime* is introduced, which focuses on whether the network as a whole can perform a given task rather than whether any individual sensor in the network is dead. Then, we studied to optimize the function-based network lifetime while meeting estimation distortion requirements. The problems involves not only the local information processing but also inter-sensor communication and networking and it was formulated as a nonlinear programming (NLP) problem of joint optimization of source coding, source throughput, and multi-hop routing. We further showed that the source coding can be optimized independently from the source throughput and multi-hop routing, and the optimal source coding is achieved by maximizing the equivalent unit-resource MSE function. Then based on the optimal source coding, the nonlinear programming (NLP) problem of network lifetime bound maximization was reformulated as a linear programming (LP) problem. Furthermore, the optimal routing for the formulated linear programming problem was shown to be *character-based routing*, where a sensor node only relays data from sensor nodes with smaller observation noise variance. Different from the traditional distance-based routing, where the routing path is selected based on the distance to the destination, *character-based routing* explicitly takes into account the heterogeneous nature of the information in wireless sensor networks.

7.1.3 Cluster-Based Distributed Estimation

In wireless sensor networks, one way to reduce energy consumption is to process the data locally, such as data compression at the local sensors. Another way to reduce energy consumption is to aggregate data at the intermediate sensor nodes, i.e., the intermediate sensor nodes make a local estimation by combining their own observations and the received messages from other sensor nodes and only send the local estimation to the fusion center (FC). Following this line, in this dissertation, cluster-based distributed estimation was also studied, where the sensor network is divided into several clusters, each cluster with a cluster head, and data aggregation is introduced at the cluster head. In this context, two clustering algorithms for fixed or rotated cluster head cases were developed to divide the general sensor networks optimally such that the total energy cost is minimized. It was shown that the clustering depends on the

the network topology, network density and the sensor node distribution and that the optimal cluster size is nonequal, specifically, the clusters farther away from the fusion center have a larger size.

7.2 Future Research Directions

The work presented in this dissertation can be extended in the following directions:

- The resource-constrained distributed estimation discussed in Chapter 3 and Chapter 4 are based on the assumption that information is from a single source and sensor observations are conditionally independent. In the case of one-dimensional or two-dimensional field source estimation, extra work needs to be done to extend the current system frameworks and algorithms to the higher-dimensional case. More specifically, node cooperation is desirable to explore the source structures such as sensor data correlation. The wireless channel fading and interference also need to be taken into account in the future system design.

- In Chapter 5, we introduced the new notions of function-based network lifetime and character-based routing. Designing a completely distributed implementation for the character-based multi-hop routing to achieve the maximum network lifetime bound and some distributed heuristic algorithms to achieve the close-to-optimal performance are interesting directions for the future work. Also, further generalization of the new concepts of function-based network lifetime and character-based routing to a broader range of wireless sensor network applications is another interesting future direction. In the study on cluster-based distributed estimation in Chapter 6, a fixed and simple data aggregation model is assumed. It is desirable to introduce more sophisticated data aggregation algorithms into the overall optimization loop.

- In this dissertation, we addressed the efficient distributed estimation in wireless sensor networks with a set of distributed sensor nodes and a fusion center by studying the correlation among resource, distortion, and lifetime, which are three major concerns for wireless sensor networks. Extending this study to ad hoc wireless sensor networks without a fusion center is an interesting future direction. Furthermore, a similar research perspective and methodology could apply to other distributed signal processing applications in both central and ad hoc wireless sensor networks, such as distributed detection, localization and tracking.

- Throughout the research conducted in this dissertation, we have fixed the sensing model for all sensors. Another research direction would be to examine adaptive sensing where the sensing components on sensors can be adjusted. Moreover, the possibility of selective and intelligent sensing can lead to significant energy savings by an integrated design of the sensing, sampling, and compression stages, where compressive sampling plays an important role and would significantly reduce the dimensionality of the data to be sensed.

REFERENCES

[1] lp_solve. [Online]. Available: http://lpsolve.sourceforge.net.

[2] Agre, J. and Clare, L., "An integrated architecture for coopertive sensing networks," *Computer*, vol. 33, pp. 106–108, May 2000.

[3] Akyildiz, I., Su, W., Sankarsubramaniam, Y., and Cayirci, E., "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, pp. 393–422, Mar. 2002.

[4] Appadwedula, S., Veeravalli, V., and Jones, D., "Energy-efficient detection in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, 2005.

[5] Aravinthan, V., Jayaweera, S., and Tarazi, K., "Distributed estimation in a power constrained sensor network," in *Proc. of IEEE 63rd Vehicular Technology Conference*, vol. 3, pp. 1048–1052, May 7-10 2006.

[6] Aysal, R. C. and Barner, K. E., "Blind decentralized estimation for bandwidth constrained wireless sensor networks," *IEEE Trans. Wireless Communications*, vol. 7, pp. 1466–1471, May 2008.

[7] Aysal, R. C. and Barner, K. E., "Constrained decentralized estimation over noisy channels for sensor networks," *IEEE Trans. Signal Processing*, vol. 56, pp. 1398–1410, Apr. 2008.

[8] Aysal, T. C., Coates, M., and Rabbat, M., "Distributed average consensus using probabilistic quantization," in *Proc. of IEEE 14th Workshop on Statistical Signal Processing*, pp. 640–644, Aug. 26-29 2007.

[9] Aysal, T. C., Coates, M. J., and Rabbat, M. G., "Rates of convergence for distributed average consensus with probabilistic quantization," in *Proc. of Allerton 2007*, 2007.

[10] Bahceci, I., AlRegib, G., and Altunbasak, Y., "Parallel distributed detection for wireless sensor networks: Performance analysis and design," in *IEEE Global Communication Conference (GLOBECOM)*, vol. 4, (St. Lousi, MO), pp. 2420–2424, Nov. 2005.

[11] Bahceci, I., AlRegib, G., and Altunbasak, Y., "Serial distributed detection for wireless sensor networks," in *IEEE International Symposium on Information Theory (ISIT)*, (Adelaide, Australia), pp. 830–834, Sep. 2005.

[12] Bhardwaj, M. and Chandrakasan, A., "Bounding the lifetime of sensor networks via optimal role assignments," in *Proc. IEEE INFOCOM*, pp. 1587–1596, 2002.

[13] Bhardwaj, M., Garnett, T., and Chandrakasan, A., "Upper bounds on the lifetime of sensor networks," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, pp. 785–790, Jun. 2001.

[14] Blum, R., Kassam, S., and Poor, H., "Distributed detection with multiple sensors: Part ii - advanced topics," *Proceedings of the IEEE*, vol. 85, pp. 64–79, Jan. 1997.

[15] Chamberland, J. and Veeravalli, V., "Decentralized detection in sensor networks," *IEEE Trans. on Signal Processing*, vol. 51, no. 2, 2003.

[16] Chamberland, J. and Veeravalli, V., "Wireless sensors in distributed detection applications," *IEEE Signal Processing Magazine*, vol. 24, no. 3, 2007.

[17] Chang, J. and Tassiulas, L., "Energy conserving routing in wireless ad-hoc networks," in *Proc. IEEE INFOCOM*, (Tel Aviv, Israel), Mar. 2000.

[18] Chang, J. and Tassiulas, L., "Maximum lifetime routing in wireless sensor networks," vol. 12, pp. 609–619, Aug. 2004.

[19] Chen, B., Tong, L., and Varshney, P., "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, 2006.

[20] Chen, B. and Varshney, P., "A bayesian sampling approach for decision fusion using hierarchical syatems," *IEEE Trans. Signal Processing*, vol. 50, pp. 1809–1818, Aug. 2002.

[21] Chen, J., K.Yao, and Hudson, R., "Source localization and beamforming," *IEEE Signal Processing Magazine*, vol. 19, pp. 30–39, Mar. 2002.

[22] Chong, C.-Y. and Kumar, S., "Sensor networks: Evolution, opportunities and challenges," *Proceedings of the IEEE*, vol. 91, pp. 1247–1256, Aug. 2003.

[23] Cui, S., Goldsmith, A. J., and Bahai, A., "Joint modulation and multiple access optimization under energy constraints," in *Proc. IEEE Global Telecomm. Conf.*, (Dallas, TX), pp. 151–155, Dec. 2004.

[24] Cui, S., Goldsmith, A. J., and Bahai, A., "Energy-constrained modulation optimizatoin," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2349–2360, Sep. 2005.

[25] Culler, D., Estrin, D., and Srivastava, M., "Overview of sensor networks," *Computer*, vol. 37, pp. 41–49, Aug. 2004.

[26] Deshpande, A., Guestrin, C., Madden, S., Hellerstein, J., and Hong, W., "Model-driven data acquisition in sensor networks," in *Proceeding of the 30th VLDB Conference*, (Toronto, Canada), 2004.

[27] Duman, T. and Salehi, M., "Decentralized detection over multiple access channels," *IEEE Trans. Aerospace Elect. Syst.*, vol. 34, pp. 469–476, Apr. 1998.

[28] Estrin, D., Girod, L., Pottie, G., and Srivastava, M., "Instrumenting the world with wireless sensor network," in *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signals Processing (ICASSP'2001)*, pp. 2675–2678, 2001.

[29] Estrin, D., Govindan, R., Heidemann, J., and Kumar, S., "Next century challenges: Scalable coordination in sensor networks," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, pp. 263–270, 1999.

[30] Gini, F., Lombardini, F., and Verrazzani, L., "Decentralised detection strategies under communication constraints," in *Proc. Inst. Elect. Eng. F: Radar, Sonar, Navigat.*, vol. 145, pp. 199–208, Aug. 1998.

[31] Giridhar, A. and Kumar, P., "Maximizing the functional lifetime of sensor networks," in *ACM Syposium on Information Processing in Sensor Networks (IPSN)*, pp. 5–12, Apr. 2005.

[32] Gubner, J., "Distributed estimation and quantization," *IEEE Transactions on Information Theory*, vol. 39, pp. 1456–1459, July 1993.

[33] Guibas, L., "Sensing, tracking, and reasoning with relations," *IEEE Signal Processing Magazine*, vol. 19, pp. 73–85, Mar. 2002.

[34] Herzog, A. and Lind, L., "Network solutions for home health care applications," *Technology and Health Care*, vol. 11, pp. 77–87, 2003.

[35] Holman, R., Stanley, J., and Ozkan-Haller, T., "Applying video sensor networks to nearshore environment monitoring," *IEEE Pervasive Computing*, vol. 2, pp. 14–21, Oct.-Dec. 2003.

[36] Hsieh, T., "Using sensor networks for highway and traffic applications," *IEEE Potentials*, vol. 23, pp. 13–16, Apr.-May 2004.

[37] Hu, J. and Blum, R., "On the optimality of finite-level quantization for distributed signal detection," vol. 47, pp. 1665–1671, May 2001.

[38] Hu, Z. and Li, B., "On the fundamental capacity and lifetime limits of energy-constrained wireless sensor networks," in *Proc. 10th IEEE Real-Time and Embedded Technology and Applications Symposium (PTAS 2004)*, 2004.

[39] Huang, Y. and Hua, Y., "Multihop progressive decentralized estimation in wireless sensor networks," *IEEE Signal Processing Letter*, vol. 14, pp. 1004–1007, Dec. 2007.

[40] Jovanov, E., Milenkovic, A., Otto, C., and de Groen, P., "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, Mar. 2005.

[41] Kalpakis, K., Dasgupta, K., and Namjoshi, P., "Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks," vol. 42, pp. 697–716, 2003.

[42] Kansal, A., Ramamoorthy, A., Srivastava, M., and Pottie, G., "On sensor network lifetime and data distortion," in *Proc. International Symposium on Information Theory (ISIT)*, pp. 6–10, Sept. 2005.

[43] Kaplan, L., Le, Q., and Molnar, P., "Maximum likelihood methods for bearings-only target localization," in *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2001)*, pp. 554–557, 2001.

[44] Kay, S., *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, New Jersey: Prentice Hall, 1993.

[45] Korhonen, I., Parkka, J., and Gils, M., "Health monitoring in the home of the future," *IEEE Engineering in Medicine and Biology Magazine*, pp. 66–73, May-Jun. 2003.

[46] Krasnopeev, A., Xiao, J.-J., and Luo, Z.-Q., "Minimum energy decentralized estimation in sensor network with correlated sensor noise," *EURASIP Journal on Wireless Communications and Networking*, vol. 5, no. 4, pp. 473–482, 2005.

[47] Kumar, S., Zhao, F., and Shepherd, D., "Special issue on collaborative information processing in microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, pp. 13–14, Mar. 2002.

[48] Lam, W. and Reibman, A., "Design of quantizers for decentralized systems with communication constraints," *IEEE Transactions on Communication*, vol. 41, pp. 1602–1605, Aug. 1993.

[49] Li, D., Wang, K., Hu, Y., and Sayeed, A., "Detection, classfication and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, pp. 17–29, Mar. 2002.

[50] Li, J. and AlRegib, G., "Energy-efficient cluster-based distributed estimation in wireless sensor networks," in *Proc. of Military Communications Conference (MILCOM 2006)*, (Washington, DC), Oct. 23-25 2006.

[51] Li, J. and AlRegib, G., "Rate-constrained distributed estimation in wireless sensor networks," in *Proc. of 15th International Conference on Computer Communications and Networks (ICCCN 2006)*, (Arlington, VA), pp. 317–322, Oct. 8-11 2006.

[52] Li, J. and AlRegib, G., "Energy-constrained distributed estimation in wireless sensor networks," in *Proc. of Military Communications Conference (MILCOM 2007)*, (Orlando, FL), Oct. 29-31 2007.

[53] Li, J. and AlRegib, G., "Rate-constrained distributed estimation in wireless sensor networks," *IEEE Transaction on Signal Processing*, vol. 55, pp. 1634–1643, May 2007.

[54] Li, J. and AlRegib, G., "Distributed estimation in energy-constrained wireless sensor networks," *IEEE Transaction on Signal Processing*, 2008. submitted.

[55] Li, J. and AlRegib, G., "Function-based network lifetime for estimation in wireless sensor networks," *IEEE Signal Processing Letters*, vol. 15, pp. 533–536, 2008.

[56] Li, J. and AlRegib, G., "Maximizing network lifetime for estimation in multi-hop wireless sensor networks," in *Proc. of the Firth International Workshop on Sensor Networks (SN2008) in conjunction with IEEE International Conference on Computer Communications and Networks (ICCCN 2008)*, (Virgin Islands, USA), Aug. 4-7 2008.

[57] Li, J. and AlRegib, G., "Network lifetime maximization for estimation in multi-hop wireless sensor networks," *IEEE Transaction on Signal Processing*, 2008. accepted.

[58] Li, J. and AlRegib, G., "Optimal weighted data gathering in multi-hop heterogeneous sensor networks," in *Proc. of Military Communications Conference (MILCOM 2008)*, (San Diego, CA), Nov. 17-19 2008. accepted.

[59] Li, W. and Dai, H., "Distributed detection in wireless sensor networks using a multiple access channel," *IEEE Trans. on Signal Processing*, vol. 55, no. 3, 2007.

[60] Lin, Y., Chen, B., and Varshney, P., "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerospace Elect. Syst.*, vol. 41, pp. 475–488, Apr. 2005.

[61] Liu, J., Chu, M., and Reich, J., "Multitarget tracking in distributed sensor networks," *IEEE Signal Processing Magazine*, vol. 24, no. 3, 2007.

[62] Liu, K. and Sayeed, A., "Type-based decentralized detection in wireless sensor networks," *IEEE Trans. on Signal Processing*, vol. 55, no. 5, 2007.

[63] Luo, Z.-Q., "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 735–744, Apr. 2005.

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

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

[66] Madan, R. and Lall, S., "Distributed algorithms for maximum lifetime routing in wireless sensor networks," vol. 5, pp. 2185–2193, Aug. 2006.

[67] Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R., and Anderson, J., "Wireless sensor networks for habitat monitoring," in *Proc. of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, vol. 3, (Atlanta, GA), pp. 88–97, 2002.

[68] Matamoros, J. and Anton-Haro, C., "Opportunistic power allocation schemes for the maximization of network lifetime in wireless sensor networks," in *Proc. of the Intl. Conf. on Acoustics, Speech, and Signal Processing*, (Las Vegas, NV), pp. 2273–2276, Mar. 30 - Apr. 4 2008.

[69] Megalooikonomou, V. and Yesha, Y., "Quantizer design for distributed estimation with communications constraints and unknown observation statistics," *IEEE Transactions on Communications*, vol. 48, pp. 181–184, Feb. 2000.

[70] Megalooikonomou, V. and Yesha, Y., "Space efficient quantization for distributed estimation by a multi-sensor fusion system," *Information Fusion*, no. 5, pp. 299–308, 2004.

[71] Nemeroff, J., Garcia, L., Hampel, D., and DiPierro, S., "Application of sensor network communications," in *IEEE Military Communi. Conf. , MILCOM 2001: Communi. for Network-Centric Operations: Creating the Information Force*, vol. 1, pp. 336–341, 2001.

[72] Niu, R., Chen, B., and Varshney, P., "Decision fusion rules in wireless sensor networks using fading channel statistics," in *Proc. of the 2003 Conference on Information Sciences and Systems*, Mar.12-14 2003.

[73] Olfati-Saber, R. and Shamma, J. S., "Consensus filters for sensor networks and distributed sensor fusion," in *Proc. of 44th IEEE Conf. Dec. Eur. Control Conf.*, (Seville, Spain), pp. 6698–6703, Dec. 2005.

[74] Otto, C., Milenkovic, A., Sanders, C., and Jovanov, E., "System architecture of a wireless body area network for ubiquitous health monitoring," *Journal of Mobile Multimedia*, vol. 1, no. 4, pp. 307–326, 2006.

[75] Papadopoulos, H., Wornell, G., and Oppenheim, A., "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Transactions on Information Theory*, vol. 47, pp. 978–1002, Mar. 2001.

[76] Pradhan, S., Kusuma, J., and Ramchandran, K., "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, Mar. 2002.

[77] Rabbat, M. G., Nowak, R. D., and Bucklew, J. A., "Generalized consensus algorithms in networked systems with erasure links," in *Proc. of 4th Int. Symp. Info. Processing Sensor Networks*, (New York), pp. 1088–1–92, Jun. 2005.

[78] Rabbat, M. and Nowak, R., "Decentralized source localization and tracking," in *Proc. of the Intl. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, (Montreal, Canada), pp. 921–924, May 2004.

[79] Ribeiro, A. and Giannakis, G., "Bandwidth-constrained distributed estimation for wireless sensor networks, part i: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, pp. 1131–1143, Mar. 2006.

[80] Ribeiro, A. and Giannakis, G., "Bandwidth-constrained distributed estimation for wireless sensor networks, part ii: Unknown pdf," *IEEE Transactions on Signal Processing*, vol. 54, pp. 2784–2796, July 2006.

[81] R.Wattenhofer, Li, L., Bahl, P., and Y.Wang, "Distributed topology control for power efficient operation in multihop wireless ad hoc networks," in *Proc. IEEE INFOCOM*, (Anchorage, AK, USA), pp. 1388–1397, Apr. 2001.

[82] Schizas, I. D., Giannakis, G. B., Roumeliotis, S. I., and Ribeiro, A., "Consensus in ad hoc wsns with noisy links - part ii: Distributed estimation and smoothing of random signals," *IEEE Trans. Signal Processing*, vol. 56, pp. 1650–1666, Apr. 2008.

[83] Schizas, I. D., Ribeiro, A., and Giannakis, G. B., "Consensus in ad hoc wsns with noisy links - part i: Distributed estimation of deterministic signals," *IEEE Trans. Signal Processing*, vol. 56, pp. 350–364, Jan. 2008.

[84] Shnayder, V., Chen, B., Lorincz, K., Fulford-Jones, T. R. F., and Welsh, M., "Sensor networks for medical care," in *ACM Sensys 2005*, Nov. 2005.

[85] Spanos, D. P., Olfati-Saber, R., and Murray, R. M., "Dynamic consensus on mobile networks," in *Proc. of the 16th IFAC World Congr.*, (Prague, Czech Republic), Jul. 2005.

[86] Tian, Q. and Coyle, E., "Optimal distributed detection in clustered wireless sensor networks," *IEEE Trans. on Signal Processing*, vol. 55, no. 7, 2007.

[87] Toh, C.-K., "Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks," *IEEE Communications Magazine*, pp. 138–147, Jun. 2001.

[88] Tsujita, W., Kaneko, S., Ueda, T., Ishida, H., and Moriizumi, T., "Air-pollution measurement system for environmental monitoring network," in *Proc. 12th IEEE Int. Conf. Solid-State Sensors, Actuators and Microsystems*, vol. 1, pp. 544–547, Jun. 2003.

[89] Viswanathan, R. and Varshney, P., "Distributed detection with multiple sensors: part i - fundamentals," *Proceedings of IEEE*, vol. 85, pp. 54–63, Jan. 1997.

[90] Wang, B., Chua, K., and Srinivasan, V., "Localized recursive estimation in energy constrained wireless sensor networks," *Journal of Networks*, vol. 1, pp. 18–26, Jun. 2006.

[91] Wu, J., Huang, Q., and Lee, T., "Energy-constrained decentralized best-linear-unbiased estimation via partial sensor noise variance knowledge," *IEEE Signal Processing Letter*, vol. 15, pp. 33–36, 2008.

[92] Wu, J., Huang, Q., and Lee, T., "Minimal energy decentralized estimation via exploiting the statistical knowledge of sensor noise variance," *IEEE Trans. Signal Processing*, vol. 56, pp. 2171–2176, May 2008.

[93] Xiao, J.-J., *Distributed Data Fusion and Information Processing in Wireless Sensor Networks.* PhD thesis, University of Minnesota, Jul. 2006.

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

[95] Xiao, J.-J. and Luo, X.-Q., "Universal decentralized detection in bandwidth-constrained sensor network," *IEEE Trans. on Signal Processing*, vol. 53, no. 8, 2005.

[96] Xiao, J.-J., Ribeiro, A., Luo, Z.-Q., and Giannakis, G. B., "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, pp. 27–41, Jul. 2006.

[97] Xiao, L. and Boyd, S., "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, pp. 65–78, Sep. 2004.

[98] Xiao, L., Boyd, S., and Lall, S., "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. of 4th Int. Symp. Inf. Processing Sensor Networks*, (Los Angeles, CA), pp. 63–70, Apr. 2005.

[99] Xiong, Z., Liveris, A., and Cheng, S., "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, pp. 80–94, Sep. 2004.

[100] Xu, N., Rangwala, S., Chintalapudi, K., Ganesan, D., Broad, A., Govindan, R., and Estrin, D., "A wireless sensor network for structural monitoring," in *ACM Sensys 2004*, Nov. 2004.

[101] Xu, Y., Heidemann, J., and Estrin, D., "Geography-informed energy conservation for ad hoc routing," in *ACM SIGMOBILE 7/01*, (Roma, Italy), pp. 70–84, 2001.

[102] Zhang, H. and Hou, J., "On deriving the upper bound of lifetime for large sensor networks," in *Proc. MobiHoc*, pp. 121–132, 2004.

[103] Zhao, F., Liu, J., Liu, J., Guibas, L., and Reich, J., "Collaborative signal and information processing: An information-directed approach," *Proceedings of the IEEE*, vol. 91, pp. 1199–1209, Aug. 2003.

[104] Zhao, F., Shin, J., and Reich, J., "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, pp. 61–72, Mar. 2002.

VITA

Junlin Li received the B.S. degree and M.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2001 and 2004, respectively. Since August 2004, he has been with the School of Electrical and Computer Engineering at Georgia Institute of Technology to pursue the Ph.D. degree. His Ph.D. research is devoted to distributed signal processing in wireless sensor networks. He received an M.S. degree from Georgia Tech in December 2006.

He spent the summer of 2006 as a Research Intern in the Microsoft Research (MSR) working on client-side echo cancellation algorithms for multi-party audio conferencing. During his Ph.D. study, he also worked on 3D graphic coding and transmission, video coding, etc. His general research interests include multimedia processing and communication, audio-visual conferencing, distributed signal processing and wireless sensor networks.