An Energy-Efficient Bit Allocation Scheme in Wireless Sensor Networks

by

Xiaoxia Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In wireless sensor networks (WSN), a large number of sensor nodes which are capable of sensing, data processing and communicating are densely deployed in an area to measure some physical phenomenon. Generally, wireless sensor nodes carry very limited irreplaceable power sources. Thus, two primary concerns in WSN are to save the overall energy consumption and to prolong the network lifetime, namely the time when all the nodes are functional.

Motivated by these two concerns, this thesis mainly focuses on the energy efficient transmission and bit allocation schemes in multi-source single-sink WSN from an information theoretic point of view. Specifically, this thesis investigates the interactions between source coding and channel coding to gain cooperation between them in terms of energy efficiency.

For transmission through additive white Gaussian noise (AWGN) channel with path loss, this work shows that the overall energy consumption can be minimized if each source transmits with minimum power and cooperates with other sensors in TDMA (time-division multiple access) mode. From the source coding perspective, the Slepian-Wolf coding theorem is applied for efficient bit allocation since sources are usually highly correlated in WSN. Combining the transmission with correlated source coding, we derive an optimal closed form bit allocation scheme to minimize the overall energy consumption. The fundamental idea is to allocate more bits to the nodes with better channel conditions and less bits to the nodes with worse channel conditions. Based on this scheme, we further maximize the network lifetime and develop a heuristic algorithm to average the distribution of energy consumption among all sensors. Both analytical and simulation results are presented to show the superiority of our schemes.

Acknowledgements

Most importantly, I would like to thank my two supervisors: Professor Liang-Liang Xie and Professor Xuemin (Sherman) Shen, for their generous and insightful guidance throughout my two-year master study at the University of Waterloo and throughout the complement of every detail in this thesis. They not only make me benefit significantly from every discussion, but also serve as a role model to conduct fundamental research.

I would also thank the readers of this thesis, Professor Mohamed Oussama Damen and Professor Zhou Wang, for taking time to read this thesis and providing constructive suggestions.

I am also grateful to my fellow students Mr. Xiugang Wu, Mr. Xinsheng Zhou and every member in BBCR group for their helpful discussions. From these discussions, I have gained a better understanding of many aspects in information theory.

I would like to thank my parents for their endless and unconditional love and for pushing me to finish this work.

Finally, I would like to thank Pierce and all my friends for their optimistic life attitude and making my graduate life joyful.

To my parents

Contents

List of Figures

Chapter 1

Introduction

The rapid evolution of hardware circuits, as well as communication theory, has prompted the realization of abundant new technologies, which were even unthinkable for engineers. Among these, one promising technology which has profound application scenarios is the utilization of low cost wireless sensors that are capable of sensing, data processing and communicating [1], [2], [3], [4].

Usually, a wireless sensor network (WSN) consists of a large number of sensors that are densely deployed in an area to measure some physical phenomenon, such as temperature, humidity, pressure, etc. After their deployment, the distributed sensor nodes collect data from their surroundings, encode the data, and transmit them to the sink node via wireless channels [5]. The sink node is a special node which is responsible of gathering all the data and thus connect sensors to the wired or wireless backbone network. Many applications for large scale WSN include habitat monitoring [6], underground coal mine monitoring [7], structural health monitoring [8], etc. [9], [10].

1.1 Motivation and Related Work

In spite of the extensive scenarios where WSN can potentially be applied to, one fundamental bottleneck is the power supply limitation of the sensor nodes. Generally, wireless sensors carry very limited irreplaceable power sources, resulting in two primary concerns in WSN: to save the overall energy consumption and to prolong the network lifetime. The network life time has numerous definitions established according to different application scenarios; however, in this thesis, we only adopt the definition that the network lifetime is the time when all the nodes are functional.

Another characteristic of WSN is that the data sensors collect are usually highly correlated, rather than independent due to the dense deployment of large number of sensors in an area. When encoding the correlated sources, Slepian-Wolf theorem remarkably indicates that if the destination knows the underlying source correlation, it is possible to achieve a fundamental limit on rate reduction without communication among the sources [11].

Therefore, in recent years, a huge surge of research activities have been dedicated in providing all kinds of protocols and algorithms to improve the energy efficiency issue in low-power WSN from several perspectives in the sensor networks protocol stack as depicted in Figure 1.1.

Traditionally, in the physical layer of a communication system, people only focus on a single part of the communication system which includes data collecting, correlated source coding and channel coding (Figure 1.2) and optimize it to improve the energy efficiency. For example, on source coding with correlated data, [12] proposes a closed form optimal rate allocation scheme to minimize some transmission cost function which is proportional to the distance from the source node to the sink, thus to decrease the overall transmitting power. As for the transmission (channel coding) in a wireless environment, [13] shows that in order to maximize the aggregate capacity with fixed power, only the user having the best channel condition is allowed to transmit at any given time, which essentially shares the same spirit as Time Division Multiple Access (TDMA). [14] examines the tradeoff between improving the sum-rate and saving energy in many aspects of designing the sensor networks. Finally,
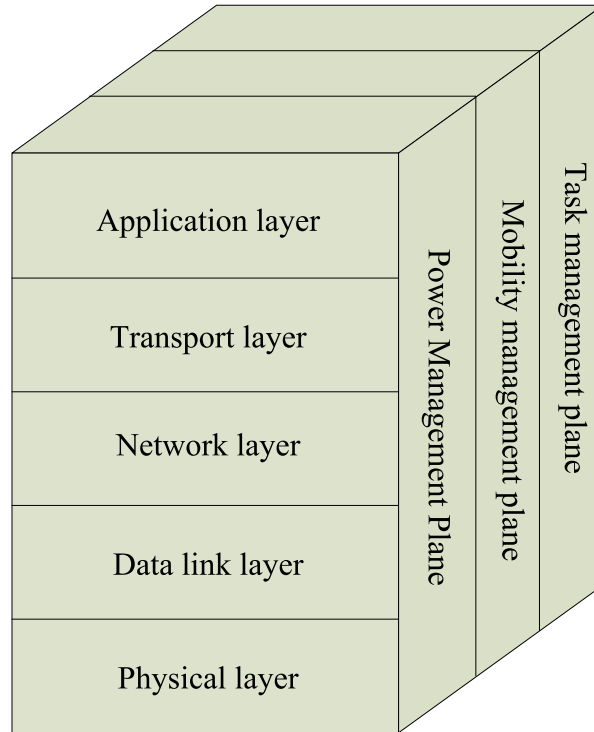
Figure 1.1: The sensor networks protocol stack [1].

[15] considers both source and channel coding to optimize the rate distortion function under a fixed power constraint. Also, [16] studies the combination to provide an algorithm to minimize the overall power consumption and to maximize network lifetime; nevertheless, it only considers a special scenario where one sensor is only correlated with another single sensor.

The data link layer is responsible for the multiplexing of data streams, data frame detection, medium access and error control. The proposed protocols mainly focus on medium access control (MAC) and error control. However, an optimal MAC design typically depends on a large range of tunable factors, such as node density, quality of service (QoS) and latency required. Thus, it is especially challenging for analysis of a MAC design. In [17] and [18], Zorzi and Rao employ a finite state machine model to study energy consumption; however, for the adaptability and flexibility to support a variety of services, a provable

3
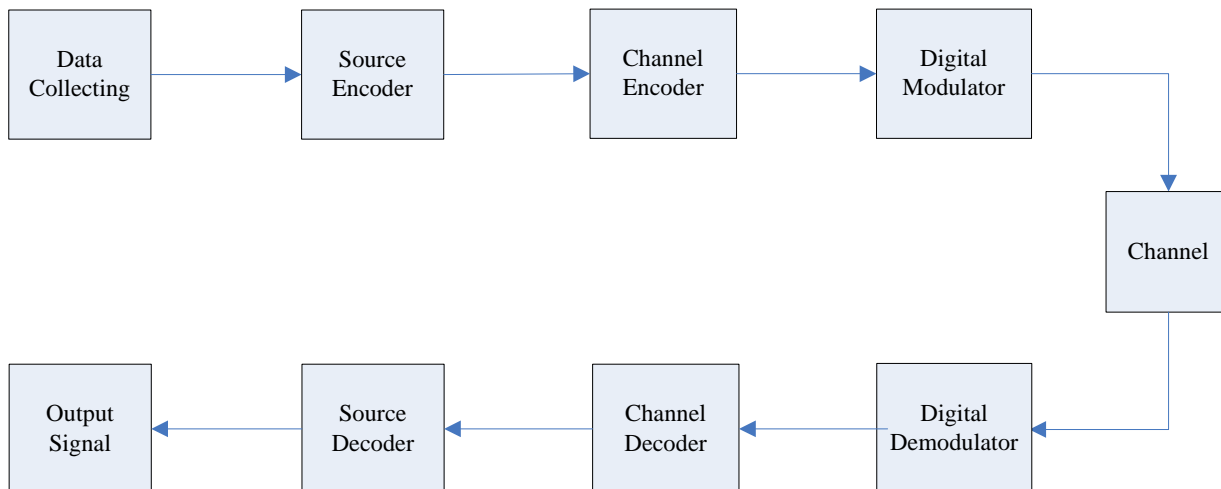
Figure 1.2: A schematic diagram of sensory communication systems.

performance without tradeoffs still remains elusive.

In the network layer, special routing protocols meeting the power limitation and other requirements of sensor networks are needed. One very common approach is to optimize a cost function where the parameters may include some combination of delay, range, hop count and battery level [19], [20]. Also, clustering algorithms naturally occur in the sensor network that supports hierarchical signal processing [21]. Another fundamental issue arises with node localization, the need for each node to know where it is in the system. This task can be tackled by deployment at known locations [22], [23]. However, in many scenarios where only random placement is available, how can we determine the relative or absolute location information, and how accurate this estimation could possibly be? Also, cooperative sensing, detection and estimation require synchronization between sensor nodes so that sensed events can be synchronized across the network.

The transport layer is needed when the system is planned to be accessed through the Internet or other external networks. However, to the best of our knowledge, there has been no attempt to propose a scheme with regard to this issue for the reason that the existing TCP matches the requirements of sensor network as well.

For the application layer, there are potentially numerous aspects of protocols that fall into three main categories: sensor management protocol, task assignment and data advertisement protocol, and sensor query and data dissemination protocol.

1.2 Objectives and Contribution

Our motivation to study this issue is to understand how source coding and channel coding can cooperate in terms of optimizing energy efficiency in general multi-source single-sink wireless sensor networks.

In this thesis, an information theoretic approach combining both source and channel coding is employed towards the energy efficiency issue in WSN where correlation exists among all sensors. A multi-source single-sink wireless sensor network is considered, where the channel is additive white Gaussian noise (AWGN) channel with path loss. The capacity is then $\frac{1}{2} \log \left(1 + \frac{P_i/d_i^\gamma}{N}\right)$, which is dependent on the transmitting power P_i and the distance d_i from source i to the sink node. We focus on the sensor network scenario where data do not change very rapidly, so that there is enough transmitting time and the transmission rate is not a major concern.

We demonstrate that when all the nodes use their respective minimum transmitting powers in the TDMA fashion to transmit, the overall consumed energy can be minimized. In this case, source-channel separation theorem [24] holds. For the source coding part, according to the distributed source coding theorem (Slepian-Wolf), we can allocate bits within the Slepian-Wolf region for each sensor to transmit avoiding loss of any information. Combining both the transmission and source coding criterion, we derive a closed form bit allocation scheme to transmit all the data while minimizing the overall energy consumption at the same time. The idea is to allocate more bits to the nodes with better channel conditions and less bits to those with worse conditions. Furthermore, based on this solution, we develop an algorithm to find the optimal bit allocation scheme to maximize the network lifetime, and meanwhile, to minimize the overall energy consumption.

Simulation results are presented after the theorems and proofs. Some comparisons are shown between our proposed algorithm and the average bit allocation scheme. The purpose is to give an intuitive impression of how efficient our scheme is.

1.3   Thesis Organization

The remainder of this thesis is organized as follows.

In Chapter 2, we review some basic concepts and provide the background knowledge in information theory and network information theory that are prerequisite for later discussion. We first call to mind the definitions of several entropies and their relationships to describe the measurement of information. Then the derivation of single user channel capacity is presented, for both discrete memoryless channel and Gaussian channel with path loss. At last, some accomplishments in network information theory including multiple access channel and Slepian-Wolf coding theorem are provided.

In Chapter 3 and Chapter 4, we study the energy efficiency issue in multi-source single-sink wireless sensor networks. Specifically, Chapter 3 focuses on minimizing the overall energy consumption. We start from single-sensor single-sink scenario and derive that it is the most energy efficient if every sensor uses its minimum power to transmit. Then we consider multi-sensor single-sink network and prove that if the network is organized in the TDMA form, the overall energy consumption will be saved. Finally, combining with correlated source coding, we propose an optimal bit allocation scheme to minimize the overall energy consumption in the network. Both proofs and simulation results are presented.

In Chapter 4, we extend the optimization problem to incorporate the network lifetime prolonging. This optimization problem then contains two objective functions. We provide an algorithm which aims to achieve lexicographical optimality, i.e, optimal in the sense that the first objective has higher priority than the second objective. This algorithm starts from the optimum solution obtained from the previous chapter that can minimize the

overall energy consumption, then try to average the energy consumption among all sensors, since the network lifetime depends on the sensor that consumes the largest energy. Thus, this algorithm can maximize the network lifetime, and under this condition, accomplish minimizing the overall energy consumption. Furthermore, simulation results are given at the end of this chapter.

Finally, the conclusion of this thesis and some discussions about possible future research are presented in Chapter 5.

Chapter 2

Background

In this chapter, we review some basic definitions and theorems in information theory and network information theory, which are the preliminaries for the subsequent of our study.

In single user information theory, we provide some definitions of the measurement of information and their properties and relationships. We also review Shannon's single user channel capacity for a discrete memoryless channel presented in his original 1948 paper and how to derive the capacity in the Gaussian environment.

Then we present some accomplishments in multi-user information theory. In particular, the multiple access channel (MAC) model and the capacity results for both the discrete memoryless and Gaussian multiple access channels are shown. Besides, for the correlated source coding, we review the distinguished Slepian-Wolf theorem presented in 1973. This theorem introduces cooperation among sources with correlated information, thus can greatly reduce source coding rate, which results in a diminution of the overall energy consumption.

2.1 Preliminaries

In this section, we rewrite some basic definitions that are fundamental yet crucial for the development of information theory. Specifically, the theory of information starts from what is the measurement of "information". In his 1948 original paper, Shannon employed the concept of entropy to measure the uncertainty of a random variable from its probability mass function, which reflects the information contained in this random variable.

Definition 2.1.1. *(Entropy):*
The entropy $H(X)$ of a discrete random variable X is defined by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{2.1}$$

where $\mathcal{X}$ is the alphabet of the random variable X and $p(x)$ is the probability mass function.

2.1.1 Joint Entropy and Conditional Entropy

Based on the concept for the single random variable, the definition of entropy can be extended to a vector of random variables. Besides, some properties with regard to the relationships between these entropies are also worth mentioning.

Definition 2.1.2. *(Joint Entropy):*
The joint entropy $H(X,Y)$ of a pair of discrete random variables (X,Y) with a joint distribution $p(x,y)$ is defined as

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y). \tag{2.2}$$

Given another random variable, the conditional entropy measures the remaining uncertainty in the random variable as defined in Definition 2.1.3.

Definition 2.1.3. *(Conditional Entropy):*
The conditional entropy $H(Y|X)$ with $(X,Y) \sim p(x,y)$ is defined as

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \tag{2.3}$$

$$= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \tag{2.4}$$

$$= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \tag{2.5}$$

The concepts of joint entropy and conditional entropy are defined especially to satisfy the fact that the entropy of a pair of random variables is the entropy of one plus the conditional entropy of the other, i.e.,

$$H(X, Y) = H(X) + H(Y|X). \tag{2.6}$$

This chain rule property can be extended to a vector of random variables as shown in Theorem 2.1.1. In this chapter, we will present the theorems without any proof. For the details, please refer to [24].

Theorem 2.1.1. *(Chain rule for entropy):*
Let $X_1, X_2, \ldots, X_n$ be drawn according to $p(x_1, x_2, \ldots, x_n)$. Then

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_1). \tag{2.7}$$

Another observation is that knowing another random variable Y, the uncertainty of random variable X should be reduced.

Theorem 2.1.2. *(Conditioning reduces entropy):*

$$H(X|Y) \leq H(X) \tag{2.8}$$

with equality if and only if X and Y are independent.

2.1.2 Mutual Information

The notion of information is so wide-ranging that one single definition is limited to capture all the intuitive interrelationships. Thus, Shannon introduced the concept of mutual information between two random variables. It is a measure of the amount of information that one random variable contains about another random variable.

Definition 2.1.4. *(Mutual Information):*
Consider two random variables X and Y with a joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$. The mutual information $I(X;Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)p(y)$, i.e.,

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \tag{2.9}$$

Actually, the mutual information is the reduction in the uncertainty of one random variable due to the knowledge of another random variable. Therefore, we can easily prove the following relationships between mutual information and entropy.

Theorem 2.1.3. *(Relationships between mutual information and entropy):*

$$I(X;Y) = H(X) - H(X|Y), \tag{2.10}$$

$$I(X;Y) = H(Y) - H(Y|X), \tag{2.11}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \tag{2.12}$$

$$I(X;Y) = I(Y;X), \tag{2.13}$$

$$I(X;X) = H(X). \tag{2.14}$$

The relationships between $H(X)$, $H(Y)$, $H(X,Y)$, $H(X|Y)$, $H(Y|X)$ and $I(X;Y)$ can be expressed in a Venn diagram depicted in Figure 2.1.

Mutual information also satisfies the chain rule.
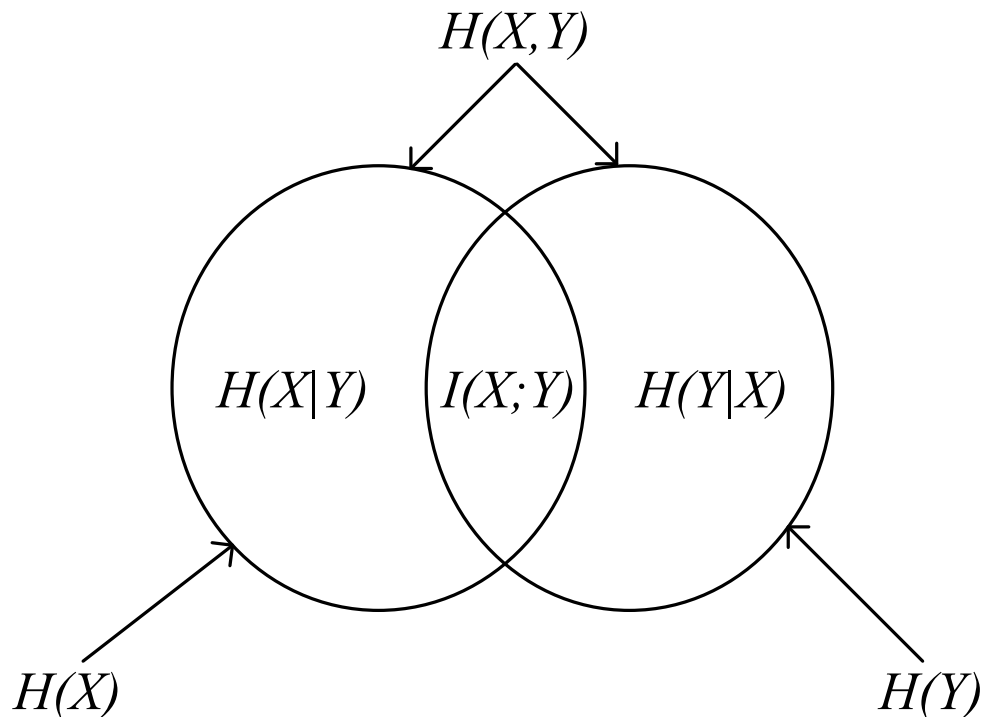
Figure 2.1: Relationships between entropy and mutual information.

Theorem 2.1.4. *(Chain rule for information):*

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_{i-1}, \ldots, X_1), \qquad (2.15)$$

where the conditional mutual information is defined by

$$I(X; Y | Z) = H(X | Z) - H(X | Y, Z). \qquad (2.16)$$

2.2   Single User Channel Capacity

One fundamental question in information theory is, in point to point communication, whether there is an ultimate limitation on how much information can be reliably trans-

mitted over a channel, instead of transmitting as much as we want. If there exists such a limitation, is there a universal expression? Fortunately, Shannon solved this problem by proposing the concept of *capacity* and hence based on which, he established the whole model of communication system thereafter.

Consider the *Discrete Memoryless Channel* (DMC) as depicted in Figure 2.2, that is the simplest channel model from which the channel capacity is derived.
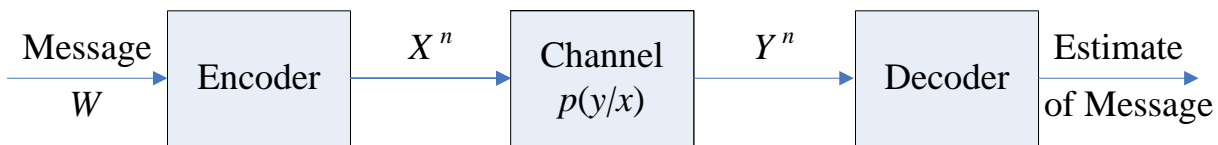
Figure 2.2: A discrete memoryless channel.

The channel consists of an input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$ and a probability transition matrix $p(y|x)$, which is the probability of observing the output y given x is sent at the source.

Definition 2.2.1. *The channel capacity of a discrete memoryless channel is defined as*

$$C = \max_{p(x)} I(X;Y), \tag{2.17}$$

where $I(X;Y)$ is the mutual information of input random variable X and output random variable Y, and the maximum is taken over all possible input distributions $p(x)$.

Shannon proved in his original 1948 paper that the channel capacity is the supremum of all the rates at which information can be sent reliably over a channel. At any rate below this capacity the reliable transmission can be implemented, and any rate above this capacity is impossible for reliable communication, no matter what coding technique is employed.

Figure 2.3: The Gaussian channel.

2.2.1 The Gaussian Channel

For most communication channels in reality, the input and output symbols are continuous and the signals are corrupted by a Gaussian noise during transmission. We usually use the Gau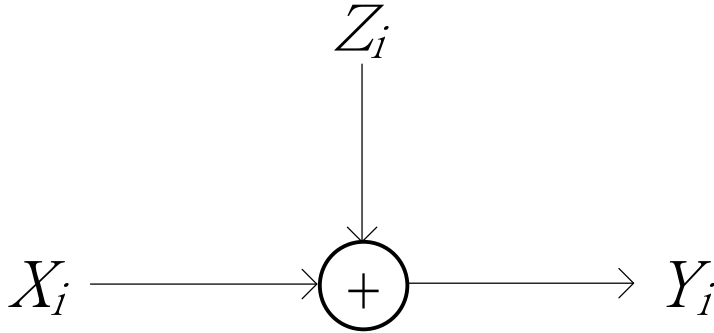ssian channel to model this type of continuous channels. The Gaussian channel is a time discrete channel with output Y_i at time i, where Y_i is the sum of the input X_i and the noise Z_i. The noise Z_i is independent of the input and is drawn i.i.d. from a Gaussian distribution with variance N. Thus,

$$Y_i = X_i + Z_i, \quad Z_i \sim \mathcal{N}(0, N) \tag{2.18}$$

Without any limitation on the input and noise, the capacity of this channel could be infinite. For example, if the noise variance is zero, the decoder could estimate the input perfectly at any arbitrary rate. Also, if there is no power constraint on the input, we can choose infinite number of inputs arbitrarily far apart to be distinguishable at the receiver. In this case, the capacity is infinite as well.

In a real Gaussian channel, the noise variance is non-zero and there is always a power constraint on the input symbols. For any codeword $(x_1, x_2, \ldots, x_n)$ transmitted over the channel, the average power is constrained by P, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \le P. \tag{2.19}$$

14

Definition 2.2.2. *The capacity of the Gaussian channel with power constraint P is*

$$C = \max_{p(x):EX^2 \leq P} I(X;Y), \qquad (2.20)$$

Expanding $I(X;Y)$, the capacity can be calculated as follows:

$$I(X;Y) = h(Y) - h(Y|X) \qquad (2.21)$$
$$= h(Y) - h(X + Z|X) \qquad (2.22)$$
$$= h(Y) - h(Z|X) \qquad (2.23)$$
$$= h(Y) - h(Z) \qquad (2.24)$$

where $h(X)$ is the differential entropy of continuous random variable X and the last equation follows from the fact that Z and X are independent.

Next, we only give the expression of differential entropy of Gaussian random variables and omit the proofs. The proofs can be found in [24].

Theorem 2.2.1. *For a single random variable X with the probability distribution $\mathcal{N}(0, \sigma^2)$,*

$$h(X) = \frac{1}{2} \log 2\pi e \sigma^2 \ \ bits. \qquad (2.25)$$

Theorem 2.2.2. *Let $X_1, X_2, \ldots, X_n$ have a multivariate Gaussian distribution with mean μ and covariance matrix K,*

$$h(X_1, X_2, \ldots, X_n) = \frac{1}{2} \log(2\pi e)^n |K| \ \ bits, \qquad (2.26)$$

where $|K|$ denotes the determinant of K.

Thus, $h(Z) = \frac{1}{2} \log 2\pi e N$. Also, since X and Z are independent and $EZ = 0$, $EY^2 = E(X + Z)^2 = EX^2 + 2EXEZ + EZ^2 = P + N$, the entropy of Y is bounded by $\frac{1}{2} \log 2\pi e (P + N)$.

Applying this result to calculate the capacity, we obtain

$$I(X;Y) = h(Y) - h(Z) \qquad (2.27)$$
$$\leq \frac{1}{2} \log 2\pi e (P + N) - \frac{1}{2} \log 2\pi e N \qquad (2.28)$$
$$= \frac{1}{2} \log(1 + \frac{P}{N}). \qquad (2.29)$$

15

Theorem 2.2.3. *The capacity of the Gaussian channel is*

$$C = \max_{p(x):EX^2 \leq P} I(X;Y)$$
$$= \frac{1}{2}\log(1 + \frac{P}{N}) \tag{2.30}$$

where the maximum is attained when $X \sim \mathcal{N}(0, P)$.

This capacity is the supremum of all the successful transmission rates and it can be achieved if the input satisfies a Gaussian distribution. In order to fully utilize the power, the source would transmit at a rate as close to the capacity as possible. Thus, in the following analysis, without loss of generality, we assume that the transmitting rate is equal to this capacity, i.e. $R = \frac{1}{2}\log(1 + \frac{P}{N})$.

The Gaussian Channel with Path Loss

In the wireless environment, signals always undergo path loss. The signal strength attenuates with some exponent of transmission distance. For example, if the distance from the sensor to the destination is d, and the path loss parameter in this area is denoted by γ, then the received signal power at the receiver end is P/d^γ, given that P is the emitting power.

Thus, the capacity of this Gaussian channel with path loss fading is represented by:

$$C = \frac{1}{2}\log\left(1 + \frac{P/d^\gamma}{N}\right) \tag{2.31}$$

2.3 Some Accomplishments in Multi-user Information Theory

In this section, we review two significant models and their corresponding results in multi-user information theory. One is the *Multiple Access Channel*, which is a channel model where there are more than one transmitters and one common receiver. The other one is *Slepian-Wolf Coding*, which is a distributed source coding scheme.

2.3.1 Multiple Access Channel

In a multiple access channel, two or more senders send information to a common receiver as depicted in Figure 2.4.
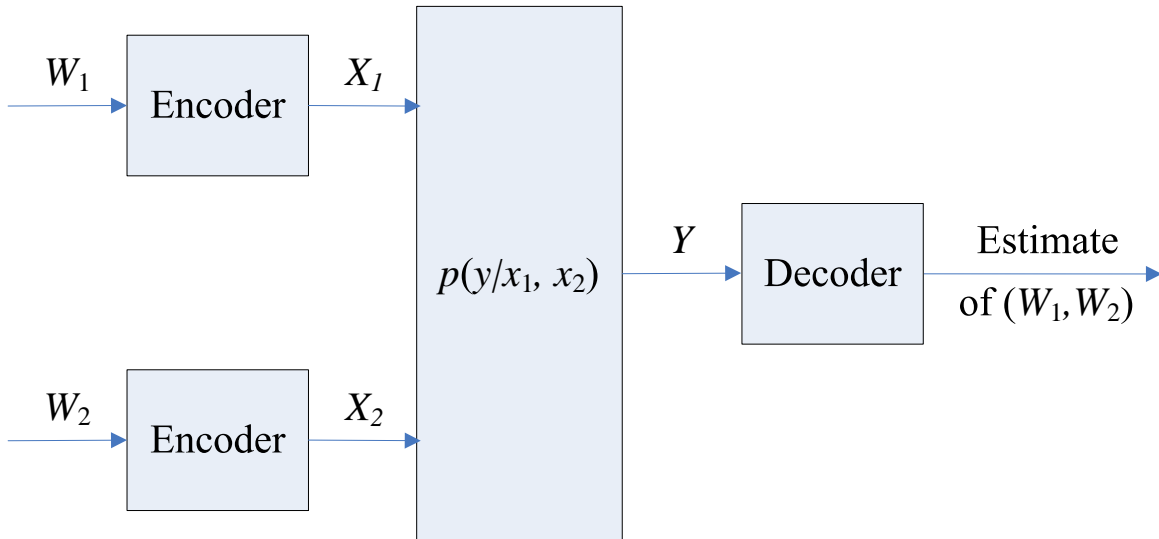
Figure 2.4: The multiple access channel.

Sender 1 sends message W_1 and employs the codeword $\mathbf{X_1}$. Sender 2 sends message W_2 and the corresponding codeword is $\mathbf{X_2}$. The *discrete memoryless multiple access channel* consists of three alphabets, $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ and a probability transition matrix $p(y|x_1, x_2)$.

The multiple access channel capacity region was found by Ahlswede [25] and Liao [26]:

Theorem 2.3.1. *The capacity of a multiple access channel is the closure of the convex hull of all (R_1, R_2) satisfying*

$$R_1 < I(X_1; Y|X_2), \tag{2.32}$$

$$R_2 < I(X_2; Y|X_1), \tag{2.33}$$

$$R_1 + R_2 < I(X_1, X_2; Y) \tag{2.34}$$

for some product distribution $p(x_1)p(x_2)$ on $\mathcal{X}_1 \times \mathcal{X}_2$.

This result can be generalized to m senders, $m \geq 2$. Let $S \subseteq \{1, 2, \ldots, m\}$ and S^c denote the complement of S. $R(S) = \sum_{i \in S} R_i$, and $X(S) = \{X_i : i \in S\}$.

Theorem 2.3.2. *The capacity region of the m-user multiple access channel is the closure of the convex hull of the rate vectors satisfying*

$$R(S) < I(X(S); Y | X(S^c)) \quad for \ all \ S \subseteq \{1, 2, \ldots, m\} \tag{2.35}$$

for some product distribution $p_1(x_1) p_2(x_2) \ldots p_m(x_m)$.

Gaussian Multiple Access Channel

There are m transmitters, each with the transmitting power P_i. In a Gaussian channel, the received signal is the sum of all the transmitted signals plus Gaussian noise, i.e.,

$$Y = \sum_{i=1}^{m} X_i + Z. \tag{2.36}$$

Then the capacity region is the closure of the convex hull of the rate vectors satisfying

$$R(S) < \frac{1}{2} \log(1 + \frac{|S|P}{N}) \quad for \ all \ S \subseteq \{1, 2, \ldots, m\}, \tag{2.37}$$

where $|S|$ denotes the cardinality of the set S.

2.3.2 Correlated Source Coding

For the problem of encoding a source X, we know that a rate $R > H(X)$ is sufficient. Now if there are two sources X and Y with the correlation $p(x, y)$ as shown in Figure 2.5, what rate vector is sufficient if they must be encoded separately? Slepian and Wolf studied this interesting problem in [11] and they found that a joint entropy of $H(X, Y)$ is sufficient, which is the same as the case when the two sources are encoded together.
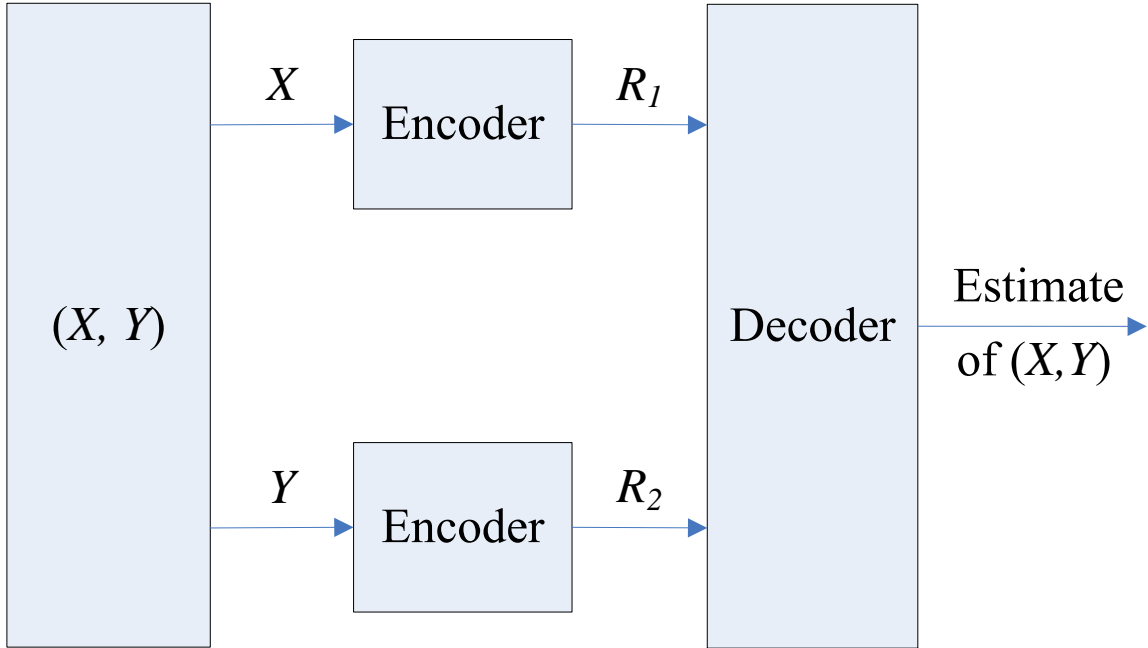
Figure 2.5: Slepian-Wolf coding.

Theorem 2.3.3. *For the distributed source coding problem for the source (X, Y) drawn i.i.d. $\sim p(x, y)$, the achievable rate region is given by*

$$R_1 \geq H(X|Y), \tag{2.38}$$

$$R_2 \geq H(Y|X), \tag{2.39}$$

$$R_1 + R_2 \geq H(X, Y). \tag{2.40}$$

This region is illustrated in Figure 2.6.

In general, if there are multiple sources $W_1, W_2, \ldots, W_n$ drawn *i.i.d.* according to $p(w_1, w_2, \ldots, w_n)$ that are encoded separately, and there is one sink node that can decode them together, then the set of achievable source coding rate vectors must lie in Slepian-Wolf region:

$$R(U) > H(W(U)|W(U^C)), \tag{2.41}$$
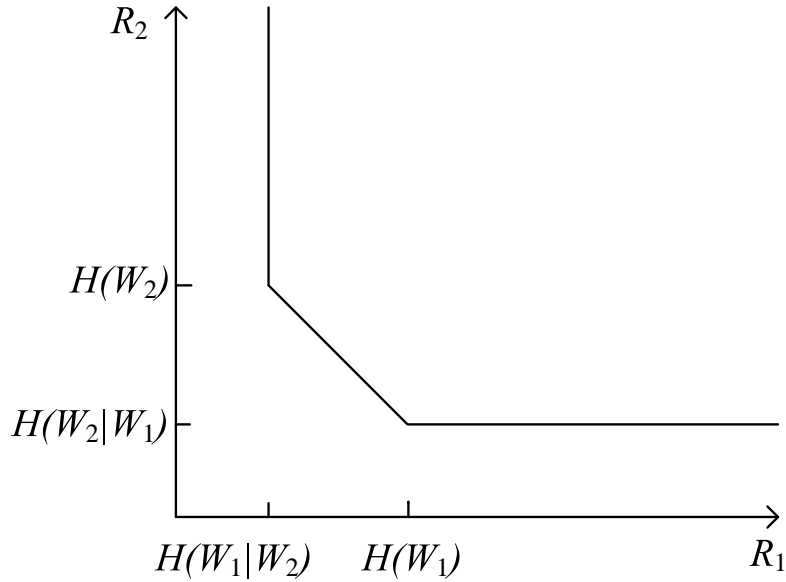
Figure 2.6: Slepian-Wolf region for two correlated sources.

for all $U \subseteq \{1, 2, \ldots, n\}$, where

$$R(U) = \sum_{j \in U} R_j, \tag{2.42}$$

and

$$W(U) = W_j : j \in U. \tag{2.43}$$

2.4 Summary

In this chapter, some basic background knowledge and preliminaries for the discussions in the next two chapters of this thesis are provided. We first review some fundamental definitions of entropies, which describe the measurement of information. Their properties and interconnections are elemental for deriving the subsequent theorems in this thesis. Then we introduce one significant triumph in Shannon's single user information theory: the channel capacity theorem, and derive the capacity expression for the Gaussian case.

Finally, two major accomplishments in multi-user information theory related to our work are presented: multiple access channel and correlated source coding.

Chapter 3

Source-Channel Communication in Minimization of Energy Consumption

3.1   Introduction

The following two chapters deal with the optimization of energy efficiency issue in wireless sensor networks. In this chapter, we mainly focus on the interactions between source coding and the channel characteristics in terms of saving the overall energy consumption. Here, the concept *energy consumption* is raised, since from the previous chapter, we know that higher rate demands higher power; However, if there are finite bits to be transmitted, higher rate also saves transmission time, thus may save energy as well. Thus, in this case energy consumption is of more study concern than power.

This chapter is organized as follows. Before arriving at the final solution, we divide the approach into several steps. In the first two sections, we only concentrate on the channel coding (signal transmission) part. For the single-sensor single sink transmission, we already know that transmission rate is a log function of the transmitting power used. The question is how much power should every sensor consume to achieve the optimal energy consumption for the transmission of fixed number of bits? Section 3.2 gives the answer to

this question. Then we extend to the scenario of multi-sensor single-sink WSN where each sensor uses its optimal power obtained in Section 3.2. From the previous chapter we know that the sensors can transmit at the same time, thus forming the multiple access channel. However, another feasible transmission mode is: the sensors transmit one after another, i.e., each uses a fraction of the total time. Now the question becomes how to coordinate the transmission among these sensors, which will be answered in section 3.3.

Based on the previous results and inspired by Slepian-Wolf coding reviewed in the previous chapter, we further come up with the bit allocation scheme that can minimize the overall energy consumption in section 3.4. We take into account the effect of different channel characteristics when encoding the sources.

3.2 Energy-Efficient Method for Single-Sensor Single-Sink Transmission

First, we consider the case where there is only one source node and one sink node. Suppose that this node has some finite non-negative bits $\mathcal{B}$ that need to be transmitted.

Generally, the transmitting power of each sensor is both lower bounded by $P_{\min}$ and upper bounded by $P_{\max}$. Still, there are many different levels of transmitting powers the sensor can adopt. In other words, the sensor can use either lower power but longer time, or higher power but shorter time to send the same amount of bits $\mathcal{B}$. Now the question is: which way is better if the overall energy consumption is the primary concern?

It is well known that for an AWGN channel, the reliable transmission rate is bounded by its capacity: $R \leq \frac{1}{2}\log\left(1 + \text{SNR}\right)$, where SNR is the received signal-to-noise ratio. Also, in a wireless environment where there exists path loss, the channel capacity is:

$$R \leq \frac{1}{2}\log\left(1 + \frac{P/d^\gamma}{N}\right), \tag{3.1}$$

where, P is transmitting power, N is the noise variance, d is the distance from the source to the sink node and γ is the path loss exponent.

Then we come up with this theorem to answer the above question.

Theorem 3.2.1. *In the single source case, given that the transmitting power can be chosen arbitrarily within the bound $P_{\min} \leq P \leq P_{\max}$, in order to transmit $\mathcal{B}$ bits in total, using $P_{\min}$ consumes the least overall energy.*

Proof. Denote the transmission time by T. Then,

$$\mathcal{B} = RT = \frac{1}{2}\log\left(1 + \frac{P/d^\gamma}{N}\right)T, \tag{3.2}$$

where $\mathcal{B}$ and N are constants. The overall energy consumption in this transmission is

$$E = PT = P\frac{\mathcal{B}}{\frac{1}{2}\log\left(1 + \frac{P/d^\gamma}{N}\right)}. \tag{3.3}$$

Taking the derivative of the overall energy consumption function over power P, we can obtain

$$\frac{dE}{dP} = 2\mathcal{B}\left[\frac{P}{\log\left(1 + \frac{P/d^\gamma}{N}\right)}\right]' \tag{3.4}$$

$$= \frac{2\mathcal{B}}{\left[\log\left(1 + \frac{P/d^\gamma}{N}\right)\right]^2\left(1 + \frac{P/d^\gamma}{N}\right)}\left[\left(1 + \frac{P/d^\gamma}{N}\right)\log\left(1 + \frac{P/d^\gamma}{N}\right) - \frac{P/d^\gamma}{N}\log e\right]. \tag{3.5}$$

Let

$$f(x) = (1 + x)\log(1 + x) - x\log e, \tag{3.6}$$

where

$$x = \frac{P/d^\gamma}{N}. \tag{3.7}$$

Since $P > 0, N > 0$, and $d > 0, x > 0$,

$$f'(x) = \log(1 + x) \geq 0, \ \forall x > 0. \tag{3.8}$$

24

Taking into account $f(0) = 0$, we can obtain

$$f(x) > 0, \ \forall x > 0, \tag{3.9}$$

which implies that

$$\frac{dE}{dP} > 0, \ \forall P > 0, \ d > 0 \text{ and } N > 0. \tag{3.10}$$

This is to say, the energy function is monotone increasing with power P. Thus, using minimum power $P_{\min}$ consumes the least energy. This result is in accordance with our observation. Since $\log(1 + x)$ approaches a linear function only when $x \to 0$, the smaller power we use, the more efficient is the transmission. □

3.3 Energy-Efficient Method for Multi-Sensor Single-Sink Transmission

For multi-source single-sink transmission, there are two types of transmission modes: source nodes transmitting one after another, i.e., every time there is only one source-sink connection (time division multiple access channel, TDMA) and more than one source-sink connections existing at some specific time (multiple access channel).

Consider a network where n sensors are distributed in an area to collect some data. All the sensor nodes transmit the data to a single sink node as depicted in Figure 3.1. Let $\mathcal{N}$ be the set of sensor indices: $\mathcal{N} = \{1, \ldots, n\}$. Then for each node $i \in \mathcal{N}$, it uses its minimum power $P_{i,min}$ and the total bits it needs to transmit are some fixed value $\mathcal{B}_i$. Suppose the distance from node i to the sink node is d_i and the path loss exponent γ is the same for all sensor nodes since they are geologically close to each other.

To answer the question of which mode is more energy efficient, we propose the following theorem.

Theorem 3.3.1. *For multi-source single-sink transmission, given that each node uses its minimum power and has finite total bits to transmit, using TDMA among all the nodes is more energy efficient.*
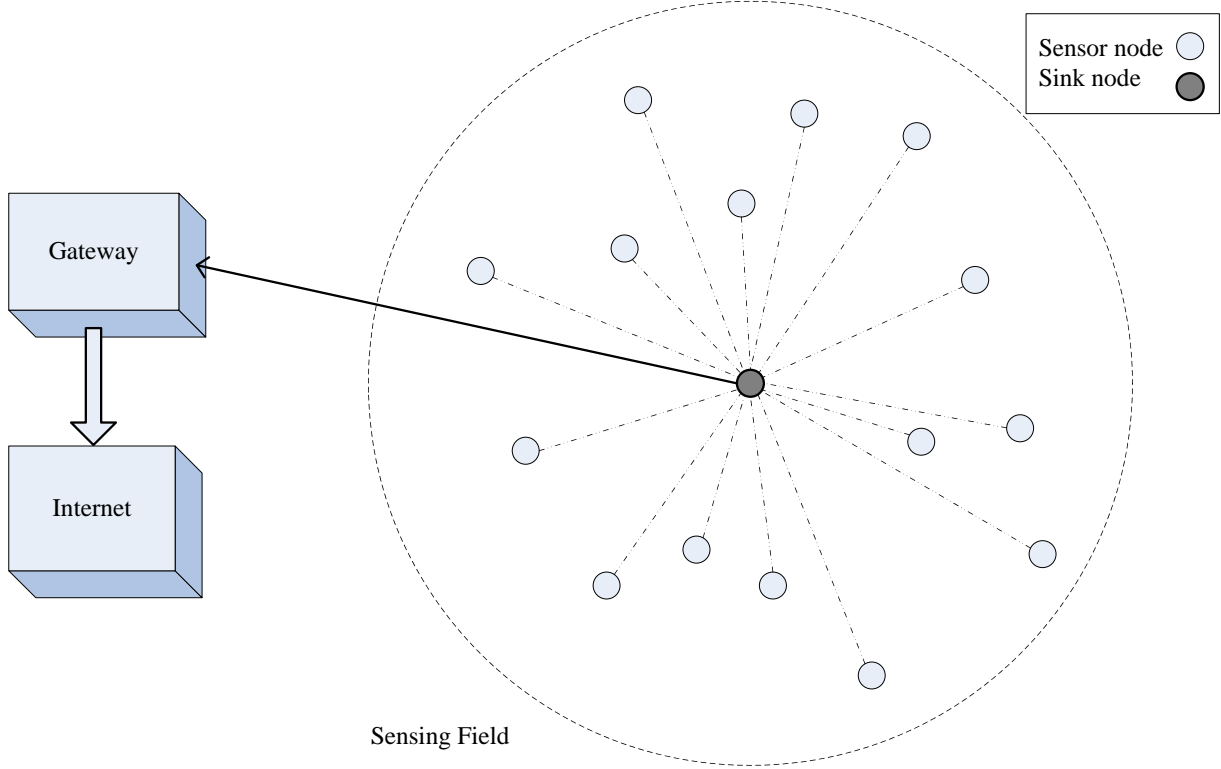
Figure 3.1: A sensor network with N sensors and one sink node.

Proof. The overall energy consumption can be calculated as $\sum_{i=1}^{n} P_{i,min} T_i$, where $T_i = \mathcal{B}_i/R_i$ is the transmission time,

$$E = \sum_{i=1}^{n} P_{i,min} \frac{\mathcal{B}_i}{R_i}. \tag{3.11}$$

1. Using TDMA, the instant rates of all source nodes can achieve their separate channel capacities, i.e.,

$$R_i = \frac{1}{2} \log \left(1 + \frac{P_{i,min}/d_i^{\gamma}}{N} \right), \ \forall i \in \mathcal{N}. \tag{3.12}$$

Traditionally, the concept "rate" usually refers to the average rate in a time frame. However, here we adopt the "instant rate" exclusively for the convenience of calculation of energy consumption.

2. Suppose at some specific time, a set $S \subseteq \mathcal{N}$ with $|S| \geq 2$ of sensors all connect to the sink node, then the rate vector should lie in multiple access channel capacity region. For any subset $S_1 \subseteq S$ (We use R_i' to differentiate):

$$\sum_{i \in S_1} R_i' \leq \frac{1}{2} \log \left(1 + \frac{\sum_{i \in S_1} P_{i,min}/d_i^\gamma}{N} \right)$$
$$< \sum_{i \in S_1} \frac{1}{2} \log \left(1 + \frac{P_{i,min}/d_i^\gamma}{N} \right)$$
$$= \sum_{i \in S_1} R_i.$$

So,

$$R_i' < R_i, \ \forall i \in S.$$

Since both $P_{i,min}$ and $\mathcal{B}_i$ are fixed, the energy consumption function is only reversely proportional to the instant rate. Thus, the overall energy consumption in Case 2 is greater than that in Case 1. Therefore, transmitting using TDMA is more energy efficient. $\square$

Remark 3.3.1. *For the multiple access channel, it is well known that CDMA is optimal in achieving the maximum rates with fixed power [24]. However, here we are interested in a different problem of sending maximum bits with fixed energy. It is shown that TDMA outperforms CDMA.*

3.4 Optimal Bit Allocation Scheme

In this section, we are trying to find the optimal bit allocation scheme to minimize the overall energy consumption based on that each sensor is utilizing its minimum power and only a fraction of the total time to transmit. In this source coding problem, we take the different channel characteristics into consideration.

This section is composed of two parts. First we explain the bit allocation problem based on the Slepian-Wolf coding theorem. Then the optimal bit allocation scheme is proposed and the proof is presented.

3.4.1 Bit Allocation Problem

In general, the measurements of sensors, e.g. temperature, humidity, sound, etc., are highly correlated. We denote the source from sensor i by W_i, $i = 1, 2, \ldots, n$. These sources are correlated with joint probability $p(W_1, W_2, \ldots, W_n)$. When encoding these correlated sources, we know from Slepian-Wolf coding theorem that a total rate of the joint entropy $R_S = H(W_1, W_2, \ldots, W_n)$, instead of the sum of separate entropies $H(W_1) + H(W_2) + \cdots + H(W_n)$, is sufficient even if these nodes are not able to communicate with each other, as long as they satisfy (Here, we use R_S to differentiate source coding rate from channel coding rate):

$$R_S(U) > H(W(U)|W(U^c)), \tag{3.13}$$

for all $U \subseteq \mathcal{N}$, where

$$R_S(U) = \sum_{j \in U} R_{Sj}, \tag{3.14}$$

and

$$W(U) = W_j : j \in U. \tag{3.15}$$

The total information generated by sensor i is $R_{Si}T$, where T is one time period for data gathering. Since the measurements of physical phenomena usually do not vary rapidly, e.g., the temperature in a day, $T > \sum_{i=1}^{n} T_i$. Under this assumption,

$$R_{Si}T = \mathcal{B}_i. \tag{3.16}$$

So $\mathcal{B}_i$ must satisfy the following set of inequalities:

$$\sum_{i \in U} \mathcal{B}_i > H(W(U)|W(U^c))T, \ \forall U \subseteq \mathcal{N}. \tag{3.17}$$

This sophisticated result has a simple interpretation. The information collected by each sensor contains two part: the unique information that can only be conveyed by this sensor and some common information which can be conveyed by other sensors. When encoding these sources, each source must first encode its unique information; then this sensor cooperate with other sensors to divide the task of common information to convey.

3.4.2 Optimal Bit Allocation Scheme

From previous section, we know that in order to collect all the information without loss, the total bits generated in each sensor must satisfy some constraints. Under these constraints, in this section, we further take the effect of channel conditions into account to consider the problem of how to allocate $\mathcal{B}_i$ among the sensor nodes to transmit in order to minimize the overall energy consumption.

Intuitively, if a sensor is very far away from the destination, channel condition is severely degraded. The transmitting rate is very slow even if high power is used, which is a waste of energy. Thus we tend to assign less task (bits) to this channel. For example, we only use this channel to transmit those information that is uniquely generated in this sensor and use other good channels to transmit those common information.

The overall energy consumed is:

$$
\begin{aligned}
E &= \sum_i P_{i,min} \frac{\mathcal{B}_i}{R_i} \\
&= \sum_i P_{i,min} \frac{\mathcal{B}_i}{\frac{1}{2} \log\left(1 + \frac{P_{i,min}/d_i^\gamma}{N}\right)} \\
&= \sum_i \frac{P_{i,min}}{\frac{1}{2} \log\left(1 + \frac{P_{i,min}/d_i^\gamma}{N}\right)} \mathcal{B}_i.
\end{aligned}
$$

Since the coefficient of $\mathcal{B}_i$ is only related to i and when $P_{i,min}$ and d_i are fixed, this coefficient is a constant. Let

$$
a_i = \frac{P_{i,min}}{\frac{1}{2} \log\left(1 + \frac{P_{i,min}/d_i^\gamma}{N}\right)}, \tag{3.18}
$$

then this problem can be formulated as follows:

$$
\min \sum_i a_i \mathcal{B}_i
$$
$$
\text{subject to } \sum_{i \in U} \mathcal{B}_i > H(W(U)|W(U^c))T, \ \forall U \subseteq \mathcal{N}. \tag{3.19}
$$

Theorem 3.4.1. *Without loss of generality, let $a_1 > a_2 > \cdots > a_n$, then the solution to the above optimization problem is $\boldsymbol{\mathcal{B}}^*$:*

$$\mathcal{B}_1^* = H(W_1|W_2\cdots W_n)T$$
$$\mathcal{B}_2^* = H(W_2|W_3\cdots W_n)T$$
$$\vdots$$
$$\mathcal{B}_{n-1}^* = H(W_{n-1}|W_n)T$$
$$\mathcal{B}_n^* = H(W_n)T. \tag{3.20}$$

Proof. We prove theorem 3.4.1 in two steps:

1. First, the solution (3.20) satisfies all the constraints in optimization problem (3.19).

2. Then, this solution can minimize the overall energy consumption.

Proof of Step 1: Define set $S_j = \{j, j+1, \ldots, n\}$, where $1 \le j \le n$, so $S_1 = \mathcal{N}$.

Then for any subset $U \subseteq \mathcal{N}$,

$$\sum_{j\in U} \mathcal{B}_j^* = \sum_{j\in U} H(W_j|W(S_{j+1}))T$$
$$= \sum_{j\in U} H(W_j|W(S_{j+1}\cap U), W(S_{j+1}\cap U^c))T$$
$$\overset{(a)}{\ge} \sum_{j\in U} H(W_j|W(S_{j+1}\cap U), W(U^c))T$$
$$\overset{(b)}{=} H(W(U)|W(U^c))T, \tag{3.21}$$

where (a) follows from conditioning reduces entropy, and (b) follows from the chain rule for entropy.

Proof of Step 2: We prove this part by induction. We start from $\mathcal{B}_n^* = H(W_n)T$.

$\mathcal{B}_1^* + \cdots + \mathcal{B}_n^* = H(W_1, \ldots, W_n)T$, which has already achieved the Slepian-Wolf bound. If there exists another set of solution, say, $\mathcal{B}_1', \ldots, \mathcal{B}_n'$, this set of solution must achieve the

bound as well. Since $\mathcal{B}_n^*$ is already the largest value that $\mathcal{B}_n$ can be, let $\mathcal{B}_n' = H(W_n)T - \delta$, where δ is a positive small value that does not affect the Slepian-Wolf conditions. Then this δ must be added to some $\mathcal{B}_i^*$, where $1 \le i \le n-1$. Then

$$
\begin{aligned}
E' &= a_1\mathcal{B}_1' + \cdots + a_n\mathcal{B}_n' \\
&= a_1\mathcal{B}_1^* + \cdots + a_{i-1}\mathcal{B}_{i-1}^* + a_{i+1}\mathcal{B}_{i+1}^* + \cdots + a_{n-1}\mathcal{B}_{n-1}^* \\
&\quad + a_i\mathcal{B}_i' + a_n\mathcal{B}_n' \\
&= E^* + (a_i - a_n)\delta.
\end{aligned}
$$

Since $a_i \ge a_n$, $(a_i - a_n)\delta \ge 0$ and thus $E' \ge E^*$. So $\mathcal{B}_n^*$ is optimum.

Given $\mathcal{B}_n^*$, $\mathcal{B}_{n-1}^* = H(W_{n-1}|W_n)T$ is the largest value that $\mathcal{B}_{n-1}$ can choose. By the same argument, it is obvious that $\mathcal{B}_{n-1}^*$ is the optimal value.

Similarly, we can prove that $\boldsymbol{\mathcal{B}}^*$ is the optimum solution. $\qquad\square$

Actually, the interpretation of this solution is quite straight-forward. The coefficient a_i is actually the energy consumption per transmitted bit. Thus, it is the most efficient if we transmit more bits to smaller a_i and less bits to larger a_i.

3.4.3   Numerical Results

From the previous section we know that $\boldsymbol{\mathcal{B}}^*$ is optimal in terms of minimizing the overall energy consumption. Furthermore, in this section, we will show how much better can this optimal scheme achieve than other schemes through the analytical results.

The reduction in the overall energy consumption using this bit allocation scheme depends significantly on the geometric and initial information distribution among all the sensors. In this section, we present some numerical results only in a specific case. Assume all the sensor nodes are uniformly distributed in a line and the sink node is at one end (Figure 3.2). All the channels are Gaussian.
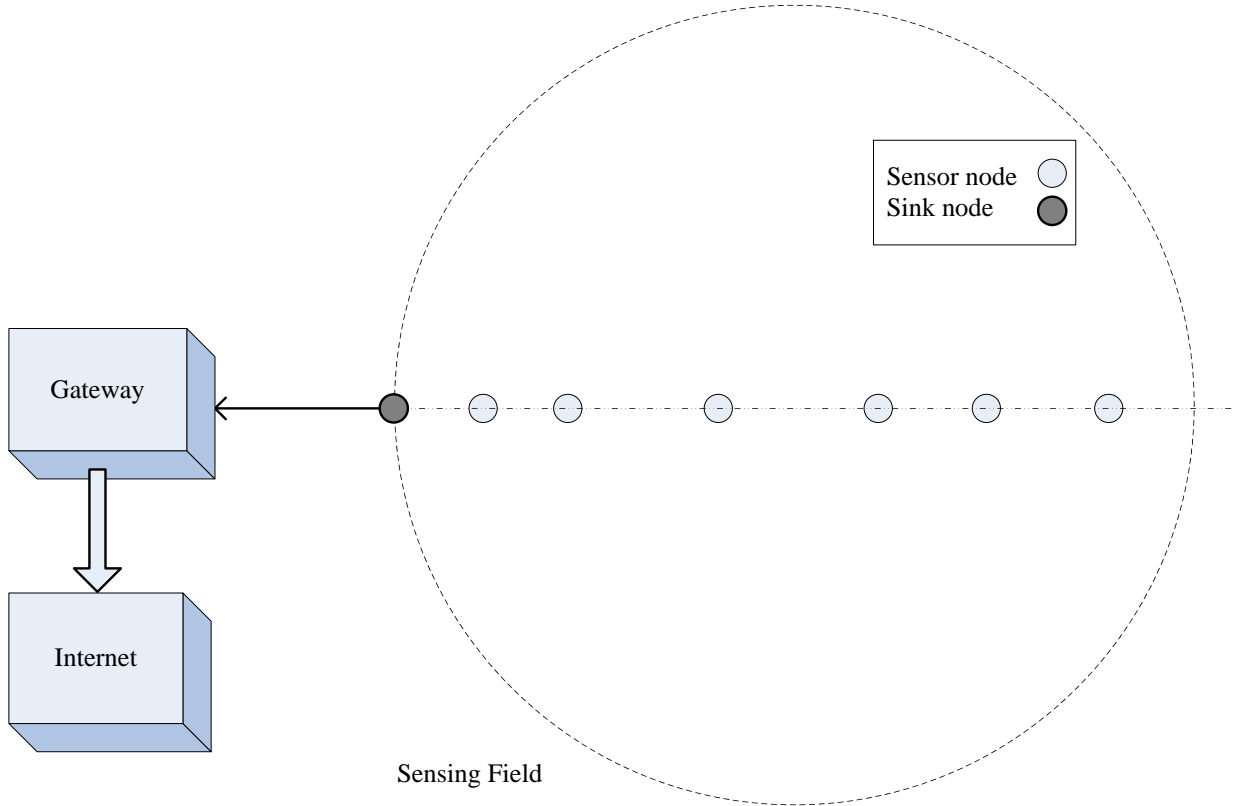
Figure 3.2: A sensor network with n sensors in a line and one sink node at one end.

The data collected by each sensor is $W_i = W + Z_i$, where $W \sim \mathcal{N}(0, N)$ is the common information and $Z_i \sim \mathcal{N}(0, N_i)$ is the noise, whose variance is quadratically proportional to the distance, i.e., $N_i = d_i^2$. Since both the sources and the channels are continuous, we use differential entropy when calculating the optimal bit in Theorem 3.4.1.

We compare the results obtained from our proposed scheme with an average bit allocation scheme, where each sensor transmits its unique information and $1/n$ of the common information. Figure 3.3 shows the overall energy consumption of our optimal bit allocation scheme and the average bit allocation scheme.

It can be seen that the proposed scheme always achieves lower overall energy consump-
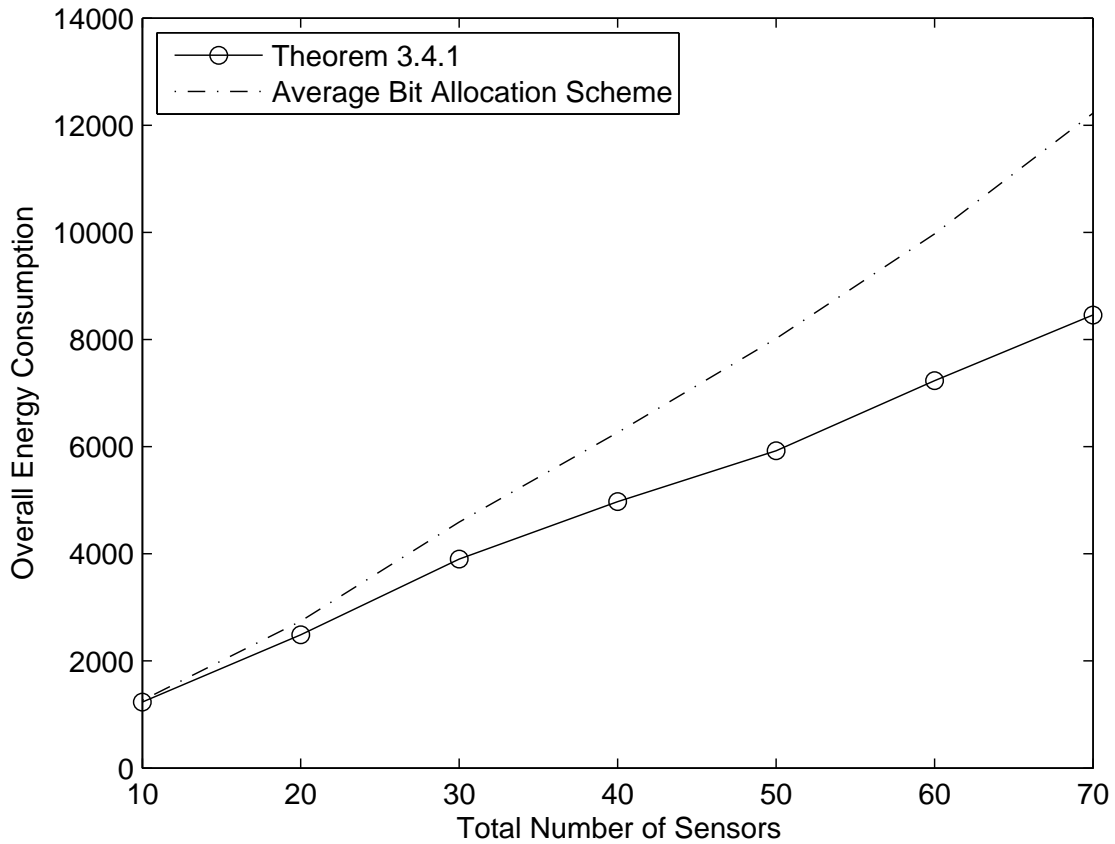
Figure 3.3: Comparison of the optimal bit allocation scheme proposed in Theorem 3.4.1 and the average bit allocation scheme in terms of overall energy consumption.

tion as has been proved. As the number of sensor nodes increases, the discrepancy of these two schemes broadens. When the number of nodes is 50, the proposed scheme is approximately 30% more energy efficient.

3.5 Summary

In this chapter, we come up with a bit allocation scheme that can minimize the overall energy consumption by combining both channel characteristics and source coding. We first prove that in AWGN channel with path loss, the optimal transmission mode is that every sensor uses its minimum transmitting power and cooperates with other sensors by TDMA. Then for the correlated source coding, the bit allocation scheme should considering the channel conditions as well in order to improve the energy efficiency. In brief, we assign higher rates (more bits) to those sensors with better channel conditions and assign lower rates (less bits) to those with worse channel conditions. In this way, we can achieve both reliable transmission and energy efficiency at the same time.

Chapter 4

Source-Channel Communication in Maximization of Network Lifetime

4.1 Introduction

In wireless sensor networks, energy consumption and network lifetime are the two primary concerns. In the previous chapter, we presented a bit allocation scheme that can minimize the overall energy consumption. However, if there is one sensor that consumes more energy than the rest of the sensors, this sensor is the first to die. And when any sensor dies, the whole network will lose a part of the information, so we say the network is not functional any more. Therefore, in this chapter, we are looking for a scheme that can average the energy consumption distribution among all the sensors, thus prolong the network lifetime, while at the same time, minimizing the overall energy consumption.

4.2 Problem Formulation

The problem can be formulated by the following lexicographic optimization:

$$\text{lexmin } (\max_i E_i, \sum_i E_i)$$

$$\text{subject to } \sum_{i \in U} \mathcal{B}_i > H(W(U)|W(U^c))T, \ \forall U \subseteq \mathcal{N}, \tag{4.1}$$

where $E_i = a_i \mathcal{B}_i$.

This optimization problem has a hierarchical structure, i.e. the first objective has the highest priority [27]. Under the condition that the highest energy consumption has been minimized, we try to further minimize the overall energy consumption to achieve lexicographic optimality.

To minimize E_i for some node i, we only need to reduce its corresponding $\mathcal{B}_i$, i.e., assign less bits that need to be transmitted by this sensor. Since we still desire the least overall energy consumption, we start from the optimal solution point $\boldsymbol{\mathcal{B}}^*$, find the node that consumes the most energy, reduce the bits of this node and reallocate these bits among the rest nodes, until the highest energy consuming node's transmitting bits cannot be reduced anymore. Thus this algorithm can achieve lexicographical optimality.

4.3 Lifetime Maximization Algorithm

Suppose $E_k, 1 \leq k \leq n$ is the largest, so $\mathcal{B}_k$ needs to be reduced. It can be seen that

$$\mathcal{B}_1^* + \mathcal{B}_2^* + \cdots + \mathcal{B}_k^* = H(W_1, W_2, \ldots, W_k | W_{k+1}, \ldots, W_n)T, \tag{4.2}$$

which is already the Slepian-Wolf bound. Thus, $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{k-1}$ need to increase the same bits in total in order to satisfy the Slepian-Wolf inequalities again.

However,

$$\mathcal{B}_k^* + \mathcal{B}_{k+1}^* + \cdots + \mathcal{B}_n^* > \text{Slepian-Wolf bound}, \tag{4.3}$$

so $\mathcal{B}_{k+1}, \ldots, \mathcal{B}_n$ may not need to change after $\mathcal{B}_k$'s decrease, which will be proved later in this section.

Suppose we first reduce $\mathcal{B}_k$ by 1 bit, so this bit needs to be added to $\mathcal{B}_1$, or $\mathcal{B}_2$, or $\ldots$, or $\mathcal{B}_{k-1}$. Since $a_1 > a_2 > \cdots > a_{k-1}$, adding this 1 bit to node $k-1$ will consume the least extra energy, thus is the most efficient. Therefore, when we need to transfer some bits from node k to the nodes before it, we prefer to add those bits to the node just before it first, and then two positions before it, and so on till the first node. [1] How many bits can we transfer at most to node $l, 1 \leq l \leq k-1$ without violating any of the Slepian-Wolf conditions?

Lemma 4.3.1. *If node k consumes the most energy, $\mathcal{B}_k^*$ needs to be reduced, and we add those bits to $\mathcal{B}_{k-1}^*$ until $\mathcal{B}_{k-1}$ is full, then to $\mathcal{B}_{k-2}^*$ until $\mathcal{B}_{k-2}$ is full, $\ldots$, and so on till $\mathcal{B}_1$ is full.[2] In this process, we can transfer to $\mathcal{B}_l, 1 \leq l \leq k-1$, at most*

$$I(W_l; W_k | W_{l+1}, \ldots, W_{k-1}, W_{k+1}, \ldots, W_n)T \tag{4.4}$$

bits before it is full.

Proof. Here, we only prove the case when $l = k-1$. For the rest of ls, the proof is trivial with the same argument. Now, suppose $\mathcal{B}_k^*$ has been reduced to $\hat{\mathcal{B}}_k$, and $\mathcal{B}_{k-1}^*$ has been increased to $\hat{\mathcal{B}}_{k-1}$, and all the rest of $\mathcal{B}_i^*$, $i \in \{1, 2, \ldots, k-2, k+1, \ldots, n\}$ remain the same. Therefore, the only possibility that Slepian-Wolf condition might not hold is that the addition of any subset of $\{\mathcal{B}_1^*, \mathcal{B}_2^*, \ldots, \mathcal{B}_{k-2}^*\}$ and $\hat{\mathcal{B}}_k$ might not be greater than the Slepian-Wolf bound. Thus, we are looking for the largest $d_k^{k-1} = \mathcal{B}_k^* - \hat{\mathcal{B}}_k$, which is the bits we can transfer at most from $\mathcal{B}_k^*$ to $\mathcal{B}_{k-1}^*$.

$\forall V \subseteq \{1, \ldots, k-2\}$,

$$\sum_{i \in V} \mathcal{B}_i^* + \mathcal{B}_k^* = \sum_{i \in V} H(W_i | W_{i+1}, \ldots, W_n)T + H(W_k | W_{k+1}, \ldots, W_n)T. \tag{4.5}$$

---

[1] Node order is based on the index.

[2] Here, $\mathcal{B}_l$ is "full" means that it achieves its largest value without violating any Slepian-Wolf condition.

However, after $\mathcal{B}_k^*$ has been reduced to $\hat{\mathcal{B}}_k$, according to Slepian-Wolf bound,

$$\sum_{i \in V} \mathcal{B}_i^* + \hat{\mathcal{B}}_k = H(W(V), W_k | W(V^c), W_{k-1}, W_{k+1}, \ldots, W_n)T. \tag{4.6}$$

Thus,

$$d_k^{k-1} = \mathcal{B}_k^* - \hat{\mathcal{B}}_k \tag{4.7}$$

$$= \sum_{i \in V} H(W_i | W_{i+1}, \ldots, W_n)T$$

$$+ H(W_k | W_{k+1}, \ldots, W_n)T$$

$$- H(W(V), W_k | W(V^c), W_{k-1}, W_{k+1}, \ldots, W_n)T \tag{4.8}$$

$$\overset{(a)}{=} \sum_{i \in V} H(W_i | W_{i+1}, \ldots, W_n)T$$

$$+ H(W_k | W(V^c), W_{k-1}, W_{k+1}, \ldots, W_n)T$$

$$+ I(W_k; W(V^c), W_{k-1} | W_{k+1}, \ldots, W_n)T$$

$$- H(W(V), W_k | W(V^c), W_{k-1}, W_{k+1}, \ldots, W_n)T \tag{4.9}$$

$$\overset{(b)}{\geq} H(W(V) | W(V^c), W_{k-1}, \ldots, W_n)T$$

$$+ H(W_k | W(V^c), W_{k-1}, W_{k+1}, \ldots, W_n)T$$

$$- H(W(V), W_k | W(V^c), W_{k-1}, W_{k+1}, \ldots, W_n)T$$

$$+ I(W_k; W(V^c), W_{k-1} | W_{k+1}, \ldots, W_n)T \tag{4.10}$$

$$\overset{(c)}{=} I(W_k; W(V^c), W_{k-1} | W_{k+1}, \ldots, W_n)T \tag{4.11}$$

$$\overset{(d)}{\geq} I(W_k; W_{k-1} | W_{k+1}, \ldots, W_n)T, \tag{4.12}$$

where (a) follows from $I(X; Y) = H(X) - H(X|Y)$, (b) follows from (3.21), (c) follows from $H(X) + H(Y|X) = H(X, Y)$, and (d) follows from the chain rule of mutual information.

Note that expression (4.12) is for the case when $V = \{1, \ldots, k-2\}$, which is the tightest bound. Thus, d_k^{k-1} is at most $I(W_k; W_{k-1} | W_{k+1}, \ldots, W_n)T$, and this completes the proof. $\qquad \square$

Remarks

There are two remarks.

1. After transferring $I(W_k; W_{k-1}|W_{k+1}, \ldots, W_n)T$ bits from $\mathcal{B}_k^*$ to $\mathcal{B}_{k-1}^*$, $\mathcal{B}_k$ becomes

$$H(W_k|W_{k+1}, \ldots, W_n)T - I(W_k; W_{k-1}|W_{k+1}, \ldots, W_n)T$$
$$= H(W_k|W_{k-1}, W_{k+1}, \ldots, W_n)T, \tag{4.13}$$

and $\mathcal{B}_{k-1}$ becomes

$$H(W_{k-1}|W_{k+1}, \ldots, W_n)T, \tag{4.14}$$

which means that $\mathcal{B}_k^*$ and $\mathcal{B}_{k-1}^*$ have switched their positions in the solution (3.20). If $\mathcal{B}_{k-1}$ is full and node k still consumes the highest energy, we need to further transfer bits from $\mathcal{B}_k$ to $\mathcal{B}_{k-2}^*$, which will result in that $\mathcal{B}_k^*$ and $\mathcal{B}_{k-2}^*$ switch positions, and so on until $\mathcal{B}_k^*$ has been moved to the first place.

2. Bits that can be reduced at most from $\mathcal{B}_k^*$ are

$$H(W_k|W_{k+1}, \ldots, W_n)T - H(W_k|W_1, \ldots, W_{k-1}, W_{k+1}, \ldots, W_n)T$$
$$= I(W_k; W_1, \ldots, W_{k-1}|W_{k+1}, \ldots, W_n)T, \tag{4.15}$$

which exactly equals to $d_k^{k-1} + d_k^{k-2} + \cdots + d_k^1$ by the chain rule.

Now we can get back to our previous question: If $\mathcal{B}_k^*$ needs to be reduced, we know that

$$\mathcal{B}_k^* + \mathcal{B}_{k+1}^* + \cdots + \mathcal{B}_n^* > \text{Slepian-Wolf bound}, \tag{4.16}$$

should $\mathcal{B}_{k+1}, \ldots, \mathcal{B}_n$ increase as well or can they remain still?

To answer this question, we propose the following lemma.

Lemma 4.3.2. *Even if $\mathcal{B}_k$ has been reduced to the bound:*

$$H(W_k|W_1, \ldots, W_{k-1}, W_{k+1}, \ldots, W_n)T, \tag{4.17}$$

$\mathcal{B}_{k+1}, \ldots, \mathcal{B}_n$ *can remain the same and satisfy Slepian-Wolf conditions.*

Proof. Among all the summations between $\mathcal{B}_k^*$ and any subset of $\{\mathcal{B}_{k+1}^*, \ldots, \mathcal{B}_n^*\}$, $\mathcal{B}_k^* + \mathcal{B}_{k+1}^*$ is the closest to the bound. Thus if $\mathcal{B}_k^*$ is reduced the most, while $\mathcal{B}_k + \mathcal{B}_{k+1}^*$ is still above the bound, so would the rest of the summations be .

$$
\begin{aligned}
\mathcal{B}_{kmin} &+ \mathcal{B}_{k+1}^* \\
=&H(W_k|W_1,\ldots,W_{k-1},W_{k+1},\ldots,W_n)T+ \\
&H(W_{k+1}|W_{k+2},\ldots,W_n)T \\
\overset{(a)}{\geq}&H(W_k|W_1,\ldots,W_{k-1},W_{k+1},\ldots,W_n)T+ \\
&H(W_{k+1}|W_1,\ldots,W_{k-1},W_{k+2},\ldots,W_n)T \\
\overset{(b)}{=}&H(W_k,W_{k+1}|W_1,\ldots,W_{k-1},W_{k+2},\ldots,W_n)T, \quad (4.18)
\end{aligned}
$$

where (a) follows from conditioning reduces entropy, and (b) follows from the chain rule of entropy. Since (4.18) is the Slepian-Wolf bound, the proof is complete. $\qquad\square$

Now we can conclude the above explanations with Algorithm 1.

Algorithm 1 Algorithm to maximize network lifetime

1: Calculate the energy consumption for each node $i = 1, 2, \ldots, n$ to find the highest one, whose index is denoted by k_1. If there are multiple nodes consuming highest energy, let k_1 be the smallest index. Let k_2 denote the node index with the second highest energy consumption;

2: If $k_1 = 1$, stop since $\mathcal{B}_1^*$ is already the smallest and cannot be reduced anymore;

3: If $k_1 \neq 1$, and $\mathcal{B}_{k_1} > H(W_{k_1}|W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T$, let

$$\delta = \min\{\mathcal{B}_{k_1} - \frac{a_{k_2}}{a_{k_1}}\mathcal{B}_{k_2}, \frac{a_{k_1}\mathcal{B}_{k_1} - a_{k_1-1}\mathcal{B}_{k_1-1}}{a_{k_1} + a_{k_1-1}},$$
$$\mathcal{B}_{k_1} - H(W_{k_1}|W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T\}, \tag{4.19}$$

transfer δ bits from $\mathcal{B}_{k_1}$ to $\mathcal{B}_{k_1-1}$;

If $H(W_{k_1}|W_{k_1-2}, W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T < \mathcal{B}_{k_1} \leq H(W_{k_1}|W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T$, let

$$\delta = \min\{\mathcal{B}_{k_1} - \frac{a_{k_2}}{a_{k_1}}\mathcal{B}_{k_2}, \frac{a_{k_1}\mathcal{B}_{k_1} - a_{k_1-2}\mathcal{B}_{k_1-2}}{a_{k_1} + a_{k_1-2}},$$
$$\mathcal{B}_{k_1} - H(W_{k_1}|W_{k_1-2}, W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T\}, \tag{4.20}$$

transfer δ from $\mathcal{B}_{k_1}$ to $\mathcal{B}_{k_1-2}$;

$\vdots$

If $\quad H(W_{k_1}|W_1, \ldots, W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T \quad < \quad \mathcal{B}_{k_1} \quad \leq$ $H(W_{k_1}|W_2, \ldots, W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T$, let

$$\delta = \min\{\mathcal{B}_{k_1} - \frac{a_{k_2}}{a_{k_1}}\mathcal{B}_{k_2}, \frac{a_{k_1}\mathcal{B}_{k_1} - a_1\mathcal{B}_1}{a_{k_1} + a_1},$$
$$\mathcal{B}_{k_1} - H(W_{k_1}|W_1, \ldots, W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T\}, \tag{4.21}$$

transfer δ from $\mathcal{B}_{k_1}$ to $\mathcal{B}_1$;

If $\mathcal{B}_{k_1} = H(W_{k_1}|W_1, \ldots, W_{k_1-1}, W_{k_1+1}, \ldots, W_n)T$, stop since $\mathcal{B}_{k_1}$ is already the smallest;

4: Go to step 1.

Remarks

1. When there are multiple nodes consuming the highest energy, we prefer to check the one with the smallest index first. This is because of the chance that this index is 1. In this case, reducing bits of other nodes will only waste more overall energy without prolonging lifetime since the first node is always the network's bottleneck.

2. The value of δ is chosen based on (4.19)-(4.21), where the first and second terms are bits that $\mathcal{B}_{k_1}$ can be reduced at most until this sensor's energy consumption is the same as the second highest one, and the third term is the total reduced bits when $\mathcal{B}_{k_1}$ can be decreased to the Slepian-Wolf bound.

3. After this algorithm, the final result will be that the node that consumes the most energy is either the one that has the least bits to transmit, i.e.,

$$\mathcal{B}_i = H(W_i | W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_n), \tag{4.22}$$

or the first node.

4.4 Numerical Results

Under the same distribution of sensor nodes and channel conditions as in Section 3.4.3. We still compare the results of the proposed algorithm with the average bit allocation scheme, where each sensor transmits its unique information and $1/n$ of the common information. Figure 3.2 illustrates the maximum energy consumption among the sensor nodes using the proposed algorithm and using the average bit allocation scheme.

The lower the maximum energy consumption is, the longer the network can live. It can be seen that the proposed algorithm effectively reduces the maximum energy consumption and hence prolongs the network life time. When the number of nodes is 50, the proposed scheme reduces the peak energy consumption by approximately 30%. Therefore, our scheme has compelling advantages.
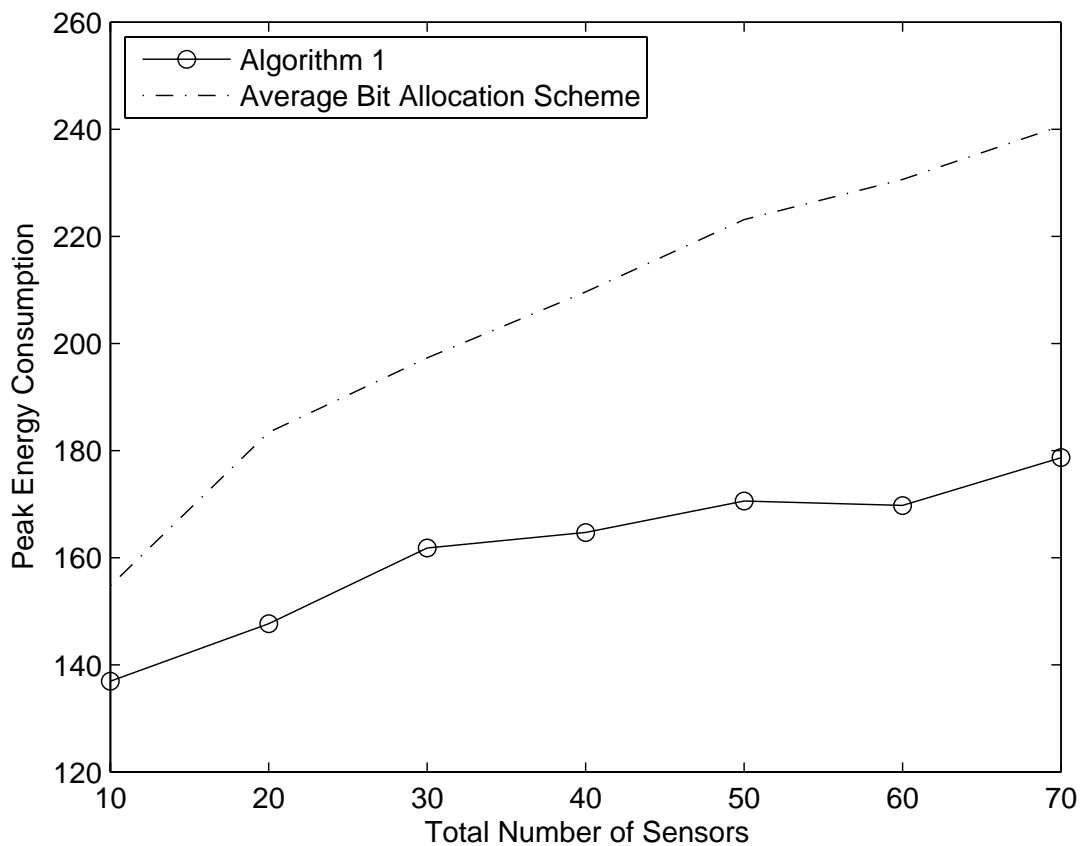
Figure 4.1: Comparison of the peak energy consumption by adopting Algorithm 1 and by adopting the average bit allocation scheme.

4.5 Summary

This chapter provides a network lifetime prolonging algorithm that can average the distribution of energy consumption among all the sensors. This algorithm is designed to meet two requirements: first extending the network lifetime and then reducing the overall energy consumption. Therefore, this algorithm proceeds as follows: we first locate the node that consumes the largest energy based on the bit allocation scheme that can minimize the overall energy consumption from the previous chapter, then reduce its transmission

bits until either this node consumes the same amount of energy as the second largest energy consumption node does or its transmission bits reach a lower bound constrained by Slepian-Wolf coding theorem, reallocate these bits to the nodes with worse channel conditions. Both proofs and analytical results are presented at last.

Chapter 5

Conclusion and Future Work

In this chapter, we conclude this thesis and give suggestions on the possible future work along this path.

5.1   Conclusion

In wireless sensor networks, energy efficiency is one of the major concerns. Towards this end, this thesis attempts to combine both source and channel coding to come up with the optimal bit allocation scheme in multi-source single-sink wireless sensor networks.

If the overall energy consumption is the goal to be minimized, in AWGN channel with path loss, this work shows that the overall energy consumption can be minimized under the condition that each source transmits with its minimum power and in TDMA mode with other sensors. In addition, considering the characteristics of correlated source coding, the Slepian-Wolf coding theorem is applied for resource efficiency. Therefore, we derive a closed form bit allocation scheme to minimize the overall energy consumption by allocating more bits to the nodes with better channel conditions and less bits to the nodes with worse channel conditions.

Based on the above solution, we further extend to maximize the network lifetime, which is the time when all the sensors are functional. In this case, the first objective is to maximize the network lifetime and upon achieving this objective, we minimize the overall energy consumption. We develop a heuristic algorithm to average the distribution of energy consumption among all sensors and demonstrate the superiority of our algorithm both by proofs and simulation results.

5.2   Future Work

According to the extensive applications of sensor networks and their intrinsic restrictions, there are tremendous fascinating topics that are still yet to be investigated and improved, e.g. time synchronization of signals, hardware design to support low power sensors, etc. Among them, some questions related to our work are of particular interest in the future.

One natural extension of our network topology is to incorporate intermediate nodes, i.e., relays. Relays can help the communication from the source to the destination [28], [29]. Then the intermediate nodes can function either purely as relays or both sources and relays. More generally, in a multi-source multi-destination multi-relay sensor network [30], [31], what is the optimal source channel coding strategy in terms of energy efficiency? Can our results be extended?

To approach the above mentioned problem, we first need to simplify the decoding method for relay channel. By far, the largest achievable rate region for the multi-source case can only be realized by the regular encoding/irregular encoding and backward decoding [32], [33]. However, backward decoding usually causes excessive delays, thus is difficult to implement. [34] proposes a new encoding scheme, namely offset encoding, to achieve the same rate region in conjunction with the use of sliding-window decoding, which can reduce the delays significantly. Unfortunately, it is proved that only in some special network topologies, can the largest rate region be achieved by this new scheme. In this regard, our subsequent research will focus on extending the offset encoding and sliding-window decoding strategy in a general network frame to achieve the rate region in [31].

If the network topology is not fixed, there would be two major issues. First, what is the optimal node placement to obtain the coverage as well as the energy efficiency [22], [23]? Then, after the sensor nodes have been deployed, what multi-hop strategy is optimal? For example, for every single sensor, should it be the pure source only or can it be a relay as well? Also, what is the best routing strategy in terms of energy efficiency [20], [35]?

Bibliography

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393 – 422, 2002. ix, 1, 3

[2] G. Pottie and W. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, 2000. 1

[3] "Special issue on collaborative signal and information processing in microsensor networks," *IEEE Signal Processing Magazine*, March 2002. 1

[4] B. Sadler, "Fundamentals of energy-constrained sensor network systems," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8 Part 2, pp. 17–35, 2005. 1

[5] K. Yuen, B. Liang, and L. Baochun, "A distributed framework for correlated data gathering in sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, pp. 578–593, 2008. 1

[6] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, (New York, NY, USA), pp. 88–97, ACM, 2002. 1

[7] M. Li and Y. Liu, "Underground coal mine monitoring with wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 5, no. 2, p. 10, 2009. 1

[8] J. Lynch, K. Loh, J. Lynch, and K. Loh, "A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring," 2008. 1

[9] A.Knaian, "A wireless sensor network for smart roadbeds and intelligent transportation systems," Master's thesis, MIT, May 2000. 1

[10] D. Cook and S. Das, *Smart environments: technologies, protocols, and applications.* Wiley-Interscience, 2005. 1

[11] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *Information Theory, IEEE Transactions on*, vol. 19, pp. 471–480, Jul 1973. 2, 18

[12] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 4, 2004. 2

[13] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC*, vol. 1, pp. 331–335, Citeseer, 1995. 2

[14] G. Mergen, Q. Zhao, and L. Tong, "Sensor networks with mobile access: Energy and capacity considerations," *IEEE Transactions on Communications*, vol. 54, no. 11, pp. 2033–2044, 2006. 2

[15] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science*, pp. 162–177, Springer, 2003. 3

[16] A. Roumy and D. Gesbert, "Optimal matching in wireless sensor networks," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, pp. 725–735, Dec. 2007. 3

[17] M. Zorzi and R. Rao, "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: multihop performance," *IEEE transactions on Mobile Computing*, vol. 2, no. 4, pp. 337–348, 2003. 3

[18] M. Zorzi and R. Rao, "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance," *IEEE transactions on Mobile Computing*, vol. 2, no. 4, pp. 349–365, 2003. 3

[19] A. Michail and A. Ephremides, "Algorithms for routing session traffic in wireless ad-hoc networks with energy and bandwidth limitations," in *Proceedings of 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2001. 4

[20] C. Schurgers and M. Srivastava, "Energy efficient routing in wireless sensor networks," in *IEEE Military Communications Conference, 2001. MILCOM 2001. Communications for Network-Centric Operations: Creating the Information Force*, vol. 1, 2001. 4, 47

[21] S. Bandyopadhyay and E. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *IEEE INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3. 4

[22] K. Kar and S. Banerjee, "Node placement for connected coverage in sensor networks," Citeseer. 4, 47

[23] Y. Hou, Y. Shi, H. Sherali, and S. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2579–2590, 2005. 4, 47

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991. 5, 10, 15, 27

[25] R. Ahlswede, "Multi-way communication channels," in *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971*, p. 23, Akadémiai Kiadó, 1973. 17

[26] H. Liao, *Multiple access channels*. PhD thesis, Hawaii University, Honolulu, 1972. 17

[27] M. Ehrgott, *Multicriteria optimization*. Springer Verlag, 2005. 36

[28] E. Van Der Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, pp. 120–154, 1971. 46

[29] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979. 46

[30] L. Xie and P. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1348–1358, 2005. 46

[31] L. Xie and P. Kumar, "Multisource, multidestination, multirelay wireless networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3586–3595, 2007. 46

[32] F. Willems and E. Van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 313–327, 1985. 46

[33] C. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems usingbackward decoding," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1160–1165, 1989. 46

[34] L. Sankar, G. Kramer, and N. Mandayam, "Offset encoding for multiple-access relay channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3814–3821, 2007. 46

[35] D. Baghyalakshmi, J. Ebenezer, and S. Satyamurty, "Low latency and energy efficient routing protocols for wireless sensor networks," pp. 1 –6, jan. 2010. 47