University of California
Los Angeles

On Fundamental Limits of Scalable Sensor Networks

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

by

Ameesh Niranjan Pandya

2004

The dissertation of Ameesh Niranjan Pandya is approved.

Kirby Baker

William J. Kaiser

Kung Yao

Gregory J. Pottie, Committee Chair

University of California, Los Angeles

2004

To my mother, Jayshree ...

who inspired me to pursue a path that

has now culminated in a Ph.D. dissertation

Table of Contents

List of Figures

List of Tables

Acknowledgments

My greatest and heartfelt gratitude goes to my advisor, mentor, and friend Professor Greg Pottie. I must have had divine blessings to have him as an advisor. As a researcher, he is brilliant communications theorist with incredible vision and has an uncanny knack for cutting to the core of any problem, separating the truly relevant wheat from the obfuscating chaff. He is always motivating, encouraging, and inspiring. It is hard to imagine having done my Ph.D. without his insights, assistance, enthusiasm, and flexibility. I also sincerely thank Professor Kung Yao, Professor Bill Kaiser, and Professor Kirby Baker for taking time to serve on my dissertation committee. Each of them provided me with invaluable feedback and encouragement.

I would like to extend special gratitude to Professor Kung Yao for his kind help and guidance in the job hunt. I am grateful to all my teachers at UCLA for providing the most rewarding intellectual experience. I pray God to provide me with sufficient strength to do justice to all that they have taught me. Without the constant support of my colleagues and friends my journey to Ph.D. would not have been smooth and entertaining. They are too numerous to mention here. In particular, I am indebted to Aman Kansal, Aditya Ramamoorthy, Arash Behzad, Hong Chen, and Jibing Wang for sharing their research experiences and insights with me.

My experiences in UCLA have been the most pleasant and memorable. I could never forget the very first day in UCLA for numerous reasons. But the most fruitful of all was meeting an intelligent and carefree guy from Mathematics Department, Kingshook Biswas. My enormous debt of gratitude can hardly be repaid to my good friend Kingshook for encouraging me to play Tabla, read

comics and above all have good food. Well jokes apart, he is true friend and we had numerous brainstorming, intellectual, and interesting discussions.

Most of all, it is with great pride that I acknowledge the fundamental role that my parents, Niranjan and Jayshree Pandya, and my dearest sister, Reena, have played in shaping all that I have accomplished in life. They provided me with abundant love and support. This dissertation would certainly have not been possible without the blessings of them, my loving grand parents - Ramshanker Pandya, Godavari Pandya, Shantilal Shah and Kanchan Shah - and Lord Hanuman.

Vita

1978	Born, Pontiac, Michigan, USA.
1998	Sangeet Visharad in Music (Tabla), Ghandharva Mahavidhyalaya (Ahmedabad), Oct. 1998. Sangeet Visharad in Music (Tabla), Akhil Bhartiya Ghandharva Mahavidhyalaya Mandal, Nov. 1998.
1999	B.E. (Electronics and Communications), Gujarat University, Ahmedabad, India.
1998 – 1999	Research Assistant, Data and Communications Technique Division, Space Applications Center - Indian Space Research Organization (SAC-ISRO), Ahmedabad, India.
1999	Lecturer, Nirma Institute of Technology, Gujarat University.
1999 – 2000	Systems Programmer–Developer, Computer Associates, Cincinnati, OH.
2000 – 2001	Teaching Assistant, Electrical Engineering Department, University of California, Los Angeles (UCLA), CA.
2001	M.S. (Electrical Engineering), UCLA.
2001 – Present	Graduate Student Researcher, Electrical Engineering Department, UCLA.
2004	Ph.D. (Electrical Engineering), UCLA.

Publications

1. Ameesh Pandya and Greg Pottie. Bounds on Achievable Rates for Cooperative Channel Coding. In *Proceedings of The* 38^{th} *Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2004.

2. Ameesh Pandya, Huiyu Luo, and Greg Pottie. Spatial Fidelity and Estimation in Sensor Networks. In *Proceedings of The* 38^{th} *Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2004.

3. Ameesh Pandya, Aman Kansal, Greg Pottie, and Mani Srivastava. Lossy Source Coding of Multiple Gaussian Sources: m-helper problem. In *Proceedings of IEEE Information Theory Workshop (ITW)*, San Antonio, TX, October 2004.

4. Ameesh Pandya, Aman Kansal, Greg Pottie, and Mani Srivastava. Fidelity and Resource Sensitive Data Gathering. In *Proceedings of the* 42^{nd} *Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2004.

5. Ameesh Pandya, Huiyu Luo, and Greg Pottie. Characterizing Sensor Networks. Recent Results Poster Session. In *International Symposium on Information Theory*, Chicago, IL. June - July 2004.

6. Ameesh Pandya and Greg Pottie. QoS in Ad Hoc Networks. In *Proceedings of IEEE Vehicular Technology Conference*, October 2003.

7. Greg Pottie, Huiyu Luo, and Ameesh Pandya. *Encyclopedia of Sensors*, chapter Sensor Network Information Theory. Edited by Craig Grimes. American Scientific Publishers. To appear, 2005.

8. Aman Kansal, Ameesh Pandya, Mani Srivastava, and Greg Pottie. Throughput and Delay in Networks with Controlled Mobility. Technical report, Center For Embedded Networked Sensing (CENS), UCLA. Submitted to IEEE International Conference on Communications (ICC) 2005. August 2004.

9. Ameesh Pandya, Aman Kansal, Greg Pottie, and Mani Srivastava. Bounds on the Rate-Distortion of Cooperative Multiterminal Gaussian Sources. Technical report # 27, Center For Embedded Networked Sensing (CENS), UCLA, 2003.

10. Huiyu Luo, Ameesh Pandya, and Greg Pottie. Detection Fidelity in Distributed Wireless Sensor Networks. Technical Report # 20, Center For Embedded Networked Sensing (CENS), UCLA, 2003.

11. Ameesh Pandya and Greg Pottie. On Scalability and Source/Channel Coding Decoupling in Large Scale Sensor Networks. Technical Report # 17, Center For Embedded Networked Sensing (CENS), UCLA, 2003.

12. Ameesh Pandya, Aman Kansal, Greg Pottie, and Mani Srivastava. Multiterminal Data Gathering Networks. Preprint. To be submitted to journal. October 2004.

13. Ameesh Pandya, Huiyu Luo, and Greg Pottie. Characterizing Sensor Networks with Spatial Fidelity. Preprint. To be submitted to journal. October 2004.

14. Behnam Rezaei, Ameesh Pandya, Vwani Roychowdhury, and Greg Pottie. A Graph Theoretical Approach to Decentralize Data Fusion. Technical report, Center for Embedded Networked Sensing, UCLA, 2004. Preprint.

15. Nima Sarshar, Behnam Rezaei, Ameesh Pandya, Vwani Roychowdhury, and Greg Pottie. Scalability of Wireless Capacity: A Renormalization Approach. Preprint. 2004.

Abstract of the Dissertation

On Fundamental Limits of Scalable Sensor Networks

by

Ameesh Niranjan Pandya

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2004

Professor Gregory J. Pottie, Chair

The goal of this thesis is to lay out the information theoretic limits and develop the cross layer understanding, from the physical layer to network layer, involved in the design of the data gathering systems such as sensor networks. Such decentralized information processing systems are finding applications ranging from habitat monitoring to entertainment.

This thesis begins with the *scalability* question in wireless ad hoc networks, i.e. behavior as network size grows. In particular, the otherwise inherently non-scalable ad hoc networks can be made scalable either by encouraging local communications or by providing extra resources such as bandwidth to each node in the network. Both static and controllably mobile networks are considered. Next, the design criteria, algorithmic techniques and information theoretic analysis for scalable sensor networks are developed.

The implications for sensor network architecture are then analyzed by introducing the concept of spatial fidelity as the design criterion. The algorithms for the decentralized data fusion to facilitate the acquisition, processing and dissem-

ination of information in sensor networks are presented. Design questions such as optimal sensor density and cooperation sensors strategies are also addressed while analyzing the distortion/density tradeoff. The suboptimal but scalable interpolation strategy is stated in order to explore network density tradeoffs in the presence of measurement error.

Fundamental performance limits are then discussed for the data gathering systems for Gaussian sources and channels. The rate distortion bounds are derived for such systems both with and without considering the sensing channel. The bounds for the m-helper, Berger-Tung and CEO coding systems are derived. An upper bound on a cooperative coding system comprised of two transmitters and two receivers is then derived. Here, the different permutations of cooperation among the pairs of transmitters and receivers are exploited, where the data from each transmitter is meant for both the receivers.

Finally, moving above physical layer, a higher layer abstraction with practical implementations in the above described networks is presented. Optimization problems and heuristic algorithms that guarantee the quality of service (QoS) in wireless sensor and ad hoc networks are formulated for this purpose.

CHAPTER 1

Introduction

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. - C. E. Shannon.

1.1 Towards Decentralized Wireless Era

Wireless communications is attributed with a rich history ranging from smoke signals to modern era Autonomous Intelligent Networks [55, 107, 207]. The world has significantly advanced from single channels to networks, and from the analog age to the digital age. Wireless networks consist of nodes communicating with each other over a wireless channel. In some networks such as cellular networks, only the last hop is wireless with the wired backbone. For others, such as multihop radio networks, ad hoc networks, and sensor networks, all links are wireless. These types of networks are the focus of the thesis. It should be noted that the ad hoc and sensor networks are resource constrained (eg., energy, bandwidth).

1.1.1 Emergence of Wireless Ad hoc networks

As described above, wireless ad hoc networks are free from having any wired links. Ad hoc networks consist of a collection of mobile and static nodes lacking any

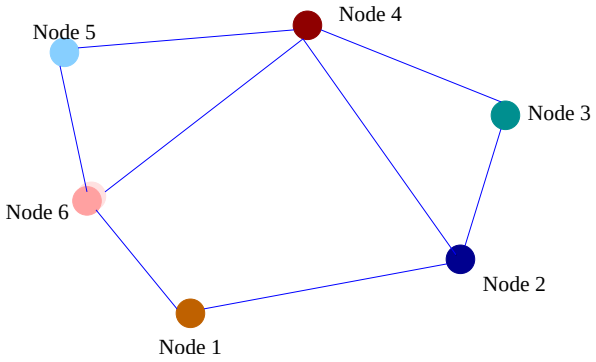


Figure 1.1: Wireless Ad hoc network.

centralized administration such as base stations. Each node in the network also acts as a router. Figure 1.1 depicts one possible topology of an ad hoc network. The traditional major issue in ad hoc networks is to cope with frequently changing topology due to addition of nodes or changes in the position of mobile nodes. One popular means of self organizing is to deal with these changes is to divide the global network into local networks by clustering [15, 91, 108, 142, 235].

Various routing protocols have been suggested for ad hoc networks such as destination-sequenced distance vector routing (DSDV) [179], dynamic source routing (DSR) [32, 118], ad hoc on-demand distance vector routing (AODV) [180], and zone routing [104]. No routing protocol outperforms others for all the likely scenarios. The performance of these protocols varies under different traffic and mobility models [30, 33, 58]. The most important criterion of ad hoc networks is to provide quality of service (QoS). This means that the routing algorithms should adapt to changes in wireless link quality, propagation path loss, fading, multiuser interference, power usage, and topological changes. Another issue that is widely studied for ad hoc networks is power control [18, 43, 121, 218, 219, 236]. This is necessary because of the limited battery power available to the nodes, and for reasons of interference management.

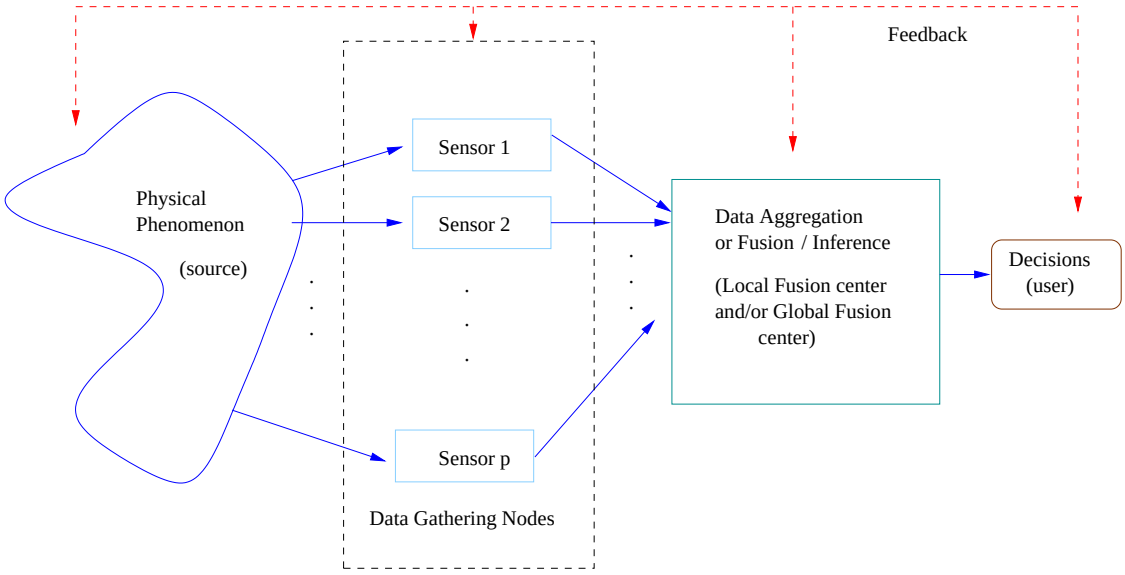


Figure 1.2: Data Gathering Sensor Network

1.1.2 Emergence of Wireless Sensor networks

Recent advances and rapid convergence of digital circuitry, wireless communications, and Micro Electro-Mechanical Systems (MEMS) have established the feasibility of sensor networks for a variety of applications [203, 181, 70, 190, 234, 224, 226, 1]. Apart from the use in defense and industrial process monitoring, sensor networks find several other applications in the area of education [224, 226], science [39, 147], arts and entertainment [34]. Sensor networks consisting of large number of sensor nodes are potentially randomly deployed either inside a physical phenomenon or near it to observe and measure it as shown in Figure 1.2. Thus, the algorithms and protocols designed for sensor networks should be self-organizing [222].

Sensor networks are distributed (in sensing and processing) [70] in nature. Distributed sensing allows one to place the sensors closer to the phenomena being monitored than if only a single sensor were used yielding higher SNR, and

improving opportunities for line of sight. This provides robustness to environmental obstacles. Finite energy budget in sensor networks motivates distributed processing. Communications is a key energy consumer as the radio signal power in sensor networks drops off with the distance as r^4 [181] due to ground reflections from short antenna heights. Therefore, one wants to process data as much as possible inside the network to reduce the number of bits transmitted, particularly over longer distances.

Compared to a single sensor platform, a network has the advantages of diversity (different sensors offer complementary viewpoints), and redundancy (reliability and increased resolution of the measured quantity) [8, 148, 149]. Sensor networks have another important property of cooperative processing. The sensors deployed to gather data have local processing capabilities to transmit only the required data to the global fusion center. This can be seen as the transmission of some function of the raw data so as to decrease the required rate [92]. The protocols and algorithms proposed and designed for wireless ad hoc networks are not necessarily suited for sensor networks. The differences between the two are pointed out in the next section. As in ad hoc networks, sensor networks are also resource constrained. The main issues, however, are scalability and sustainability. The data gathering system in Figure 1.2 could be summarized as follows. The sensor nodes observe and detect the phenomenon and locally process the observed data. They then transmit the processed data to the central or global fusion unit. During the whole process the power is consumed during sensing, communications, and data processing. However, the communication cost typically dominates the data processing in micro sensor networks. Some power efficient protocols proposed for the sensor networks are found in [251, 45, 220, 106, 105, 227]. Embedded sensor networks are a relatively new research topic and hence, there are many open research problems, as discussed in [10]. It is also possible for the fusion

center to provide the feedback so as to improve the quality of reconstruction.

The consequences of mobility in sensor networks are considered in [122]. Mobility in sensor networks could be either imparted by introducing robots or providing infrastructure. As will be discussed, the presence of mobile agents has a large impact on network scalability.

1.2 Why Sensor networks are not Ad hoc networks?

Although the sensor networks are inspired from wireless ad hoc networks, they are very much different from each other.The key difference between sensor networks and ad hoc networks is the ability to have local collaborative processing in the former. Sensor networks are deployed to observe a certain physical phenomenon. In such cases, a group of sensors end up with spatially correlated data. Also, since the receiver is interested in detecting and/or estimating the phenomenon, local collaborative signal processing is both feasible and desirable. That is, the data gathered from the multiple sensors is fused before transmission.

Several other differences are summarized in [10]. However, counter examples to these differences can be provided. For instance, [10] argued that the sensor networks are densely deployed compared to ad hoc networks. That is, as compared to the number of nodes in ad hoc networks, the sensor nodes could be several orders of magnitude higher. However, most deployments of the sensor networks to date involve relatively few nodes. Also, there have been proposals for dense ad hoc networks on metropolitan scales [3, 48].

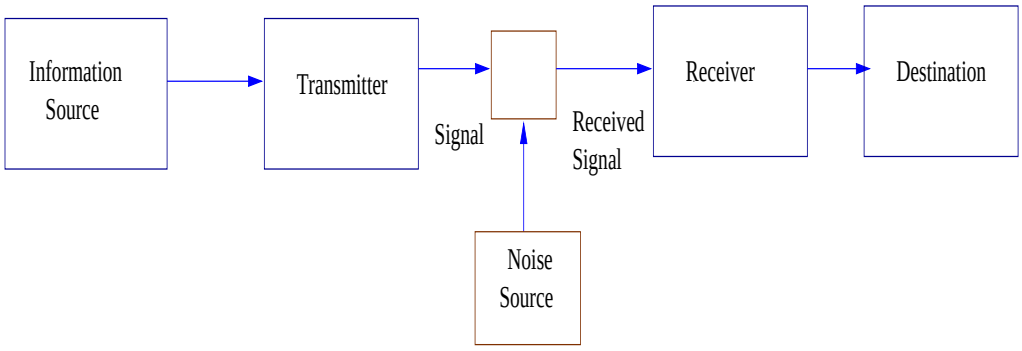
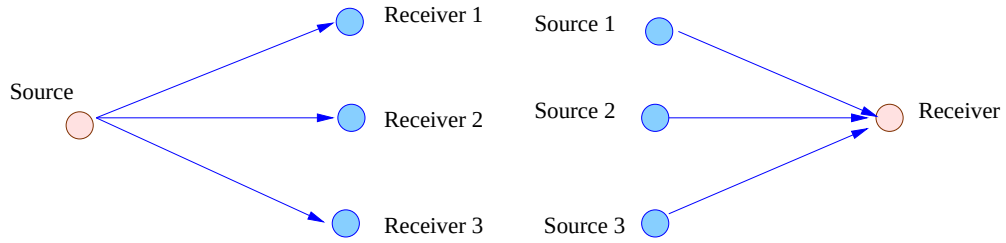


Figure 1.3: Shannon's schematic diagram of a general communication system.

1.3 Network Information Theory

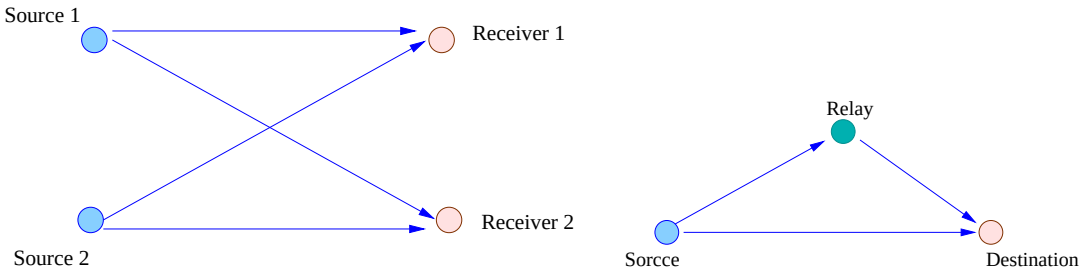
The foundations of information theory were laid by C. Shannon in 1948 through his landmark paper [214]. Shannon considered the point-to-point communication system as shown in Figure 1.3. Information theory proposes and suggests the ways of achieving data compression and transmission limits using source coding and error correcting codes respectively [54, 75]. These limits are necessary as it suggests the transmission rate for error-free communications or source-coding rate for achieving the desired fidelity. Apart from electrical engineering, information theory finds useful applications in mathematics, computer science, physics, statistics, probability theory, and economics [54].

In recent years, the application of information theory to networks has been a major focus. Shannon showed for a communication system with single sender-receiver pair as in Figure 1.3 that the source and channel coding can be treated independently of each other. However, this result does not apply to many senders and many receivers simultaneously communicating. This makes the ongoing studies on networks more interesting. In this section, we shall briefly discuss the information theory as applied to networks.



(a) Broadcast Channel (b) Multiple Access Channel

Figure 1.4: Information theoretically solved channel models.



(a) Interference Channel (b) Relay Channel

Figure 1.5: Information theoretically unsolved channel models.

Although network information theory has been around for more than half a century, there are lots of open problems in this field. For instance, the capacity of the broadcast (Figure 1.4(a)) and multiple access (Figure 1.4(b)) channels are known, whereas the simplest interference (Figure 1.5(a)) and relay channels (Figure 1.5(b)) are yet to be completely solved. However, the rate regions for their special cases have been solved [54, 75].

Slepian and Wolf considered the separate coding of the correlated sources. They considered the correlated coding system as depicted in Figure 1.6 [54, 221]. The correlated sources, say X and Y, are independently encoded and transmitted to the receiver at rates R_1 and R_2 respectively. At receiver, the received data is jointly decoded. To summarize their result, recall the Theorem 14.4.1 in [54]:

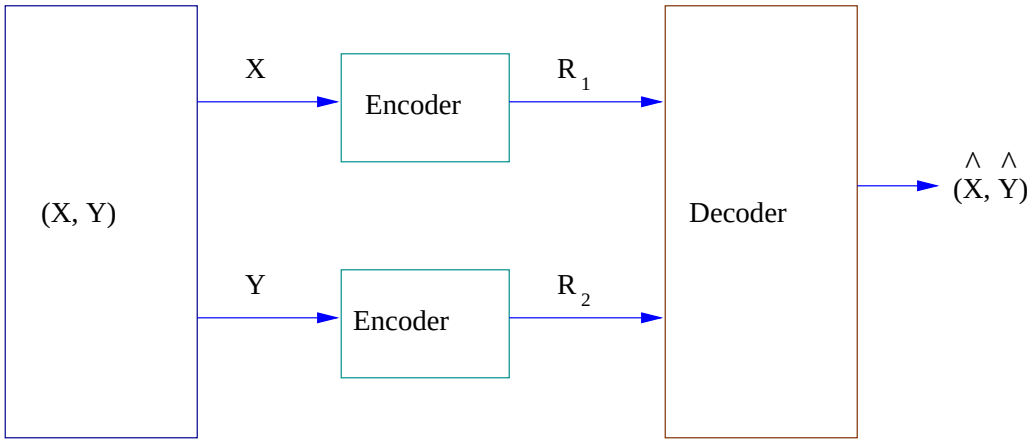


Figure 1.6: Slepian-Wolf Coding System.

Theorem 1.3.1 (Slepian-Wolf) *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:*

$$R_1 \geq H(X|Y) \tag{1.1}$$

$$R_2 \geq H(Y|X) \tag{1.2}$$

$$R_1 + R_2 \geq H(X, Y) \tag{1.3}$$

Several variations of Slepian-Wolf coding system have been studied. Of these, an important one is source coding with side information (Section 14.8, [54]). Here, the random variables X and Y are encoded separately but only one of the sources, say X, is to be recovered as shown in Figure 1.7. Such systems can be also seen in practice, for example, if the sensors are measuring the same phenomenon then we may not need to reproduce the readings of all the sensors but only one (generally the one with high signal-to-nose ratio (SNR)) using others as the side information. The result for such a system is cited as Theorem 14.8.1 in [54]:

Theorem 1.3.2 *Let* $(X, Y) \sim p(x, y)$*. If* Y *is encoded at rate* R_2 *and* X *at rate* R_1*, then* X *can be recovered with an arbitrarily small probability of error if and*

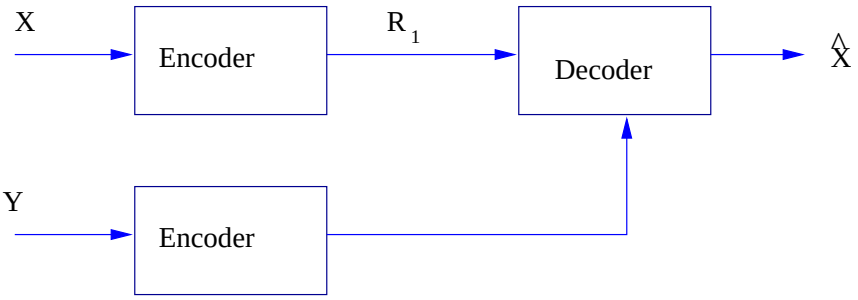


Figure 1.7: Encoding with side information.

only if:

$$R_1 \geq H(X|U) \tag{1.4}$$

$$R_2 \geq I(Y;U) \tag{1.5}$$

for some joint probability mass function $p(x,y)p(u|y)$*, where* $|\mathcal{U}| \leq |\mathcal{Y}| + 2$*.*

The design of distributed source codes that achieve Slepian-Wolf bounds are extensively studied. A few examples of such codes are linear codes, turbo codes [93, 158], and low density parity check codes (LDPC) [74].

So far, the lossless reproduction of the sources has been considered. However, for physical phenomena, the sources are continuous and thus measurement and source reconstruction are subject to distortions. Hence, the rate-distortion problem has to be considered [19, 54, 215].

$$R(D) = \min_{f(\hat{x}|x):E[(\hat{X}-X)^2]\leq D} I(X;\hat{X}) \tag{1.6}$$

The rate-distortion problem for networks was first considered by Wyner and Ziv [252, 253]. They considered a problem of rate distortion with side information. Let (X_i, Y_i) be i.i.d. $\sim p(x,y)$. The coded information X and the uncoded information Y is available at the decoder. Here, the objective is to reproduce

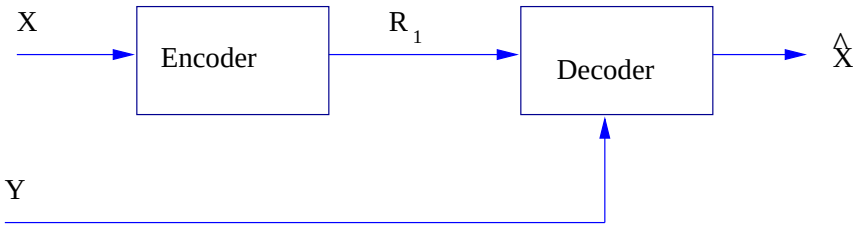


Figure 1.8: Wyner-Ziv Coding system.

X within desirable distortion, D. The problem is to find the rate-distortion function. Such a system is considered in Figure 1.8.

The result is summarized in the following theorem (Theorem 14.9.1, [54]):

Theorem 1.3.3 (Rate distortion with side information) *Let (X, Y) be drawn i.i.d. $\sim p(x, y)$ and let $d(x^n, \hat{x}^n) = \frac{1}{n} d(x_i, \hat{x}_i)$ be given. The rate distortion function with side information is*

$$R_Y(D) = \min_{p(w|x)} \min_{f} (I(X; W) - I(Y; W)) \tag{1.7}$$

where the minimization is over all functions $f : \mathcal{Y} \times \mathcal{W} \to \hat{\mathcal{X}}$ and conditional probability mass functions $p(w|x)$, $|\mathcal{W}| \leq |\mathcal{X}| + 1$, such that

$$\sum_x \sum_w \sum_y p(x, y) p(w|x) d(x, f(y, w)) \leq D$$

The reproduction of multiple sources at the receiver was considered by Berger-Tung in [20, 241]. Various other coding systems were then explored such as m-helper system [167, 172], Sgarro's problem [20, 212], Körner-Marton's Zig-Zag problem [20, 132, 133], Berger-Chang extension of Wyner-Ziv theory [20], Wyner-Ziv problem with multiple sources [87, 88, 85], etc. Many of these systems can be mapped to a sensor network model neglecting the sensing channel. On considering the sensing-channel the sensor network can be best explained by a CEO system [243, 168, 41, 183, 64, 42, 25, 171]. Note that for analytical tractability the

sources are generally assumed to be Gaussian. Other significant contributions to rate-distortion theory is found in [22, 29, 197, 195, 196, 198, 255, 265]. A comprehensive summary on lossy source coding is found in [22].

1.4 Classification of Information Theoretic Issues in Sensor Networks

In this section, we shall discuss some information theoretic problems in sensor networks. Depending on the services required by the network application, the constraints could be either on delay and/ or rate [182]. Table 1.1 summarizes these problems.

In general, information theory problems assume perfect time and position synchronization in networks. However, this in practice is not true. For instance, in the time domain, it is highly unlikely that the crystal oscillators of all the nodes in the networks are completely synchronized. This leads to a relative time error, Δt, between the nodes [67, 66, 111]. Here, the trade-off is in minimum use of resources (rate) to satisfy the desired time distortion. Hence, this forms a rate-distortion problem.

The data transmission from different nodes to the destination involves the reception of the messages with the minimum possible error probability. Here, the limiting factors are bandwidth and power and the problem is to calculate the maximum rate with these constraints. Hence, the problem formulation here is of channel capacity.

The (possibly processed) measurements recorded by sensors are eventually transmitted to the central unit. At the decoder, the issue is the quality of reconstruction of the measured readings in both spatial and temporal dimensions.

Services	**Quality Measure**	**Constraints**	**Problem Type**
Time and Position synchronization	Δt, Δr	delay, rate	Rate-Distortion
Data Transport	Probability of error, $P(e)$	bandwidth, power	Channel capacity
Measurements	Δr, Mean squared error	delay, rate	Rate Distortion and Channel capacity

Table 1.1: Classification of Network Information Theoretic Problems.

Here, the main concern is the resources or rate. Based on the applications, even delay could be the limiting factor. This type of service combines the problem of rate distortion and channel capacity due to the involvement of data processing and communications.

Table 1.2 provides the cause and possible solution for the constraints described above. We observe that the two main limiting factors in network services are rate and delay. Note, that the rate here defines the resources and hence, also includes energy and bandwidth.

Delay Constraint Latency (delay) in networks is largely due to the multihop transmission and congestion. We neglect the other possible latency factors such as node failures. With the use of higher bandwidth, longer transmission hops, or reduction in interference by the use of greater directionality in antennas, the latency could be ameliorated.

Rate Constraint This constraint arises from a combination of the limited bandwidth and signal to noise ratio (SNR). Low SNR could be due to low trans-

Constraints	Causes	Possible Remedies
Delay	Multihop Transmission, Network Congestion	higher bandwidth, longer transmission hops, greater directionality in antennas
Rate	Limited bandwidth, Low SNR	cooperative processing, high node density, increased power range

Table 1.2: Summary of Network service constraints and their possible remedies.

mission power and/or the noise sources. The constraint becomes more relaxed with cooperative processing such as the use of multiple antennas as in multiple-input-multiple-output (MIMO) systems or local processing at source, provision of higher power range, and increasing the node density.

1.5 Organization and Contributions of the Thesis

1.5.1 Focus of the Thesis

This thesis considers some of the rate-distortion and channel capacity problems described in Section 1.4. As seen before, the deployment of sensor networks may involve a large number of sensor nodes. Also, the futuristic applications of ad hoc networks such as smart house [2] could result in the large deployment of ad hoc nodes. This leads to an interesting question of scalability, that is, the feasibility of network services as the number of nodes tends to infinity. This is the staring point of our study.

The majority of our research is concentrated on sensor networks and the rate-

distortion limits when various models of such networks are considered. The main advantage of these networks is to have collaborative capabilities. Thus, by cooperative techniques an otherwise disconnected cluster of nodes could have normal communications. Although the focus of the thesis is largely on the physical layer algorithms of the wireless sensor and ad hoc networks, some higher layer abstractions with practical approacher are also considered.

1.5.2 Organization of the Thesis

The rest of the thesis is organized as follows.

Chapter 2 This chapter discusses the scalability issues in wireless ad hoc networks. Prior results have shown that for ad hoc networks with uniform source-destination probabilities, where each node generates traffic, the transport capacity for each node in the network declines with the network size [98]. However, we show that the scalability in such networks is possible if the local communication predominates or extra resources such as bandwidth are provided leading to a hierarchial architecture. We further consider the relay network model as considered in [89] and infer interesting scalability results when combined with the point-to-point coding model. Next, mobility in ad hoc networks is considered. The use of mobility has been shown to be beneficial in wireless ad hoc and sensor networks, for improving communication performance and other functionality. The communication throughput and delay trade-offs are considered when a set of mobile nodes are used as relays to transfer data among multiple static nodes. While previous work has considered randomly mobile nodes, we consider controlled mobile agents. The results for the worst case delay and throughput with controllably mobile relays are derived.

Chapter 3 In here, the scalability for sensor networks have been considered. The scalability issue in sensor networks transforms to the problem of information extraction at the fusion center under a fidelity constraint. However, as discussed in Section 1.4, this needs the application of joint source and channel coding to achieve information theoretic limits. However, on increasing the sensors and relays densities, the source and channel coding could be decoupled and still achieve scalability. This approach, although simple, is not practical. We consider the notion of spatial fidelity for such networks observing either point or distributed phenomena. Based on the source density, spatial fidelity, sensing strength of the sensors, and desired distortion, the analysis for local cooperative processing and required sensor density could be carried out. This chapter also summarizes and proposes algorithms for the data fusion mechanisms that might be used [239, 242]. Further, the distortions at various stages of sensing and transmitting distributed phenomena are identified. It is shown that these errors are bounded and do not propagate to distant nodes.

Chapter 4 This chapter considers the network information theoretic problem of finding the rate distortion bound when multiple correlated Gaussian sources are present. One of these is the source of interest but some side information from other sources is also transmitted to help reduce the distortion in the reproduction of the first source. The other sources are treated as helpers and are also coded. Special cases of this problem have been solved before, such as when the reproduction is lossless, when the sources are conditionally independent given one of them, or when the number of helpers is limited to one. We consider a generalized version and show that the previously derived expressions fall out as special cases of our bound. Our results

can be directly utilized by designers to choose not only how many of the available sources should actually be communicated but also which sources have the highest potential to reduce the distortion. Also, based on this result, the rate region for the Berger-Tung multiterminal system is derived where all the sources are reproduced at the receiver.

Chapter 5 Sensor networks collect data at multiple distributed nodes and transfer the acquired information to points of interest. The raw data collected by each individual sensor is typically not of interest. Instead, a reduced representation of the measured phenomenon is to be generated. Multiple readings, however, add to the information about the phenomenon by providing its description at multiple points in space for distributed phenomena and multiple perspectives for a localized phenomenon. We also note that sensor readings have noise, and multiple readings can help mitigate the effect of this noise. Thus, while all the sensor readings need not be communicated, enough data must be exchanged to reliably reproduce the phenomenon. Considering the above effects, it becomes important to determine how much data should be transmitted from multiple sensors such that only useful information is exchanged and energy or bandwidth are not wasted on redundant data. This question is addressed in Chapter 5 using information theoretic techniques. The effects of sensor noise and correlation in the sensor readings are explicitly modelled.

Chapter 6 This chapter addresses the problem of cooperative coding among pairs of transmitters and corresponding pairs of receivers. This may be used, for example, to overcome gaps in a multihop network. We derive upper bounds on the achievable rates for several scenarios and show that the transmitter cooperation provides significantly more improvement in rate

than receiver cooperation.

Chapter 7 This chapter considers the higher layer abstraction of wireless networks as compared to physical layer discussions in the previous chapters. For mobile ad-hoc networks, nodes must balance a variety of tasks including sensing and communications relays. Mobile nodes might thus change location or trajectory for sensing purposes, subject to constraints on disruption of network QoS. This chapter maximizes the non-communication application QoS (node motion to facilitate sensing) with communication QoS constraints (packet delay, etc.). It also gives the formulation for maximizing the throughput for the newly formed links at the new position of the node, taking link capacities into consideration. Each link is shared by multiple streams of traffic from different QoS classes, and each stream traverses many links. Although these formulations are non-linear, they can be posed as geometric programs, which can be solved efficiently. Heuristics to implement the above algorithms in ad hoc networks are also considered.

Chapter 8 The conclusions of the thesis is presented in this chapter. Along with the conclusions, some future research directions are also suggested.

CHAPTER 2

Scalability in Wireless Ad Hoc Networks

For 50 *years, people have worked to get to the channel capacity he said was possible. Only recently have we gotten close.* - R. Lucky (on C. E. Shannon), quoted in Technology Review, July 2001.

Wireless networks often include multihop communications over a wireless channel. There are few wireless networks that have a wired backbone, such as cellular voice and data networks [27, 231]. In contrast, ad hoc networks [178] have all wireless links and do not possess any centralized control. In recent years, engineers are looking forward to the dense deployment of networks to knit the world and beyond together [2, 247]. Whether this is feasible or not is answered by knowing their fundamental (information theoretic) limits, including scalability.

This chapter focusses on the scalability issues of ad hoc networks. By scalability, it means that the per node throughput does not go to zero with the increase in network size (number of nodes in the network). The starting point of the discussion is the capacity of the resource (bandwidth, power) constrained static ad hoc network model as considered by Gupta and Kumar in [98]. Their results suggest that ad hoc networks are inherently non-scalable. In particular, for the network size of n nodes in the unit disk with uniform traffic pattern, the per-node transport capacity of $O\left(1/\sqrt{n}\right)$ [1] bit-meter per second is obtained. The

[1] $f(n) = O(g(n))$ implies that there exists a constant c and integer N such that $f(n) \leq cg(n)$ for $n > N$; $f(n) = \Theta(g(n))$ means $f(n) = O(g(n))$ and $g(n) = O(f(n))$

limitation arises from the infeasibility of long range communications which would otherwise cause interference. Hence, the communications is limited to the nearest neighbor which is known to be at distance of $O\left(1/\sqrt{n}\right)$. The packet is routed to the destination through other nodes that are typically $O(\sqrt{n})$ serving as relays. However, it was shown in [139, 174] that these networks can be made scalable if the local communications predominates. A second approach to get a scalable network is to provide extra resources. The planar network model consisting of multilevel relays is considered in [254]. They allow for the node cooperation and also suggest the appropriate cooperation approach. More interestingly, for a particular case of network model they show the network to be scalable. Another type of relay network is considered by Gastpar and Vetterli in [89, 86]. The model considered by them is very much different than that in [98] and obtained the capacity in bits/sec. for that network. It should be noted that the results of Gupta and Kumar in [98] are model specific and could be applied only to models similar to that they considered. Lots of work on capacity of network has been done in [238, 134, 162, 63, 258], but the best known throughput result for the resource constrained ad hoc network model is that given by Gupta and Kumar in [98]. Note, that these scalability results do not strictly apply to wireless sensor networks. Chapter 3 presents the rather surprising and encouraging scalable nature of sensor networks. The issues of scalability that arise in ad hoc wireless networks, take a back seat when extended to sensor networks as unlike here, the problem is of information extraction under a fidelity criterion.

The above discussion, so far, did not consider mobility in the ad hoc networks. Grossglauser and Tse in [96], assumed random mobility of all the nodes in the network. Mobility provides an extra resource to the network and results in [96] confirmed that by showing the network to be scalable. However, the delay was unconstrained in their model and this leads to a scalable network with infinite

delay. Later, the models in [237, 47, 14, 81, 124] accounted for both mobility and delay suggesting trade-offs between throughput and delay. Here, the controlled mobility among nodes has been considered. This chapter summarizes the previous known scalability results and presents the approaches to attain scalability in ad hoc networks.

Prior known results on capacity of static ad hoc network are summarized in the next section followed by the methods to attain scalable network in Section 2.2. The relay network model and interesting inferences from relay model is considered in Section 2.3. Section 2.4 briefly describes the known mobility models and capacity results. The throughput-delay tradeoffs for a controlled mobile network is presented in Section 2.5. Practical methods for routing the mobile agents are discussed in Section 2.6. The Chapter ends with conclusions in Section 2.7.

2.1 Prior Results for Static Ad Hoc Wireless Networks

Most recent studies on scalable networks have been motivated by the results in [98]. For the sake of simplification and also for analytical tractability, the capacity derived by Gupta and Kumar in [98] is in bits-meter/sec. rather than bits/sec. They term this as transport capacity, which in words is defined as the total number of meters travelled by all the bits per time unit. To set the discussion, their result is presented here. Consider the disk of unit area with uniformly distributed nodes $\{X_i\}_{i=1}^{n}$ with their positions being i.i.d. Each source node has an independently and randomly chosen destination node. Such a model is depicted in Figure 2.1. Each node transmits at W bits per second over a common wireless channel. Packets are transmitted to the destination in a multihop fashion. Also, radios that are sufficiently distant can transmit concurrently and the total amount of data that can be simultaneously transmitted for one hop increases

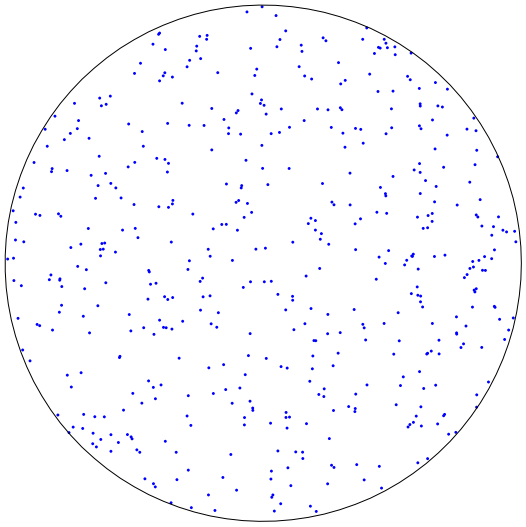


Figure 2.1: n nodes randomly located in the disk of unit area with uniform traffic pattern

linearly with the total area of the ad hoc network. For this model, the following theorem can be stated [98, 96]:

Theorem 2.1.1 (main result 4 in [98] and theorem 3.1 in [96]) *There exists constants c and c' such that*

$$\lim_{n\to\infty} Pr\left\{\lambda(n) = \frac{cW}{\sqrt{n\log n}} \qquad \textit{is feasible}\right\} = 1 \tag{2.1}$$

and,

$$\lim_{n\to\infty} Pr\left\{\lambda(n) = \frac{c'W}{\sqrt{n}} \qquad \textit{is feasible}\right\} = 0 \tag{2.2}$$

From the above theorem, it is evident that, for a fixed ad hoc network, within a factor of $\sqrt{\log n}$, the throughput per source-destination pair goes to zero in $W/\sqrt{n}$ fashion. This theorem is under the physical model for interference [98]:

$$\frac{\frac{P}{|X_i - X_j|^\alpha}}{N + \sum\limits_{k\in\mathcal{T}, k\neq i} \frac{P}{|X_k - X_j|^\alpha}} \geq \beta$$

where P is the common transmission power level for all nodes, and $\{X_k, ; k \in \mathcal{T}\}$ is the subset of nodes simultaneously transmitting at some time instant over a certain subchannel. The above mentioned model is for the successful receipt of transmitted message from node X_i, $i \in \mathcal{T}$ to X_j.

Results in [98] can be easily understood from geometry and intuition as described in [96]. As discussed earlier, all the nodes in the network transmit at the same power level P. From the physical model, it is clear that successful transmission of the packet from node X_i to node X_j occurs only if there is no other transmitting source in a disk of radius proportional to $|X_i - X_j|$ i.e. no interference. Hence, a successful transmission over d meters will incur a cost of d^2 by excluding other interfering transmissions. So to maximize the transport capacity of the network, scheduling a large number of smaller transmissions is advisable. Hence, the communications is limited to the nearest neighbors. The neighbors are typically at distance $1/\sqrt{n}$ [97]. Thus, transport capacity is at most $\sqrt{n}$ b-m/s. Since there are n sessions, with an expected distance of $\Theta(1)$, the throughput per session is $O(1/\sqrt{n})$. Also, this is the best that one can do in terms of throughput.

This suggests that the ad hoc networks are inherently non-scalable. However, the next section demonstrates the situation where these networks are scalable.

2.2 Towards Scalable Static Ad Hoc Networks

Successful communication is still possible with almost constant node bandwidth, if the distribution of the source-destination (S-D) pairs is such that the average hops per communication is small enough.

Consider the 2-D framework as in [98], described in Section 2.1. Servetto in

[211] observes that the scalability in sensor networks of arbitrary size is achievable as long as the rate at which nodes generate information decays faster then the throughput of the network. However, we note that it is not the fact of correlated sources which is most fundamental to this result, but rather the S-D pair distribution.

Observation 2.2.1

For the 2-D geometric model mentioned in Section 2.1, the average distance between source and destination should be $O(\frac{1}{\sqrt{n \log n}})$.

Clearly, in order for the average throughput per node to be constant, the average number of hops between source and destination should grow as $O(1)$. This follows immediately by observing that this criterion is necessary for transport capacity meeting the upper bound of $\sqrt{n \log n}$.

2.2.1 Scalability achieving S-D pair distribution

Consider a network with n nodes randomly distributed over an unbounded area. The x and y coordinates of the node locations are independent Gaussian distributions with zero mean and variance σ^2 as illustrated in Figure 2.2. Defining R as the transmission range of each node, the probability density function (pdf) of the link distance, r, between any two arbitrary nodes is given by:

$$p_r(r) = \frac{r}{2\sigma^2} e^{-r^2/4\sigma^2} \tag{2.3}$$

The derivation for the pdf of the link distance can be found in [152].

Now, the probability of a 2-hop connection between an arbitrary source and

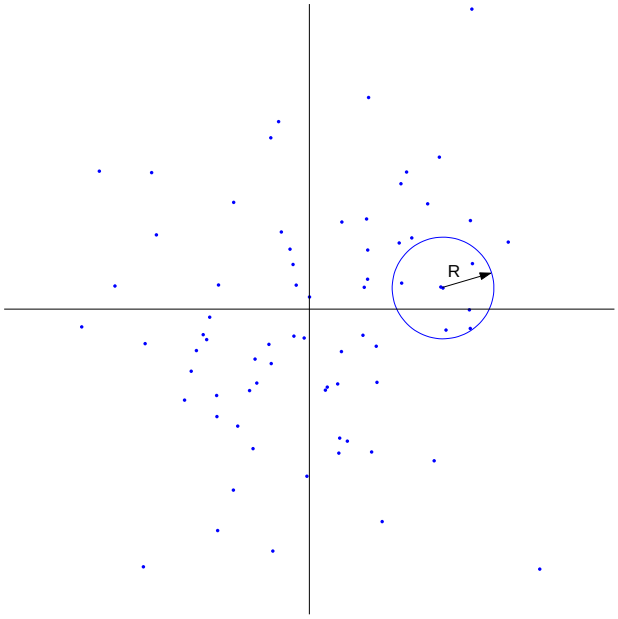


Figure 2.2: Random Network with dispersion parameter σ, and transmission range R.

destination pair is given as in [153]:

$$\begin{aligned} P_2 &= Pr\{1 \rightarrow 2 \text{ in 2 hops}\} \\ &= Pr\{R < r < 2R \text{ and at least one other node in the area of intersection} \\ &= \underbrace{\iiint\!\!\iint}_{R<r<2R} p_{x,y}(x_1, y_1, x_2, y_2) \\ &\quad \times \left[1 - \left[1 - \underbrace{\iint}_{A(x_1,y_1,x_2,y_2)} p_{x,y}(x_3, y_3) dx_3 dy_3\right]^{n-2}\right] dx_1 dy_1 dx_2 dy_2 \end{aligned} \tag{2.4}$$

As $n \rightarrow \infty$, (2.4) can be approximated with the following upper bound [153]:

$$P_2 < P_{2\infty} = \int_{R^2/4\sigma^2}^{R^2/\sigma^2} e^{-\nu} d\nu = e^{-R^2/4\sigma^2} - e^{-R^2/\sigma^2} \tag{2.5}$$

Similarly, the asymptotic probability of an m-hop connection is given by:

$$P_m < P_{m\infty} = e^{(m-1)^2 R^2/4\sigma^2} - e^{m^2 R^2/4\sigma^2} \tag{2.6}$$

Based on this, that the average number of hops between node pair can be calculated as in [153],

$$E\{h\} = \sum_{m=1}^{n-1} mP_m < \sum_{m=1}^{\infty} mP_{m\infty} \tag{2.7}$$

$$E\{h\} < \sum_{m=0}^{\infty} e^{-\frac{m^2}{4\gamma^2}} \doteq \overline{h}_+(\gamma) \tag{2.8}$$

where γ is the mobile dispersion given by $\gamma = \sigma/R$. Using the non-linear regression techniques and subsequent linearization, the asymptotic average number of hops for random source destination pair is,

$$\overline{h}_+(\gamma) \approx 0.5 + 1.772(\sigma/R) \tag{2.9}$$

In terms of actual distance, multiply by the transmission range of each node R, to upper bound the average hop distance per transmitted bit:

$$E\{h\} < R\left(0.5 + 1.772\frac{\sigma}{R}\right) \tag{2.10}$$

The average hop distance in (2.10) is for an unbounded disk. Since the nodes in the scenario described earlier are zero-mean Gaussian distributed, 99.7% of all the nodes are expected to lie within a 3σ radius of the center. The resulting area of the disk of radius 3σ is $9\pi\sigma^2$. Hence, scale the result in (2.10) by this factor. Also note, that now the average hop distance is consistent with the framework of Gupta and Kumar [98],

$$H = \frac{R}{9\pi\sigma^2}\left(0.5 + 1.772\frac{\sigma}{R}\right) \tag{2.11}$$

The above equation, as observed, is independent of n.

Hence, the zero-mean truncated Gaussian is one of the many distributions that achieve the finite per node capacity for the geometric model of [98].

An approach mentioned so far is based on results in [211], but one can find similar results with the alternative approach in [139]. J. Li, et. al. in [139] show that the traffic pattern determines whether an ad hoc network's per-node capacity will scale to large networks. In particular, for the total capacity to scale up with network size the average distance between source and destination nodes must remain small as the network grows. In other words, this coincides with our conclusion of encouraging local communications.

In practice, scalability can be achieved in two basic ways:

1. Local cooperative processing to produce decisions (e. g. in sensor networks).
2. Adding communications hierarchy so that communications in each level is local (e. g. telecommunication network).

The latter of course requires additional resources, but typically also provides latency benefits.

We shall now briefly describe the scalability attained by providing extra resources as mentioned in [200].

2.2.2 Hierarchial approach to Scalability

If the latency is considered to be important along with scalability, then Gupta and Kumar's approach in [98] is certainly not recommended. If the bandwidth available to node scales with at least $O(\sqrt{n})$ then scalability in [98] is achieved. Hence, providing extra resource in terms of bandwidth to each node can do the job as far as throughput per node is concerned. However, the latency still scales as $O(\sqrt{n})$.

A dynamic but stable structure is considered in [200] by the introduction of a stochastically self-similar, multi-resolution structure on a homogeneous set of

n nodes, i.e., by providing hierarchial structure to nodes. The model assumed is a square of area A with randomly distributed nodes. Each node is assumed to have access to K non-interfering additive white Gaussian noise (AWGN) channels. Also, at each time, a node is assumed to be able to communicate with at most one other node at any given channel. Furthermore, the nodes are assumed to each know their position X_i as well as the position of the nodes they need to communicate with and the position of the destination. For uniform traffic pattern, it can be shown that on an average, the latency of $O(\log^2 n)$ and per-node throughput of $O(1/\log^3 n)$ can be obtained if bandwidth scales as $O(n \log n)$. The idea behind multi-resolution structure is that similar to skip-lists [188]. The cost of bandwidth is relatively high for the $\log^2 n$ latency. This may not find lot of admirers but considering that even the telecommunication networks pays this cost, it might be worth to pursue for future applications.

An optimal throughput-delay tradeoff for the model in [98] is given in [81] as $D(n) = \Theta(nT(n))$, where $T(n)$ and $D(n)$ are throughput and delay respectively.

2.3 Capacity of Relay Modelled Ad Hoc Networks

In previous sections, bounds on the throughput for point-to-point network model having multiple transmission sessions is considered. This section deals with relay network model as considered by Gastpar and Vetterli in [89].

Consider a network of n nodes in the disk of unit area. The source and destination are randomly chosen. It is to be noted that the traffic pattern here is different than that of point-to-point coding model. Here, there is only one active link. So apart from source and destination all the remaining nodes act as relays. The relay nodes follow the amplify-and-forward relaying technique.

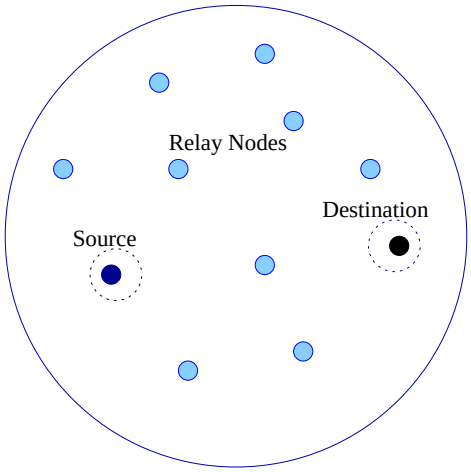


Figure 2.3: A wireless network under the relay traffic pattern

Hence, arbitrary cooperation is allowed between the nodes, for example multiple access and broadcast. The received signal at a node is a sum of the faded signals from the other nodes plus additive white Gaussian noise (AWGN). Such a model is depicted in Figure 2.3. The assumptions for this model are that the there are no other nodes within non-zero radius around the source, and the source transmits the information for only half of the time. These assumptions, however, are necessary to facilitate capacity analysis as mentioned in [89]. In contrast to Gupta and Kumar, the capacity bounds derived here are in bits/sec.

Before proceeding further to capacity bounds, it should be noted that question of scalability should not be asked in this context. This is because, there is only one active transmission session as compared to multiple sessions in point-to-point coding model. So even if the node density increases, there will not be any effect on the nature of the transmission session. However, this model is interesting as it shows that the direct application of results in [98] does not work. The direct application of Gupta and Kumar results would yield $O(\sqrt{n})$ b/s or with more careful application $O(1)$. This is certainly in contrast to $O(\log n)$ b/s that will

be obtained for this model as shown below.

The upper bound on the capacity of the considered relay network is given by [89]:

$$C \leq C_{upper} = \frac{1}{4}\log_2\left(1 + \frac{||\alpha||^2 P}{N}\right) \tag{2.12}$$

where α denotes the vector of length $n-1$ of all the fading coefficients α_i's, that is $\alpha = (\alpha_2, \ldots, \alpha_n)$, P is the power constraint on transmitting node, and N is the variance of zero mean AWGN. Consequently, $||\alpha||$ denotes the magnitude of that vector. The fading coefficients from the source node to the node k is defined by $\alpha_k = 1/d_{1k}^r$, for $k = 2, , n$. d_{1k} is the Euclidean distance between the source node and node k, and $r \in \Re^+$ is the path loss.

The lower bound follows from a consideration of (almost) uncoded transmission of a particular source across the Gaussian relay network [89]:

$$C \geq C_{lower} = \frac{1}{4}\log_2\frac{P}{D_1} \tag{2.13}$$

where D_1 is the mean square error. The details on mean square error and power constrains can be found in [86, 89].

To summarize, as $n \to \infty$, the capacity for relay network model will follow $O(\log n)$ b/s. Interestingly, with fusing a relay model to a point-to-point model, a scalable network can be obtained.

In addition to relay network model constraints, assume that nodes have their own position information and as well as the position information of their neighbors and final destination. Also assume that the nodes are equipped with the omnidirectional antenna. The source node transmits the information to all the nodes in the circular region of radius r, the transmission range of a node. For simplicity, suppose all nodes are identical. See Figure 2.4 for the network model.

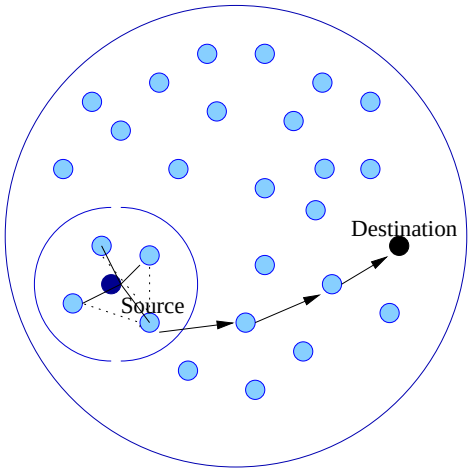


Figure 2.4: Fusing relay network model in point-to-point coding model

The nodes lying in the circular region of radius r will cooperate and transmit a message to the node which is nearest to the destination. In every such circular disk of radius r, there will be $n_r = \sqrt{\pi} nr$ uniformly distributed nodes. This forms a local relay network model having a local source and destination. The other nodes in the local network act as relays. The message will be decoded at the local destination node which in turn will open another session for transmission. The decoding of a message at an intermediate node is essential to maintain the quality of the information. If the distance to the destination from the source is L then the message can be delivered to the destination in L/r hops. Also, from a sphere packing solution, a maximum of $1/\pi r^2$, $r \leq 1/\sqrt{\pi}$, sessions at any time instant t is possible. The upper and lower bound on capacity, C, for one session will be similar to that derived above for the relay model. If we have m transmitting sessions in parallel, then network capacity will be mC bits/sec. Here, however, instead of n source-destination pairs, we have m pairs. But in any case, if $r > 0$, then as $n \to \infty$, per session throughput will be $O(\log n)$ b/s. This certainly is better than the scalability results in [98]. The reduction in the

number of sessions in a way acts as provisioning of extra resources. Also, in this case the delay will be $O(\sqrt{m})$. Hence, the network model suggested can achieve scalability and also optimal m and hence r, for any given application could be calculated. This shows that the physical interference model has a significant impact on capacity.

The other notable work in this field is presented in [100] that shows that for a certain class of wireless networks, transport capacity of $\Theta(n)$ b-m/s is feasible. An extension of two dimensional capacity analysis to three dimensions is presented in [99].

2.4 Mobility in Ad hoc Networks and Related Work

This chapter, so far, considered various models of fixed or static networks. Now, the concentration will be on mobility in ad hoc networks. Mobility has long been considered as an overhead for networks because the protocol stack had to provide additional functionality to handle mobile users [143, 127, 13, 36, 178]. More recently, however, mobility has been found to be useful for wireless ad hoc networks for increasing their throughput capacity [96, 14, 81, 61]. These advantages are accrued by utilizing the mobility of the nodes for transferring data. These provide additional capacity to the network over and above the wireless channel amongst static nodes alone.

Mobility can be classified into three categories:

1. **Random Mobility**: The nodes are assumed to move in an arbitrarily random pattern typically modelled as uniform Brownian motion for analytical convenience [96, 61, 14, 81]. This model has also been used in data mule work [213, 116].

Figure 2.5: Mobility in Sensor Networks - Networked Infomechanical Systems (NIMS)

2. **Predictable Mobility**: This model assumes that the pattern of mobility of the mobile nodes is known and this knowledge can be exploited to route data [40, 115, 177]. The mobile agents are not moving for the purpose of data transfer and hence their paths may not coincide with the routing requirements.

3. **Controlled Mobility**: Here the mobility pattern of the mobiles is completely under the control of the network. Prototypes of such networks have been considered [125, 263, 264, 122]. It should be noted that the controlled mobility could also be implemented by providing infrastructure to move the nodes. In some situations, such as in sensor networks deployed for habitat monitoring, provision of infrastructure could be cost effective. Figure 2.5 considers the controlled mobility in sensor networks through the provision

of infrastructure (Networked Infomechanical Systems (NIMS)). In case of NIMS, overhead cables are mounted to move the sensor nodes.

We consider the network with controlled mobile agents and characterize the delay and throughput properties. Our analysis shows that there are fundamental differences in these properties when the mobility is controlled as opposed to when the mobility is random or predictable. Also note that using controlled mobility, we are able to guarantee performance for any arbitrary topology, while the results derived for random models only provide the expected throughput and delay averaged over multiple topologies.

2.4.1 The Need for Controlled Mobility

Random mobility is not a valid model in all classes of wireless ad hoc networks. In particular, sensor networks are autonomous embedded systems and do not involve user carried nodes. Most of their nodes are either static or mounted on robots which can be controlled as per application requirements. Further, the use of controlled mobility has several advantages in wireless ad hoc and sensor networks.

First, the mobile nodes can help save energy in static embedded nodes. This is because if the mobile nodes are used for carrying data, then the static nodes need not relay data from other nodes over multi-hop wireless routes. The extra energy overhead of mobility is not a major concern as these nodes are mobile and can thus periodically recharge themselves. Self recharging robots have been prototyped such as the robot tortoise described in [245] and the commercially available Sony Aibo [223]. Second, mobile agents can be used to connect sparse and disjoint networks. Particular network components can get disconnected due to deterioration in channel conditions and these can be connected using mobile

components. Third, the number of wireless hops travelled by a data packet are reduced and this reduces the possibility of packet error, helping enhance goodput performance, and delays due to retransmission.

There are other advantages of using mobile components for improving network sustainability. It was shown in [68] that the time synchronization error increases with an increasing number of hops between two nodes. Using the mobiles for time synchronization reduces the hop distance between nodes, and hence much finer time synchronization is possible than in a multi-hop case. Controlled mobility also helps improve the performance of localization systems [208]. Mobile components can support other system activities such as delivering required resources [137, 191]. Thus, the use of control mobility is helpful in several situations, and it is worthwhile to study the performance considerations for such a scenario.

2.4.2 Related Work

As seen in Section 2.1, the capacity of wireless networks was first evaluated in [98]. They assumed a network of n randomly deployed nodes in a unit area disc and found the average throughput if sources and destinations are chosen randomly across the network. The throughput per node was found to be $\Theta\left(1/\sqrt{n \log n}\right)$. For the model considered in [98], if the traffic pattern is such that the average distance between source and destination nodes remains small as the network grows the throughput per node was derived to be $O(1)$ in [139]. The use of mobility was considered in [96]. They assumed all nodes to be randomly mobile within a unit disc area. The data traffic pattern was assumed to be random as in [98]. Data travelled over only two wireless hops, from the source node to a mobile node which acted as relay and then from the relay to the destination. With this model the throughput was found to be $\Theta(1)$. Later, [61] showed that if

the mobility was restricted to one dimension only, the constant order throughout can still be achieved. The delay for the above scenarios was found in [81]. They also discussed algorithms for improving throughput at the cost of delay and vice versa. They found the delay $D(n)$ to be related to the throughput $T(n)$ as $D(n) = \Theta(nT(n))$ for the wireless network scenario of [98]. For the model in [96] when the nodes are randomly mobile with average velocity $v(n)$, they found the delay to scale as $\Theta(\sqrt{n}/v(n))$.

Another scenario for a network with mobile nodes was considered in [14]. The network consisted of n static nodes which acted as sources and destinations for data. However, the network also had m randomly mobile nodes which were used as relays. For this model, using the routing scheme proposed in [14], the throughout is $O(m/n \log^3 n)$ with an average delay of $2d/v$ where v is the velocity of the mobile nodes. We consider a similar model, with n static and m mobile nodes but the mobility is controlled instead of random. The problem is defined in greater detail in the next section.

2.5 Capacity of the Mobile Channel with Controlled mobile agents

Let us first consider the network scenario as considered in [96] where all nodes are mobile except that instead of considering the nodes to be randomly mobile, we assume their motion is controlled. For this case an obvious communication strategy is for each source to move to its destination and communicate at almost zero range. Hence interference among simultaneous transmissions is zero and each sender-receiver pair could utilize the full available bandwidth W. The per node throughput is W with constant delay. The delay here depends on the travelling

time of mobile nodes to reach their destinations which is constant as network area is constant. This is significantly better than the worst case delay in [96] which is infinite. Clearly, controlled mobility has the potential to yield fundamentally different throughput and delay limits compared to those achieved with random mobility in [96, 14]. We now consider a more practical scenario wherein all nodes are not mobile but a small number of controllably mobile agents are available for routing data.

2.5.1 Problem Statement

Assume n nodes are deployed randomly in a unit area disc. Any node may have data to be sent to any other node, resulting in up to n transmitter-receiver pairs. Further, m mobile agents are available whose paths can be controlled as per the data requirements. Here m could potentially be much smaller than n. The mobile travels close to the source location, collects its data over a wireless link, and then travels to the destination location and delivers the data to the destination node, again using a wireless link. The path from the source to destination may not be direct but may involve servicing other data transfer requests in between. This is because, using one mobile to serve only one source destination pair at one time does not necessarily yield optimal path planning for the mobiles and we allow the mobiles to collect data from multiple sources for multiple destinations, depending on the node locations. Figure 2.6 shows the problem scenario. The data rate on the wireless channel is W. Note that W could potentially be very high as the mobile node can travel very close to the static node. However, we perform the analysis for any available W. The velocity of the mobile nodes is denoted by v. For ease of exposition, we assume that each source has k bits for its destination. Some nodes may have less than k bits, however using k for all nodes suffices for

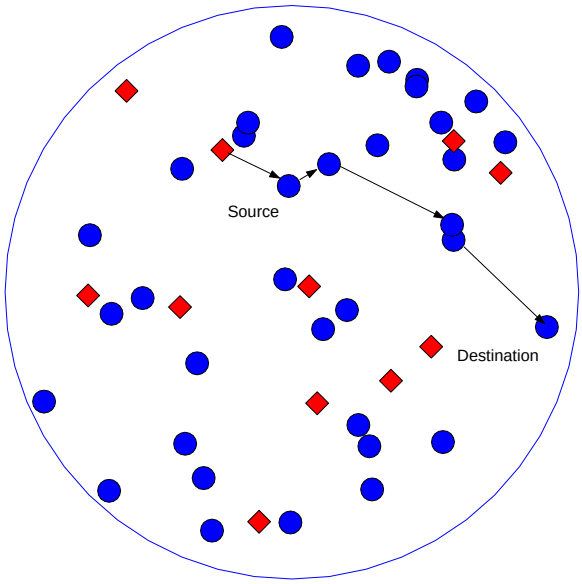


Figure 2.6: Network with mobile and static nodes. The red diamond shaped nodes represent the mobile nodes and the blue circular nodes are the static nodes.

evaluating worst case limits.

We consider the problem of determining the achievable throughout and delay experienced by any data packet in the worst case. We later also discuss the problem of allotting the optimal routes to the mobile nodes for routing data.

2.5.2 Solution

We now solve the problem laid out above. Let us calculate the delay suffered by a k-bit packet in travelling from its source to destination after it is ready to be transmitted at the source. For calculating the worst case achievable delay, we assume that each mobile serves n/m sender-receiver pairs. Depending on what metric is used for optimizing the paths of the mobile nodes, the actual allocation may deviate from this equitable allotment; we consider practical methods for such an allotment in Section 2.6. The delay and throughput derived below

are definitely achievable, using the (potentially sub-optimal) equitable allotment. The sender-receiver pairs need not be addressed one after the other, rather multiple such pairs will be simultaneously served. Instead of calculating the worst case delay individually, we evaluate the delay and throughout for a group of n/m nodes served by a single mobile. Since not all packets suffer the worst case delay, this method allows us to calculate the achieved throughput for all the data, rather than just for the node which happens to get the worst case delay. This total delay in serving n/m requests is also the worst case delay suffered by a packet, since the worst case packet would be the one which is served last in the group. The delay can be calculated as follows:

1. Mobile collects data from source: This time is the summation of the following terms:

   (a) Contention near source: When the mobile arrives at a source, there may be other mobiles too at this node delivering data destined for this node. Assume that the traffic pattern is such that at most γ nodes could send data to a particular node. Then, the maximum number of mobiles which can be present at a node, Γ, to deliver can be

   $$\Gamma = \min(\gamma, m) \tag{2.14}$$

   Hence the time spent waiting for these transmissions becomes $d_1 \leq \Gamma k/W$, where W is the bandwidth used for communication among static and mobile nodes. This communication occurs at very small range and does not interfere with other communication in the network.

   (b) Wireless data communication delay: The time taken to send k bits from the source to the mobile agent will be $d_2 = k/W$.

(c) Motion delay: The distance to a source from the previous node served can be at most $1/\pi$, the diameter of the network with unit area. Hence this delay is $d_3 \leq 1/(\pi v)$ where v is the speed of the mobile.

2. Time to travel from a source to a destination. This can be at most $d_4 = 1/(\pi v)$ as explained for the travel time taken to reach a source. The source and destination need not correspond. The delay, d_4, is only the time for one leg of the mobile's journey. These will be added together for the n/m requests being served by this mobile.

3. Contention at destination: This is $d_5 = \Gamma k/W$ as was the contention at source.

4. Time to deliver at destination: The transmission time at destination is $d_6 = k/W$.

Thus, the total delay, D, for the worst case packet among the n/m sender-receiver pairs served by one mobile agent is the summation of the delays evaluated above:

$$D \leq \frac{n}{m}\left[d_1 + d_2 + d_3 + d_4 + d_5 + d_6\right] \tag{2.15}$$

On substituting the values of $\{d_i\}_{i=1}^{6}$ in the above equation, we obtain:

$$D \leq \frac{2n}{m}\left[\frac{k}{W}(\Gamma + 1) + \frac{1}{\pi v}\right] \tag{2.16}$$

The throughput per node can be defined as the bits transmitted by a node divided by the time taken for those bits. Thus, the throughput, T, for the node with the worst case delay is:

$$T = \frac{k}{D} \geq \frac{mkW\pi v}{2n[k\pi v(\Gamma + 1) + W]} \tag{2.17}$$

Note that the network throughput is nT since each mobile agent has serviced n/m nodes within D and there are a total of m mobiles. Thus, we observe that the per node throughput and delay are $T = O(m/n)$ and $D = O(n/m)$.

2.5.3 Scalability Concerns

Here, the effect of the wireless bandwidth available among static and mobile nodes is considered. This bandwidth can be very large as the communication range is small and fast technologies such as UWB or a contact based transfer may be used. Let $W \to \infty$ in equations (2.16) and (2.17):

$$D_{W\to\infty} = \frac{2n}{m\pi v} \tag{2.18}$$

$$T_{W\to\infty} = \frac{mk\pi v}{2n} \tag{2.19}$$

As expected, the delay and throughput are limited by the mobile velocity, v, in this case. However, the delay and throughput can be very high here as the value of k used can be very large when $W \to \infty$.

Consider next the case when the number of mobiles used is a linear function of the number of static nodes in the network, i.e., $m = an$ where a is a positive constant, potentially much less than 1. In this case $\Gamma = \min(\gamma, m) = \gamma$ since we expect n and hence an to be much greater than γ. Here, the delay and throughput can be seen to be:

$$D_m \leq \frac{2}{a}\left(\frac{k}{W}(\gamma + 1) + \frac{1}{\pi v}\right) \tag{2.20}$$

$$T_m \geq \frac{akW\pi v}{2[k\pi v(\Gamma + 1) + W]} \tag{2.21}$$

These are both independent of n and m. Hence we achieve $T = O(1)$ and $D = O(1)$ for this case. Observe that compared to the case of random mobility [96] where $O(1)$ throughput is achieved when all n nodes are mobile, using controlled mobility such throughput can be achieved with less than n mobile nodes. Also, while the worst case delay in [96] can be infinite, here the worst case delay is also $O(1)$.

2.5.4 Practical Constraints on Mobile Nodes

In a practical scenario the mobile agents are not aware of all the data transfer requirements before beginning their journey. They get to know new requests while serving the previously known ones. Further, the mobiles may have only a limited buffer which is not sufficient to store the data for n/m requests simultaneously as was assumed in the above analysis. Both these factors do not change the order of either throughput or delay as mobiles can address each request one after the other still achieving $O(m/n)$ throughput per node and $O(n/m)$ delay. However, the constants calculated above will change as the time for each trip may have to be accounted for separately.

2.5.5 Mobile and Wireless Channels

We calculated the capacity when data is transferred using only the mobile relays and multihop wireless transmission is not exploited. Let us now consider the capacity when the multihop relaying is used in addition to mobile agents. Since the mobile router is controllably mobile it can come very close to the static node it is communicating with and thus cause negligible interference to any other communication which may be simultaneously taking place. Hence, the wireless multihop relaying channel can be used simultaneously along with the mobile channel. The

capacity of wireless channel has already been shown to be $\Theta\left(\sqrt{n \log n}\right)$ in [98].

The delay for the wireless multihop relaying channel was calculated in [81] while the delay for the mobile channel was shown in section 2.5.2. Thus, data travelling on the two channels experiences the respective delays. Data may also choose the hybrid route using multi-hop wireless channel in part and relaying over a mobile node for the remaining journey, achieving a trade-off in delay and throughput among the two channels.

2.5.6 Trade-offs in the Throughput-Delay Space

The trade-offs for throughput and delay for the models in [98, 96] are considered in [81]. The curve PRS in Figure 2.7[2] depicts that trade-off. The segment PR is valid for the model in [96] where all the nodes are mobile. The segment RS holds for the model in [98] using only static nodes. The per node throughput can be varied from $O(1/n)$ to $O(1)$ by paying a penalty in delay. The use of controlled mobility introduces new possibilities, shown along the curve AB. Different points on the curve AB correspond to different values of m. As m increases delay is reduced. The point A is achieved with $m = 1$ and point B is achieved with only a small increase to $m = 0.05n$.

Now consider the point P in Figure 2.7. This point corresponds to the constant per node throughput achieved in [96] using the routing scheme given in [81]. However the delay is $O(n)$. The point B on the other hand shows constant per node throughput at constant delay using the controlled mobility model.

[2]The relative magnitudes of the curve as shown in the graph are valid for large n, since the results from [81, 96, 98] are valid only at large n even though our results are valid for finite m and n also.

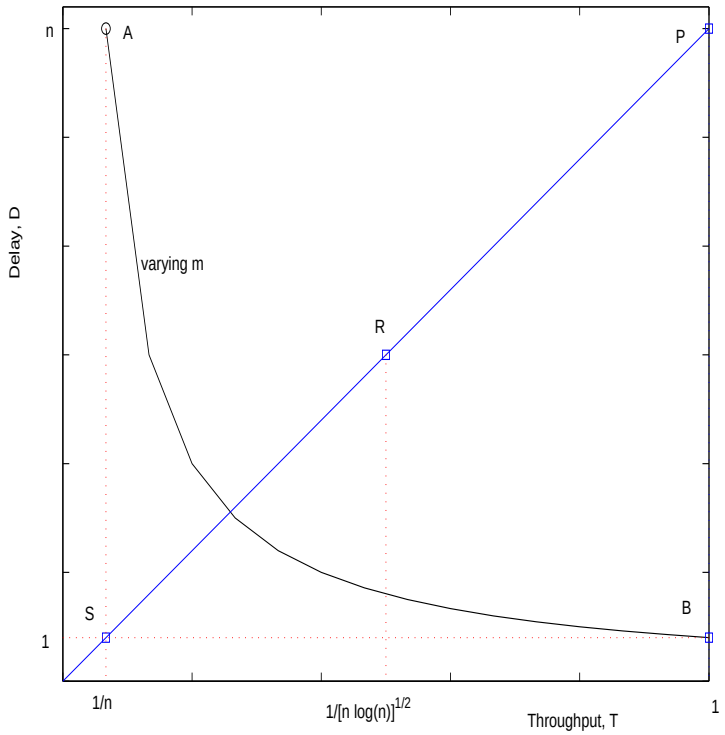


Figure 2.7: Comparison of throughput-delay trade-offs in controlled mobility and other models.

2.6 Mobile Routing Protocols

We now consider practical methods to control the paths of the mobile agents. This problem can be reduced to a known optimization problem, namely, the general pickup and delivery problem (GPDP) stated in [202, 65] as follows. Define the following variables:

N**:** the set of all sender-receiver pairs, where the cardinality of N is n.

M**:** the set, with cardinality m, of all the mobiles.

N_i^+**:** the i^{th} data source, $i \in \{1, 2, \ldots, n\}$.

N_i^-**:** the i^{th} data destination corresponding to N_i^+.

s^+**:** the start location of mobile s, where $s \in M$.

s^-**:** the end location of mobile s, where $s \in M$.

M^+**:** set of start locations of all the mobiles.

M^-**:** set of end locations of all the mobiles.

Z**:** $M^+ \cup M^-$.

Now, $\forall\ i, j \in \{N^+ \cup N^- \cup Z\}$, let d_{ij} denote the travel distance from location i to j and t_{ij} is the time taken.

Definition 2.6.1 (Mobile Path)

A path R_s for mobile s is an ordered subset $Z_s \in Z$ such that:

1. **Docking Constraint:** R_s starts in s^+ and ends in s^-.
2. **Pairing Constraint:** $(N_i^+ \cup N_i^-) \cap Z_s = \emptyset$ or $(N_i^+ \cup N_i^-) \cap Z_s = N_i^+ \cup N_i^-$. (This implies that if a data source is visited by mobile s, then s must also visit the corresponding destination.)
3. **Precedence Constraint:** If $N_i^+ \cup N_i^- \subseteq Z_s$, then N_i^+ is visited before N_i^-.

Definition 2.6.2 (Routing Plan for Mobile Relays)

A routing plan is a set of routes $\mathcal{R} = \{R_s | s \in M\}$ such that:

1. R_s is a valid path for mobile s, for each $s \in M$.

2. $\{Z_s | s \in M\}$ is a partition of Z.

3. $f(\mathcal{R})$ denotes the cost of routing plan $\mathcal{R}$.

The problem of finding the routing plan for mobile relays is then:

$$\min_{\mathcal{R} \text{ is a routing plan}} f(\mathcal{R}) \tag{2.22}$$

The cost function $f(\cdot)$ depends on the required objective. For instance, in our problem f could quantify the worst case delay suffered by any packet.

Our problem, which is also a GPDP, is known to be NP-complete problem [128, 38]. Hence, heuristic approaches need to be considered. Several applicable heuristics exist. For example, the approximations for the static multiple vehicle pickup and delivery problem in [202, 57] employs the decomposition of the problem in clusters and chains. This algorithm is based on set partitioning and column generation.

So far we did not explicitly consider the buffer limit in the mobile relays and that the sender-receiver requests may not all be known at the start of the mobile relays' journey. However, these constraints could be incorporated into the GPDP formulation and practical algorithms are also known [202, 60, 187].

Thus, practical methods are available to control the trajectories of the mobile relays and can be employed with the appropriate objective function.

2.7 Conclusions

This chapter discusses the issues of scalability, and throughput-delay tradeoffs in wireless ad hoc networks for both fixed as well as mobile models. Scalability can be controlled by making the S-D pair distribution peaked to local. When

this occurs, the local interactions will dominate resources. Thus, the cooperative signal processing and communication problems are most profitably considered in these local domains - the typical interactions (that are application specific) thus may involve relatively small numbers of nodes. Further, even though the optimization may be intricate, it is feasible because of the small numbers. The adjustment in S-D pair distribution for scalability can be done directly through cooperative signal processing to bias high volume traffic to local destinations. Average delay may also be controlled by adjusting the S-D pair distribution by biasing towards closer nodes, but this does not help with peak delay. For relay models it was shown that in conjunction with point-to-point model interesting scalability results are obtained.

We also considered the fundamental limits on throughput and delay for a network with mobile nodes using a new model for mobility. While previous work considered random mobility patterns, we discussed the scenario with controllably mobile data relays. We saw that this scenario is applicable for an emerging class of wireless network applications, such as sensor networks. Our analysis showed that controlled mobility can significantly change the throughput and delay trade-offs. In particular, networks scalability can be ensured using only a small fraction of mobile nodes and still achieve constant per node throughput at constant delay. The mobile nodes are equivalent to provisioning of extra resources. We reduced the problem of routing mobile nodes to a special case of GPDP. This formulation allows optimizing for several practical concern that network designer may have for her particular application, such as optimal number of mobile agents, total travelling time, or the worst case data transfer delay. Further work is required to determine the optimal combination of wireless multihop routing and the use of mobile relays to obtain a required performance for a given network size.

CHAPTER 3

Scalability and Source-Sensor Relations in Wireless Sensor Networks

"Embedded networks help make sense of nature, life and the world"- San Diego Union-Tribune.

3.1 Overview of the Chapter

In the previous chapter, the scalability issues for the wireless ad hoc networks were considered. As seen in Chapter 1, sensor networks are different than ad hoc networks and hence, need a different set of rules for their architecture and protocols design. Thus, the issues regarding scalability in sensor networks could not be addressed akin to ad hoc networks. Unlike ad hoc network nodes, sensor nodes are deployed to gather data from either point or distributed phenomena. This leads to a cluster of sensors having nearly the same information, leading to correlation among themselves. Individual nodes will have some combination of sensing, signal processing and communications capability and may self-organize for a variety of cooperative sensing and communication tasks, subject to resource constraints such as energy and bandwidth [69].

Basic information theoretic questions for such networks include the minimum resources required to extract information about some source to some level of fi-

delity - a rate distortion problem, and whether the network capacity per node is scalable subject to bandwidth constraints [98]. Fidelity encompasses such concepts as spatial or temporal resolution, misidentification probability or other accuracy measures, and network quality of service related measures such as latency from initial observation. Resource constraints can include signal processing cycles, energy consumption, and information rate. Even given that source-channel coding separability fails for network information theory problems, there remain many long-standing unsolved problems for source and channel coding treated individually. However, it is possible to derive rate regions in order to answer questions about whether the required information can be extracted or the network can scale [150, 211] given resource constraints. Hence, the scalability question here, is the information extraction at the desired fidelity.

Based on the application or interest, many different information theoretic questions could be asked for data gathering networks. Examples include the many-to-one model considered in [150, 82] and the many-to-many model considered in [211]. For the many-to-many correlated coding problem considered in [211], it was concluded that the information can be extracted at the desired level of fidelity for dense networks. However this is possible only if the information rate in each sample decays at least as fast as the throughput of the network. On the other hand, [150] discusses a many-to-one problem having significantly different set of objectives and concludes that successful extraction of information (asymptotically) is not possible. Their result is however for non-bandlimited[1] field and lossless coding. Hence, every sample will contribute to some extra information. Even on increasing the sampling rate, the problem cannot be solved as for perfect reconstruction, an infinite number of samples is needed. For this particular

[1]Non-bandlimited field implies that the power spectral density (psd) of the field is non-bandlimited. If the psd of the field is bandlimited, then it is called to a bandlimited field.

model, the efficiency with which the sensor network functions, degrades with the increase in the density of the sensors. Thus, the model of the sensor network and the objectives are critical to the conclusions regarding scalability.

In this chapter, the sensor network problem consists of extracting information about sources in some region to some desired level of fidelity, and transmitting this information to some gateway(s). Firstly, point phenomena such as the heat or acoustic sources are considered. For analytical simplicity, the point sources are assumed to be Gaussian. The problem of feasible rates for such networks is considered. An approach involves the dense deployment of sensors and communication relays. A sub-optimal decoupling of source and channel coding can be shown to be sufficient for achieving scalability in this context. Based on these results, we explore how the issues of scalability, source separation, and information extraction can be dealt with by altering the relative densities of sources, communication relays and the sensors. This is presented in Section 3.2.

The above mentioned technique of allowing dense deployment is, in general, not practical [174], although interesting inferences are obtained. To completely characterize the sensor network, the notion of *spatial fidelity* [173] is considered. In brief, spatial fidelity depends on the distance separating the point sources in Euclidean space. Section 3.3.1 considers the spatial fidelity for the point phenomena. Based on the spatial fidelity, source density, and desired distortion, the sensor density can be evaluated. The required sensor density for local cooperation can be also be calculated. These interesting source-sensors density relations are discussed in Section 3.3.2. The data fusion algorithms that might be employed are considered in Section 3.10(b).

The discussion is then extended for distributed phenomena such as a temperature field. The distributed phenomenon could be either bandlimited [206] or

non-bandlimited [150]. In practice, distributed continuous processes are, however, never fully observable. A typical approach in sensing is to sample the processes in time and space, in which the distributed phenomena are reasonably modelled as sets of correlated point sources.

The spatial fidelity criterion for distributed phenomena as considered in Section 3.5 transforms into a problem of sampling resolution. In other words, it dictates the rate at which the field should be sampled with the sensors. Obviously, for better reproduction of the field the resolution should be as fine as possible which leads to denser sampling of the field. But this, however, is highly unlikely in a real time situation. Hence, based on the field gradient or change in the sensor readings, the sensors could be either turned on or off resulting in energy conservation. For example, sensors deployed on highly smooth field [164, 249] such as C^0 or C^1 need only a few sensors functioning. On the other hand, an abruptly or frequently changing field needs a large number of sensors to have the desired reproduction quality.

For scalability in sensor networks observing distributed phenomena, one of the design principles is to have the bounded error. In other words, the errors should not propagate to distant nodes. Here, these are identified and then the techniques to bound them are discussed. In particular, a suboptimal but scalable interpolation strategy is considered in order to explore network density tradeoffs [146]. Section 3.6 considers these errors involved in sensing, transmitting and reproducing a distributed phenomenon.

Similar problems have also been considered in image processing [114, 199]. Many results from the image processing field may be applied here. However, the distributed nature of sensor networks must be considered. This results in communication cost, and hence a set of constraints on the rate. Additionally, the

sensors may be irregularly deployed and may be heterogeneous.

This chapter provides several insights into the information extraction problem and answers several vital questions regarding sensor density, cooperation, and sampling rate. This is done by considering an additional and significant parameter, spatial fidelity, that allows for complete characterization of the sensor networks. Key to above mentioned problems is the exploitation of spatial correlation that is inherent in sensor networks.

3.2 Consequences Of Decoupling Source, Sensor And Relay Densities

As discussed in Section 3.1, the basic problem in sensor network is to extract measurements of some physical phenomenon, to some desired level of fidelity, subject to constraints on energy consumption and bandwidth (resources). Nodes may also have explicit limits on signal processing and storage, which we will neglect here. By considering source, sensor and communications relay densities separately, we show that extraction of such information can be achieved without the requirement of complicated joint source-channel coding schemes, in the limit of high sensor and relay densities. Further, this formulation admits simple classification of a broad set of network information theory problems.

3.2.1 Spatial Source Separation

Consider a two-dimensional bounded region, $\mathcal{R}$. Assume a finite number of zero mean Gaussian point sources, say m, randomly located in $\mathcal{R}$. Unless stated otherwise, the point sources considered in this chapter are homogeneous [2]. These

[2]Homogeneous source means that all the sources are of the same type

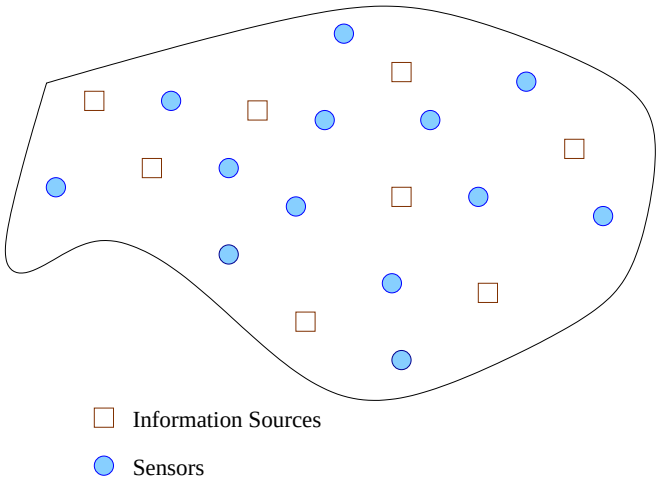


Figure 3.1: Sensor Network deployed in bounded region $\mathcal{R}$. Here, $p = O(m)$ where, m are the number of sources and p represents number of sensors.

sources need to be measured and reproduced at the fusion center within some desired fidelity. For sensing purposes, p sensors are randomly deployed with positive density over $\mathcal{R}$. Note that the locations of these sensors are i.i.d. Furthermore, power decays with an exponent $\alpha \geq 2$ with distance. Such a network model is depicted in Figure 3.1. The problem is now to find the rate distortion region for such a network assuming mean squared error distortion.

For now, suppose only one ($m = 1$) zero mean Gaussian point source, X, with variance σ_X^2 is present, sensed by p sensors. Let the observation available at sensor i be Y_i. Hence, the data available at p sensors can be represented by $\{Y_i\}_{i=1}^{p}$. The data available at each sensor can be mathematically modelled as:

$$Y_i = X + N_i, \qquad i = 1, 2, \ldots, p \tag{3.1}$$

where N_i, $i = 1, 2, \ldots, p$ represents the i.i.d. zero-mean Gaussian noise with variance σ_N^2 at the sensors. Note, that (3.1) can also be modelled as $Y_i = k_i X + N_i$ with non-identical independent noise at the sensors. The assumption of zero-mean allows the scaling to obtain (3.1). The encoding of these observations is

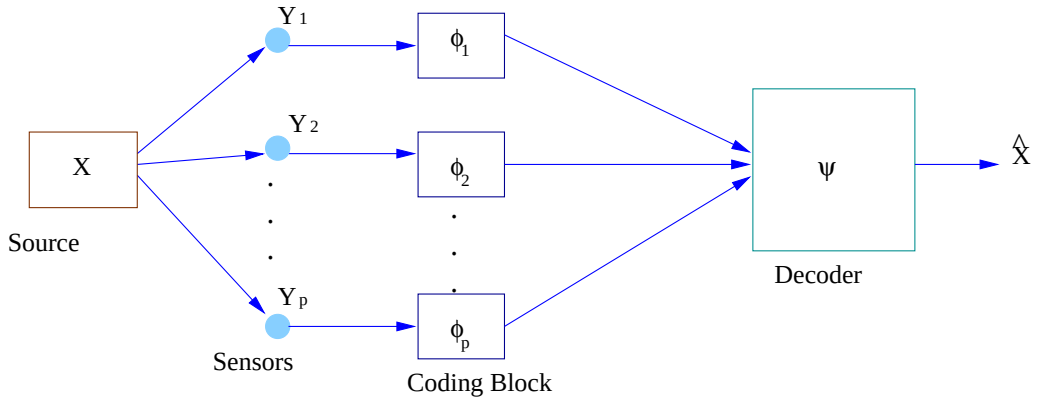


Figure 3.2: Gaussian CEO Model

done independently and are jointly decoded at the receiver. If $E(X - \hat{X})^2 \leq D$ where $E(\cdot)$ represents expectation, then this system reduces to the well known Gaussian CEO system described in [25, 41, 168]. Figure 3.2 depicts the Gaussian CEO coding model.

The sensors are correlated with the covariance matrix $\Lambda \in \Re^{p \times p}$. However, Y_i's are conditionally independent given X. For the above system, the sum-rate distortion function is given by:

$$\sum_{i=1}^{p} R_i(D) = \frac{1}{2} \log_2^+ \left\{ \frac{\sigma_X^2}{D} \left(\frac{D\sigma_X^2 p}{D\sigma_X^2 p - \sigma_X^2 \sigma_N^2 + D\sigma_N^2} \right)^p \right\} \tag{3.2}$$

where $\log^+ x = \max(0, \log x)$.

The Gaussian CEO system, however, cannot model the network represented in Figure 3.1. Since there is more than one source, the sensors must deal with interference from the different sources and hence the CEO model is not obvious.

When the number of sensor nodes in the network is of $O(m)$, not all distortions are achievable regardless of the rate constraints. This is because sensors may not be close enough to sources. Also, not all the rates are achievable because capacity may not be sufficient. Here, there is an interference among information streams

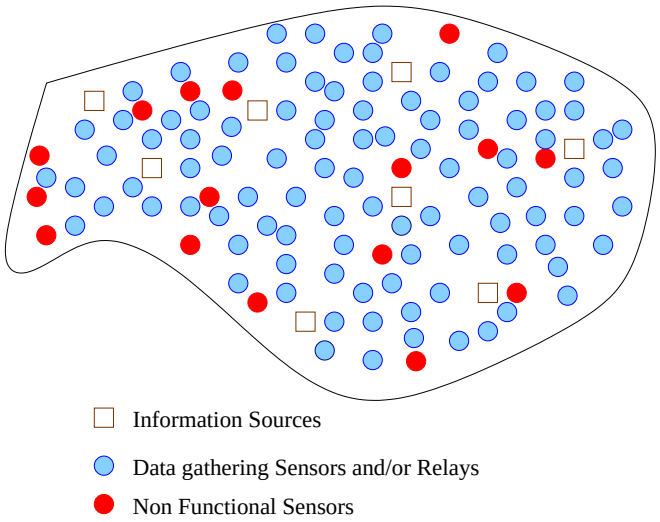


Figure 3.3: Gaussian sensor network with $p \gg m$. Sensors marked in red are turned off to conserve energy.

and in the signals received by each sensor and so a joint source-channel coding approach would be needed to achieve the largest rate region. There is little prospect of actually implementing such a system for large m, although clearly it is a rich regime for future research. More on joint source-channel coding is considered in Chapter 8.

Hence, for the network in Figure 3.1, the rate distortion bound is a hard problem to solve even for the Gaussian case. However, it can be readily converted to analytically tractable formulation by increasing the density of sensor nodes. This is depicted in Figure 3.3.

As the density of sensors is increased in a way that $p \to \infty$, there exists at least one sensor node in very close vicinity to the point source. In other words, there exists at least one sensor with no interfering signal for every source. This can be shown mathematically as in Lemma 3.2.1.

Lemma 3.2.1 *For the network in Figure 3.3, if $p \to \infty$ then, there exists at*

least one sensor node in a very close vicinity of the point sources with probability tending to 1.

Proof. Sensor nodes are distributed with positive density at every point in $\mathcal{R}$. So, for any $\delta > 0$, $\exists X_1$ s.t. $P(\|S_i - X_1\| < \delta) = \epsilon > 0$, where S_i is the sensor node and X_1 is the point source. For $i = 1, 2, \ldots, p$:

$$\begin{aligned} P(\|S_i - X_1\| < \delta) &= 1 - P(\|S_i - X_1\| \geq \delta) && (3.3) \\ &= 1 - P(\|X_1 - B_1\| \geq \delta)^p && (3.4) \\ &\rightarrow 1 \text{ as } p \rightarrow \infty \quad \blacksquare && (3.5) \end{aligned}$$

Also, not all sensors are required to gather information. The sensor closest to the source along with the highly correlated sensors could gather information depending on the rate or distortion constraints, while others could act as relays that purely carries traffic. A variety of algorithms exist that could be used to elect these nodes [84]. In addition, a few sensors which are not required as relays could be turned off saving energy consumption. In this scenario, the rate distortion bound for the network reduces to the ensemble of individual rate distortion bounds for each point source. Every point source corresponds to an independent Gaussian CEO system. Hence, we have the separation of point sources in the network. The clear distinction of this approach from [211, 150] is that not every node need to gather information, only the closest. In most cases, coding data from only two sensors locally is sufficient to achieve desired distortion level.

The above discussion is summarized in the following theorem:

Theorem 3.2.2 (Source Separation) *A network with finite zero-mean Gaussian point sources, say* m*, and number of sensor nodes,* p*, such that* $p \gg m$*, can be modelled with* m *separate Gaussian CEO coding systems, assuming that power decays at least as the square of the distance and* $p \rightarrow \infty$*.*

Proof. The proof follows from Lemma 3.2.1. Since $p \to \infty$, the frequency reuse distance goes to zero and thus there is no interference between information pathways through the network. Hence, a simple relay is adequate for carrying the traffic even if it is not capacity achieving. For the nodes that are in very close vicinity to the Gaussian point sources, the signal-to-interference noise ratio (SINR) is very large and could be considered as tending to infinity. So, any value of distortion, D, is achievable and there is no substantial interference among sources. Hence, local processing will be sufficient. In the scenario considered here, each Gaussian point source can be separated from the rest and source coding can be posed as a Gaussian CEO system. This is because Gaussian point sources are independent of each other. Since we have m point sources in the network of Figure 3.3, it can be seen as m separate Gaussian CEO systems.

The assumption that the signal decays with distance faster than some particular rate is required to avoid interference growing without bound for large fields of sources. This assumption is quite reasonable for typical deployments of sensor networks for many physical phenomena of interest. ∎

From Theorem 3.2.2 and limiting cooperation to only q sensors, $q << p$, it is evident that the data rate, R_{X_i}, $i = 1, 2, , \ldots, m$, associated with each point source is given by (3.2). In most cases, $q = 1$ or 2 would be sufficient. The rate distortion bound for the network will be the ensemble over the position of all the point sources. Similar to the rates, the achievable distortion for the network will be $D = \sum_{i=1}^{m} D_{X_i}$.

Now, it is also possible that for given values of m, p and D, the capacity of the network may be inadequate. However, by allowing the number of communication relays $n \gg p$ then the information can be extracted. Large over-provisioning will enable decoupling of source and channel coding.

The sensor network model in this chapter maps closely to the distributed Gaussian CEO model considered by Pandya, et. al. in [171]. However, the assumptions and constraints considered in [171] limits an application of their result for this case. The model assumed the same density of sources and sensors. This limits the relative densities and also channel coding was not considered. Hence, the relaying of traffic cannot be addressed in this context. Moreover, the observations available at a sensor were considered helpful information rather than interference. Also, the channel matrix is constrained to be non-singular. However, in a practical system, some sources will certainly act as interference. Also, the data is required to be transmitted to a local fusion center where only a few sources need to be reproduced. Note that even for the system in Figure 3.2, only source coding was performed. Thus, the issues of decoupling source and channel cannot be answered by CEO systems.

3.2.2 Comparison of Scalability Results

Decoupling of source and channel for sensor networks in this section could not be considered practical and may be trivial. But based on the relative densities of sources, sensors and relays interesting comparisons with the previous known results could be sketched. This is summarized in Table 3.1.

The scalability question considered in [98] assumed the same number of sources, sensors and relays. Their conclusion is that the network is non-scalable. However, if local communication predominates, then scalability could be achieved (Chapter 2). The problem of information extraction considered in [211] had a model with a fixed number of sources. For these sources, the same number of relays and sensors were considered. It should be noted that sensor and relay densities are assumed to be greater than source density. Their conclusion was that in-

Question Asked	**Source Density**	**Sensor Density**	**Relay Density**	**Solution**
Scalability?	x	x	x	if S-D pair distribution local
Information Extraction?	Fixed, y	x	x	if correlation increases fast enough [211]
Scalability and Information Extraction?	Fixed, m	Distortion Dependent, p	Traffic Dependent, n	Source Separation

Table 3.1: Comparison of Scalability Results. x, y, m, n, and p represent the densities.

formation could be extracted as long as correlation among sensors increases fast enough. In this chapter, both scalability and information extraction problems are considered. From the source separation principle, both these questions could be positively answered as long as the sensor density is based on the desired distortion and relay density is traffic dependent for the fixed number of sources.

3.3 Spatial Fidelity and Collaborative Processing for Point Phenomenon

Continuing the discussion on point phenomena, the previous section considered the decoupling of source and channel coding in sensor networks by altering the relative densities of sources, sensors and relays, thus avoiding the otherwise unsolved problem of joint source-channel coding in the networks. However, deployment of

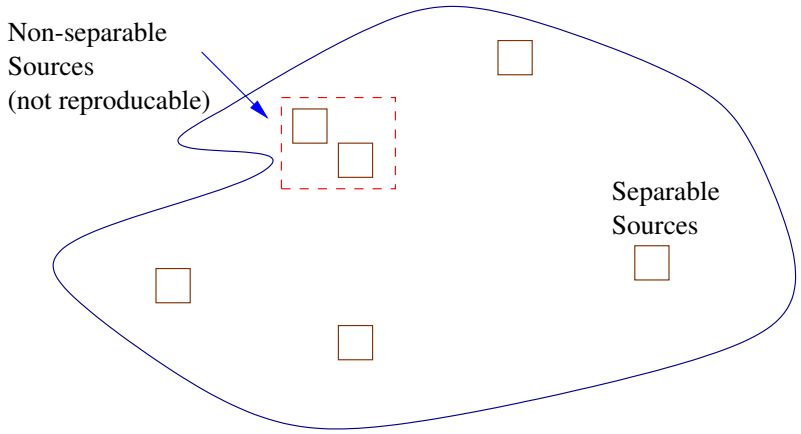


Figure 3.4: Separable and non-separable point sources.

infinite sensors is not practical. The sensor density is usually finite and hence the sources are not always separable. If the point sources are near to each other as in Figure 3.4, in some cases even overlapping, there would be significant interference at a sensor, making difficult any effective sensing and thus reproduction of either of the sources. Hence, the sensor network design should incorporate these situations and for this purpose the notion of spatial fidelity is considered.

3.3.1 Spatial Fidelity

Assume that the power decays with the distance according to an exponent $\alpha \geq 2$. In order to have any reasonable sensing of the point sources there should be a minimum separation between sources in space. This will either completely or partially suppress the interference from other sources. Based on this, the following definition of spatial fidelity is presented.

Definition 3.3.1 (Spatial Fidelity for Point Phenomenon)

Spatial fidelity, δ_{ij}, is the minimum separation between any two homogenous sources i and j, $i \neq j$, in Euclidean space to attain the desired detection proba-

bility or reproduction quality. Thus,

$$\|X_i - X_j\| \geq \delta_{ij}, \qquad \forall\ i, j\ (i \neq j). \tag{3.6}$$

where $\|\cdot\|$ denotes the Euclidean norm. Hence, spatial fidelity, by analogy to image processing, represents resolution. This chapter assumes the same value of spatial fidelity for any two homogeneous sources, $\delta_{ij} = \delta$, $i \neq j$. The value of δ depends on the applications and the desired distortion in the reproduced source.

In sensor network applications, if the sources do not satisfy spatial fidelity constraints then they are considered to be non reproducible. For instance, if two heat sources are less than some distance away to have any reliable reading, the sources are considered to be indistinguishable. However, there could be a scenario where the sources are heterogenous and even if they are not satisfying spatial fidelity constraints, the reproduction is required. In such scenarios, the sensors deployed are also heterogenous and each sensor has the specific function to accomplish and hence the signals from other non-intended sources will be treated as noise instead of interference. For instance, say an image source and sound source are sitting next to each other in a way that would not satisfy spatial fidelity constraints. For this application, the sensors deployed would be a mix of image and acoustic sensors. So for an image sensor, sound source is not an interference but noise. Hence, it is very important to consider spatial fidelity only for the homogeneous sources being sensed by the homogeneous sensors.

It should be noted that on considering spatial fidelity, the scalability issue is inherently solved. This is because, only those sources are now considered that are reproducible.

3.3.2 Collaborative processing

Spatial fidelity allows for the deployment of finite sensors for the purpose of sensing. The number of sensors required to observe the sources depends on the number of sources. Since a cluster or group of sensors is involved in gathering data from the same source, the sensors can cooperate by locally fusing their data, hence limiting the number of transmitting sensors and also exploiting spatial correlation. Here, the problem of required sensor density for given spatial fidelity, number of sources, and desired distortion is considered. Also, the number of sensors needed for local fusion is also evaluated.

3.3.2.1 Density of Locally Cooperating Sensors

Consider a circular region of radius R_s containing m randomly located point sources. Also deployed are p uniformly distributed sensors as shown in Figure 3.5. Assume that the number of sources and sensors along with the spatial fidelity is known. For the purpose of this chapter, assume that the number of sensors per unit area, ρ, remains constant. Mathematically, ρ, is defined as:

$$\rho = \frac{p}{\pi R_s^2}. \tag{3.7}$$

As discussed earlier in the chapter, each source has a group of sensors observing it rather than one dedicated sensor. Hence, multiple copies of the same information are available. However, based on the distance between source and sensor, the observation quality at each sensor will differ. In any case, it is advisable to have only one transmission per source. Ideally, the sensor with the best (or desired) signal-to-interference noise ratio (SINR) should be considered. However, if no sensor has the desired signal-to-interference noise ratio then the nodes can cooperate locally. Local fusion is also encouraged for reliable detection [242].

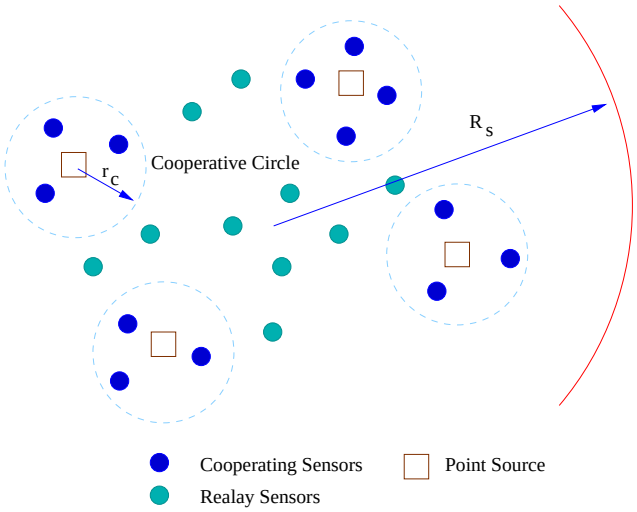


Figure 3.5: Data Gathering system multiple point sources and their respective cooperative region.

For the purpose of local fusion, there exists a cooperating circular region of radius r_c around each source. The radius of this region certainly varies from source to source, but for the sake of simplicity assume it to be the same for all sources. Also, assume that the source localization has been already performed and hence the sensors are aware of the relative distance between them and any given source. This assumption is only for analytical tractability. Later, this assumption will be relaxed for calculating sensor density. The cooperating regions could either be non-overlapping or overlapping for any particular source. Consider a source X_m being observed by q sensors lying within a cooperative region as shown in Figure 3.6. The interference at the sensors is assumed only due to sources and not the communication interference from the other sensors. This is a valid assumption as the appropriate choice of medium access control (MAC) layer protocol can always prevent the communication interference. The SINR at sensor i lying in

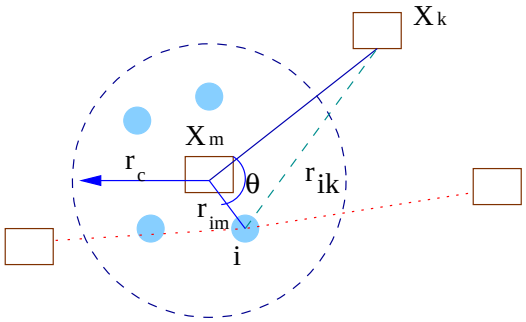


Figure 3.6: Source-sensor positions for SINR calculations.

the cooperative region of X_m is then:

$$\text{SINR}_i = \frac{P_{X_m} \left(r_{im}/r_0\right)^{-\alpha}}{\sigma^2_{N_i} + \sum_{k=1}^{m-1} P_{X_k} \left(r_{ik}/r_0\right)^{-\alpha}} \tag{3.8}$$

where α is the path-loss coefficient, r_{ij} is the Euclidean distance between the sensor i and source j, r_0 is the constant of proportionality, P_{X_j} is the transmitted power by source X_j and $\sigma^2_{N_i}$ is the variance of zero-mean Gaussian noise, N_i, at sensor i. The denominator of (3.8) is the interference at sensor i:

$$I_i = \sigma^2_{N_i} + \sum_{k=1}^{m-1} P_{X_k} \left(r_{ik}/r_0\right)^{-\alpha} \tag{3.9}$$

From geometry, see Figure 3.6, the distance between the sensor i and source k, $k = 1, 2, \ldots, m-1$, can be calculated to be:

$$r_{ik} = \sqrt{r^2_{X_m X_k} + r^2_{im} - 2 r_{X_m X_k} r_{im} \cos\theta} \tag{3.10}$$

where $r_{X_m X_k}$ is the distance between the source X_m and X_k, and θ is the measure of $\angle i X_m X_k$. If the spatial fidelity constraint, δ, is satisfied by the source X_k with X_m, then $r_{X_m X_k} \geq \delta$. If the constraint is not being satisfied then both the source X_m and X_k would not be considered for the reproduction at the fusion center. Hence, it is clear that the spatial fidelity will influence SINR_i. The larger the source separation, the higher the SINR at the sensor.

Firstly, consider the statistical approach. The probability that the sensor node will be in radius r_c is:

$$P_{r_c} = \frac{\pi r_c^2}{\pi R_s^2} = \frac{r_c^2}{R_s^2} \tag{3.11}$$

Using (3.7), the above equation can be rewritten as:

$$\mathrm{P}_{r_c} = \frac{\rho \pi r_c^2}{p} \tag{3.12}$$

Therefore, the probability that the number of sensors, q, in the cooperative region of X_m is k:

$$\Pr(q = k) = \left(\begin{array}{c} p \\ k \end{array} \right) \mathrm{P}_{r_c}^k (1 - \mathrm{P}_{r_c})^{p-k} \tag{3.13}$$

If $p \to \infty$, then

$$\lim_{p \to \infty} \Pr(q = k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{3.14}$$

where $\lambda = \rho \pi r_c^2$. Hence, asymptotically the number of sensors in the cooperative region follows a Poisson distribution with intensity $\rho \pi r_c^2$.

Suppose X_m to be a zero-mean Gaussian with the variance $\sigma_{X_m}^2$. Consider the q sensors observing X_m within the cooperative region. The observation available at each sensor could be modelled as:

$$Y_i = \gamma_i X_m + Z_i \tag{3.15}$$

where i ranges from $1, \ldots, q$ and $\mathrm{SINR}_i = \frac{\gamma_i^2 \sigma_{X_m}^2}{\sigma_{Z_i}^2}$. Here, Z_i also considers the interference from other sources. The interference, I_i, in (3.9) is certainly not Gaussian for a finite number of sources. But, if $m \to \infty$, then from the central limit theorem $\sum_{k=1}^{m-1} P_{X_k} (r_{ik}/r_0)^{-\alpha}$ tends to Gaussian. γ_i's are the inhomogeneous coefficients, modelling the practical scenario of not being identical. Based

on this, the mean squared estimate of X_m is given by [205]:

$$\hat{X}_m = \left[\left(\Gamma\sigma_{X_m}^2\Gamma' + R_Z\right)^{-1}\Gamma\sigma_{X_m}^2\right]' \mathbf{Y} \tag{3.16}$$

where $\mathbf{Y}$ denotes column vector $\{Y_i\}_{i=1}^{q}$, Γ is a column vector $\{\gamma_i\}_{i=1}^{q}$ and R_Z denotes the covariance matrix of noise $\{Z_i\}_{i=1}^{n_c}$. Since Z_i's are independent of each other R_Z is a diagonal matrix. If Z_i's are also identical with variance σ_Z^2, then the distortion in reproduction of source X_m is given by:

$$D = \frac{\sigma_{X_m}^2\sigma_Z^2}{\left(\Gamma'\Gamma\sigma_{X_m}^2 + \sigma_Z^2\right)} \tag{3.17}$$

The distortion, D, clearly depends on SINR which is a function of δ. This implies that $D = f(\delta)$. Using this and (3.17), the number of sensors needed for cooperation can be calculated in terms of spatial fidelity. For instance, assume Γ to be a vector of identical elements γ. In this case, $\Gamma'\Gamma = q\gamma^2$. Hence from (3.17),

$$q = \frac{\sigma_Z^2}{\gamma^2\sigma_{X_m}^2}\left[\frac{\sigma_{X_m}^2}{D} - 1\right]^+ \tag{3.18}$$

where $x^+ = \max(0, x)$.

The above considered the case of one source having q sensors locally cooperating. If m sources are present satisfying the fidelity constraint, then the sensor network will have a total of mq sensors locally cooperating for those m sources within their assigned cooperative region.

3.3.2.2 Calculating Sensor Density

The analysis for calculating the sensor density involving m point sources is highly complex and to some extent non-tractable. For instance, assuming the localization of sources as in the previous discussion results in a non-practical scenario and even that does not help in simplifying things. This is because calculating

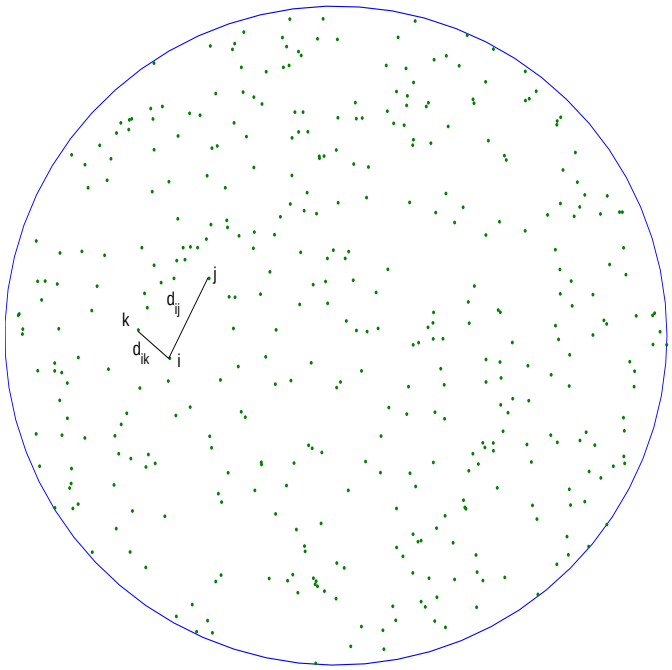


Figure 3.7: Simulation Setup for evaluating sensor density.

the cooperating sensors could be done but evaluating the total number of relays is highly traffic dependent. Even the statistical analysis is equally difficult. For example, consider the computation of Pr(distance between any two sources $\geq \delta$). Although the sensor nodes are i.i.d. in location, the Euclidean distance between them is dependent. Hence, calculating this probability for m sources itself is complicated. Hence, simulation is performed to evaluate the sensor density for a given spatial fidelity, number of sources, and desired mean squared distortion.

The simulation setup assumes a unit area circular region with m Gaussian point sources randomly deployed. The setup is as shown in Figure 3.7. The spatial fidelity is here the user specified parameter. For the purpose of the simulations, the area of the circular region is kept constant.

Firstly, the simulation is carried out to calculate the probability that the sources do not satisfy spatial fidelity criteria. That is, the number of sources

are calculated that are spaced at distance less than δ. For this purpose, m is varied from 10 to 200. The simulation for each value of m is executed for 1000 iterations and the value of probability is calculated averaging over those values. The probability plot for different values of δ is shown in Figure 3.8.

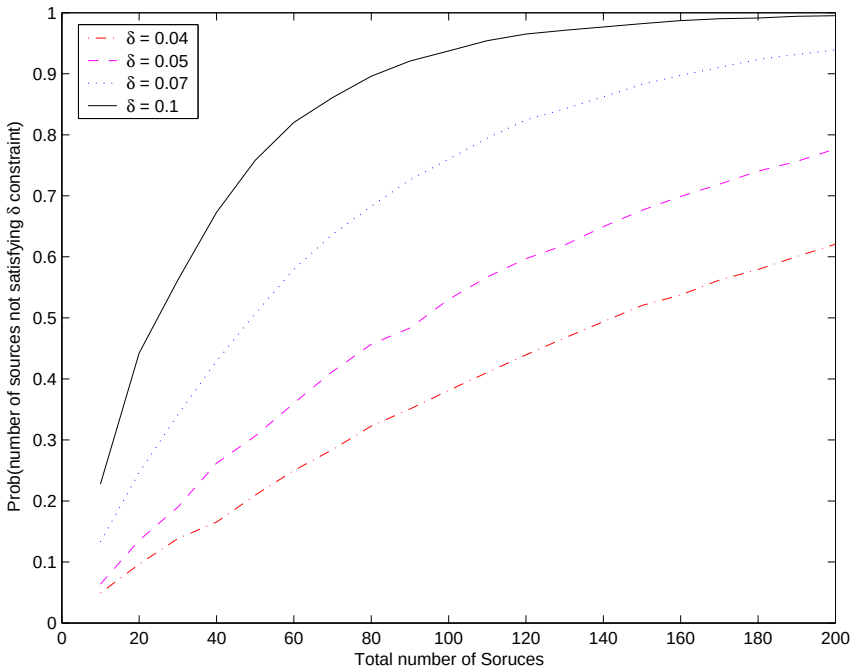


Figure 3.8: Probability plot for the number of sources not satisfying spatial fidelity constraints.

With the increase in number of sources, the spatial separation between them will certainly reduce as the area of the region is constant. Hence at higher node density, the resolution should be finer for the desired reproduction. As seen in Figure 3.8, at higher source density, the probability of sources disobeying spatial fidelity constraint, δ, increases.

Next, simulation for the sensor density is considered. For this purpose the number of sources, m, is fixed. The goal in this simulation is to know the mean squared distortion in the estimates. This will give the relation between the re-

quired number of sensors and distortion. The transmission power is assumed to decay in second order with the distance. The sensors observing the source are locally fused until the desired signal strength is obtained. More on fusion will be considered in next section. The simulation plot is given in Figure 3.9. Similar

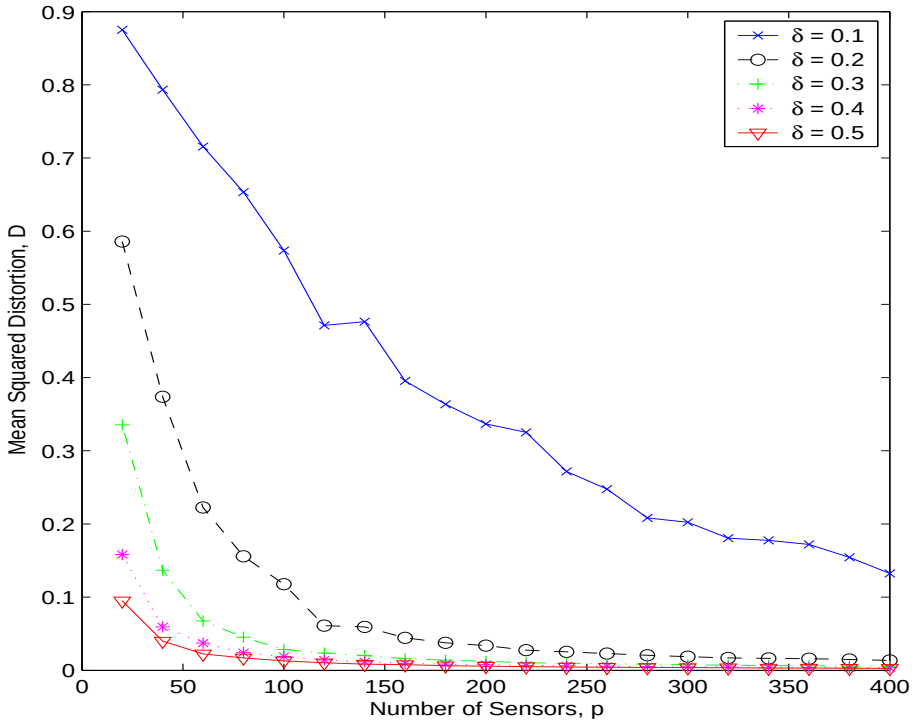


Figure 3.9: Simulation Plot depicting the relation between the required number of sensors and mean squared distortion for different values of δ. The path-loss coefficient is assumed to be 2.

to the probability simulation, each point in the plot is the ensemble average over 1000 iterations. For any particular δ, with the increase in number of sensors, the distortion should decrease. This is obvious as the dense deployment of sensors provides better signal strength at the sensor along with high spatial correlation and sensor diversity. Now, as the δ increases, then the number of sources that satisfy the spatial fidelity constraint decreases. Hence, the total number of sources

to consider for reproduction decreases. This leads to the deployment of fewer sensors for higher δ as seen from Figure 3.9.

To summarize this section, the sensor networks are not well described unless spatial fidelity is considered. Introduction of spatial fidelity, inherently solves the issues regarding scalability. In addition, it also makes the calculation of sensor density possible for the desired distortion.

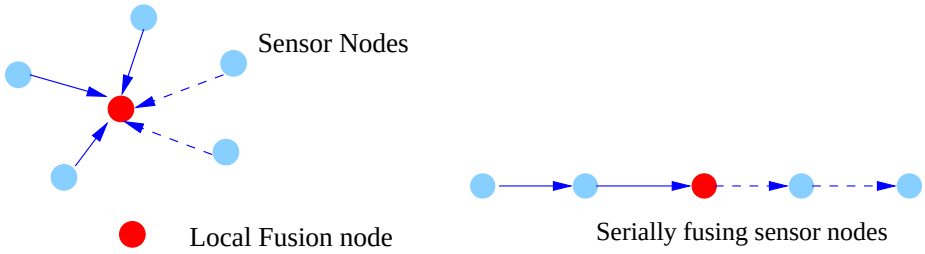
3.4 Decentralized Data Fusion Algorithms

Data fusion is encouraged in sensor networks to increases the reliability [242]. This section discusses the likely choice of data fusion structure and suggests an algorithm for decentralized fusion. Here, the fusion algorithm is presented for locally cooperating nodes. The data fusion could either include all the sensors in the cooperative region (Figure 3.5) or only a few depending on the requirements. For example, if the criterion is to have the best detection possible then it is advisable to have all the available sensors to participate in fusion. But, if only a certain level of detection is desired then not all the sensors need to participate in fusion. It is assumed that all the deployed sensor nodes have the capabilities to perform data fusion and have the knowledge of the SINR available to them.

The data aggregation in sensor network could take either of the following two forms:

1. Sensors transmit and fuse their information to a sensor leading to the lowest communication cost or with the best SINR among them. This is depicted in Figure 3.10(a). Note, that the selection of fusion node depends on the cost function.

2. Information is transmitted through chain of nodes and at every node data



(a) Fusion at one node (b) Chain or Serial Fusion

Figure 3.10: Data Fusion at the sensor nodes. The dashed arrow indicates the transmission only if required.

is fused. Once the desired SINR level is obtained, further transmission along the chain ceases. For instance, say A, B, and C form the chain as $A \rightarrow B \rightarrow C$. Node A transmits to node B where their data is aggregated. If the SINR threshold is attained then further fusion stops, otherwise B transmits to C. The transmission, in general, begins from the node with maximum SINR among the group. Figure 3.10(b) shows the chain (or serial) fusion.

It is typically preferred to have chain fusion. However, it depends on applications. In this chapter, the focus is on chain fusion rather than one node fusion. Extensive research on data fusion has been done [130, 131, 145, 165, 192, 242] and still a lot has to be done yet in terms of optimality.

Previous work on the chain fusion algorithms, considered the latency as the optimizing parameter [145]. However, delay is not a constraint for local fusion. This is due to the fact that all the sensors within a local region are not distantly placed to cause notable delay and also generally the sensors locally form a complete graph (Figure 3.11), i.e. they are within one-hop distance. Here, the algorithms are proposed for two cases:

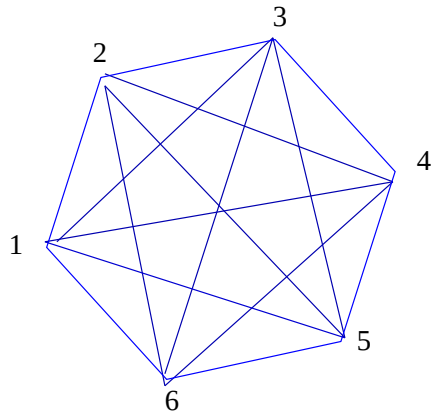


Figure 3.11: Nodes forming Complete Graph.

1. when all the sensor nodes in a cluster carry out fusion so as to increase estimation quality i.e. to reduce the mean square error involved in reproduction, and

2. when not all the nodes are required to participate in fusion, e.g. habitat monitoring may not need all nodes to locally fuse their data.

The solution for the first case where all the nodes cooperate, is formulated as a decentralized travelling salesman problem (TSP) [16, 129] and minimum weight spanning tree problem [77] on a complete graph. The objective is to find a minimum power cost fusion chain and tree respectively. The TSP is a well known NP hard problem. For the second case, the problem is formulated as a generalized Steiner tree [230, 250]. We propose a heuristic that prunes the nodes to get a minimal power spanning tree achieving the desired SINR.

3.4.1 All Nodes participating in cooperation.

Consider an undirected complete graph G, of size M, with distinct labels $i = 1, \ldots, M$. The communications are assumed to be symmetrical so that the edges

are undirected with communication power as the associated weight, That is,:

$$e_{i,j} = e_{j,i} = P_{i,j} \tag{3.19}$$

where, $e_{i,j}$ defines cost for an edge between nodes i and j and $P_{i,j}$ is the communication power required between the nodes. Also, note that $P_{i,j}$ depends on the distance between the two nodes and hence, $P_{i,j} = f(d_{i,j})$. The goal is now to find the Hamiltonian path, i.e. a path containing all the nodes only once, or spanning tree with minimum total weight. The total weight in this case is power. Since TSP is NP hard, there does not exist any polynomial time algorithm unless $P = NP$. It should be clear that the sensor nodes considered here belong to a local cluster observing some phenomenon. Based on this, we now propose the algorithms.

Algorithm 3.4.1 (Forming Fusion Chain)

1. Start from random node $v = u$
2. For current sensor v, go to the next sensor w in such a way that

$$e_{v,w} = \min_{s \in \text{neighbors of u and not in the chain yet}} (e_{v,s}).$$

3. repeat step 2 till all the sensors are added to the chain.
4. Each sensor pair (m, n) change their position in the chain, if the swapping strictly reduces the total cost.

This algorithm is based on the Hamiltonian path finding for disk scheduling proposed by Gallo et. al. [78]. The complexity involved in this approach is $O(M^3)$ and reaches a constant factor times the optimal result. More sophisticated algorithms also exist that can reach $1+\epsilon$ times optimal, $\forall \epsilon > 0$, within polynomial

time but are not suitable for distributed applications like sensor networks. Note that, Algorithm 3.4.1 can also be extended to the asynchronous case, where the weight from i to j is different than the weights from j to i. Also, the asynchronous case is more practical for many wireless sensor networks due to multi-path fading.

Algorithm 3.4.2 (Forming Minimum Span Tree (MST))

For this purpose, the distributed algorithm by Gallager et. al. [77] is considered. Each sensor operates in three states: *Sleeping*, *Find*, and *Found*. Sensors also keep the states of their edges as *Basic*, *Branch*, and *Rejected*. It is possible for the two neighboring nodes to have temporarily inconsistent state of an edge. A set of procedures is defined for different states of the system. A random sensor is awakened and starts the algorithm. For the sake of brevity, the complete routines of the algorithm are skipped. The following lemma summarizes the result:

Lemma 3.4.3 *For the complete graph G of size M, the total number of messages required is at most* $5M\log(M) + 2 \times M(M-1)$.

3.4.2 Fraction of the nodes involved in fusion.

Here, not all the sensors are considered for fusion. The number of sensor nodes is limited by the desired SINR threshold. Once the desired SINR is achieved, no further fusing of data is done. This problem can be formulated as finding a Steiner tree on the complete graph, G. Since not all the sensors take part in fusion, the set of sensors could be divided into two categories - set of required sensors (vertices), and set of Steiner sensors (vertices). The set of required sensors is the ones that are actually needed to achieve a certain SINR but may not be optimal in terms of weight. The other set consists of the rest of the sensors. For example, if we have a local group of 10 sensors and we need only 4 to achieve the

desired SINR then these 4 sensors form the required set, while the remaining 6 are in the Steiner set. The problem is then defined as finding a minimum cost tree in G that contains all the required sensors and any subset of Steiner sensors. A decentralized heuristic is proposed to find the suboptimal tree in scalable time and communication cost. There are various heuristics available for the Steiner tree [250]. One of the widely used heuristics is minimum-cost paths heuristic (MPH) [230]. This heuristic has worst-case cost performance of twice optimal cost solutions and no heuristics with a better worst-case are known [250].

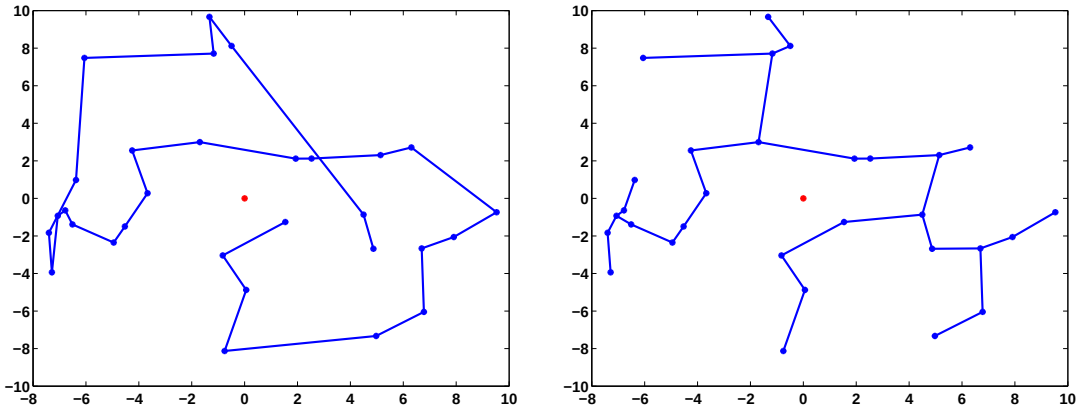
Algorithm 3.4.4

Consider a graph $G(V, E)$ with non-negative costs e (communication power) of the edges, and a subset of V ($U \subseteq V$). This is assumed to be known. Now, the problem is to find a subgraph, T, of G, such that there exists a path between any pairs of U, and the total power cost T is minimized.

1. Start from a random sensor $u \in U$. Add u to T.
2. Add the closest neighbor of u, say w, that is not in T but is in U. Add w to T and set $u = w$
3. Repeat step 2 till $U \subseteq T$ or $u \in T$ has a neighbor only in U (and not in T).

3.4.3 Simulation Results

Here, the simulation of the proposed algorithms in sections 3.4.1 and 3.4.2 is presented. We consider Monte Carlo simulations for a single source with $M = 30$ sensors, uniformly distributed in a circle of radius 10 meter. The source is assumed to be at position $(0, 0)$.



(a) Suboptimal Minimum Weight fusion chain using decentralized algorithm 3.4.1

(b) Optimal minimum weight fusion tree using decentralized algorithm 3.4.2

Figure 3.12: Simulation plots for the Algorithms 3.4.1 and 3.4.2. Red circular node in the center indicates the source.

The simulation for Algorithms 3.4.1 and 3.4.2 is depicted in Figures 3.12(a), and 3.12(b).

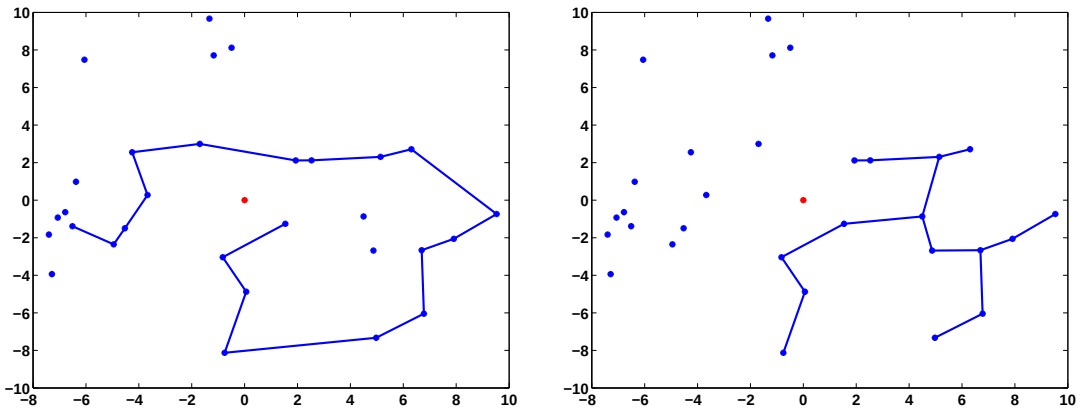
Figures 3.13(a) and 3.13(b) show the results of Algorithm 3.4.4 where we assume some desired SINR and restrict the number of nodes involved in fusion based on that. Table 3.2 summarizes the result for the proposed algorithms. Note that total power cost decreases dramatically (almost three times) for the case when not all the sensors cooperate as expected. This serves as motivation to limit the number of cooperating sensors till the desired SINR is achieved.

The transport of locally fused data to the assigned fusion center is the problem of global routing in the network. For this purpose the routing algorithm suggested in Chapter 2 (Section 2.2.2), [200] could be employed.

Bayesian hypothesis testing for chain or serial networks as in Figure 3.10(b) is summarized in [242]. Along with the serial networks, various other topologies such as parallel fusion networks, tree networks, and networks with feedback are

Proposed Algorithms (Heuristics)	Number of Sensors	Normalized Transmitted Power (Watts)	Aggregated SINR(dB)
Algorithm 3.4.1	30	0.29	10.02
Algorithm 3.4.2	30	0.29	10.02
Algorithm 3.4.4(chain)	18	0.12	3.62
Algorithm 3.4.4(tree)	15	0.07	3.84

Table 3.2: Comparisons of proposed heuristics for different values of desired SINR



(a) Suboptimal Minimum Weight fusion chain achieving desired SINR using decentralized algorithm 3.4.4

(b) Suboptimal Minimum Weight fusion tree achieving desired SINR using decentralized algorithm 3.4.4

Figure 3.13: Simulation plots for the Algorithm 3.4.4. Red circular node in the center indicates the source.

also considered. In practise, it is advisable to use the detection rule that is also valid for non-ergodic phenomena. For instance, if sensors are deployed to detect fire then they are sensing a non-ergodic process.

3.5 Spatial Fidelity for Distributed Phenomenon

The prior sections considered the data gathering system with point phenomena. This and the next section considers the sensing of distributed phenomena. Distributed phenomena such as fields could be either bandlimited or non-band limited. The sampling of these fields is one of the most challenging areas of research. In an ideal case, sampling above the Nyquist rate avoids any aliasing effect. The assumption of super Nyquist sampling also holds for this chapter. However, determining whether the field is under sampled, over sampled, or critically sampled still poses an interesting problem. These sampling issues have been considered in [83, 113, 136]. Here, sampling of the phenomenon is considered in a different context. Note that while collectively sensors are observing a distributed source, each of them is collecting data from a point source. All together, these highly correlated point sources form the distributed phenomena.

Section 3.3.1 discussed the notion of spatial fidelity for point phenomena. We now extend that definition for distributed phenomena. The idea behind the spatial fidelity, however, remains the same. That is, it represents resolution of the field.

Definition 3.5.1 (Spatial Fidelity for Distributed Phenomenon)

Spatial fidelity, δ, for the distributed phenomenon represents the cut-off or sampling rate. It is governed by the gradient or the change in the amplitude of the field with respect to time. Consider the distributed field as in Figure 3.14. Sup-

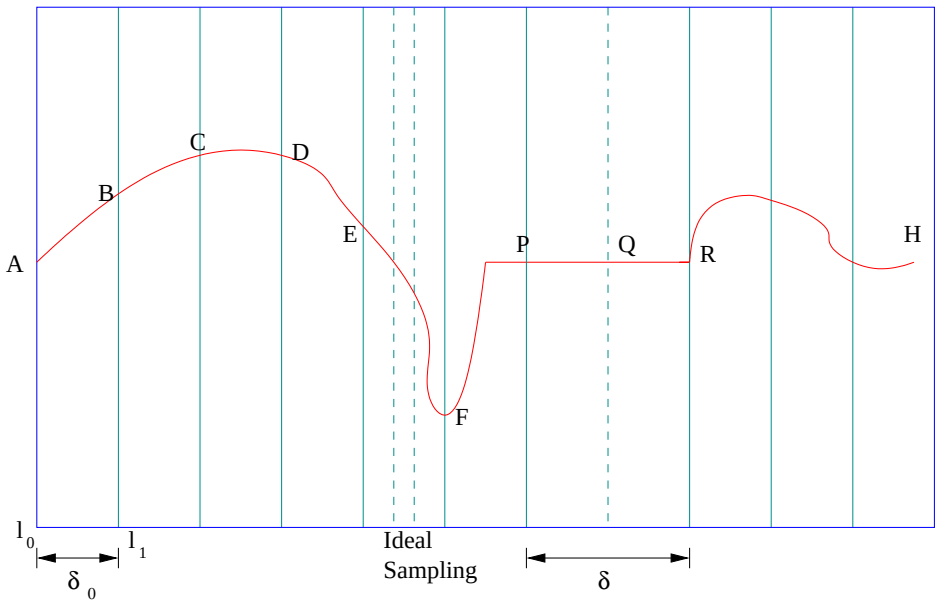


Figure 3.14: Distributed field with sampling interval, δ_0 (sampling frequency $= 1/\delta_0$. The dashed lines between points E and F represent an ideal situation.

pose the field is sampled at every δ_0 units. Hence, the spatial sampling frequency of the field is $\frac{1}{\delta_0}$. If change in the amplitude is represented by ΔX, then the slope of a curve within a sample is given by (Figure 3.14):

$$\zeta = \lim_{\Delta l \to 0} \frac{\Delta X}{\Delta l} = \frac{\partial X}{\partial l} = \frac{\partial X}{\delta}$$

However, this frequency may not result into true reproduction of the field. Consider the curve EF. It abruptly changes from E to F and hence, it is desirable to have more samples (or higher sampling rate) for the better reproduction of the field. This situation is an example of under sampling. In contrast, the region PR is smooth and flat. For this region, the sampling rate could be low, that is the sample at point Q is not required. With the readings at P and R, the region could be accurately reproduced. This situation corresponds to over sampling. The critically sampled region is AB in Figure 3.14. All these contribute to energy cost in sensor networks. Hence, the sampling should be adaptive. This

adaptive nature of sampling is represented by spatial fidelity. To summarize, let δ_0 be the user specified spatial fidelity criterion. If i and j represent two points on the field, then the following three conditions for spatial fidelity, δ, could be possible:

Under Sampling

$$\left|\frac{\Delta X}{\delta}\right| > \epsilon \qquad \text{and} \qquad \delta < \delta_0 \tag{3.20}$$

Critical Sampling

$$\left|\frac{\Delta X}{\delta}\right| \in [0, \epsilon_1] \qquad \text{and} \qquad \delta = \delta_0 \tag{3.21}$$

Over Sampling

$$\left|\frac{\Delta X}{\delta}\right| \in (\epsilon_1, \epsilon] \qquad \text{and} \qquad \delta > \delta_0 \tag{3.22}$$

where, ϵ and ϵ_1 are positive quantities. The above three conditions define spatial fidelity for distributed sources. Note, that the above sampling criteria is proposed as heuristics.

In most cases, the user provides the lower bound on the spatial fidelity. If the lower bound is δ_0, then $\delta \geq \delta_0$. With this constraint then we can reproduce the field with the desired distortion. The error in reconstruction may occur from the situations such as the region between C and D in Figure 3.14. The readings at both these points are the same but there is change in the readings in between which cannot be captured. However, this error does not propagate and is limited to local region.

Since the distributed source could be observed as the collection of correlated point sources, the source-sensor relations derived for point sources could be extended for distributed phenomena also.

3.6 Errors in Data Gathering System observing Distributed Phenomenon

Consider a sensor network deployed to observe a certain distributed phenomenon such as a temperature field. The sensed data are transmitted to the fusion center. Due to stringent constraints on the available data rate, the correlation among sensors should be exploited to bring down the communication cost. At the fusion center, an interpolation or approximation algorithm is used to reconstruct the source. The following errors are encountered during the process of observing the phenomenon through to reconstruction.

1. **Sensing Error:** Sensing errors are due to the ambient and circuit noise of sensor nodes, and can be reduced by increasing the number of independent observations.

2. **Quantization Error:** This error occurs during the process of quantizing (source coding) the observed data. Quantization error is constrained by the limited data rate of the network.

3. **Interpolation Error:** For a distributed source with a continuous sample path, the source is usually reconstructed at the fusion center by interpolating the measured points. This causes an interpolation error. The interpolation error is affected by both the spatial sampling rate (or mesh size) and the former two types of errors.

This section shows that the local processing in sensor networks need not result in error propagation to distant nodes.

3.6.1 Error due to Sensing

As seen, the observations at different sensors for a distributed phenomena can be considered as data generated from correlated point sources. For reasons of network sustainability, the data rate and energy is constrained. Hence, the correlation among sensors ought to be exploited to cut the rate needed to transmit the data samples.

Consider sensor i, observing the zero-mean Gaussian source with variance σ_X^2. The observation at the sensor could be modelled as:

$$Y_i = a_i X + Z_i, \qquad i = 1, 2, \ldots, p \tag{3.23}$$

where Z_i is the i.i.d. zero-mean Gaussian noise with variance σ_Z^2, and a_i is the attenuation factor given by:

$$a_i = \frac{1}{1 + \kappa r_i^2}$$

where r_i is the distance between source and the sensor with constant κ that controls the dependence of attenuation on distance.

Now, we consider the effect of sensor density on the accuracy of sensing. This is depicted in Figure 3.6.1. For simulation purposes, different quantization techniques are employed. The sensing accuracy is studied under these schemes. It can be seen that for a relatively sparse sensor network ($n \leq 50$), the sensing error is the major contribution to the total distortion. Hence, the increase in sensing accuracy (by deploying more sensors) leads to a significant drop in distortion. As the sensor network gets denser, quantization error due to insufficient rate starts to dominate, and a rate increase reduces the distortion.

To reduce the sensing error, the data gathered by the sensors should be locally fused as discussed in the earlier sections.

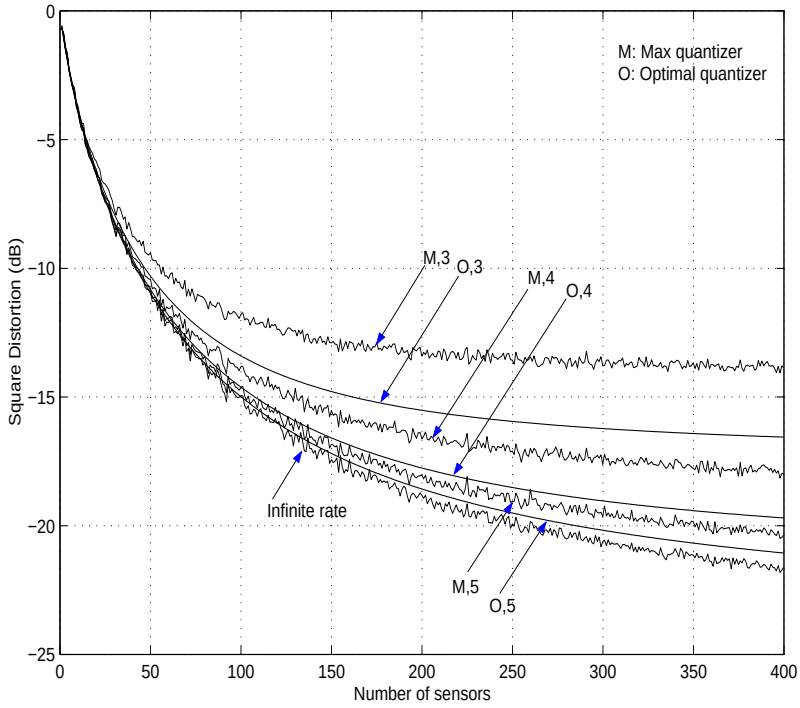


Figure 3.15: Comparison of the distortion of a Max quantizer with the optimal quantizer. Here, $\sigma_X^2 = 1$, $\sigma_Z^2 = 0.0032$, $\kappa = 500$. $M/O, r \equiv$ Max/optimal quantizer with quantization rate r bits per sample; Infinite rate $\equiv$ quantizer with infinite rate.

3.6.2 Error due to Quantization

When observing a single point source, the sensing error often dominates since a relatively small amount of information needs to be transmitted to the global fusion center. However, there is a great amount of information embedded in a distributed source. Thus, there exists a trade-off between the reduction of quantization error and the constrained network resource. Also, due to the spatial correlations of a distributed source, it is possible to significantly reduce the rate at which sensors transmit to the global fusion center. Usually attaining this

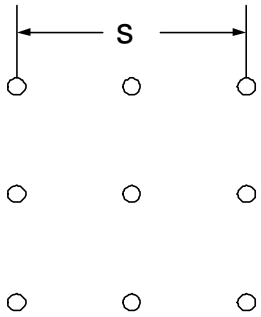


Figure 3.16: Placement of nine sensors on a square grid in random field.

reduction entails intensive local interactions among sensors. This kind of trade-off is desirable because the energy (capacity) constraint on global transmissions (between sensors and global fusion center) is far more severe than local transmissions (among sensors and local fusion centers), whose range is bounded [174]. The data rate limits on coding correlated sources have long been studied by the image processing community.

Consider the rate-distortion problem for a two dimensional isotropic random field with correlation function $e^{-|r|/d_c}$, where d_c is the coherence distance. Nine sensors are placed in a square grid, and s is defined as in Figure 3.16.

The minimum total rates required for transmitting the data collected at these nine sensors to the global fusion center are plotted in Figure 3.17. It can be seen that as the sensors become closer (with s decreasing), the correlation among sensors increases. As a result, the data rate needed to transmit to the global fusion center decreases. Moreover, this rate drop can be accomplished by local fusion.

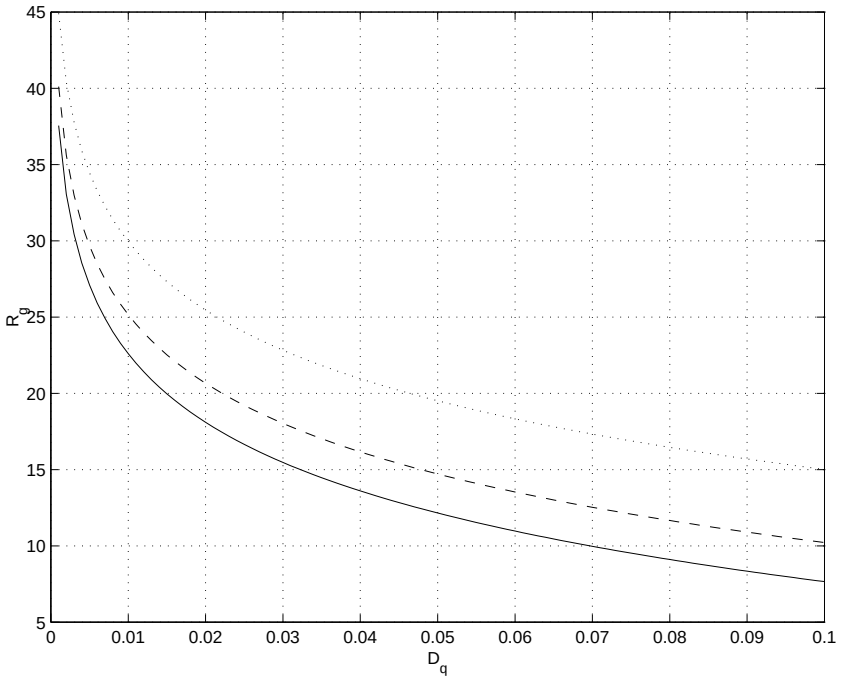


Figure 3.17: The comparison of rate-distortion with different coherence distance. Here, $a_i = 1, \sigma_X^2 = 1, \sigma_Z^2 = 0.01$, solid line $\equiv s/d_c = 0.4$, dashed line $\equiv s/d_c = 0.7$, dotted line $\equiv s/d_c = \infty$.

3.6.3 Error due to Interpolation

The distributed source is assumed to be continuous or at least piecewise continuous in space (for the latter, we consider one continuous piece of the sample path). The global fusion center considers the use of local interpolation algorithms such as spline fitting for the reconstruction of the distributed source. Usually, the measurements at prescribed points are assumed to be error free. In contrast, we consider the effect of sensing and quantization errors on the result of interpolation.

For reconstruction, the source is implicitly considered to be deterministic. However, it is considered to be a random process for exploiting the limits of source

coding rate. The reconciliation of these two models is reached as follows. First, minimum source rates are obtained only by considering jointly coding the i.i.d. blocks of source realizations. Second, evaluating the minimum rate demands some knowledge of the distribution of the source field. Spline fitting takes advantage of the correlation among local observations embedded in the continuity of the source. Third, information embedded in a distributed source is never completely conveyed to the fusion center. For a continuous sample path, spline fitting makes reasonable estimates of the missing data, provided that the sample points are closely spaced.

The derivation of this section is based on the cubic spline theory [4]. For the sake of simplicity, we first consider the absolute error, $d(X, \hat{X}) = |X - \hat{X}|$, and then shall extend the arguments for the mean squared error, $d(X, \hat{X}) = \|X - \hat{X}\|^2$.

Cubic Spline Fitting

Consider a one dimensional cubic spline. Given the locations of $(N+1)$ points and a set of associated ordinates $\Delta : \quad a = x_0 < x_1 < \cdots < x_N = b$, and $Y : \quad y_0, y_1, \cdots, y_N$; the spline function on $[x_{j-1}, x_j]$, $(j = 1, 2, \ldots, N)$ is defined as:

$$\begin{aligned} S_\Delta \;=\; & M_{j-1}\frac{(x_j - x)^3}{6h_j} + M_j\frac{(x - x_j)^3}{6h_j} + \left(y_{j-1} - \frac{M_{j-1}h_j^2}{6}\right)\frac{x_j - x}{h_j} \\ & + \left(y_j - \frac{M_j h_j^2}{6}\right)\frac{x - x_{j-1}}{h_j} \end{aligned} \tag{3.24}$$

where $h_j = x_j - x_{j-1}$. Also, $M_j = S''_\Delta(x_j)$, the moments of the spline, satisfying the following equations [4]:

$$\begin{bmatrix} 2 & \lambda_0 & 0 & \dots & 0 \\ \mu_1 & 2 & \lambda_1 & \dots & 0 \\ 0 & \mu_2 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ M_2 \\ \vdots \\ M_N \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \tag{3.25}$$

Now, consider that the appropriate end conditions and converging meshes Δ_k ($\lim_{k\to\infty} \parallel \Delta_k \parallel = 0$) are given. If the curve belongs to $C^n[a,b]$, $n = 0, 1, 2$ and 3, having n^{th} continuous derivative ($n = 0$ is for the continuous smooth curve), and satisfying Hölder's condition in the order of α ($0 < \alpha \leq 1$), then the interpolation error uniformly converges with respect to x in $[a,b]$ (Theorem 2.3.1, 2, 3 and 4 [4]):

$$e_1 = |f(x) - S_\Delta(x)| \leq K_1 \parallel \Delta_k \parallel^{n+\alpha}, \quad \text{for some constant } K_1 \tag{3.26}$$

However, the spline reconstructed at the fusion center is not $S_\Delta(x)$ but a shifted spline $S^e_\Delta(x)$ due to the sensing and quantization error at measured points. Next, we determine the magnitude by which interpolation deteriorates. For this, we are given that the noise at prescribed points is bounded by: $E|\delta y_i| \leq e_{sq}$ and $E(\delta y_i)^2 \leq D_{sq}$.

First, consider an absolute error.

$$e = E|f(x) - S^e_\Delta(x)| \leq |f(x) - S_\Delta(x)| + E|S_\Delta(x) - S^e_\Delta(x)| = e_1 + e_2 \tag{3.27}$$

Note that since both $f(x)$ and $S_\Delta(x)$ are considered deterministic, $E(e_1) = e_1$. For e_2:

$$e_2 = E\left|\sum_{i=0}^{N} \frac{\partial S_\Delta}{\partial y_i}\,\delta y_i\right| \leq e_{sq} \sum_{i=0}^{N} \left|\frac{\partial S_\Delta}{\partial y_i}\right| \tag{3.28}$$

Appendix I in Section 3.8 shows that for proper end conditions $\lambda_0, \mu_N < 2$ and evenly distributed meshes (with bounded h_j/h_{j+1}), the following holds:

$$\beta = \sum_{i=0}^{N} \left| \frac{\partial S_\Delta}{\partial y_i} \right| \leq K_2, \quad \text{for some finite number } K_2$$

Hence the total absolute error is bounded by:

$$e \leq e_1 + \beta e_{sq} \tag{3.29}$$

Now, consider the mean squared error. Since data is locally fused before it is transmitted to the global fusion center, the correlation between the noise δy_i at different sensors is not necessarily zero, but is bounded as:

$$E(\delta y_i \delta y_j) \leq \sqrt{E(\delta y_i)^2 E(\delta y_j)^2} \leq D_{sq}$$

We first find a bound on the mean square error between the original and noise corrupted splines with respect to $x \in [a, b]$.

$$\begin{aligned} E\left[S_\Delta(x) - S_\Delta^e(x)\right]^2 &= E\left(\sum_i \frac{\partial S_\Delta}{\partial y_i}\,\delta y_i\right)^2 \\ &= \sum_i \left(\frac{\partial S_\Delta}{\partial y_i}\right)^2 E(\delta y_i)^2 + 2\sum_{i\neq j}\left(\frac{\partial S_\Delta}{\partial y_i}\frac{\partial S_\Delta}{\partial y_j}\right) E\delta y_i \delta y_j \\ &\leq \left(\sum_i \frac{\partial S_\Delta}{\partial y_i}\right)^2 D_{sq} \\ &\leq \beta^2 D_{sq} \end{aligned}$$

where, D_{sq} is the distortion due to sensing and quantization.

The total mean square error is now given by:

$$\begin{aligned} D &= E\left[(f(x) - S_\Delta(x)) + (S_\Delta(x) - S_\Delta^e(x))\right]^2 \\ &= \left[f(x) - S_\Delta(x)\right]^2 + 2\left[f(x) - S_\Delta(x)\right] E\left[S_\Delta(x) - S_\Delta^e(x)\right] \\ &\qquad + E\left[S_\Delta(x) - S_\Delta^e(x)\right]^2 \\ &\leq e_1^2 + 2\beta e_1 e_{sq} + \beta^2 D_{sq} \end{aligned}$$

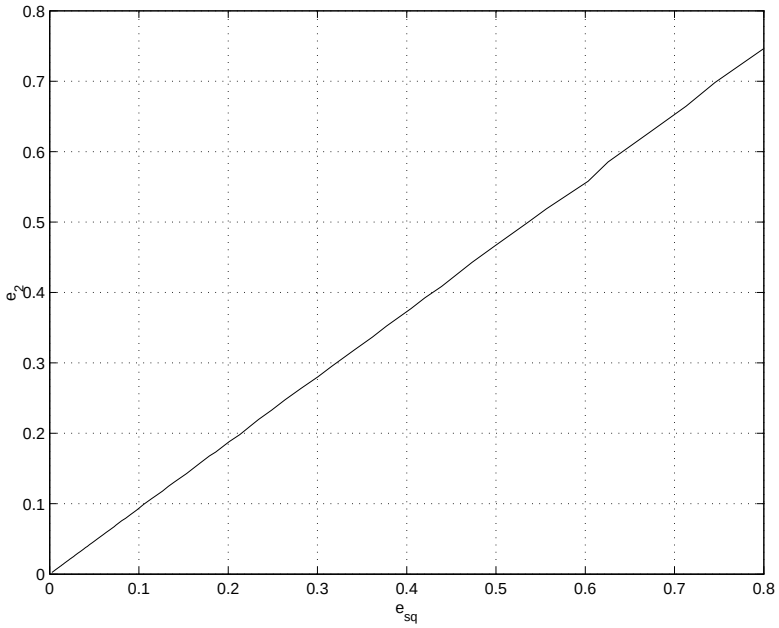


Figure 3.18: Simulation Plot of e_2 versus sensing and quantizing error e_{sq}.

Hence, the mean squared error is also bounded.

Consider the following simulation where cubic a spline is used to fit a sinusoid function based on noise corrupted data. The simulation plot of error e_2 and the noise due to sensing and quantization e_{sq} is as shown in Figure 3.18. Simulation is done for relatively small mesh size and hence, $e \approx e_2$. For this setup, the relation between e_2 and e_{sq} appears to be approximately linear.

This result is readily extensible to a two-dimensional doubly cubic spline defined on a rectangular grid, $\Delta_t : a = t_0 < t_1 < \cdots < t_N = b$, $\Delta_s : c = s_0 < s_1 < \cdots < s_M = d$, by noticing that a doubly cubic spline can be obtained by partial splines on t and s (pp. 238 [4]). The resulting error after twice one-dimensional interpolation is thus bounded by:

$$e \leq e_{1s} + \beta_s(e_{1t} + \beta_t e_{sq}) \tag{3.30}$$

where, $\beta_i, e_{1i}, (i = s, t)$ are the corresponding parameters on s and t coordinates.

A similar derivation applies to the mean squared error.

We now summarize this section. At the final reconstruction, two types of errors are to be distinguished. One is an interpolation error due to the discretizing process, and the other is an error propagated from the sensing and quantization process. The former is determined by sampling the mesh sizes.

Many local interpolation algorithms are available such as cubic Hermite, cubic Bessel, cubic B-splines [4], etc. Due to the nature of local algorithms, the error does not propagate to distant nodes. Therefore, the error would be be bounded. Thus, with finite spatial sampling, the distributed source can be reconstructed with bounded error, even with relatively simple techniques.

3.7 Conclusions

Various issues such as scalability, information extraction under fidelity constraint, local cooperation, optimal sensor density, and propagation of error in sensor networks have been considered. Discussion began with the issues of scalability, and extraction of the measurements of a physical phenomenon to a desired level of fidelity. The solution to these problems is in the relative densities of sources, communication relays and sensors as summarized in Table 3.1.

For sensor networks, we believe that the main objective is information extraction to some level of fidelity. The objective of scalability and information extraction under a distortion criterion is attainable by the independent play of the relative densities of sensors, sources and relays. This can be shown by allowing the sensor density to be much greater than that of sources. This we term as source separation principle. Hence, the network now consists of independent Gaussian CEO systems. The source separation leads to the useful abstraction

that the scale for specialized communications and signal processing for linkage to the physical world need not be large in the limit of high sensor density relative to the source density. At the networking and higher layers, the standard approaches for large scale networks can be utilized.

The cases where sources, communication relays and sensors density are approximately the same are much more difficult problems. This is because as we approach the critical sampling density (such as Nyquist sampling), larger scale interactions are required, and the communication and source coding become tightly coupled. If the number of sensors is limited, certain distortion values are not achievable regardless of the data rate available. Notice that both distortion and capacity must be considered jointly since the data rate must also be achievable, and if we use some efficient scheme, we might use fewer resources in extracting the necessary information. Such problems remain both open and interesting.

However, the source separation principle leads to non-practical solutions. So the notion of spatial fidelity is considered to completely characterize the sensor network. In a broad sense, spatial fidelity is the desired separation between the point sources. For a distributed phenomenon, it translates into desired and/or adaptive sampling frequency. This is analogous to the resolution of pixels in image processing. For reliability, the data is considered to be fused locally. Based on the spatial fidelity and SINR available at the sensors, sensor density for local cooperation is calculated. Further, the relation between sensor density and desired distortion was evaluated, assuming the knowledge of sources and spatial fidelity. The ideas for a point phenomenon were then extended to distributed phenomena.

In particular, the distortions due to sensing, quantization and interpolation in sensor networks observing distributed phenomena are identified. The distributed

phenomenon is modelled as correlated point sources. Sensing error is determined by measurement noise and signal attenuation, which can be reduced by improving sensor coverage and increasing the number of observations. When observing a distributed source, network capacity may become strained due to the large raw information rate. In this case, local cooperation and fusion based on correlations among nearby observations can be used to bring down the quantization rate, which is bounded by rate-distortion theory. Here, we considered cubic splines as the local algorithm to reconstruct the continuous sources. The total error was found to converge at least on the same order of sensing and quantization error given appropriate mesh sizes.

There remain a large set of open problems with varying ratios of sources, relays, and sensors densities for point phenomena. There are many resource optimization problems that will differ in character according to communication resources and source densities. Among the optimization parameters are energy, bandwidth and latency. The extension to non-Gaussian sources may also be challenging. Many topics for future research in this general area suggest themselves. Different sensing models can be proposed, which will lead to different behaviors for the sensing error. Alternative practical local fusion algorithms can be designed and compared to the rate bounds. In this chapter, only source coding was considered, but channel coding enters the picture either by setting the limits on quantization rate or joint source-channel coding. The convergence of distortion with other interpolation schemes (besides the cubic spline) in the presence of sensing and quantization noise can also be studied.

3.8 Appendix I - Boundedness of $\sum_{i=0}^{N} \left| \frac{\partial S_\Delta}{\partial y_i} \right|$

Rearrange the cubic spline function for $x_{j-1} \leq x \leq x_j$ as:

$$S_\Delta(x, y_0, y_1, \ldots, y_N) = \alpha_j M_{j-1} + \beta_j M_j + \gamma_j y_{j-1} + (1-\gamma_j) y_j$$

$$\alpha_j = \left[\frac{(x_j - x)^3}{6h_j} - \frac{h_j(x_j - x)}{6} \right], \quad |\alpha_j| \leq \frac{h_j^2}{9\sqrt{3}}$$

$$\beta_j = \left[\frac{(x - x_j)^3}{6h_j} - \frac{h_j(x - x_{j-1})}{6} \right], \quad |\beta_j| \leq \frac{h_j^2}{9\sqrt{3}}$$

$$0 \leq \gamma_j = \frac{(x_j - x)}{h_j} \leq 1$$

On differentiating the spline function about y_i:

$$\frac{\partial S_\Delta}{\partial y_i} = \alpha_j \frac{\partial M_{j-1}}{\partial y_i} + \beta_j \frac{\partial M_j}{\partial y_i} + \gamma_j \delta_{i,j-1} + (1-\gamma_j)\delta_{i,j}$$

$$\delta_{i,j} = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{otherwise} \end{cases}$$

Taking the absolute values on both sides and summing all the equations over $i = 0, 1, \ldots, N$:

$$\sum_{i=0}^{N} \left| \frac{\partial S_\Delta}{\partial y_i} \right| \leq \frac{h_j^2}{9\sqrt{3}} \sum_{i=0}^{N} \left[\left| \frac{\partial M_{j-1}}{\partial y_i} \right| + \left| \frac{\partial M_j}{\partial y_i} \right| \right] + 1. \tag{3.31}$$

Now, differentiate both sides of Equation (3.25) about y_i:

$$\begin{bmatrix} \frac{\partial M_0}{\partial y_i} \\ \frac{\partial M_1}{\partial y_i} \\ \vdots \\ \frac{\partial M_N}{\partial y_i} \end{bmatrix} = \begin{bmatrix} 2 & \lambda_0 & \ldots & 0 \\ \mu_1 & 2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 2 \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial d_0}{\partial y_i} \\ \frac{\partial d_1}{\partial y_i} \\ \vdots \\ \frac{\partial d_N}{\partial y_i} \end{bmatrix} = \mathbf{B}^{-1} \begin{bmatrix} \frac{\partial d_0}{\partial y_i} \\ \frac{\partial d_1}{\partial y_i} \\ \vdots \\ \frac{\partial d_N}{\partial y_i} \end{bmatrix} \tag{3.32}$$

Observe that,

$$\frac{\partial d_j}{\partial y_i} = \frac{6\delta_{i,j+1}}{h_{j+1}(h_j + h_{j+1})} - \frac{6\delta_{i,j}}{h_j h_{j+1}} + \frac{6\delta_{i,j-1}}{h_j(h_j + h_{j+1})}$$

On taking the absolute values on both sides of Equations (3.32), and summing over $i = 0, 1, \ldots, N$, it follows that:

$$\sum_{i=0}^{N} \begin{bmatrix} |\frac{\partial M_0}{\partial y_i}| \\ |\frac{\partial M_1}{\partial y_i}| \\ \vdots \\ |\frac{\partial M_N}{\partial y_i}| \end{bmatrix} \leq \| \mathbf{B}^{-1} \| \sum_{i=0}^{N} \begin{bmatrix} |\frac{\partial d_0}{\partial y_i}| \\ |\frac{\partial d_1}{\partial y_i}| \\ \vdots \\ |\frac{\partial d_N}{\partial y_i}| \end{bmatrix} \leq \frac{12}{h_j h_{j+1}} \| \mathbf{B}^{-1} \|$$

where $\| \mathbf{B}^{-1} \|$ is the row-max norm of matrix $\mathbf{B}^{-1}$ (p. 20 [4]). For proper end conditions ($\lambda_0, \mu_N < 2$), [4] shows that the following is true:

$$\| \mathbf{B}^{-1} \| \leq \max \left[(2 - \lambda_0)^{-1}, (2 - \mu_N)^{-1}, 1 \right]$$

Therefore,

$$\sum_{i=0}^{N} \left| \frac{\partial S_\Delta}{\partial y_i} \right| \leq 1 + \frac{8\sqrt{3}\eta}{9},$$

$$\eta = \max \left[(2 - \lambda_0)^{-1}, (2 - \mu_N)^{-1}, 1 \right] \left(\frac{h_j}{h_{j+1}} \right)$$

Thus, for meshes with bounded $\left(\frac{h_j}{h_{j+1}} \right)$, the quantity $\sum_{i=0}^{N} \left| \frac{\partial S_\Delta}{\partial y_i} \right|$ is bounded.

CHAPTER 4

Rate-Distortion Bounds on the Multiple Gaussian Sources

A thorough understanding of the mathematical foundation and its communication application is surely a prerequisite to other applications. - C. E. Shannon, "The Bandwagon", March 1956.

A network information theoretic problem that arises when information from multiple distributed correlated sources is to be communicated to a single receiver has been dealt with in this chapter. Distributed detection of phenomena is an important problem in sensor networks [204, 70, 181]. It is known that higher reliability and lower probability of detection error can be achieved using multiple observations from a distributed set of sensors and intelligent fusion algorithms [242]. We consider the source coding problem for such a multiterminal scenario.

The multiterminal coding theory problem for two correlated memoryless sources with separate encoders was first addressed by Slepian and Wolf [221]. A related problem of source coding with side information when only one of the sources is reproduced was considered in (Section 14.8, [54]). However, both the above problems considered lossless reproduction. Han and Kobayashi [102], and Csiszar and Korner [56] have also focused on special extensions of Slepian and Wolf. We consider the related problem when multiple correlated sources are available and only one of them is reproduced but instead of lossless coding, the rate-distortion

version is considered.

Wyner and Ziv [253, 252] had solved the rate distortion coding problem with uncoded side information, summarized in (Section 14.9, [54]). Related problems have also been considered by Berger [20], Kaspi and Berger [126], Berger et. al. [23], Tung [241] and Berger and Yeung [24]. Oohama [167] solved an extension of the Wyner-Ziv problem when the side information is also coded, for the case of two sources. The extension to more than two sources was considered in [169] when the sources are conditionally independent given one of them. We consider the general problem when sources are correlated and the conditional independence does not hold.

Such a problem is of practical importance in coding when multiple sensors are measuring correlated data. For instance, a network of multiple sensors may be installed to monitor a physical environment, consisting of some resource constrained wireless sensors and some wired sensors. The wireless sensors may be deeply embedded into the environment and hence close to the phenomenon of interest while the wired sensors are farther off. Networked Infomechanical System (NIMS) [122, 123] is an example of such a system. In this case, the wired sensors can provide side information to reduce the coding rate or distortion in the data stream produced by the sensor of interest. We refer to the sensors providing side information as helpers while the sensor whose measurements are to be reproduced is referred to as the main source. The rate required at the main source is of concern since this source is wireless and resource constrained.

The objective is to find the rate distortion relation between the rate of the main source and the distortion in reproduction, for any arbitrary positive set of rates available for the helpers [172]. The helper rates are assumed to be free (unconstrained) and may be very large in some situations, such as when the

helper information is transmitted over high bandwidth wired channels. The exact problem, referred to as the *m-helper* problem, is specified in the next section.

The m-helper region is then used to obtain the outer region of the multiterminal problem proposed by Berger and Tung [20, 241, 170]. The interest here is to reproduce all the sources at the fusion center.

We shall assume that the sources are Gaussian. A Gaussian source represents the maximum information content and would lead us to the worst case rate-distortion bound in case of the single source problem. Hence, considering it for multiple sources seems to be intuitively justifiable. A further attraction is analytical tractability.

4.1 Analytic Formulation of m-helper Problem

Let $X, Y_1, \ldots, Y_m$ be correlated random variables such that $\{X_t, Y_{1t}, \ldots, Y_{mt}\}_{t=1}^{\infty}$ are jointly normal, stationary and memoryless sources. For each observation time $t = 1, 2, 3, \ldots$, the random $(m+1)$-tuplet $(X_t, Y_{1t}, \ldots, Y_{mt})$ takes a value in the $(m+1)$-dimensional real space $\mathcal{X} \times \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_m$. The probability density function $p_{X,Y_1,\ldots,Y_m}(x, y_1, \ldots, y_m)$ is $\mathcal{N}\ (0, \Lambda)$ where, the covariance matrix, Λ, is given by

$$\begin{pmatrix} \sigma_X^2 & \rho_{XY_1}\sigma_X\sigma_{Y_1} & \ldots & \rho_{XY_m}\sigma_X\sigma_{Y_m} \\ \rho_{XY_1}\sigma_X\sigma_{Y_1} & \sigma_{Y_1}^2 & \ldots & \rho_{Y_1Y_m}\sigma_{Y_1}\sigma_{Y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{XY_m}\sigma_X\sigma_{Y_m} & \rho_{Y_1Y_m}\sigma_{Y_1}\sigma_{Y_m} & \ldots & \sigma_{Y_m}^2 \end{pmatrix}$$

with $-1 < \rho_{ij} < 1$, $(i, j) \in (X, Y_1, \ldots, Y_m)$. Let n independent instances of $\{X_t\}_{t=1}^{\infty}$ be $X^n = \{X_1, X_2, \ldots, X_n\}$ and similarly $Y_i^n = \{Y_{i1}, Y_{i2}, \ldots, Y_{in}\}$, for $i = 1, 2, \ldots, m$. Consider the system depicted in Figure (4.1). Data sequences X^n and Y_i^n are separately encoded to $\varphi_0(X^n)$ and $\{\varphi_i(Y_i^n)\}_{i=1}^{m}$. The encoder

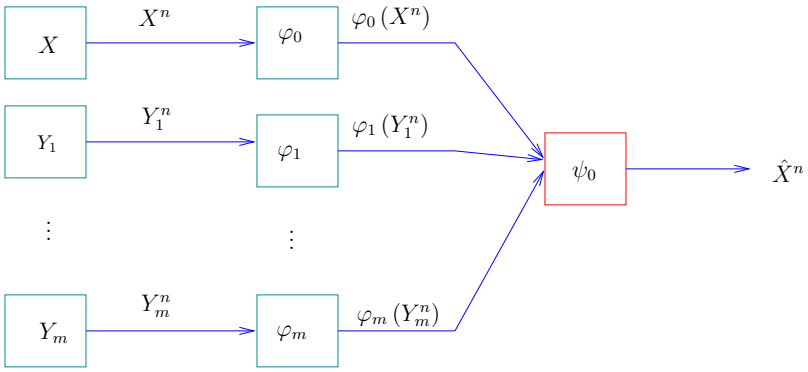


Figure 4.1: m-helper coding system

functions φ_0 and φ_i's are defined by

$$\begin{aligned} \varphi_0 &: \mathcal{X}^n \to \mathcal{C}_0 = \{1, 2, \ldots, C_0\} \\ \varphi_i &: \mathcal{Y}_i^n \to \mathcal{C}_i = \{1, 2, \ldots, C_i\} \end{aligned}$$

The coded (compressed) sequences are sent to a fusion center, and the rates are

$$\frac{1}{n} \log C_i \leq R_i + \delta, i = 0, 1, 2, \ldots, m$$

where δ is an arbitrary positive number. Note that, all logarithms in this paper are to the base 2. The decoder function observes the $(m+1)$-tuplet

$$\left(\varphi_0\left(X^n\right), \varphi_1\left(Y_1^n\right), \ldots, \varphi_m\left(Y_m^n\right)\right)$$

to estimate the main source as $\hat{X}^n$. The decoder function ψ_0 is given by

$$\psi_0 : \mathcal{C}_0 \times \mathcal{C}_1 \times \ldots \times \mathcal{C}_m \to \mathcal{X}^n$$

Note that the goal is to reproduce only X, the main source. The other sources, $\{Y_i\}_{i=1}^m$, are used as helpers, and are not reproduced. Hence, there is no distortion constraint on the helpers. Any available rate can be used for coding the helper information. Let

$$d_0 : \mathcal{X}^2 \to [0, \infty)$$

be the squared distortion measure. The average distortion, Δ_0, for

$$\hat{X}^n = \psi_0\left(\varphi_0\left(X^n\right), \varphi_1\left(Y_1^n\right), \ldots, \varphi_m\left(Y_m^n\right)\right)$$

is defined by,

$$\Delta_0 = E\frac{1}{n}\sum_{t=1}^{n} d_0\left(X_t, \hat{X}_t\right) \leq D_0$$

The m-helper problem is to find the rate-distortion relation between R_0 and D_0 for the above coding system.

An attempt to derive the rate-distortion region for the general Gaussian case was made in [9, 8], but that derivation was not for the most general situation. Specifically, consider equation (3.5) in [8]:

$$\begin{aligned} n\left(R_x + \delta\right) &\geq \log C_1 \\ &\geq H\left(W_x\right) \\ &\geq I\left(X^n; \hat{X}^n\right) - \frac{1}{n}\sum_{k=1}^{n} I\left(X^n; W_k\right) \end{aligned} \tag{4.1}$$

where $W_x = \varphi_x\left(X^n\right)$ and $W_i = \varphi_i\left(Y_i^n\right)$. The second term in (4.1) i. e.

$$\frac{1}{n}\sum_{k=1}^{n} I\left(X^n; W_k\right)$$

is the problem term. This term does not account for the correlation between W_i's.

4.2 Solving the m-helper problem

With the constraints and definitions as described in Section 4.1, we state the following theorem:

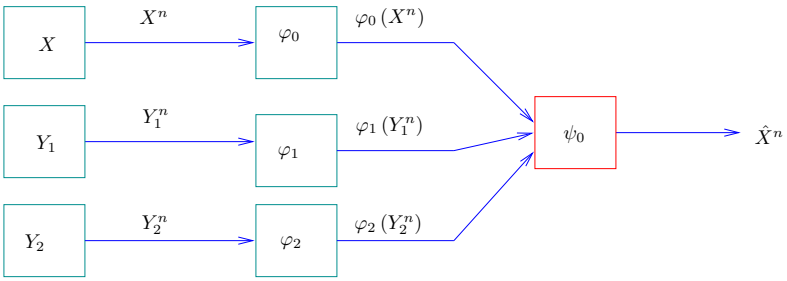


Figure 4.2: The coding system for 2-helper case

Theorem 4.2.1 *For the m-helper coding system, data streams from correlated Gaussian sources can be fused to reduce the data rate, R_0, required for source X. R_0 satisfies the lower bound:*

$$R_0(D_0) \geq \frac{1}{2}\log^+\left[\frac{\sigma_X^2}{D_0}\prod_{i=1}^{m}(1-\rho_i^2)\Gamma_i\right] \tag{4.2}$$

$$\textit{where } \Gamma_i = 1-\rho^2_{XY_i|(Y_1,\ldots,Y_{i-1})}+\rho^2_{XY_i|(Y_1,\ldots,Y_{i-1})}\cdot 2^{-2R_i},$$

$\rho_i^2 = 1-\frac{\sigma^2_{X|Y_1,\ldots,Y_{i-1}}}{\sigma_X^2}$, *and* $\log^+ x = \max\{\log x, 0\}$.

Proof. To simplify the presentation we derive the rate-distortion region for the 2-helper case (Figure 4.2) and then generalize it for the m-helper system. Set $W_0 = \varphi_0(X^n)$, $W_1 = \varphi_1(Y_1^n)$ and $W_2 = \varphi_2(Y_2^n)$. Then

$$\begin{aligned}
n(R_0+\delta) &\geq \log C_0 \\
&\geq h(W_0) \\
&\overset{(a)}{\geq} h(W_0|W_1,W_2) \\
&\overset{(b)}{=} I(X^n;W_0|W_1,W_2) \\
&\overset{(c)}{=} I(X^n;W_0,W_1,W_2)-I(X^n;W_1,W_2) \\
&\geq I(X^n;\hat{X}^n)-I(X^n;W_1)-I(X^n;W_2|W_1)
\end{aligned} \tag{4.3}$$

Here, (a) holds because conditioning reduces entropy, (b) is obtained using the fact that W_0 is a function of X^n, and (c) follows from the chain rule of mutual

information. Now we express the rate of the second helper, accounting for the correlation among the helpers:

$$\begin{aligned} n(R_2+\delta) &\geq \log C_2 \\ &\geq h(W_2) \\ &\geq h(W_2|W_1, Y_1^n) \\ &= I(Y_2^n; W_2|W_1, Y_1^n) \end{aligned} \tag{4.4}$$

Observe that for $i = 1, 2$,

$$\begin{array}{ccccc} W_i & \rightarrow & Y_i^n & \rightarrow & X^n \\ (W_1, W_2) & \rightarrow & (Y_1^n, Y_2^n) & \rightarrow & X^n \end{array} \tag{4.5}$$

are Markov chains. We use (4.3) to derive a lower bound on R_0. For this, let

$$\begin{aligned} F_n(D_0) &= \inf_{\hat{X}^n : \Delta_0 \leq D_0} \frac{1}{n} I(X^n; \hat{X}^n) \\ G_{n1}(R_1) &= \sup_{\substack{W_1 \,:\, \frac{1}{n} I(Y_1^n; W_1) \leq R_1 \\ W_1 \rightarrow Y_1^n \rightarrow X^n}} \frac{1}{n} I(X^n; W_1) \\ G_{n2}(R_2) &= \sup_{\substack{W_2 \,:\, \frac{1}{n} I(Y_2^n; W_2) \leq R_2 \\ (W_1, W_2) \rightarrow (Y_1^n, Y_2^n) \rightarrow X^n}} \frac{1}{n} I(X^n; W_2|W_1) \end{aligned} \tag{4.6}$$

Therefore,

$$R_0 + \delta \geq F_n(D_0 + \delta) - G_{n1}(R_1 + \delta) - G_{n2}(R_2 + \delta) \tag{4.7}$$

A lower bound on $F_n(D_0)$ can be derived as in [167]:

$$F_n(D_0) \geq \frac{1}{2} \log \frac{\sigma_X^2}{D_0} \tag{4.8}$$

An upper bound on $G_{n1}(R_1)$ is as derived in [167]:

$$G_{n1}(R_1) \leq \frac{1}{2}\log\left(\frac{1}{1-\rho_{XY_1}^2+\rho_{XY_1}^2\cdot 2^{-2R_1}}\right) \tag{4.9}$$

Now we consider $G_{n2}(R_2)$. This is different from the calculation of $G_{n1}(R_1)$ as the correlation between Y_1 and Y_2 also needs to be considered. To evaluate $G_{n2}(R_2)$, define the random variables $X(y_1) = X|(Y_1 = y_1)$ and $Y_2(y_1) = Y_2|(Y_1 = y_1)$. Now, $E\left[X(y_1)|Y_2(y_1)\right] = aY_2(y_1)$, where $a = \rho_{XY_2|Y_1}\frac{\sigma_{X|Y_1}}{\sigma_{Y_2|Y_1}}$. Hence, we can write

$$X(y_1) = aY_2(y_1) + N$$

where where, N is a zero-mean Gaussian random variable with variance $\sigma_N^2 = \sigma_{X|Y_1}^2\left(1-\rho_{XY_2|Y_1}^2\right)$ and is independent of $Y_2(y_1)$.

Since the sequences are memoryless, this leads to

$$X^n(y_1^n, w) = aY_2^n(y_1^n, w) + N^n \tag{4.10}$$

where $w = (w_1, w_2)$, $X^n(y_1^n, w)$ is the conditional random variable X^n conditioned on $Y_1^n = y_1^n$ and $W = w$, and $Y_2^n(y_1^n, w)$ is similarly defined.

Using the entropy power inequality [54, 28] in (5.12),

$$\begin{aligned} 2^{\frac{2}{n}h(X^n(y_1^n,w))} &\geq 2^{\frac{2}{n}h(aY_2^n(y_1^n,w))} + 2^{\frac{2}{n}h(N^n)} \\ &= 2^{2h(N)} + a^2 2^{\frac{2}{n}h(Y_2^n(y_1^n,w))} \end{aligned} \tag{4.11}$$

The entropy of N can be substituted in the above expression. This entropy is given by:

$$h(N) = \frac{1}{2}\log\left\{2\pi e\left(\sigma_{X|Y_1}^2(1-\rho_{XY_2|Y_1}^2)\right)\right\}$$

where $\sigma_{X|Y_1}^2 = \sigma_X^2(1-\rho_{XY_1}^2)$. With this, (4.11) becomes

$$\begin{aligned} 2^{\frac{2}{n}h(X^n(y_1^n,w))} \geq\ & a^2 2^{\frac{2}{n}h(Y_2^n(y_1^n,w))} \\ & +2\pi e\left[\sigma_X^2(1-\rho_{XY_1}^2)(1-\rho_{XY_2|Y_1}^2)\right] \end{aligned}$$

Taking the logarithm of the above equation, we get:

$$\frac{1}{n}h(X^n(w, y_1^n)) \geq T\left(\frac{1}{n}h(Y_2^n(y_1^n, w))\right) \tag{4.12}$$

where:

$$T(x) = \frac{1}{2}\log\left[a^2 2^{2x} + 2\pi e\sigma_X^2\left\{(1-\rho_{XY_1}^2)(1-\rho_{XY_2|Y_1}^2)\right\}\right] \tag{4.13}$$

Next, we take expectations on both sides of (5.15) with respect to $W = (W_1, W_2)$ and Y_1^n. Note that from the definition of our conditional random variables, it follows that

$$\begin{aligned} E_{W,Y_1^n}[h(X^n(w, y_1^n))] &= h(X^n|W, Y_1^n) \\ E_{W,Y_1^n}[h(Y_2^n(w, y_1^n))] &= h(Y_2^n|W, Y_1^n) \end{aligned}$$

where $E_Z[\cdot]$ denotes expectation w.r.t. Z. Observe that $T(x)$ is a convex function of x. Applying Jensen's inequality, we get

$$\frac{1}{n}h\left(X^n|Y_1^n, W\right) \geq T\left(\frac{1}{n}h(Y_2^n|Y_1^n, W)\right) \tag{4.14}$$

Since $T(x)$ is monotone increasing with respect to x, the inequality is preserved. From the definition of mutual information, (5.16) can be rewritten as,

$$\frac{1}{n}h\left(X^n|Y_1^n, W_1, W_2\right) \geq T\left(\frac{1}{n}h(Y_2^n|Y_1^n) - \frac{1}{n}I(Y_2^n; W_1, W_2|Y_1^n)\right)$$

By the chain rule of mutual information,

$$I(Y_2^n; W_1, W_2|Y_1^n) = I(Y_2^n; W_1|Y_1^n) + I(Y_2^n; W_2|W_1, Y_1^n)$$

Also, from the definition of W_1 it follows that $I(Y_2^n; W_1|Y_1^n) = 0$. Using this in (4.15), we get

$$\frac{1}{n}h\left(X^n|W_1, W_2, Y_1^n\right) \geq T\left[\frac{1}{n}h(Y_2^n|Y_1^n) - \frac{1}{n}I(Y_2^n; W_2|W_1, Y_1^n)\right]$$

Now, using (4.4) in the above equation:

$$\frac{1}{n}h\left(X^n|W_1,W_2,Y_1^n\right) \geq T\left(\frac{1}{n}h(Y_2^n|Y_1^n) - R_2\right) \tag{4.15}$$

This can be used to derive $G_{n2}(R_2)$, defined in (4.6), as follows:

$$\begin{aligned}
\frac{1}{n}I(X^n;W_2|W_1) &= \frac{1}{n}h(X^n|W_1) - \frac{1}{n}h(X^n|W_1,W_2) \\
&\overset{(a)}{\leq} \frac{1}{n}h(X^n) - \frac{1}{n}h(X^n|W_1,W_2,Y_1^n) \\
&\overset{(b)}{\leq} \frac{1}{n}h(X^n) - T\left(\frac{1}{n}h(Y_2^n|Y_1^n) - R_2\right) \\
&= \frac{1}{2}\log(2\pi e\sigma_X^2) - T\left(\frac{1}{2}\log(2\pi e\sigma_{Y_2|Y_1}^2) - R_2\right)
\end{aligned}$$

where (a) holds because conditioning reduces entropy, and (b) follows from (4.15).

Now, expressing $T(x)$ using (4.13) with $x = \frac{1}{2}\log(2\pi e\sigma_{Y_2|Y_1}^2) - R_2$ we obtain

$$\begin{aligned}
G_{n2}(R_2) &\leq \frac{1}{2}\log\left(\frac{1}{(1-\rho_{XY_1}^2)\Gamma_2}\right) \\
\text{where,} \qquad \Gamma_2 &= 1 - \rho_{XY_2|Y_1}^2 + \rho_{XY_2|Y_1}^2 \cdot 2^{-2R_2}
\end{aligned} \tag{4.16}$$

Finally, using (4.8), (4.9) and (4.16) in (4.7) we get:

$$\begin{aligned}
R_0 + \delta \geq \frac{1}{2}\log\frac{\sigma_X^2}{D_0+\delta} + \frac{1}{2}\log\left(1 - \rho_{XY_1}^2 + \rho_{XY_1}^2 \cdot 2^{-2R_1}\right) \\
+\frac{1}{2}\log\left((1-\rho_{XY_1}^2)(1 - \rho_{XY_2|Y_1}^2 + \rho_{XY_2|Y_1}^2 \cdot 2^{-2R_2})\right)
\end{aligned}$$

Letting $\delta \to 0$, the outer region for the two helper case becomes:

$$R_0(D_0) \geq \frac{1}{2}\log^+\left[\frac{\sigma_X^2}{D_0}\prod_{i=1}^{2}(1-\rho_i^2)\Gamma_i\right]$$

$$\text{where } \Gamma_i = 1 - \rho_{XY_i|(Y_1,\ldots,Y_{i-1})}^2 + \rho_{XY_i|(Y_1,\ldots,Y_{i-1})}^2 \cdot 2^{-2R_i},$$

and $\rho_i^2 = 1 - \frac{\sigma_{X|Y_1,\ldots,Y_{i-1}}^2}{\sigma_X^2}$ leading to $\rho_1^2 = 0$, and $\rho_2^2 = \rho_{XY_1}^2$. Since the joint distribution $p_{X,Y_1,Y_2}(x,y_1,y_2)$ is known, the correlation coefficient $\rho_{XY_2|Y_1}$, required for

evaluating Γ_2, can be calculated to be:

$$\rho_{XY_2|Y_1} = \frac{\rho_{XY_2} - \rho_{XY_1}\rho_{Y_1Y_2}}{\sqrt{1-\rho_{XY_1}^2}\sqrt{1-\rho_{Y_1Y_2}^2}}$$

Generalizing the two helper case to m-helpers using exactly the same arguments as above, we obtain Theorem 4.2.1. ∎

Note that when there is no helper, (4.2) collapses to the classic Gaussian rate-distortion expression [54]:

$$R_0(D_0) \geq \frac{1}{2}\log\frac{\sigma_X^2}{D_0} \tag{4.17}$$

We now consider some examples for which our derived rate-distortion region for the m-helper system collapses to previously known cases.

4.2.1 One-helper System.

On substituting $m = 1$ in (4.2), we obtain:

$$R_0 \geq \frac{1}{2}\log\left[\frac{\sigma_X^2}{D_0}\left(1-\rho^2+\rho^2\cdot 2^{-2R_1}\right)\right]$$

where ρ is the correlation between the main source X and the helper Y. This is same as the result stated in [167].

4.2.2 Two-helpers with $R_2 = 0$.

Consider a main source X and two helpers Y_1 and Y_2. Since $R_2 = 0$, there is no help obtained from Y_2. This is equivalent to the one helper case and thus we should obtain the rate for one-helper. On substituting $m = 2$ and $R_2 = 0$ in (4.2), we obtain:

$$R_0 \geq \frac{1}{2}\log\left[\frac{\sigma_X^2}{D_0}\left(1-\rho_{XY_1}^2+\rho_{XY_1}^2\cdot 2^{-2R_1}\right)\right]$$

which is indeed the expected rate.

4.3 Significance of Helpers

We now consider the potential benefit that may be derived from using the helper rates to reduce the distortion in reproduction.

Suppose a sensor is able to report on source X. Also, suppose two other sensors, Y_1 and Y_2 (helper sources), are able to sense the source. However they are farther from the source than X and hence their measurements are not worth reproducing. The algorithm for selecting which sensors act as helpers may depend on the quality of measurement at each sensor.

Assume now that the correlation of the helpers with the source X depends on their distances from X. Let the correlation, ρ, follow an inverse power law with distance, d:

$$\rho = \frac{\rho_0}{d^\alpha}$$

where ρ_0 is a constant of proportionality. Let us take $\alpha = 2$ and evaluate $R_0(D_0)$. Other correlation models can also be used, such as exponential in distance [150, 206].

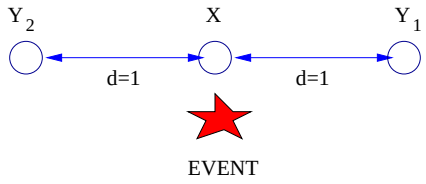


Figure 4.3: Source Placement

Suppose that each of the helpers is located at distance $d = 1$ from the main source X, and if the three sensors lie along a straight line, the distance between the two helpers becomes $d = 2$ as depicted in Figure 4.3.

$R_0(D_0)$ is plotted for this scenario in Figure 4.4. The figure shows that the rate R_0 is reduced when the helpers are used. While the maximum potential

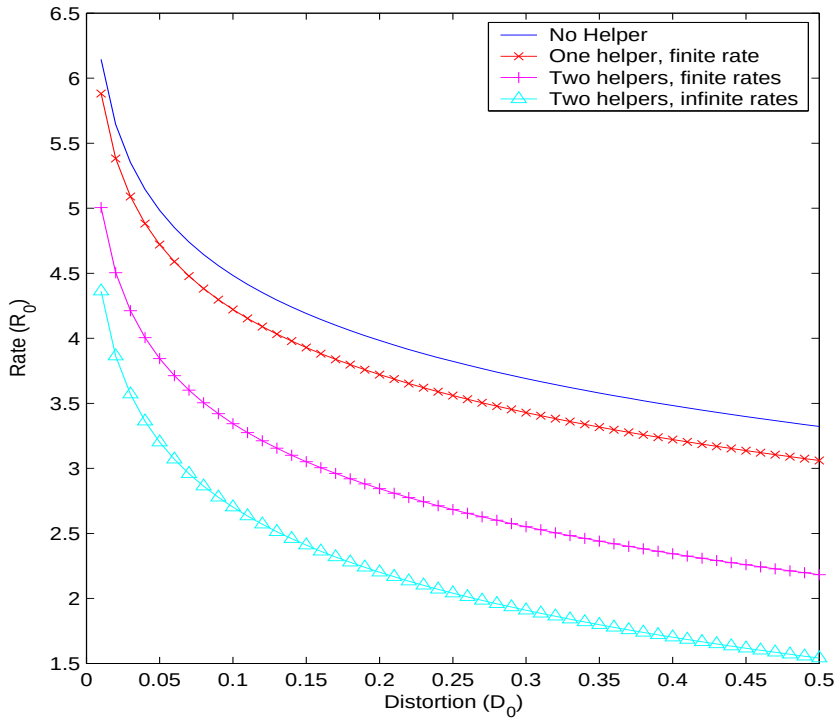


Figure 4.4: Rate-distortion bounds with multiple helpers. The correlations values are: $\rho_{XY_1} = 0.7, \rho_{XY_2} = 0.7$, and $\rho_{Y_1Y_2} = 0.175$.

benefit is shown by the curve using infinite rates for both the helpers, it can be observed that even with a finite rate, the helpers improve the distortion in the reproduction of X. Also, the graph shows that using more helpers reduces the rate further. However, it may be noted that in practice the distance between the sources will increase as more and more sources are added and hence the correlation will fall. This will make more sources yield diminishing improvement in the rate-distortion performance.

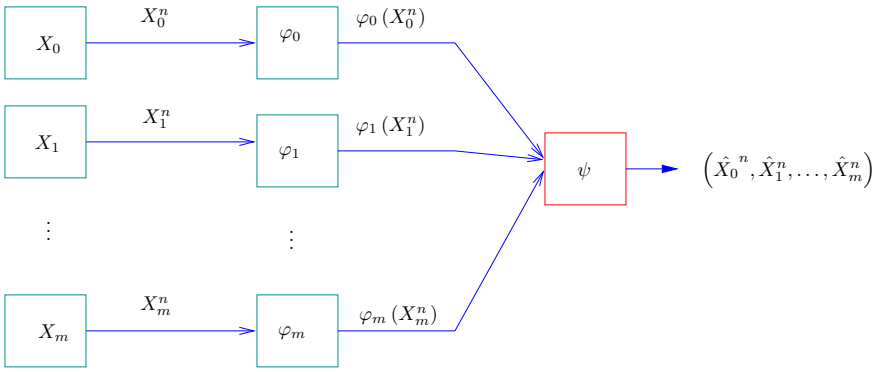


Figure 4.5: The separate coding system for $m+1$ correlated sources

4.4 Multiterminal Coding System

So far we discussed the performance bounds of the m-helper systems and its implications. The solution of m-helper could be efficiently used to derive the outer region of the system proposed by Berger and Tung [20, 241] as depicted in Figure 4.5. We shall now call this the *"Separate Coding System"*. The solution for the most general case of the separate coding system is not yet known. However, the special cases of separate coding system such as two sources in Gaussian setup [20, 167, 110, 166], and source coding with high resolution [144, 259], are solved and well understood. Here, we shall derive the outer bounds on such systems considering $m+1$ correlated sources rather than two sources as solved previously using our m-helper results.

4.4.1 Analytic Formulation : $m+1$-Separate Correlated Sources

Similar to $R_0(D_0)$ derived in Section 4.2, we have $R_1(D_1), \ldots, R_m(D_m)$. The rate and distortion constraints of X_0 can be extended to other sources $X_1, \ldots, X_m$. It is evident that regions $R_0(D_0), R_1(D_1), \ldots, R_m(D_m)$ are the outer regions of

$R(D_0, D_1, \ldots, D_m)$.Define,

$$\begin{aligned}
R_{[1]}(D) &= R_0(D_0) \cap R_1(D_1) \cap \ldots \cap R_m(D_m) \\
R_{[2]}(D) &= R_{01}(D_0, D_1) \cap R_{02}(D_0, D_1) \cap \ldots \cap R_{0m}(D_0, D_m) \cap \\
&\quad R_{12}(D_1, D_2) \cap \ldots \cap R_{1m}(D_1, D_m) \cap \ldots \cap R_{(m-1)m}(D_{m-1}, D_m) \\
\vdots &\quad \vdots \quad \vdots \\
R_{[m-1]}(D) &= R_{01\ldots m-1}(D_0, D_1, \ldots, D_{m-1}) \cap \ldots \cap R_{1\ldots m}(D_1, D_2, \ldots, D_m) \\
R_{[m]}(D) &= R_{01\ldots m}(D_0, D_1, \ldots, D_m)
\end{aligned}$$

where,

$$\begin{aligned}
&R_{01\ldots m}(D_0, D_1, \ldots, D_m) = \\
&\left\{(R_0, R_1, \ldots, R_m) : R_0 + R_1 + \ldots + R_m \geq \tfrac{1}{2}\log\left[(1-\rho^2)\tfrac{\sigma^2_{X_0}\sigma^2_{X_1}\ldots\sigma^2_{X_m}}{D_0^2 D_1^2 \ldots D_m^2}\right]\right\}
\end{aligned} \tag{4.18}$$

and $\rho = f\left(\rho_{X_0X_1}, \rho_{X_0X_2}, \ldots, \rho_{X_0X_m}, \rho_{X_1X_2}, \ldots, \rho_{X_1X_m}, \ldots, \rho_{X_{m-1}X_m}\right)$. Further discussion on ρ will be presented later in the section. The outer region for separate coding problem is then given by,

$$R_{\text{out}}(D_0, D_1, \ldots, D_m) = R_{[1]}(D) \cap R_{[2]}(D) \cap \ldots \cap \mathcal{R}_{[m]}(D) \tag{4.19}$$

4.4.2 Outer Region of Separate Coding System

With the background provided in section 4.4.1, we obtain the following theorem

Theorem 4.4.1 *For every* $D_0, D_1, \ldots, D_m > 0$

$$R(D_0, D_1, \ldots, D_m) \subseteq R_{out}(D_0, D_1, \ldots, D_m) \tag{4.20}$$

Proof Similarly to the proof for m-helpers, we will consider the case with 3 separate sources, X_0, X_1 and X_2 (Figure 4.6). As observed earlier, $R_0(D_0)$, $R_1(D_1)$, and $R_2(D_2)$ are outer regions of $R(D_0, D_1, D_2)$. Therefore, it suffices to show that if $(R_0, R_1, R_2) \in R(D_0, D_1, D_2)$,

$$R_1 + R_2 + R_3 \geq \frac{1}{2}\log\left\{\left(1-\rho^2\right)\frac{\sigma^2_{X_0}\sigma^2_{X_1}\sigma^2_{X_2}}{D_0 D_1 D_2}\right\} \tag{4.21}$$

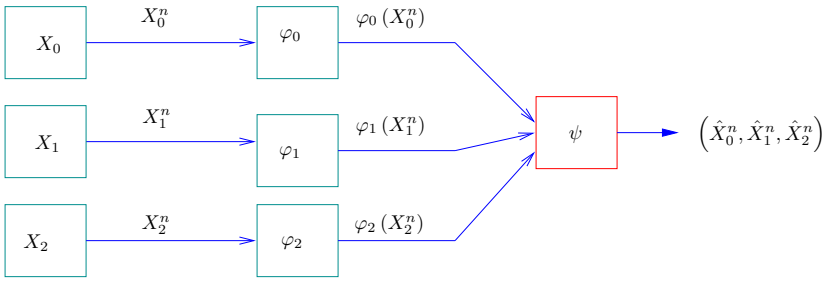


Figure 4.6: The coding system for 3 correlated sources

where $\rho^2 = \rho^2_{X_0X_1} + \rho^2_{X_0X_2} + \rho^2_{X_1X_2} - 2\rho_{X_0X_1}\rho_{X_0X_2}\rho_{X_1X_2}$. Assume that $(R_0, R_1, R_2) \in R(D_0, D_1, D_2)$. Therefore,

$$\begin{aligned}
n(R_0 + R_1 + R_2 + 3\delta) &\geq \log C_0 + \log C_1 + \log C_2 \\
&\geq h(W_0) + h(W_1) + h(W_2) \\
&\overset{(a)}{\geq} h(W_0, W_1, W_2) \\
&\overset{(b)}{=} I(X_0^n, X_1^n, X_2^n; W_0, W_1, W_2) \\
&\overset{(c)}{=} I(X_0^n; W_0, W_1, W_2) + I(X_1^n, X_2^n; W_0, W_1, W_2 | X_0^n) \\
&\overset{(d)}{=} I(X_0^n; W_0, W_1, W_2) + I(X_1^n; W_0, W_1, W_2 | X_0^n) \\
&\quad + I(X_2^n; W_0, W_1, W_2 | X_0^n, X_1^n) \qquad (4.22)
\end{aligned}$$

where, (a) is due to the fact that $h(X, Y) \leq h(X) + h(Y)$ (equality if and only if x and Y are independent random variables), (b) follows from the definition of mutual information, and (c) and (d) are the result of the chain rule of mutual information. Now consider $I(X_1^n; W_0, W_1, W_2 | X_0^n)$,

$$\begin{aligned}
I(X_1^n; W_0, W_1, W_2 | X_0^n) &= I(X_1^n; X_0^n, W_0, W_1, W_2) - I(X_0^n; X_1^n) \\
&= I(X_1^n; X_0^n, W_0, W_1, W_2) - nI(X_0; X_1) \qquad (4.23)
\end{aligned}$$

The above follows from the chain rule of mutual information. Similarly,

$$I(X_2^n; W_0, W_1, W_2 | X_0^n, X_1^n) = I(X_2^n; X_0^n, X_1^n, W_0, W_1, W_2) - I(X_2^n; X_0^n, X_1^n) \qquad (4.24)$$

From (4.22), (4.23), (4.24) we have:

$$\begin{aligned} n\left(R_0+R_1+R_2+3\delta\right) &\geq I(X_0^n;W_0,W_1,W_2)+I(X_1^n;X_0^n,W_0,W_1,W_2) \\ &\quad +I(X_2^n;X_0^n,X_1^n,W_0,W_1,W_2) \\ &\quad -nI(X_0;X_1)-nI(X_2;X_0,X_1) \\ &\geq I(X_0^n;\hat{X}_0^n)+I(X_1^n;\hat{X}_1^n)+I(X_2^n;\hat{X}_2^n) \\ &\quad -nI(X_0;X_1)-nI(X_2;X_0,X_1) \end{aligned} \tag{4.25}$$

Now, $I(X_2;X_0,X_1)=h(X_2)-h(X_2|X_0,X_1)$. From the definition of differential entropy, we know that $h(X_2)=\frac{1}{2}\log 2\pi e\sigma_{X_2}^2$ and

$$\begin{aligned} h(X_2|X_0,X_1) &= h(X_0,X_1,X_2)-h(X_0,X_1) \\ &= \frac{1}{2}\log 2\pi e\sigma_{Y_2}^2\frac{1-\rho^2}{1-\rho_{X_0X_1}^2} \end{aligned}$$

where $\rho^2=\rho_{X_0X_1}^2+\rho_{X_0X_2}^2+\rho_{X_1X_2}^2-2\rho_{X_0X_1}\rho_{X_0X_2}\rho_{X_1X_2}$. Therefore,

$$I(X_2;X_0,X_1)=\frac{1}{2}\log\frac{1-\rho_{X_0X_1}^2}{1-\rho^2} \tag{4.26}$$

Also,

$$\begin{aligned} \frac{1}{n}I(X_0^n;\hat{X}_0^n) &\geq \frac{1}{2}\log\frac{\sigma_{X_0}^2}{D_0} \\ \frac{1}{n}I(X_1^n;\hat{X}_1^n) &\geq \frac{1}{2}\log\frac{\sigma_{X_1}^2}{D_1} \\ \frac{1}{n}I(X_2^n;\hat{X}_2^n) &\geq \frac{1}{2}\log\frac{\sigma_{X_2}^2}{D_2} \end{aligned}$$

From the above equations, (4.25) and (4.26):

$$\begin{aligned} &R_0+R_1+R_2+3\delta \geq \\ &\frac{1}{2}\left\{\log\left(\frac{\sigma_{X_0}^2}{D_0+\delta}\frac{\sigma_{X_1}^2}{D_1+\delta}\frac{\sigma_{X_2}^2}{D_2+\delta}\right)+\log(1-\rho_{X_0X_1}^2)+\log\frac{1-\rho^2}{1-\rho_{X_0X_1}^2}\right\} \end{aligned}$$

On letting $\delta\to 0$,

$$R_1+R_2+R_3\geq\frac{1}{2}\log\left\{\left(1-\rho^2\right)\frac{\sigma_{X_0}^2\sigma_{X_1}^2\sigma_{X_2}^2}{D_0D_1D_2}\right\}$$

where $\rho^2 = \rho^2_{X_0X_1} + \rho^2_{X_0X_2} + \rho^2_{X_1X_2} - 2\rho_{X_0X_1}\rho_{X_0X_2}\rho_{X_1X_2}$, which is (4.21).

Generalizing the above proof for the general case of $m+1$ correlated sources we have:

$$R_1 + R_2 + \ldots + R_m \geq \frac{1}{2}\log\left\{\left(1-\rho^2\right)\frac{\sigma^2_{X_0}\sigma^2_{X_1}\ldots\sigma^2_{X_m}}{D_0D_1\ldots D_m}\right\}$$

where $\rho = f\left(\rho_{X_0X_1}, \rho_{X_0X_2}, \ldots, \rho_{X_0X_m}, \rho_{X_1X_2}, \ldots, \rho_{X_1X_m}, \ldots, \rho_{X_{m-1}X_m}\right)$. We derived the value of ρ for 3 sources case. It is a tedious but straightforward to derive the value of ρ for m sources. ∎

4.5 Conclusions

We considered a multi-terminal network information theory problem when several correlated sources are fused to reproduce a source of interest under a distortion constraint. We presented a generalized solution to this rate-distortion problem with side information and showed that previously known results can be viewed as special cases of the derived expression. The close match for special cases also suggests that the derived lower bound is close to the rate distortion function; however the derivation of the inner region is still an open problem. We also discussed the significance of helper rates and the correlation between them for reducing the distortion in reproduction. These results are an essential building block in the development of a complete information theory for the sensing coverage and communication performance of sensor networks.

The results can be useful for the choice of correlated sources to be used in a real implementation. Sensor networks are typically deployed at reasonably high densities for fault tolerance and for ensuring sufficient coverage in random deployments. This means that several sources may be available for measuring the same phenomenon. Due to severe energy constraints in such systems, it is extremely

important to utilize resources to their maximum. Hence, it is not desirable to code and transmit extra sources if it does not help to reduce distortion. Our results can be directly used with the relevant sensor specific measurement correlation models for estimating the number of sensors that should actively transmit.

The problem can be extended to the more practical case of non-Gaussian sources. The bound will help definitively compare various data fusion and network coding schemes for wireless networks, with regards to their performance and efficiency.

CHAPTER 5

Information Processing in Data Gathering Networks

Let's start by specifying a model of communication system to which the theory to be developed shall apply This model should be sufficiently general to include, as special cases, most of the communication systems of practical interest, yet simple enough to lend itself to a detailed quantitative study. - R. Fano lecture notes, MIT Archives.

In this chapter we consider the process of data gathering in sensor networks, which is the key functionality of such systems. The common process, which underlies most sensor networking applications, can be viewed as follows. A phenomenon of interest exists in the environment within the sensing range of the deployed system. Multiple sensors collect readings about this phenomenon, which are subject to noise in the sensor transducers. The sensed data is now communicated to points of interest. The sensors may communicate among themselves and with the destinations to transmit this data, or parameters of interest derived from this data, in the most efficient manner. In-network processing may take place as the data travels through the network. Determination of the most efficient communication techniques for this process leads to multi-point network information theoretic problems.

We can however model the problem at a reduced complexity by considering

those aspects which occur more commonly in practice. For instance, with the current technology [159, 225], communication cost dominates processing costs and hence, all processing should be performed locally, as close to the source as possible, such that only relevant data needs to be communicated over longer network paths [174]. We develop such a model for the data gathering process and derive the optimal data rates required to communicate an estimate of the phenomenon at the required fidelity. We assume that the network does not communicate the complete set of raw measurements collected by the sensors but just enough information to meet the fidelity requirements. Such an assumption is valid for sensor networks as the data sinks are typically interested only in the measured phenomenon or its location and not the identities or raw measurements from the individual sensor nodes, enabling the network to fuse sensor data close to the source.

Similar problems have been solved before [150, 206] and are summarized in section 5.1. However, the effect of sensor noise has not been explicitly considered. This is significant since the presence of noise changes the desired communication strategy. For instance, if noise at different sensors is independent, readings from two sensors places at approximately the same location, can be used to average out the effect of noise leading to a better estimate. Data from such co-located sensors might have been considered redundant in the absence of noise, but in a practical implementation noise cannot be ignored and it becomes relevant to communicate this data.

5.0.1 Key Contributions

We model the data gathering process with multiple noisy sensors as an information theoretic problem and derive the optimal data rate required to communicate

an estimate of the phenomenon at the desired distortion level [171]. Since these multiple sensors will share the same wireless channel, the total data rate required, by all the sensors together, is considered. This problem is a variation of the Gaussian CEO problem [243, 168, 41] as discussed later. The sensors communicate compressed data to a local fusion center and then the fused estimates are communicated over the network.

5.0.2 Outline

The next section summarizes the prior work in this field and shows how our work builds upon it. Section 5.2 specifies our problem formulation and the abstractions used. Section 5.3 derives a lower bound on the optimal rate-distortion relationship and compares it to a previously known special case. An upper bound is derived in Section 5.4 and shown to lie close to the lower bound, establishing that the lower bound is close to the rate-distortion function. Section 5.6 concludes.

5.1 Related Work

The data carrying capacity of multi-hop wireless networks was estimated in [98] when each node generated data independently. This clearly does not model the data as generated in sensor networks, which typically comes from a set of common sources and hence is likely to be correlated. The correlated data model was considered in [206] and a rate-distortion relationship was derived to show how data generated from multiple sensors could be used to reduce distortion. The data gathering problem for a correlated source was also considered in [150], with a non band-limited field phenomenon.

We begin with the same model for the phenomenon as used in the above

papers, i.e., a multivariate stochastic process with non-zero correlation among its spatial components. However, we also account for sensor noise. The noise itself is modelled as a Gaussian stochastic process.

Slepian and Wolf [221] had calculated the rate required when multiple sources transmit correlated data. That problem does not model sensor noise. Also, they consider lossless reproduction of sources. In practice, we do not need to (or cannot afford to, in view of limited energy and bandwidth resources) reproduce the complete raw data. Rather, only a fused version, such as a feature of interest is reproduced, and hence lossy coding suffices. The case when only one of the sources is reproduced and the others are treated as helpers, was studied in [167, 172].

Rate-distortion bounds when the sources code and transmit a noisy version of the phenomenon and the receiver fuses information from multiple sources to reproduce the phenomenon with non-zero distortion were first considered in [25]. Rate-distortion bounds for that problem were derived in [243, 168, 41, 183, 64, 42, 72] for Gaussian sources. However, the above solutions assume that all sensors are measuring exactly the same value of the phenomenon. This does not address the case of a distributed phenomenon (such as a temperature field) or the case when the different sensors are measuring multiple perspectives of a single phenomenon. We extend the problem formulation to the distributed case.

5.2 Problem Description

The data gathering problem (Figure 5.1) can be abstracted to the following mathematical formulation. Let $\{X_1(t), \ldots, X_L(t)\}_{t=1}^{\infty}$ represent an L-dimensional source. The source sequences are assumed temporally memoryless and stationary.

The vector $\{X_1(t), \ldots, X_L(t)\}$ is modelled as a zero mean Gaussian random

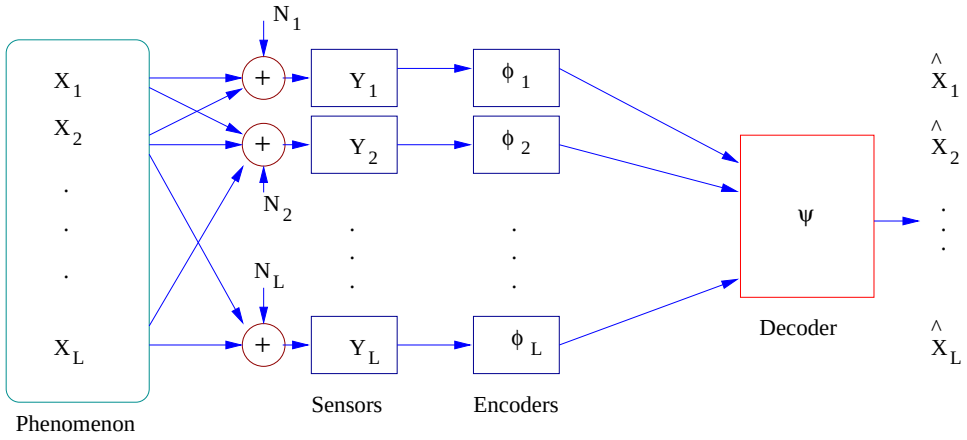


Figure 5.1: The distributed Gaussian CEO System.

variable with a non-singular covariance matrix $R_{\mathbf{X}}$ for all t. The sensor readings $\{Y_1(t), \ldots, Y_L(t)\}$ are noisy versions of $\{X_1(t), \ldots, X_L(t)\}$. Each of $X_i(t)$ and $Y_i(t)$ take values on the real lines $\mathcal{X}_i$ and $\mathcal{Y}_i$ respectively. We use boldface letters to represent L-dimensional vectors; for instance $\mathbf{X} = \{X_1(t), \ldots, X_L(t)\}$. The observed readings $Y_i(t)$'s are modelled as:

$$\mathbf{Y} = H\mathbf{X} + \mathbf{N}$$

where $H \in \Re^{L \times L}$ is a positive definite attenuation matrix and $\mathbf{N}$ is an additive white zero mean Gaussian noise vector with covariance matrix $R_{\mathbf{N}}$. N_i's model the noise in the sensor transducers and hence are independent of each other.

Data sequences, Y_i^n, of block length n, are separately encoded to $\{\varphi_i\left(Y_i^n\right)\}_{i=1}^{L}$ and sent to a fusion center. The encoder functions φ_i's are defined by

$$\varphi_i \quad : \quad \mathcal{Y}_i^n \rightarrow \mathcal{C}_i = \{1, 2, \ldots, |\mathcal{C}_i|\}$$

and the transmitted rates are

$$\frac{1}{n} \log |\mathcal{C}_i| \leq r_i, \qquad i = 1, \ldots, L$$

The sum of the rates, i.e., the total bandwidth used, is denoted by r,

$$r = \sum_{i=1}^{L} r_i$$

Here, the decoder receives the codewords, $\mathbf{C}$, to produce the estimate $\hat{\mathbf{X}}^n$. The decoder function ψ_L is given by

$$\psi_L : \mathcal{C}_1 \times \mathcal{C}_2 \times \ldots \times \mathcal{C}_L \rightarrow \mathcal{X}^n$$

The distortion in the reproduction is defined as:

$$D^n(\mathbf{X}^n, \hat{\mathbf{X}}^n) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{tr}(E||\mathbf{X} - \hat{\mathbf{X}}||^2) \tag{5.1}$$

(where $\mathrm{tr}(A)$ represents the trace of matrix A) and is subject to the distortion constraint:

$$D^n(\mathbf{X}^n, \hat{\mathbf{X}}^n) \leq D \tag{5.2}$$

The distributed Gaussian CEO problem is to find the rate distortion relationship between r and D. This will quantify the minimum rate required to achieve a required distortion.

5.3 Deriving the Rate-Distortion Relationship

We first derive the lower bound on the rate-distortion function. Then, we numerically compare it with an upper bound to evaluate how close our derived bound is to the exact function.

5.3.1 Outer Region

Theorem 5.3.1 *For a given distortion D, the sum of the rates of the coded sensor data streams is bounded as*

$$r(D) \geq \frac{1}{2}\log^{+}\left[\frac{|\Theta H' R_{\mathbf{N}}^{-1}|^2 |R_{\mathbf{N}}| \left(\prod \lambda_i\right) |R_{\mathbf{X}}|}{|\Theta| \left[\frac{D}{L}|R_{\mathbf{X}}|^{\frac{1}{L}} - \left(\prod \lambda_i\right)^{\frac{1}{L}} |\Theta|^{\frac{1}{L}}\right]^{L}}\right]$$

where i ranges over $1, \ldots, L$, $\Theta = (R_{\mathbf{X}}^{-1} + H' R_{\mathbf{N}}^{-1} H)^{-1}$, λ_i are the eigenvalues of $R_{\mathbf{X}}$ and $\log^{+} x = \max\{\log x, 0\}$.

Note that all logarithms are to base 2. Theorem 5.3.1 holds for $\mathrm{tr}(R_{\mathbf{N}}) \leq D \leq \mathrm{tr}(R_{\mathbf{X}})$, otherwise $r(D) = 0$. Intuitively, if the noise in sensor readings is higher than the acceptable distortion, D, then there is no need to transmit any data. The result assumes that the required distortion is greater than that achievable using an optimal estimator, i.e. $D \geq \Theta$.

Proof

The mutual information between the source $\mathbf{X}$ and $\hat{\mathbf{X}}$ is related to the distortion as [54]:

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}^n; \hat{\mathbf{X}}^n) &\geq \sum_{i=1}^{L} \frac{1}{2} \log \frac{\lambda_i}{D_i} \\ &= \frac{1}{2} \log \left(\prod_{i=1}^{L} \frac{\lambda_i}{D_i} \right) \end{aligned} \tag{5.3}$$

where $\{\lambda_i\}_{i=1}^{L}$ are the eigenvalues of the covariance matrix $R_{\mathbf{X}}$ of the phenomenon $\mathbf{X}$ and

$$D_i = \begin{cases} K & \text{if } K < \lambda_i \\ \lambda_i & \text{otherwise} \end{cases}$$

and K is such that $\sum_{i=1}^{L} D_i = D$. The D_i's can be calculated by reverse water filling [54].

From (5.3) we have,

$$\frac{2}{nL} I(\mathbf{X}^n; \hat{\mathbf{X}}^n) \geq \log \left(\prod_{i=1}^{L} \frac{\lambda_i}{D_i} \right)^{1/L} \tag{5.4}$$

We now need to relate the mutual information $I(\mathbf{X}^n; \hat{\mathbf{X}}^n)$ with the compressed data rate, r, at which the codewords are sent. Since $\hat{\mathbf{X}}$ is estimated from the compressed data $\mathbf{C}$ it follows that $\mathbf{X}^n \rightarrow \mathbf{C}^n \rightarrow \hat{\mathbf{X}}^n$ is a Markov chain and hence $I(\mathbf{X}^n; \hat{\mathbf{X}}^n) \leq I(\mathbf{X}^n; \mathbf{C}^n)$. Using this in (5.4), raising both sides to the power 2 and rearranging terms we get:

$$\left[\prod_{i=1}^{L} D_i \right]^{\frac{1}{L}} \geq \left[\prod_{i=1}^{L} \lambda_i \right]^{\frac{1}{L}} \exp \left[\frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right] \tag{5.5}$$

where $\exp(z)$ represents 2^z. Next, we express $I(\mathbf{X}^n; \mathbf{C}^n)$ in terms of r:

Lemma 5.3.2

$$\exp \left[\frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right] \geq \frac{\left(\frac{|\Theta|}{|R_{\mathbf{X}}|} \right)^{\frac{1}{L}}}{1 - \Gamma \left[\prod_i \frac{\lambda_i}{D_i} \right]^{\frac{1}{L}} \exp \left[\frac{-2}{L} r \right]}$$

where $\Gamma = \frac{|\Theta H' R_{\mathbf{N}}^{-1}|^{\frac{2}{L}} |R_{\mathbf{N}}|^{\frac{1}{L}}}{|\Theta|^{\frac{1}{L}}}$ *and* i *ranges from* $1, \ldots, L$.

(The proof of the above lemma has been moved to Appendix I in Section 5.7 to maintain continuity.)

Also, since the arithmetic mean is greater than or equal to the geometric mean (AM-GM inequality):

$$\left[\prod_{i=1}^{L} D_i \right]^{\frac{1}{L}} \leq \frac{1}{L} \sum_{i=1}^{L} D_i = D \tag{5.6}$$

Substituting Lemma 5.3.2 and (5.6) in (5.5) we obtain Theorem 5.3.1. ∎

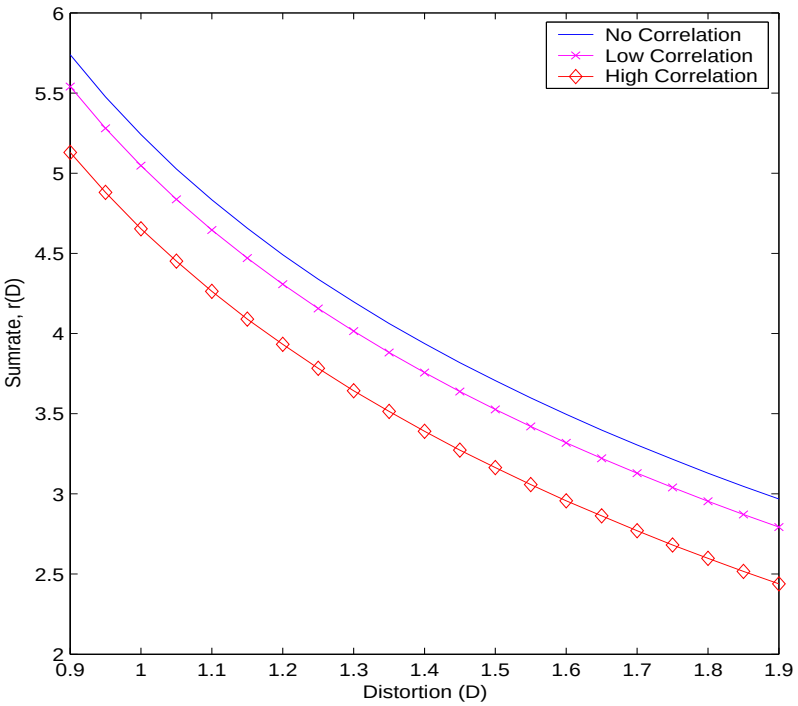


Figure 5.2: Evaluating the derived expression from Theorem 5.3.1.

As a verification we consider a previously known special case. Consider $L = 1$. Substituting $L = 1$ and $H = 1$ in Theorem 5.3.1, $R_{\mathbf{X}} = \sigma_X^2$, and noise variance $R_{\mathbf{N}} = \sigma_N^2$, we get:

$$r(D) \geq \frac{1}{2} log^{+} \left(\frac{\sigma_X^4}{D\sigma_X^2 - \sigma_X^2\sigma_N^2 + D\sigma_N^2} \right)$$

which in fact is same as the $r(D)$ function given in [168, 41]. Thus, the bound yields the exact function for the special case.

The derived expression is plotted for some values of correlation among the sources in Figure 5.2. We can see that the data rate required is lower when there is more correlation among the data sources which is intuitive since when the sources are correlated, the total information content is lower.

This theorem provides the minimum amount of non-redundant data that must be transmitted by the set of distributed sensors for reproducing within the distortion constraint. Thus, if a compression algorithm implemented at the sensors transmits only the amount of data mentioned in the above theorem, it will be an optimal scheme. However, the theorem does not claim that the above rate can actually be achieved by a practical compression algorithm. We address this issue next.

5.4 Achievability of the rate-distortion bound

We now state a theorem which provides an achievable data-rate for reproducing at the given distortion, i.e., an upper bound on the rate required. We will show that this upper bound is close to the minimum bound derived in the previous section, which means that practical compression algorithms can in fact operate close to the minimum bound of Theorem 5.3.1, and hence it can be used as an approximate measure of the practical data rate required.

Theorem 5.4.1 *Consider the random variables* $(X_1, X_2, \ldots, X_L, Y_1, Y_2, \ldots, Y_L)$ *with the joint distribution given by* $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$. *Let* $\mathcal{C}_{in}(D)$ *be the set of compressed data vectors,* $\mathbf{C} = (C_1, C_2, \ldots, C_L)$, *such that:*

1. $C_i \rightarrow Y_i \rightarrow (\mathbf{X}, \mathbf{Y}\backslash Y_i, \mathbf{C}\backslash C_i)$ *form a Markov chain for* $i = 1, \ldots, L$ *where* $\mathbf{Y}\backslash Y_i$ *refers to* $(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_L)$ *etc.*

2. *There exists a decoding function*

$$f : \mathcal{C}_1 \times \mathcal{C}_2 \ldots \times \mathcal{C}_L \rightarrow \mathcal{X}_1 \times \mathcal{X}_2 \ldots \times \mathcal{X}_L$$

such that $Ed(\mathbf{X}, \hat{\mathbf{X}}) \leq D$ *where* $d(\mathbf{X}, \hat{\mathbf{X}})$ *is as defined in equation (5.1) and* $\hat{\mathbf{X}} = f(\mathbf{C})$.

Let $\mathcal{R} = \{(r_1, r_2, \ldots, r_L) : \sum_{i \in Z} r_i \geq I(C_Z; Y_Z | C_{Z^c}),\ \forall\ Z \subseteq (1, \ldots, L)\}$ *where* Z^c *is the complement of* Z*. Then,*

$$r_{in}(D) \stackrel{\triangle}{=} \textit{convex hull of } \left\{ \bigcup_{\mathbf{C} \in \mathcal{C}_{in}(D)} \mathcal{R} \right\}$$

Here $r_{in}(D)$ *is the inner region for the rate-distortion relationship.*

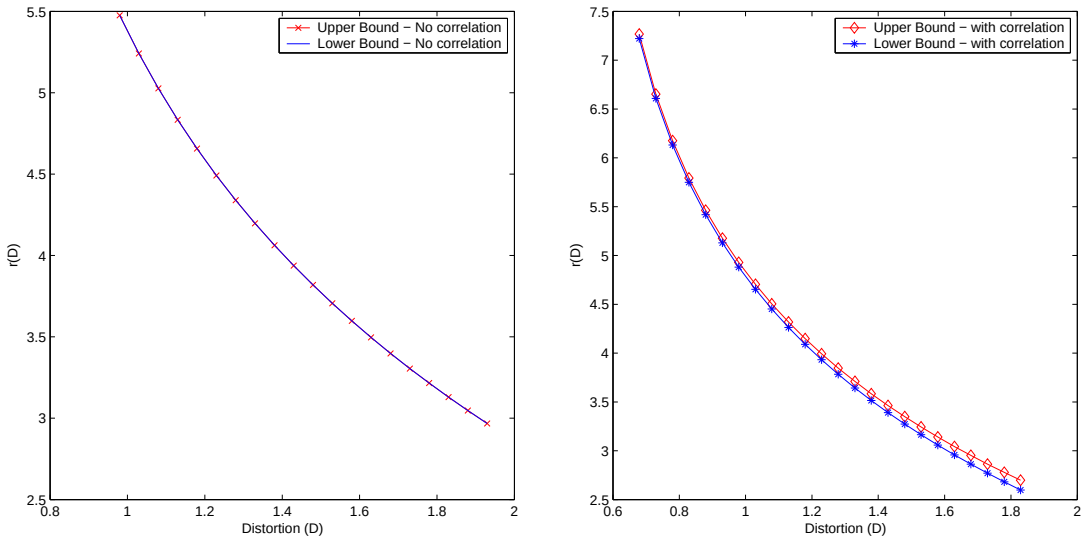
The proof of this theorem is an extension of the proof for Theorem 1 in [41], from the case of a scalar source to that of a vector source, $\mathbf{X}$. The proof is based on the joint typicality of codewords from different sensors [51, 20, 102, 241]. The encoding and decoding procedures used in the proof in [41] extend directly to the vector case. The proof of Theorem 5.4.1 has been moved to Appendix II in Section 5.8.

To compare the inner and outer regions, and thus verify that the derived outer region is close to the actual rate-distortion relationship, we now evaluate both the bounds. The expression for the inner region of the rate-distortion relationship in the above theorem can be evaluated numerically for the jointly Gaussian source used in Theorem 5.3.1 as follows.

We numerically minimize the value of $I(C_Z; Y_Z | C_{Z^c})$ for a Gaussian model of the source $\mathbf{X}$ and noise $\mathbf{N}$. From the known expression for mutual information in the Gaussian case:

$$I(C_Z; Y_Z | C_{Z^c}) = \frac{1}{2} \log^+ \frac{|R_{\mathbf{Y}_Z \mathbf{C}_{Z^c}}||R_{\mathbf{C}}|}{|R_{\mathbf{Y}_Z \mathbf{C}}||R_{\mathbf{C}_{Z^c}}|}$$

To find the optimal $\mathbf{C}$, we begin with $\mathbf{C} = L\mathbf{Y} + \mathbf{T}$ where L is a diagonal matrix and $\mathbf{T}$ is a vector Gaussian random variable. The values of L and the variances of the components of $\mathbf{T}$ are found using numerical optimization, thus leading to the optimal value of $I(C_Z; Y_Z | C_{Z^c})$.



(a) No correlation

(b) With correlation

Figure 5.3: Evaluating the inner and outer regions (a) when there is no correlation among the source components, and (b) with correlation among the source components

The evaluated upper and lower bounds are plotted in Figure 5.3 for $L = 3$. The first figure shows that when there is no correlation among the L sources, then the upper and lower bounds exactly overlap. When the correlation among the sources in $\mathbf{X}$ is non-zero, even then the upper and lower bounds are close. This shows that the derived lower bound is close to the actual rate-distortion function. Thus, the derived lower bound can be used as an estimate of the minimum data rate required to achieve the desired fidelity. The analytical derivation of the achievable bound, or the exact rate-distortion function, however, is still an open problem.

5.4.1 Discussion

The sensor nodes now need compression algorithms which can operate at the optimal rates derived above. The only statistics used in the above derivations which are required for generating a compression scheme is the source covariance matrix $R_{\mathbf{X}}$. The noise covariance matrix and the attenuation matrix H are required at the decoder for finding the optimal linear estimator but not at the sensors for compression. These matrices can be learned by training. The source correlation can be learned at the fusion center on the fly from the received data using methods such as suggested in [46] and then a distributed source coding scheme can be used to achieve the optimal rates [184, 185, 161, 244, 160].

One of the important issues in sensor network is the coverage. Much research has been done on the coverage problem, addressing the issues like sensor placement for optimal coverage, maintaining the sensor coverage, power optimal coverage, etc. [246, 151, 140, 213, 261, 112, 257, 71]. The distributed Gaussian CEO studied in this chapter is also useful in providing the fundamental solutions to the coverage problem. Based on the results of Gaussian CEO, the design engineer can calculate the sensors needed for the coverage.

5.5 Comparing Data Gathering Systems

Chapter 4 discussed the data gathering system modelled as m-helper coding model. That problem is fundamentally different than the Gaussian CEO problem discussed in this chapter. For instance, the m-helper system uses the information from helper to reduce the data rate of the main source but that does not necessarily decreases the total rate (i.e. sum of the main and helper rates) of the network. This increase is acceptable as the helper rates are assumed to be free.

In case of Gaussian CEO system, there is a cost on data transmission from each sensor. Hence, use of additional sensors for observing the same source must not increase the sum rate.

The above mentioned difference, however, is neither obvious nor evident from the rate equations. Hence, consider the m-helper system with only a helper, Y, along with the main source, X and CEO model with couple of sensors, Y_1 and Y_2, observing point source, say, S. The rate-distortion function, $R_X(D_X)$, for one helper system was derived in [167] and is given by:

$$R_X(D_X) = \frac{1}{2}\log^+\left[\frac{\sigma_X^2}{D_X}\left(1-\rho^2+\rho^2\cdot 2^{-2R_Y}\right)\right] \tag{5.7}$$

where, R_Y is the free rate of the helper and σ_X^2 is the variance of the zero-mean Gaussian main source X.

For Gaussian CEO system with L agents (sensors) observing the source S, the sum-rate distortion function, $r(D_S)$, is given in [168, 41] as:

$$r(D_S) = \sum_{i=1}^{L} R_i(D_S) = \frac{1}{2}\log^+\left[\frac{\sigma_S^2}{D_S}\left(\frac{D_S\sigma_S^2 L}{D_S\sigma_S^2 L - \sigma_S^2\sigma_N^2 + D\sigma_N^2}\right)^L\right] \tag{5.8}$$

where, σ_S^2 is the variance of the zero-mean Gaussian source S and σ_N^2 is the noise variance assuming that the nose at each sensor is identical Gaussian noise with zero mean.

Example 5.5.1 (1-helper System)

Let the correlation between the main source, X, and helper, Y, be modelled as $\rho = 0.8$ with $\sigma_X^2 = 11$. Suppose $D_X = 1$ and free rate, $R_Y = \{0, 0.5, 1\}$, then R_X and $R = R_X + R_Y$, are given by:

R_Y	R_X	$R = R_X + R_Y$
0	1.73	1.73
0.5	1.45	1.95
1	1.26	2.26

This example clearly shows that the use of helper does decrease the rate required to transmit main source information but increases the total rate of the network.

Example 5.5.2 (CEO System)

Consider the CEO coding model with $L = \{1, 2\}$, $\sigma_S^2 = 11$ and $D_S = 1$. Here, the two cases could be summarized: firstly when the nose variance is low, $\sigma_N^2 = 0.01$ and secondly for the high noise variance, $\sigma_N^2 = 0.8$. For this values, the sum-rate could be calculated as:

	σ_N^2	r	
		$L = 1$	$L = 2$
Low Noise Variance	0.01	1.74	1.74
High Noise Variance	0.8	2.69	2.38

At low noise variance, the sensors receive high quality measurement from the source and not much distortion is introduced by the sensing channel. Hence, there is no help obtained from the second sensor. However, the sum rate is not increased but remains the same. In case of high noise variance, there is significant distortion introduced by the sensing channel. The data measured by the sensors is far from accurate and hence, additional sensors certainly help in this case. Thus for high noise variance situations, additional sensors help in reducing the sum rate.

From the above examples 5.5.1 and 5.5.2, it is clear that the m-helper system

reduces the rate of the main source by using the extra information from the helpers but the net rate of the network does not decrease. On the other hand, the CEO system does not allow the increase of sum rate.

5.6 Conclusions

We derived the optimal data rate required to communicate when a set of sensors measure a distributed phenomenon. The effect of sensor noise was explicitly considered. This is useful for determining the minimum amount of resources required for achieving a desired level of fidelity and for evaluating the performance of data compression schemes used in the network.

In this work, we considered the fused estimate to be a reproduction of the phenomenon. However, the data rate will change if other estimates were considered, such as functions of the reproduced phenomenon. We showed that the derived lower bound on the rate-distortion function is close to the exact function, but an analytical derivation of the exact rate-distortion function is still an open problem.

5.7 Appendix I - Proof of Lemma 5.3.2

From the definition of mutual information we have

$$I(\mathbf{X}^n;\mathbf{C}^n) = h(\mathbf{X}^n) - h(\mathbf{X}^n|\mathbf{C}^n) \tag{5.9}$$

Substituting the known entropy measure for the multivariate Gaussian vector $\mathbf{X}$ as $h(\mathbf{X}) = \frac{1}{2}\log(2\pi e)^L|R_{\mathbf{X}}|$ in (5.9) and rewriting:

$$\exp\left[\frac{-2}{nL}I(\mathbf{X}^n;\mathbf{C}^n)\right] = \frac{1}{2\pi e|R_{\mathbf{X}}|^{\frac{1}{L}}}\exp\left[\frac{2}{nL}h(\mathbf{X}^n|\mathbf{C}^n)\right]$$

(5.10)

Now, we find a bound on $h(\mathbf{X}^n|\mathbf{C}^n)$ in terms of the required quantity r.

Let $S = E(\mathbf{X}|\mathbf{Y})$ where $E(\cdot)$ denotes expectation. Therefore, $S = A\mathbf{Y}$ where $A \in \Re^{L\times L}$ and is given by [205],

$$A = (R_{\mathbf{X}}^{-1} + H' R_{\mathbf{N}}^{-1} H)^{-1} H' R_{\mathbf{N}}^{-1}$$

If $\tilde{\mathbf{N}}$ be a zero-mean Gaussian with variance $R_{\tilde{\mathbf{N}}} = (R_{\mathbf{X}}^{-1} + H' R_{\mathbf{N}}^{-1} H)^{-1}$ [205] then we have,

$$\mathbf{X} \;\; = \;\; \mathbf{S} + \tilde{\mathbf{N}} \tag{5.11}$$

and $\tilde{\mathbf{N}}$ is independent of $\mathbf{Y}$. Since the sequences are memoryless, this leads to

$$\mathbf{X}^n(\mathbf{c}^n) = \mathbf{S}^n(\mathbf{c}^n) + \tilde{\mathbf{N}}^{\mathbf{n}} \tag{5.12}$$

where $\mathbf{X}^n(\mathbf{c}^n)$ is the conditional random variable $\mathbf{X}^n$ conditioned on $\mathbf{C}^n = \mathbf{c}^n$, and $\mathbf{S}^n(\mathbf{c}^n)$ is similarly defined. Also, note that since $\tilde{\mathbf{N}}^n$ is independent of $\mathbf{Y}^n$, it is also independent of $\mathbf{C}^n$.

Using the entropy power inequality [54, 28] in (5.12),

$$\exp\left[\frac{2}{nL} h(\mathbf{X}^n(\mathbf{c}^n))\right] \geq \exp\left[\frac{2}{nL} h(\mathbf{S}^n(\mathbf{c}^n))\right] + \;\; \exp\left[\tfrac{2}{nL} h(\tilde{\mathbf{N}}^n(\mathbf{c}^n))\right] \tag{5.13}$$

Substituting the value of $h(\tilde{\mathbf{N}})$ in (5.13),

$$\exp\left[\frac{2}{nL} h(\mathbf{X}^n(\mathbf{c}^n))\right] \geq \exp\left[\frac{2}{nL} h(\mathbf{S}^n(\mathbf{c}^n))\right] + 2\pi e |R_{\tilde{\mathbf{N}}}|^{1/L} \tag{5.14}$$

Taking the logarithm of the above equation, we get:

$$\frac{1}{nL} h(\mathbf{X}^n(\mathbf{c}^n)) \;\; \geq T\left(\tfrac{1}{nL} h(\mathbf{S}^n(\mathbf{c}^n))\right) \tag{5.15}$$

where:

$$T(z) = \frac{1}{2}\log\left[2^{2z} + 2\pi e|R_{\tilde{\mathbf{N}}}|^{1/L}\right]$$

Next, we take expectations on both sides of (5.15) with respect to $\mathbf{C}^n$. Note that from the definition of our conditional random variables, it follows that

$$\begin{aligned} E_{\mathbf{C}^n}[h(\mathbf{X}^n(\mathbf{c}^n))] &= h(\mathbf{X}^n|\mathbf{C}^n) \\ E_{\mathbf{C}^n}[h(\mathbf{S}^n(\mathbf{c}^n))] &= h(\mathbf{S}^n|\mathbf{C}^n) \end{aligned}$$

Observe that $T(z)$ is a convex function of z. Applying Jensen's inequality, we get

$$\frac{1}{nL}h(\mathbf{X}^n|\mathbf{C}^n) \geq T\left(\frac{1}{nL}h(\mathbf{S}^n|\mathbf{C}^n)\right) \tag{5.16}$$

Since $T(z)$ is monotone increasing with respect to z, the inequality is preserved. From (5.14), (5.15) and (5.16) we obtain,

$$\exp\left\{\frac{2}{nL}h(\mathbf{X}^n|\mathbf{C}^n)\right\} \geq \exp\left\{\frac{2}{nL}h(\mathbf{S}^n|\mathbf{C}^n)\right\} + 2\pi e|R_{\tilde{\mathbf{N}}}|^{1/L} \tag{5.17}$$

Now, we evaluate $h(\mathbf{S}^n|\mathbf{C}^n)$. From the definition of mutual information,

$$\frac{1}{n}h(\mathbf{S}^n|\mathbf{C}^n) = \frac{1}{n}h(\mathbf{S}^n|\mathbf{C}^n, \mathbf{X}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{S}^n|\mathbf{C}^n) \tag{5.18}$$

To evaluate $h(\mathbf{S}^n|\mathbf{C}^n)$, we shall first bound $\frac{1}{n}I(\mathbf{X}^n; \mathbf{S}^n|\mathbf{C}^n)$.

$$\begin{aligned} \frac{1}{n}I(\mathbf{X}^n; \mathbf{S}^n|\mathbf{C}^n) &= \frac{1}{n}I(\mathbf{X}^n; \mathbf{S}^n, \mathbf{C}^n) - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &\geq \frac{1}{n}I(\mathbf{X}^n; \mathbf{S}^n) - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{2}\log\frac{|R_{\mathbf{X}}|}{|R_{\tilde{\mathbf{N}}}|} - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \end{aligned} \tag{5.19}$$

Next, we derive a lower bound on $h(\mathbf{S}^n|\mathbf{C}^n, \mathbf{X}^n)$. For this let $\mathbf{S}^n(\mathbf{x}^n, \mathbf{c}^n)$ be a conditional random variable $\mathbf{S}^n$ conditioned by $(\mathbf{X}^n, \mathbf{C}^n) = (\mathbf{x}^n, \mathbf{c}^n)$. We have a

similar definition for $\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)$. Also observe that $Y_i^n \rightarrow \mathbf{X}^n \rightarrow \mathbf{C}^n \backslash C_i^n$ forms a Markov chain. The conditional independence of (Y_i^n, C_i^n), $i = 1, 2, \ldots, L$ given $\mathbf{X}^n$ also extends to $\mathbf{S}^n(\mathbf{x}^n, \mathbf{c}^n)$. Hence, from the definition of $\mathbf{S}^n$ we have:

$$\mathbf{S^n}(\mathbf{x}^n, \mathbf{c}^n) = A\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)$$

The above equation implies:

$$\begin{aligned} h(\mathbf{S^n}(\mathbf{x}^n, \mathbf{c}^n)) &= h(A\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)) \\ &= \log|A| + h(\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)) \end{aligned}$$

Taking expectation on both the sides of above equation with respect to $(\mathbf{X}^n, \mathbf{C}^n)$ we get,

$$\frac{1}{n}h(\mathbf{S^n}|\mathbf{X}^n, \mathbf{C}^n) = \log|A| + \frac{1}{n}h(\mathbf{Y}^n|\mathbf{X}^n, \mathbf{C}^n) \tag{5.20}$$

From the definition of sum rate r, we have

$$\begin{aligned} r &\geq \sum_{i=1}^{L} \log|C_i|^n \geq \sum_{i=1}^{L} \frac{1}{n}h(C_i^n) \geq \frac{1}{n}h(\mathbf{C}^n) \\ &\geq \frac{1}{n}h(\mathbf{C}^n|\mathbf{X}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{n}I(\mathbf{Y}^n; \mathbf{C}^n|\mathbf{X}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{n}h(\mathbf{Y}^n|\mathbf{X}^n) - \frac{1}{n}h(\mathbf{Y}^n|\mathbf{X}^n, \mathbf{C}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{2}\log(2\pi e)^L|R_{\mathbf{N}}| - \frac{1}{n}h(\mathbf{Y}^n|\mathbf{X}^n, \mathbf{C}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \end{aligned}$$

From the above equation we have:

$$\frac{1}{n}h(\mathbf{Y^n}|\mathbf{X}^n, \mathbf{C}^n) \geq \frac{1}{2}\log(2\pi e)^L|R_{\mathbf{N}}| - \left[r - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n)\right] \tag{5.21}$$

From (5.20) and (5.21), we have:

$$\frac{1}{n}h(\mathbf{S^n}|\mathbf{X}^n, \mathbf{C}^n) \geq \log|A| + \frac{1}{2}\log(2\pi e)^L|R_{\mathbf{N}}| - \left[r - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n)\right] \tag{5.22}$$

Substituting (5.19) and (5.22) in (5.18), and dividing both the sides by L we have:

$$\frac{1}{nL}h(\mathbf{S^n}|\mathbf{C}^n) \geq \frac{1}{L}\log|A| + \frac{1}{2L}\log(2\pi e)^L|R_{\mathbf{N}}| - \frac{1}{L}\left[r - \frac{1}{n}I(\mathbf{X}^n;\mathbf{C}^n)\right] + \frac{1}{2L}\log\frac{|R_{\mathbf{X}}|}{|R_{\tilde{\mathbf{N}}|}} - \frac{1}{nL}I(\mathbf{X}^n;\mathbf{C}^n) \quad (5.23)$$

Substituting (5.23) in (5.17) we obtain :

$$\exp\left[\frac{2}{nL}h(\mathbf{X}^n|\mathbf{C}^n)\right] \geq \frac{|A|^{\frac{2}{L}}2\pi e|R_{\mathbf{N}}|^{\frac{1}{L}}|R_{\mathbf{X}}|^{\frac{1}{L}}}{|R_{\tilde{\mathbf{N}}}|^{\frac{1}{L}}}\exp\left\{\frac{-2}{L}\left[r - \frac{1}{n}I(\mathbf{X}^n;\mathbf{C}^n)\right]\right\}\exp\left[\frac{-2}{nL}I(\mathbf{X}^n;\mathbf{C}^n)\right] + 2\pi e|R_{\tilde{\mathbf{N}}}|^{\frac{1}{L}}$$

where, $A = \Theta H' R_{\mathbf{N}}^{-1}$ and $R_{\tilde{\mathbf{N}}} = \Theta$.

From the above equation and (5.10), we obtain the required relation between the rate r and $I(\mathbf{X}^n;\mathbf{C}^n)$:

$$\exp\left[\frac{-2}{nL}I(\mathbf{X}^n;\mathbf{C}^n)\right] = \frac{|A|^{2/L}|R_{\mathbf{N}}|^{1/L}}{|R_{\tilde{\mathbf{N}}}|^{1/L}}\exp\left\{\frac{-2}{L}\left[r - \frac{1}{n}I(\mathbf{X}^n;\mathbf{C}^n)\right]\right\}\exp\left\{\frac{-2}{nL}I(\mathbf{X}^n;\mathbf{C}^n)\right\} + \frac{|R_{\tilde{\mathbf{N}}}|^{1/L}}{|R_{\mathbf{X}}|^{1/L}}$$

The above equation can be rearranged to obtain:

$$\exp\left[\frac{-2}{nL}I(\mathbf{X}^n;\mathbf{C}^n)\right]\left[1 - \frac{|A|^{2/L}|R_{\mathbf{N}}|^{1/L}}{|R_{\tilde{\mathbf{N}}}|^{1/L}}\eta\right] \geq \frac{|R_{\tilde{\mathbf{N}}}|^{1/L}}{|R_{\mathbf{X}}|^{1/L}}$$

where, $\eta = \exp\left\{\frac{-2}{L}\left[r - \frac{1}{n}I(\mathbf{X}^n;\mathbf{C}^n)\right]\right\}$.

Noting that $R_{\tilde{\mathbf{N}}} = \Theta$ and $I(\mathbf{X}^n;\mathbf{C}^n) \geq \frac{1}{n}I(\mathbf{X}^n;\hat{\mathbf{X}}^n) \geq \frac{1}{2}\log\left(\prod_{i=1}^{L}\frac{\lambda_i}{D_i}\right)$ (Section 5.3.1) in the above equation we get Lemma 5.3.2. ∎

5.8 Appendix II - Proof of Theorem 5.4.1

Let $(C_1, C_2, \ldots, C_L)$ and function f satisfy the conditions given in the theorem 5.4.1. Construct the random codebooks $\{\mathcal{W}^n = (\mathcal{W}_1^n, \mathcal{W}_2^n, \ldots, \mathcal{W}_L^n)\}$ (where $\mathcal{W}_i^n$ denotes the codebook of encoder i) as follows:

At encoder i, generate M_i i.i.d. codewords C_i^n according to $\prod_{j=1}^{n} p\left(c_i(j)\right)$ and index them $C_i^n(j)$, $j = 1, 2, \ldots, M_i$. Let $\mathcal{W}_i^n = \{C_i^n(j)\}_{j=1}^{M_i}$. Randomly assign the indices of the codewords to one of the 2^{nR_i} bins using a uniform distribution over the indices of the bins such that every bin contains $N_i = M_i 2^{-nR_i}$ codewords.

Suppose that every generated C_i^n satisfies strong typicality. By the Weak Law of Large Numbers, this assumption holds with probability close to 1 when n is large enough. The other way to ensure this assumption is by simply drawing C_i^n from the strongly typical set.

Encoding Scheme

At encoder i, given observations y_i^n, if its typical, map it onto the $c_i^n(j) \in \mathcal{W}_i^n$ with the smallest index j such that $(y_i^n, c_i^n(j))$ are jointly typical. Let $c_i^n(y_i^n)$ denotes the c_i^n onto which y_i^n is mapped.

For coding, the index of the bin which contains $c_i^n(y_i^n)$ is sent. If y_i^n is not typical or there does not exist $c_i^n(j) \in \mathcal{W}_i^n$ such that $(y_i^n, c_i^n(j))$ are jointly typical, then special error symbol is sent. This special error symbol does not increase the rate R_i in the limit of large n, so we may safely ignore it.

Decoding Scheme

Let $b_i(y_i^n)$ denotes the bin index and $\mathcal{B}_i(b_i)$ denotes the bin with index b_i at the encoder i. Given $(b_1, b_2, \ldots, b_L)$, if there exists a unique $(c_1^n, c_2^n, \ldots, c_L^n)$ such that $w_i^n \in \mathcal{B}_i(b_i)$ and $(c_1^n, c_2^n, \ldots, c_L^n)$ are jointly typical, then call it $(\hat{c}_1^n, \hat{c}_2^n, \ldots, \hat{c}_L^n)$; otherwise declare an error an incur maximum distortion $d_{\max}$. If the received vector contains special error symbol, declare an error and incur the maximum distortion

$d_{\max}$. Assuming no error, produce the estimate $\hat{x}(k) = f(\hat{c}_1(k), \hat{c}_2(k), \ldots, \hat{c}_L(k))$ for $k = 1, 2, \ldots, n$.

Analysis of the Probability of Error

Consider the following exhaustive error events:

$E_1 : (\mathbf{X}^n, \mathbf{Y}^n)$ are not jointly typical.

$E_2 : \bigcup_{i=1}^{L} E_{2,i}$ where, $E_{2,i} = E_1^c \bigcap F_i$ for $i = 1, 2, \ldots, L$; $F_i : (Y_i^n, W_i^n)$ not typical for all $C_i^n \in \mathcal{W}_i^n$.

$E_3 : E_1^c \bigcap E_2^c \bigcap F_{L+1}$ where F_{L+1} : there does not exist $(C_1^n, C_2^n, \ldots, C_L^n)$ such that $C_i^n \in \mathcal{B}_i(b_i)$ and $(C_1^n, C_2^n, \ldots, C_L^n)$ are jointly typical.

$E4 : E_1^c \bigcap E_2^c \bigcap E_3^c \bigcap F_{L+2}$ where $F_{L+2} : (\hat{C}_1^n, \hat{C}_2^n, \ldots, \hat{C}_L^n)$ not unique.

$E_5 : E_1^c \bigcap E_2^c \bigcap E_3^c \bigcap E_4^c \bigcap F_{L+3}$ where $F_{L+3} : \frac{1}{n} d(\mathbf{X}^n; \hat{\mathbf{X}}^n) > D + \epsilon$.

Let P_e denotes the probability of decoding error averaged over the ensemble codebooks. Therefore,

$$P_e = P\left(\bigcup_{i=1}^{5} E_i\right) \leq \sum_{i=1}^{5} P_i(E_i).$$

Now,

1. $P(E_1) \to 0$ as $n \to \infty$ by weak law of large numbers.

2. $$P(E_2) \leq \sum_{i=1}^{L} P(E_{2,i})$$

$$P(E_{2,i}) = P\left(E_1^c \bigcap F_i\right) = P(F_i|E_1^c)P(E_1^c) \leq P(F_i|E_1^c)$$
$$= P\left\{(Y_i^n, C_i^n) \text{ not typical for all } C_i^n \in \mathbf{W}_i | (Y_1^n, Y_2^n, \ldots, Y_L^n) \text{ typical}\right\}$$
$$= \left[1 - P\left\{(Y_i^n, C_i^n) \text{ typical for randomly chosen } C_i^n | Y_i^n \text{ typical}\right\}\right]^{M_i}$$

The above equation can be further reduced to,

$$P(E_{2,i}) \leq \left[1 - 2^{-n[I(Y_i;C_i)+\epsilon_i]}\right]^{M_i} \leq \exp\left(-M_i 2^{-n[I(Y_i;C_i)+\epsilon_i]}\right) \qquad (5.24)$$

(5.24) $\rightarrow 0$ as $n \rightarrow \infty$ if $M_i \geq 2^{-n[I(Y_i;C_i)+2\epsilon_i]}$ where ϵ_i could be made arbitrary small as $n \rightarrow \infty$.

3. $P(E_3) \rightarrow 0$ as $n \rightarrow \infty$ by the Markov Lemma (Lemma 14.8.1 [54])on typicality.

4.
$$P(E_4) \leq P\left(F_{L+2} | \bigcup_{j=1}^{3} E_j^c\right) = P\left\{(\hat{C}_1^n, \hat{C}_2^n, \ldots, \hat{C}_L^n) \text{ is not unique } | \bigcup_{j=1}^{3} E_j^c\right\}$$
$$\leq \sum_{Z \subseteq (1,2,\ldots,L)} \left[\left(\prod_{i \in Z} N_i\right) 2^{-n(\Sigma_Z - \epsilon_Z)}\right] \rightarrow 0 \text{ if } \prod_{i \in Z} N_i \leq 2^{n(\Sigma_Z - \epsilon_Z)}$$
$$\forall Z \subseteq (1, 2, \ldots, L)$$

The above follows from the following reasoning: Suppose $(\hat{C}_1^n, \hat{C}_2^n, \ldots, \hat{C}_L^n)$ is the correct decoding. This implies that $\hat{C}_Z^n$ are jointly typical for all $Z \subseteq (1, 2, \ldots, L)$. Then the probability that there exist $Z \subseteq (1, 2, \ldots, L)$ and $\hat{\hat{C}}_Z^n \neq \hat{C}_Z^n$ (i.e. $\forall i \in Z$,$\hat{\hat{C}}_i^n \neq \hat{C}_i^n$) such that $(\hat{\hat{C}}_Z^n, \hat{C}_{Z^c}^n$ are jointly typical is upper bounded by:

$$\left(\prod_{i \in Z}(N_i - 1)\right) 2^{-n(\Sigma_Z - \epsilon_Z)} \leq \left(\prod_{i \in Z} N_i\right) 2^{-n(\Sigma_Z - \epsilon_Z)}$$

Without the loss of generality (w.l.o.g.) suppose $Z = \{i_1, i_2, \ldots, i_k\}$, then $\sum_Z \triangleq \sum_{j=1}^{k} I\left(C_{Z_{k+1-j}^c}; C_{i_j}\right)$; $Z \triangleq \{i_{k-j+1}, i_{k-j+2}, \ldots, i_k\}$, i.e. the last j elements of Z; ϵ_Z can be made arbitrarily small as $n \rightarrow \infty$.

5. $P(E_5) \rightarrow 0$ as $n \rightarrow \infty$ by the jointly strong typicality of $(\mathbf{X}^n, \hat{\mathbf{C}}^n)$, the definition of f and the boundedness of d (i.e. $d_{\max} < \infty$).

To summarize, if $M_i \geq 2^{n[I(Y_i;C_i)+2\epsilon_i]}$ and $\prod_{i\in Z} N_i \leq 2^{n(\Sigma_Z - \epsilon_Z)}$, $\forall i \in \{1, 2, \ldots, L\}$ and $Z \subseteq (1, 2, \ldots, L)$, then $P_e \rightarrow 0$ as $n \rightarrow \infty$. So we have:

$$2^{n\sum_{i\in Z} R_i} = \frac{\prod_{i\in Z} M_i}{\prod_{i\in Z} N_i} \geq 2^{n[I(Y_i;C_i)-\Sigma_Z-\Delta_Z]}$$

where, $\Delta_Z \triangleq 2\sum_{i\in Z} \epsilon_i + 2\epsilon_Z$. Since Δ_Z can be arbitrary small, it follows that:

$$\sum_{i\in Z} R_i \geq \sum_{i\in Z} I(Y_i; C_i) - \Sigma_Z$$

Since, $C_i \rightarrow Y_i \rightarrow (Y_{\{i\}^c}, C_{\{i\}^c})$, we have

$$\begin{aligned}
I(Y_Z; C_Z | C_{Z^c}) &= h(C_Z|C_{Z^c}) - h(C_Z|C_{Z^c}, Y_Z) \\
&= \sum_{j=1}^{k} h(C_{i_j}|C_{Z^c_{k+1-j}}) - \sum_{j=1}^{k} h(C_{i_j}|Y_{i_j}) \\
&= \sum_{j=1}^{k} \Big[h(C_{i_j}|C_{Z^c_{k+1-j}}) - h(C_{i_j})\Big] - \sum_{j=1}^{k} \big[h(C_{i_j}|Y_{i_j}) - h(C_{i_j})\big] \\
&= -\sum_{j=1}^{k} I(C_{i_j}; C_{Z^c_{k+1-j}}) + \sum_{j=1}^{k} I(C_{i_j}; Y_{i_j}) \\
&= \sum_{j\in Z} I(Y_j; C_j) - \Sigma_Z
\end{aligned}$$

Suppose $Z = (i_1, i_2, \ldots, i_k)$, we can write the above equation as:

$$\sum_{i\in Z} R_i \geq \sum_{i\in Z} I(Y_Z; C_Z | C_{Z^c}) \ \forall Z \subseteq (1, 2, \ldots, L) \quad \blacksquare$$

CHAPTER 6

Cooperative Channel Coding

Doing things for others always pays dividends. - C. M. Bristol.

Sensor networks are mostly energy constrained and hence it is essential to adopt the various available techniques to conserve energy. To address this very goal, the data fusion, local cooperative signal processing, scalability, controlled mobility, sensor density and exploitation of correlation through source coding have been so far discussed. All of these techniques rely on sensor cooperation. In this chapter, the energy concern in sensor networks is further considered for communications among clusters of sensor nodes.

Consider a wireless sensor network deployed to measure some phenomenon [181, 70, 204, 190]. The sensed data by some group (or cluster) of sensors has to be relayed to the destination. It is very well possible that the clusters of sensors are widely separated, demanding cooperation to achieve communications at the desired rate. Such a scenario is depicted in Figure 6.1 and is easy to envision for sensor networks.

The transmitting cluster, in Figure 6.1, senses the phenomenon and the measured data needs to be transmitted to the destination. In many cases, the data from more than one sensor needs to be transmitted. For example, the sensors could be sensing different perspectives of the phenomenon and hence the information stream contained by each sensor could be different. The most likely scenario

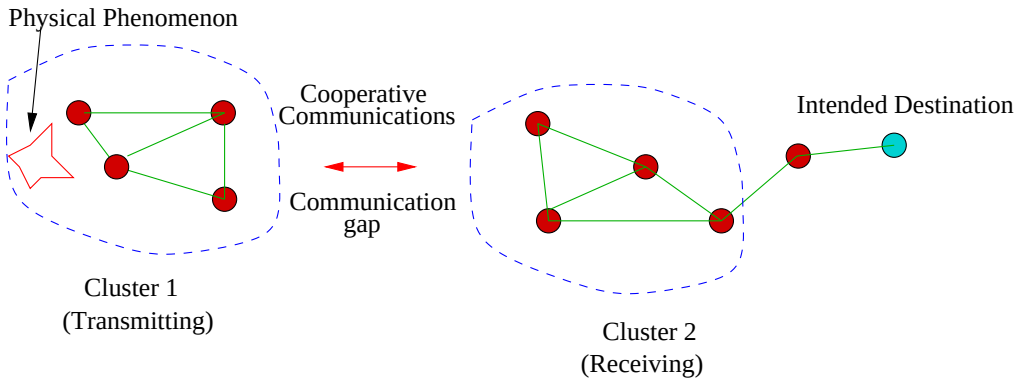


Figure 6.1: Cooperative Communications in Sensor Network

for sensor network is to have 4-node cooperative situation, i.e. two independent transmitters and two independent receivers. Note, that the distance between the two clusters is much larger than the nodes within a cluster. This chapter deals with the information theoretic aspects of such a channel model and explicitly derives the data rates for each transmitting node and considers the effect of transmitter and/or receiver cooperation on the rate region.

Apart from overcoming the gap in the sensor networks, the sensors cooperate with each other to achieve more reliable and higher rate communications. Note, that power rather than bandwidth is the main constraint. Also, multiple sensors occupy the same channel and hence, standard multiplexing techniques like TDMA, FDMA, CDMA, and OFDM may not be readily employed.

6.1 Relaying Techniques

The common functionality in any type of cooperation is the employment of a relay. This can be done in either of the following two methods:

Decode-and-Forward. Here, relay nodes decode the received information and

encode it into a fresh signal before transmitting. By this approach, multiple information streams can be supported [98, 209, 256, 138] in a network. Secondly, by decoding the data at the relay node, the interference is also removed in the freshly coded signal. However, depending on the node locations, the interference can either combine coherently or destructively causing high or no interference detection respectively.

Amplify-and-Forward. In this approach, the relay node amplifies the received signal and then transmits it [89, 138]. The relay nodes, along with the information stream, amplify any noise associated with the received signal. The advantage gained by this technique is the data from the nodes other than the transmitter is not treated as interference. Hence, in the scenarios where information streams are correlated, the receiver can use this information as helper or side information to decode the data [167, 172, 41, 171, 252, 253]. This is in contrast to the decode-and-forward approach where correlation exploitation is not easy. The drawback of this technique is that it allows only one level of relay nodes. Supporting multiple information streams may not be easy as relay nodes do not distinguish between the signals of different nodes.

Although both the above mentioned approaches are not spotless, the decode-and-forward method is more intuitive from an information theory perspective. Section 6.2 discusses the related work in this field followed by the problem statement and solution in sections 6.3 and 6.4 respectively. A practical approach to implement cooperation is presented in section 6.5 with the conclusions in section 6.6.

6.2 Related Work

Network information theory has always sprung surprises and excitement among information theorists. Although it has been around for half a century, it has a large set of unsolved problems [54, 80]. For instance, the capacity for the relay channel [59] is still unknown, although some special cases such as the degraded relay channel [52] are solved. Various other channels such as multiple access channels, [54, 5, 141, 73] broadcast channels, [54, 49, 26, 76, 50, 79] and interference channels [54, 216, 6, 201, 17, 103, 37] along with their variations (such as introducing feedback) are either solved or tightly bounded.

In recent times, the notion of cooperative transmission has been considered. The system consisting of three nodes was considered in [209, 210, 229, 228, 256]. The achievable rates for the channel model with two cooperative transmitters and a receiver is derived in [209, 256]. Various permutations of the channel model that could be possible with three nodes, like two cooperating receivers and a single transmitter, are considered in [228, 229]. The four nodes scenario with two nodes acting purely as relays is considered in [155]. A channel with two cooperating transmitters and non-cooperating receivers is considered in [138, 109]. However, the concentration was on outage and diversity. The behavior for fading channel is considered in [138] and for non-fading channel but with a complicated transmitter cooperation scheme involving dirty paper coding in [109]. In more recent work, two cooperating receivers along with the two cooperating transmitters are considered in [117]. However, the model did not consider the transmission of information from a transmitter to both the receivers. In contrast to this, the system with two cooperating transmitters and two cooperating receivers is considered here. The data stream from each transmitter is intended for both the receivers.

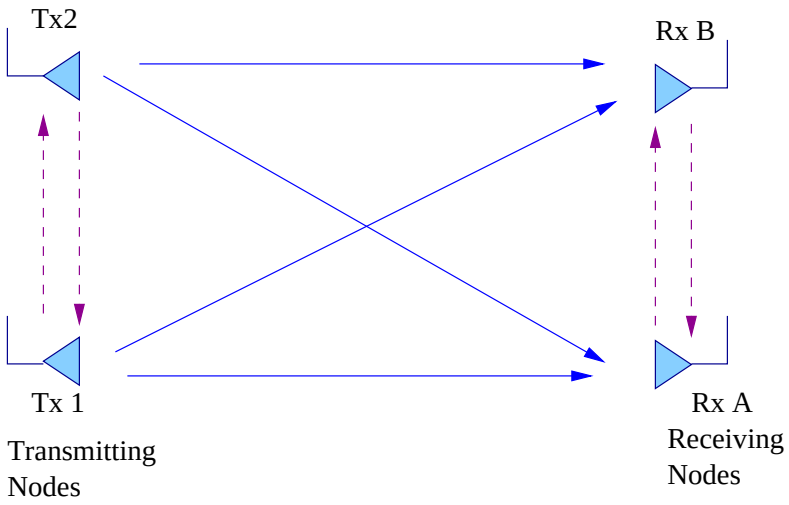


Figure 6.2: Channel model for 4-node cooperation

6.3 Channel Model and Problem Statement

Consider the nodes 1 and 2 of transmitting cluster, and nodes A and B of the receiving cluster. Such a channel model is depicted in Figure 6.2. The transmitters send the information cooperatively and similarly receiving nodes decode the transmitted data on cooperation. Each node receives an attenuated and noisy version of the transmitted data. Note that the transmission from receiver to transmitter is not allowed for the obvious reasons of not being practical given the limited energy budget. Let $Y_i(t)$ be the received baseband signals for the nodes $i = 1, 2, A, B$. The channel model can then be mathematically expressed as,

$$\begin{aligned}
Y_1(t) &= K_{21}X_2(t) + Z_1(t) \\
Y_2(t) &= K_{12}X_1(t) + Z_2(t) \\
Y_A(t) &= K_{1A}X_1(t) + K_{2A}X_2(t) + K_{BA}X_B(t) + Z_A(t) \\
Y_B(t) &= K_{1B}X_1(t) + K_{2B}X_2(t) + K_{AB}X_A(t) + Z_B(t)
\end{aligned}$$

where $X_i(t)$ is the transmitted signal by node i, $Z_i(t) \sim \mathcal{N}(0, N_i)$ is zero mean additive Gaussian channel noise with variance N_i, and K_{ij} is the path gain from node i to node j corresponding to Rayleigh fading. The nodes are assumed to follow the decode-and-forward [138, 209] form of relaying. Let the power constraint on $X_i(t)$, $\{i = 1, 2, A, B\}$ be P_i and W_j, $j = 1, 2$, be the information content of the transmitters.

Assume a discrete time version of the channel model. Node 1 transmits information $W_{1\to 2}$, $W_{1\to A}$ and $W_{1\to B}$ to nodes 2, A and B respectively. Hence, the rates can be divided as:

$$R_{1\to A} = R_{1A} + R_{1BA} + R_{12A} \tag{6.1}$$

$$R_{1\to B} = R_{1B} + R_{1AB} + R_{12B} \tag{6.2}$$

$$R_{2\to A} = R_{2A} + R_{2BA} + R_{21A} \tag{6.3}$$

$$R_{2\to B} = R_{2B} + R_{2AB} + R_{21B} \tag{6.4}$$

For simplicity, consider the signal transmitted by node 1:

$$\begin{aligned} X_1 = {} & X_{1A} + X_{1B} + X_{12A} + X_{12B} + X_{1BA} + X_{1AB} \\ & + U_{12A} + U_{12B} + U_{1BA} + U_{1AB} \end{aligned}$$

where X_{1A}, X_{1B} are direct path signals, X_{12A}, X_{12B}, X_{1BA}, X_{1AB} are the relay path signals, and U_{12A}, U_{12B}, U_{1BA}, U_{1AB} are the coherently combined signals.

The direct path signal conveys the message W_{1A} and W_{1B} with rates R_{1A} and R_{1B} respectively. The relay path signal transmits W_{12A} to node 2 (conveyed by X_{12A}) at rate R_{12A}, W_{12B} to node 2 at rate R_{12B}, W_{1BA} to node B at rate R_{1BA}, and W_{1AB} to node A at rate R_{1AB}. The coherently combining signals involve both transmitter and receiver cooperation. For example, U_{12A} is intended to combine coherently with the signal from node 2 (transmitter cooperation), U_{1BA}

is intended to combine with the signals from node B (receiver cooperation) and in both instances the signal is meant for node A. A similar explanation holds for other coherently combining signals.

Expressions similar to that for X_1 can be found for other nodes. Even though the nodes A and B do not have any information of their own to convey they may relay information in the form of coherently combined signals. Based on the composition of X_1, the power is divided into:

$$\begin{aligned} P_1 &= P_{1A} + P_{1B} + P_{12A} + P_{12B} + P_{1BA} + P_{1AB} + P_{U12A} + \\ &\quad P_{U12B} + P_{U1BA} + P_{U1AB} \end{aligned}$$

We assume that the B blocks of length n are transmitted and the block size B and n are sufficiently large for the perfect decoding. The nodes cooperate based on the information stream received in the previous block. The direct path signal depends on the direct path message in the current block and relay path messages transmitted or decoded in the previous block. The relay path signal depends on the relay messages transmitted in the current block and the relay messages transmitted or decoded in the previous block. Similarly, the coherent combining signal from a node to the destination depends on the relay messages transmitted or decoded in the previous block. For example,

$$\begin{aligned} X_{1A} &= \sqrt{P_{X_{1A}}}\tilde{X}_{1A}\left[W_{1A}(b), W_{12A}(b-1), W_{21A}(b-1), W_{1BA}(b-1), W_{1AB}(b-1)\right] \\ X_{12A} &= \sqrt{P_{X_{12A}}}\tilde{X}_{12A}\left[W_{12A}(b), W_{12A}(b-1), W_{21A}(b-1)\right] \\ U_{12A} &= \sqrt{P_{U_{12A}}}\tilde{U}\left[W_{12A}(b-1), W_{21A}(b-1)\right] \end{aligned}$$

The other direct-path, relay-path and coherently combined-path signals have analogous expressions.

With the above background, the goal is then to carry out an information

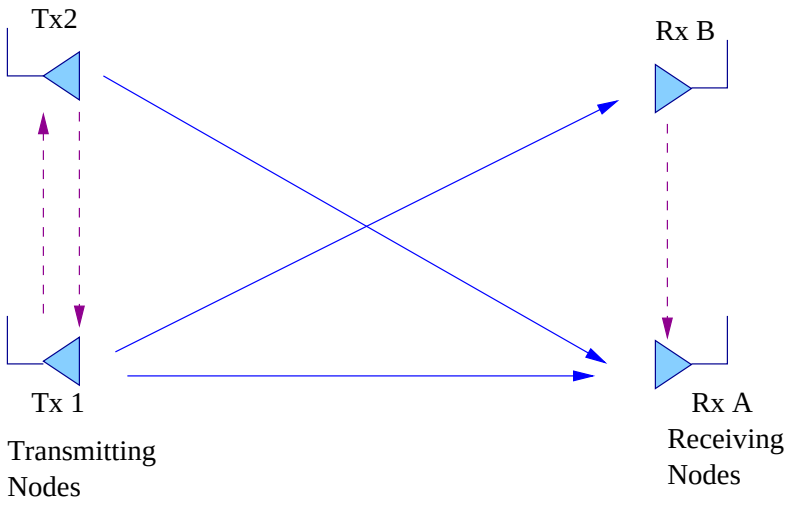


Figure 6.3: Channel model with respect to node A

theoretic analysis of this channel model to derive the upper bound or inner region of the rate region [176].

6.4 Deriving the Rate Region

From the above mentioned channel model, the upper bounds on the achievable rate region can be derived using the technique of forward and backward decoding [53, 248, 260] employed in [228, 229]. This technique is the revised version of the one employed in [209]. In contrast to [209], where the relay-path signal is decoded in the forward stage, the revised technique explores the idea of simultaneously decoding as many messages as possible irrespective of their immediate need. Further details on the revised technique are available in [228, 229].

For the sake of brevity and simplicity, consider the rates from the point of view of transmitting node A. Figure 6.3 illustrates this channel if only node A is transmitting the information. We use the forward and backward decoding approach as in [209, 229]. Also, all the direct path signals are intended to be

decoded in the backward decoding stage. In the forward decoding stage node A has to decode W_{1AB} and W_{2AB} if these messages are employed. Note that, if node A can decode all the data contained in the signals, the channel is like a multiple access channel [54] from the point of view of node A. Based on this, we obtain a set of rate constraints defining the upper bound on the achievable region with respect to node A. For instance,

$$\begin{aligned}
R_{12} &< E\left\{C\left(\frac{K_{1A}^2 P_{12}}{N_A}\right)\right\} && (6.5)\\
R_{1A} &< E\left\{C\left(\frac{K_{1A}^2 P_{1A}}{N_A}\right)\right\} && (6.6)\\
R_{1B} &< E\left\{C\left(\frac{K_{1A}^2 P_{1B}}{N_A}\right)\right\} && (6.7)\\
R_{BA} &< E\left\{C\left(\frac{K_{BA}^2 P_{BA}}{N_A}\right)\right\} && (6.8)\\
R_{1AB} &< E\left\{C\left(\frac{K_{1A}^2 P_{1AB}}{N_A}\right)\right\} && (6.9)\\
R_{2AB} &< E\left\{C\left(\frac{K_{2A}^2 P_{2AB}}{N_A}\right)\right\} && (6.10)\\
R_{1A}+R_{1B} &< E\left\{C\left(\frac{K_{1A}^2 (P_{1A}+P_{1B})}{N_A}\right)\right\} && (6.11)
\end{aligned}$$

where, $E(\cdot)$ denotes an expectation and $C(x) = (1/2)\log(1+x)$. Node A has to decode W_{1AB} and W_{2AB} to facilitate a relay in the next block. Thus, constraints on R_{1AB} and R_{2AB} are required. The above rate constraints (6.5)-(6.11) are only few examples of a large set of constraints that could be penned. The other rate constraints, such as $\{R_{21}, R_{2A}, R_{2B}, R_{12A}, R_{1BA}, \ldots, R_{21A}, R_{21B}, \ldots, R_{1A} + R_{12}, \ldots, R_{1A} + R_{1B} + R_{12} + R_{12A} + R_{12B} + R_{1AB} + R_{1BA}, \ldots\}$, could be derived for node A.

Following the same approach as for node A, the rate constraints for nodes B,1 and 2 could also be derived. Hence, in the forward decoding stage node B has to decode W_{1BA} and W_{2BA}, node 1 needs to decode W_{21A} and W_{21B}, and node 2

will need to decode W_{12A} and W_{12B}.

In the backward decoding stage, node A has to decode W_{1A}, W_{2A}, W_{1BA}, W_{2BA}, W_{12A} and W_{21A}. Therefore,

$$R_{1A} < E\left\{C\left(\frac{K_{1A}^2 P_{1A}}{N_A}\right)\right\} \tag{6.12}$$

$$R_{2A} < E\left\{C\left(\frac{K_{2A}^2 P_{2A}}{N_A}\right)\right\} \tag{6.13}$$

But, if say W_{1B} can also be decoded rather than being considered as noise then,

$$R_{1B} < E\left\{C\left(\frac{K_{1A}^2 P_{1B}}{N_A}\right)\right\} \tag{6.14}$$

Similarly, if W_{2B}, W_{12} and W_{21} can also be decoded instead of being considered as noise, then:

$$R_{1B} < E\left\{C\left(\frac{K_{1A}^2 P_{1B}}{N_A}\right)\right\} \tag{6.15}$$

$$R_{2B} < E\left\{C\left(\frac{K_{2A}^2 P_{2B}}{N_A}\right)\right\} \tag{6.16}$$

$$R_{12} < E\left\{C\left(\frac{K_{1A}^2 P_{12}}{N_A}\right)\right\} \tag{6.17}$$

$$R_{21} < E\left\{C\left(\frac{K_{2A}^2 P_{21}}{N_A}\right)\right\} \tag{6.18}$$

The above mentioned rate constraints are only a few from the large set of constraints. As in forward decoding, a huge set of rate constraints can be derived for nodes A, B, 1, and 2. The messages that the nodes have to decode in forward and backward decoding have been summarized in Table 6.4. Note, that some of the rate constraints in the set describing them might be redundant depending on the scenarios.

The rate constraints derived, so far, handle the cooperation implicitly. We shall now explicitly concentrate on transmitter and/or receiver cooperation and examine their consequences. Consider the rate required to transmit from node 1

Decoding Stage	Transmitting Nodes		Receiving Nodes	
	1	2	A	B
Forward	W_{21A}, W_{21B}	W_{12A}, W_{12B}	W_{1AB}, W_{2AB}	W_{1BA}, W_{2BA}
Backward	W_{21}	W_{12}	W_{1A}, W_{2A}, W_{1BA}, W_{2BA}, W_{12A},W_{21A}	W_{1B}, W_{2B}, W_{1AB}, W_{2AB}, W_{12B}, W_{21B}

Table 6.1: Messages required by nodes $\{1, 2, A, B\}$ to decode in forward and backward decoding

to A exploiting cooperation, $R_{1\rightarrow A}$, as described in (6.1):

$$R_{1\rightarrow A} \;=\; R_{1A} + R_{1BA} + R_{12A} \tag{6.19}$$

The rates R_{1A}, R_{1BA} and R_{12A} were derived during the backward decoding. Now, by employing backward decoding for node A and considering cooperation, the rate, $R_{1A} + R_{12A}$ is given by:

$$R_{1A} + R_{12A} < E\left\{C\left(\frac{K_{1A}^2(P_{1A} + P_{12A} + P_{U12A}) + K_{2A}^2 P_{U21A} + 2K_{1A}K_{2A}\sqrt{P_{U12A}P_{U21A}}}{N_A}\right)\right\}$$

Similarly, the other rates in (6.1) could be evaluated to yield:

$$R_{1\rightarrow A} \;<\; E\left\{C\left(\frac{Z}{N_A}\right)\right\} \tag{6.20}$$

where,

$$Z = K_{1A}^2 P_{1\rightarrow A} + K_{2A}^2 P_{U'} + K_{BA}^2 P_{UBA} + 2K_{1A}K_{2A}P_{U''}$$

with,

$$\begin{aligned} P_{1\rightarrow A} &= P_{1A} + P_{1BA} + P_{12A} + P_{U1BA} + P_{U12A} \\ P_{U'} &= P_{U2BA} + P_{U21A} \\ P_{U''} &= \sqrt{P_{U12A}P_{U21A}} + \sqrt{P_{U1BA}P_{U2BA}} \end{aligned}$$

In (6.20), U_{BA} represents the receiver cooperative signal coherently combined with the signals transmitted from nodes 1 and 2. Although U_{1BA} and U_{2BA} are coherent signal meant to combine at node B, they do consist of transmitter cooperation as the signals transmitted to node B also results from cooperative coding at the transmitters. If no transmitter cooperation is allowed, then the signals available at B will only be the relay signals from nodes 1 and 2 coherently combining. Hence, clearly the effect of transmitter cooperation is higher than the receiver cooperation. Also, by appropriate selection of power values in (6.20), transmitter and/or receiver cooperation can be explored. Similarly to (6.20), the other rate constraints described in (6.2), (6.3), and (6.4) can also be derived. Based on these, define:

$$R_1 = R_{1\rightarrow A} + R_{1\rightarrow B} \tag{6.21}$$

$$R_2 = R_{2\rightarrow A} + R_{2\rightarrow B} \tag{6.22}$$

Using the above definition for R_1 and R_2, the advantages gained by using cooperation over not using any are summarized in Figure 6.4. Here, the distance between transmitters (within a cluster) is assumed to be much less than that between the transmitter and receiver. The distance between the receivers is assumed to be same as that between the transmitters. From Figure 6.4, it is quite evident that there are for sure advantages of using cooperation. However, employing only receiver cooperation does not yield significant gain compared to employing only transmitter cooperation. Using both transmitter and receiver cooperation is

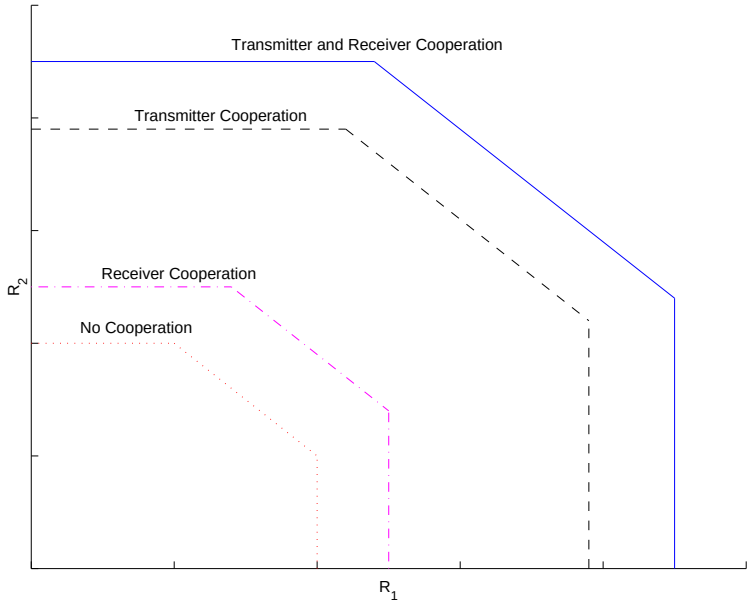


Figure 6.4: Rate Region for 4-node channel model with different cooperating scenarios

superior to using either alone. This result coincides with that in [117], although they had different channel models. Hence, it is sufficient to use only transmitter cooperation if the rate is not a hard constraint. It should be noted that if the distance between the nodes within a cluster is increased, then the gain obtained from cooperation will certainly reduce. If the distance between the nodes is large enough then the rate region will collapse to that without cooperation.

6.5 Discussion and Practical Implementation

The above sections discussed the cooperative rate regions for a four nodes channel model. From the derivations, it is evident that there certainly is gain in using cooperation. However, as few good things come for free, there is a high processing cost to pay. Hence, cooperative transmission should be used only when necessary.

For instance, if a relay node is available, then it is always advisable to opt for it.

Also note that the cooperative transmission channel is very much different from the known channels such as multiple access, broadcast, relay [54], and multiple input multiple output (MIMO) [232]. However, these channels can be considered as the special cases of our formulation. Another channel model similar to cooperative transmission is distributed MIMO [62] but this involves high cost in node synchronization and also involves only one source and destination.It should be noted that the application of a cooperative strategy is highly dependent on geographical locations. For example, if two transmitters are collocated, then it can essentially behave as one node [228, 229]. Hence, the decision to implement a cooperative scheme also depends on the geographical constraints.

The theoretical analysis of the 4-node channel model suggested that the transmitter cooperation outperforms receiver cooperation significantly. Hence, in practice it is wise to consider only a transmitter cooperative strategy. This will also reduce the processing and implementation cost which otherwise is very high. One of the approaches is to implement the algorithms suggested in [138, 109] depending on the channel conditions. Another approach is superposition of the nodes. For instance, consider the channel model as in Figure 6.2. The four nodes, $\{1, 2, A, B\}$, can be divided into two sets of three nodes, for example $\{1, 2, A\}$ and $\{1, 2, B\}$. The partitioning of the sets depend on the geographic location. For instance, consider the channel model as in Figure 6.5.

In this case, the transmitting nodes 1 and 2 cooperatively transmit a message to node B, which in turn relays the information stream to node A. Once the set is partitioned into a 3 node channel model, the algorithm suggested in [209, 210] could be implemented.

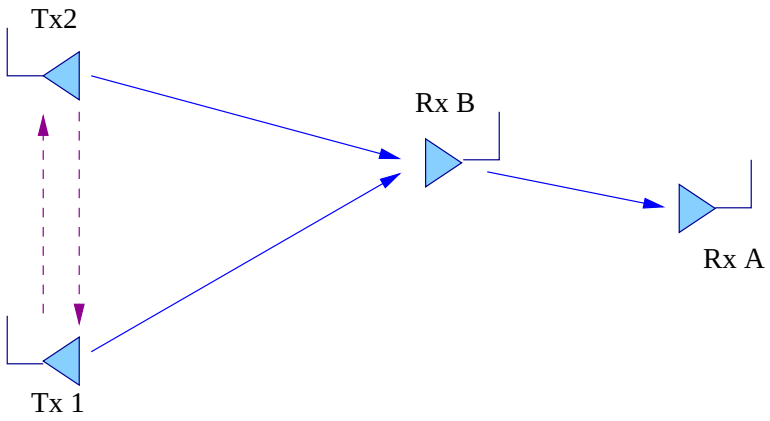


Figure 6.5: Channel model for superposition

6.6 Conclusions

This chapter considered the problem of communications between two distant clusters of nodes. A cooperative transmission strategy was considered to overcome the gap in the multihop networks. The outer bound on the rates was derived. From the information theoretical analysis, it is evident that the transmitter cooperation is more significant than the receiver cooperation. Some practical implementation techniques were outlined. It may be noted that due to the high cost in implementation involved in cooperative coding, whenever relays are available they should be used.

CHAPTER 7

Resource Allocation and QoS in Wireless Ad Hoc Networks

quality of service (QoS) - *The performance properties of a network service, possibly including throughput, transit delay, priority.*
- Webster's Dictionary.

Wireless ad hoc networks, also called wireless multi-hop networks, are formed by multiple nodes, each possessing a wireless transceiver, communicating amongst themselves. An ad hoc network can be used to exchange information between the nodes and to allow nodes to communicate with remote sites that they otherwise would not have the capability to reach. Wireless ad hoc networks can be either static, e. g. sensor networks, or mobile, e. g. Unmanned Aerial Vehicle (UAV) networks. The most important design criterion for any type of networks is guaranteeing Quality of Service, QoS [154].

QoS has become an important issue in various kinds of data networks as some users are no longer satisfied with resource allocation based on service provisioning. QoS measures include bandwidth, delay and delivery guarantees. Different classes of traffic (e.g. voice, data, image, video, etc.) have different bandwidth and delay requirements. Many issues of resource allocation for QoS provisioning are discussed in [44, 120, 119].

In previous chapters, physical layer implications of the ad hoc and sensor networks are presented. This chapter deals with the higher layer abstraction of these networks. Here, the possibility of communication cooperation is not considered and links are considered as aggregation of communicating groups.

We consider mobile ad hoc networks, in which nodes must balance a variety of tasks including sensing and communications. Nodes might thus change location or trajectory for sensing purposes, with constraints on disruption to the QoS of the network. For example, a sensor node may be required to have its position change for improved source identification without disrupting the pre- existing communications. Similarly, in a UAV network, a node might have to move to replace a failed backbone node. This chapter proposes a solution to the following two problems [175]:

1. Maximization of non-communication application QoS (node motion to facilitate sensing) with communication QoS constraints (packet delay, etc.).

2. Maximizing throughput for the newly formed links with communication QoS and link capacities constraints at the new position of the node.

Because the mobile radio channel is fast varying and the number of user nodes is large, a fast and robust decision making algorithm is needed that accommodates a large number of variables for dynamic resource allocation to be feasible [44, 120]. We also propose a heuristic that produces an algorithm that approaches the theoretical limits.

7.1 Convex Optimization and Geometric Programming

Convex optimization refers to minimizing a convex objective function over convex constraint sets. The problems considered in this chapter are non-linear and hence needs some efficient algorithms for solution. However, these problems can be turned into a special class of convex optimization problems called *geometric programming* [31, 240]. Geometric programs focus on monomial and posynomial functions [31].

Definition 7.1.1

A monomial is a function $f : \mathbf{R}^n \rightarrow \mathbf{R}$, where the domain contains all real vectors with positive components:

$$f(x) = cx_1^{a_1}x_2^{a_2}\cdots x_n^{a_n},\ c \geq 0 \text{ and } a_i \in \mathbf{R} \tag{7.1}$$

The exponents a_i of a monomial can be any real numbers, including fractional or negative, but the coefficient c must be nonnegative.

Definition 7.1.2

A sum of monomials, i. e. a function of the form

$$f(x) = \sum_k c_k x_1^{a_1 k}x_2^{a_2 k}\cdots x_n^{a_n k},\ c_j \geq 0 \tag{7.2}$$

is called a posynomial.

Posynomials are closed under addition, multiplication, and nonnegative scaling. Monomials are closed under multiplication and, when the denominator is nonzero, division. If a posynomial is multiplied by a monomial, the result is a posynomial; similarly, a posynomial can be divided by a nonzero monomial, with the result a posynomial.

Geometric programming is an optimization problem of the form:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 1, \quad i = 1, \ldots, m \\ & h_i(x) = 1, \quad i = 1, \ldots, p \end{aligned} \tag{7.3}$$

where $f_0, \ldots, f_m$ are posynomials and $h_1, \ldots, h_p$ are monomials. Geometric programming in the above form is not a convex optimization problem. However, with a change of variables: $y_i = \log x_i$ and $b_{ik} = \log c_{ik}$, the geometric programming form is put into convex form:

$$\begin{aligned} \text{minimize} \quad & \tilde{f}_0(y) = \log \textstyle\sum_k \exp\left(a_{0k}^T y + b_{0k}\right) & \\ \text{subject to} \quad & \tilde{f}_i(y) = \log \textstyle\sum_k \exp\left(a_{ik}^T y + b_{ik}\right) & \leq 0 \\ & \tilde{h}_i(y) = g_i^T y + h_i & = 0 \end{aligned} \tag{7.4}$$

It can be verified that the log sum of exponentials is a convex function [31]. Since the functions $\tilde{f}_i$ are convex and $\tilde{h}_i$ are affine, the problem is a convex optimization problem. Convex optimization problems can be solved globally and efficiently through interior point and primal dual methods [163], with running times that usually scale to the square root of the problem size.

7.2 Problem Formulation

Similar to [44], optimization variables include powers, the number of packets in each traffic class, bandwidth, delay and delivery guarantee required for each QoS class, and capacity for each link. Consider a network with J links with the link capacity of C_j packets per second for each link j. The link capacity can be calculated from Shannon's capacity [214, 186] as,

$$C_j = w_j \log_2 \left(1 + \frac{P}{x_i^4 N_0}\right) \tag{7.5}$$

where w_j Hertz is the bandwidth available on link j and P/x_i^4 Watts is the power available to the receiver i at distance x meters from the transmitter.

There are K classes of traffic with different QoS requirements to be transported over the network. For each QoS class k, the bandwidth required on link j is b_{kj} Hertz and the delay guarantee in the service level agreement is $d_{k,UB}$ seconds. Also, a minimum probability of delivering a packet across the unreliable network is required in the service level agreement, SLA, denoted by $p_{k,LB}$. Here, propagation delay is ignored as it is constant for optimization parameters and only delay due to the transmission line is taken into account.

Each stream of traffic from source s to destination d will traverse certain specific links as dictated by the routing protocol. Denote by K_j the set of traffic using link j and by J_k the set of links traversed by QoS class k. Denote by n_{kj} the number of packets dynamically admitted in the k^{th} class of traffic on link j.

The possibility of link failure should also be taken into consideration. Let p_j be the probability that a link will be maintained during the transmission. The link may fail either due to power shutdown or deep fading causing an outage. The probability p_j can be increased by increasing the transmission power on link j while keeping other parameters of the network constant.

The above constraints on link capacity, bandwidth requirement, delay and delivery probability guarantee do not form a linear program. However, these non-linear optimizations can be turned into a geometric programming problem.

7.2.1 Displacement maximization

Here, we assume that the nodes know the location of their 1-hop neighbors. i.e. the nodes have the distance and angular information of their immediate neighbors. In addition, we make the assumption of common radio range R, for

all the nodes in the network. We will first consider the case where the node which has to move from its original location or deviate from its original path has only one nearest neighbor i.e. has only one link attached to it. Then, we shall deal with the general case of n links. Let X denote the location of a node; we will also use X to refer to the node itself.

7.2.1.1 Single Link Formulation

Consider a simple wireless network with 3 nodes and 2 links as shown in Figure 7.1. A node C needs to move to C'. This has to be done without disrupting

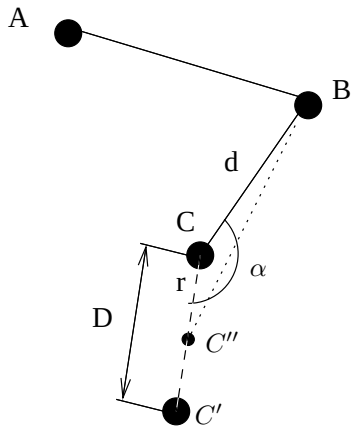


Figure 7.1: A Simple Network with 2 links

ongoing communications. Here, any change in the position of node C will only affect the node B. Hence, the aim is to find the largest possible displacement for node C along the path CC' (consider only the geodesics).

Let the distance between C and C' be denoted by D and the displacement of node C by r, then the objective function is to minimize $D - r$. As the node's position changes, the link distance of that node with its 1-hop neighbors changes. Hence, the link distance BC changes with the node's position. The change in

link distance will cause the change in the link capacity as given by (7.5). Let C'' be the new position of node at distance r from C. If the link distance BC is denoted by d_i then the distance BC'' is given by,

$$x_i = \sqrt{r^2 + d_i^2 - 2d_i r\cos\alpha} \tag{7.6}$$

where α is $\angle BCC''$ and subscript i generalizes for any link i. In general, from geometry it follows,

$$x_i > d_i \text{ for } \begin{cases} 0^o \le \alpha < 90^o & \text{and } r > 2d_i\cos\alpha \\ 90^o \le \alpha \le 270^o & \text{and } r > 0 \\ 270^o < \alpha \le 360^o & \text{and } r > 2d_i\cos\alpha \end{cases} \tag{7.7}$$

$$x_i \le d_i \text{ for } \begin{cases} 0^o \le \alpha < 90^o & \text{and } 0 \le r \le 2d_i\cos\alpha \\ 270^o < \alpha \le 360^o & \text{and } 0 \le r \le 2d_i\cos\alpha \end{cases}$$

The optimization problem for any link i can thus be formulated as:

$$\begin{array}{lll} \text{minimize} & D - r & \\ \text{subject to} & & \\ \text{(i)} & w_i \log_2\left(1 + \frac{P}{x_i^4 N_0}\right) & = C_i \\ \text{(ii)} & \sqrt{r^2 + d_i^2 + 2d_i r\cos\alpha} & = x_i \\ \text{(iii)} & x_i & \le R \\ \text{(iv)} & r & \le D + \delta \\ \text{(v)} & \sum_{k\in K_j} b_{kj} n_{kj} & \le C_j \\ \text{(vi)} & \sum_{j\in J_k}\left(\frac{\sum_{i\in K_j} n_{ij}}{C_j}\right) & \le d_{k,UB}, \quad \forall k \\ \text{(vii)} & \prod_{j\in J_k} p_j & \ge p_{k,LB}, \quad \forall k \\ \text{(viii)} & b_{kj} n_{kj} & \ge R_k, \quad \forall k \\ \text{(ix)} & n_{kj} & \ge N_{kj}, \quad \forall k \\ \text{(x)} & p_j & \le p_{j,UB} \\ \text{(xi)} & b_{kj}, C_j, p_j, d_{k,UB}, p_{k,LB} & \ge 0 \end{array} \tag{7.8}$$

The constraints (i) and (ii) in 7.8 define link capacity and link distance respectively. The constraint (iii) is the radio range constraint. The constraint (iv) bounds the displacement, r, of the node. The constraint (v) is the link capacity constraint, constraint (vi) is the delay guarantee constraint and (vii) the delivery probability constraint. The constraint (viii) delivers a guaranteed data rate to each class of traffic. The constraint (ix) guarantees the minimum data packets, N_{kj} to each link. The other constraints are positivity constraints on the variables, and an upper bound constraint on p_j.

The following parameters are all potential optimization variables: b_{kj}, n_{kj}, p_j, C_j, $d_{k,UB}$ and $p_{k,LB}$. Variables b_{kj}, $d_{k,UB}$ and $p_{k,LB}$ are terms in service level agreement (SLA). The link capacities C_j and probability of maintaining a link p_j are network resources to be optimized over. Admission control and throughput is reflected in n_{kj}.

The optimization problem in (7.8) is of the non-linear form as the optimization variables are multiplied together. Also, it cannot be solved as any standard optimization problem because of the presence of logarithmic and root functions. But, note that, by maximizing $D - r$, we are minimizing r and hence x_i as long as α satisfies the conditions given in (7.7). This in turn will minimize C_j. Also, note that, if capacity increases then the above constraints will be satisfied and the solution becomes trivial. So, we shall consider the case where link distance increases with r. With this background, the optimization problem for the link i

attached with the node to be displaced can be reformulated as,

$$
\begin{aligned}
\text{minimize} \quad & C_i \\
\text{subject to} \quad & \textstyle\sum_{k \in K_j} b_{kj} n_{kj} && \leq C_j \\
& \textstyle\sum_{j \in J_k} \left(\frac{\sum_{i \in K_j} n_{ij}}{C_j} \right) && \leq d_{k,UB}, \qquad \forall k \\
& \textstyle\prod_{j \in J_k} p_j && \geq p_{k,LB}, \qquad \forall k \\
& b_{kj} n_{kj} && \geq R_k, \qquad \forall k \\
& n_{kj} && \geq N_{kj}, \qquad \forall k \\
& p_j && \leq p_{j,UB} \\
& b_{kj}, C_j, p_j, d_{k,UB}, p_{k,LB} && \geq 0
\end{aligned}
\tag{7.9}
$$

In the formulation (7.9), the constraint on the minimum number of data packets on the link ensures that the throughput is not vastly affected while minimizing the link capacity. The formulation is still a non-linear optimization but all the constraints are in posynomial form. Thus, the optimization problem in (7.9) is a form of geometric programming and can be easily solved. On obtaining the value of the feasible link capacity, the link distance and hence the displacement r can be calculated from (7.5), (7.6) and the fact that $r \leq D + \delta$ where $\delta > 0$. We can now extend the optimization problem for the general case of multiple links attached to the node to be displaced.

7.2.1.2 Multi-link Formulation

Consider a network as depicted by Figure 7.2. Here, there are 3 links attached to the node C to be displaced. Similarly to the 1-link case, the interesting case is where the angle formed by the links with CC'' is obtuse. The constraints for the n-link case remain the same as the single link. Instead of minimizing capacity of a single link, we need to minimize the capacity of all the links attached with node at C. Let L be the number of links attached to the node that needs to change its

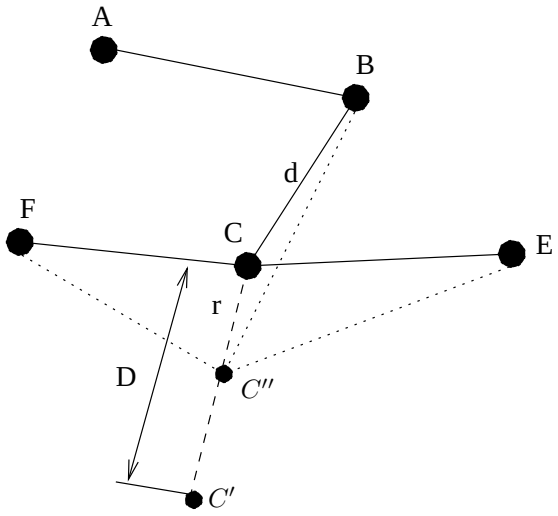


Figure 7.2: Network with multiple links attached to node to be displaced

position or trajectory and γ_j be the weighting factor or priority assigned to each link, then we have the following lemma:

Lemma 7.2.1 *The displacement maximization optimization problem can be reduced to a geometric program.*

$$
\begin{aligned}
&\textit{minimize } \textstyle\sum_{j=1}^{L} \gamma_j C_j \\
&\textit{subject to } \textstyle\sum_{k \in K_j} b_{kj} n_{kj} && \leq C_j \\
&\qquad \textstyle\sum_{j \in J_k} \left(\frac{\sum_{i \in K_j} n_{ij}}{C_j} \right) && \leq d_{k,UB}, \qquad \forall k \\
&\qquad \textstyle\prod_{j \in J_k} p_j && \geq p_{k,LB}, \qquad \forall k \\
&\qquad b_{kj} n_{kj} && \geq R_k, \qquad \forall k \\
&\qquad n_{kj} && \geq N_{kj}, \qquad \forall k \\
&\qquad p_j && \leq p_{j,UB} \\
&\qquad b_{kj}, C_j, p_j, d_{k,UB}, p_{k,LB} && \geq 0
\end{aligned}
\tag{7.10}
$$

The formulation presented above has all the constraints as posynomials and hence can be solves as a geometric program [31, 240, 163].

We can also minimize the link capacity for the links forming acute angles with CC'' as the link distance is bound to increase from $2d_i \cos\alpha$. Similar to the single link case, once the feasible capacity for the individual link (attached to node C) is obtained, we can calculate if the displacement r is feasible for the individual links using (7.5), (7.6) and $r \leq D + \delta$. i.e., if link i and link j are attached to the node to be displaced and associated feasible link capacities are C_i and C_j respectively then, we can calculate the individual displacements r_i and r_j for the links i and j respectively. On knowing the node displacement that the individual links can afford, the final displacement that is in agreement with all the associated links, here links i and j, is $r = \min\{r_i, r_j\}$.

7.2.2 Throughput Maximization

The node which has moved from its original position to a new one may form a new set of 1-hop neighbors. Figure 7.3 depicts such a scenario. This calls for the

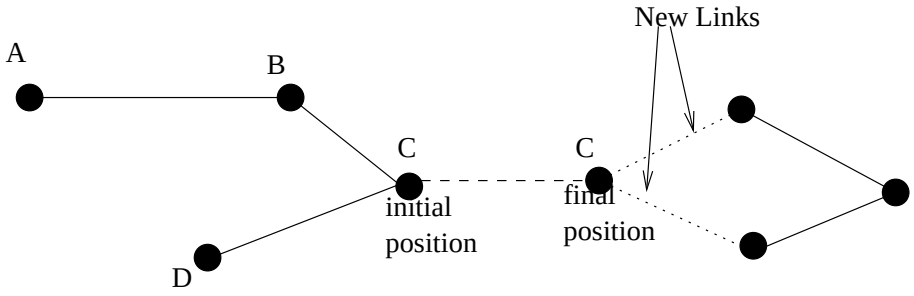


Figure 7.3: Displaced node forming new links

reassignment of traffic on the links attached to the displaced node. The reassignment of traffic can be done by solving the optimization problem for throughput maximization as described in Lemma 7.2.2.

Lemma 7.2.2 *Throughput maximization for the newly formed links*

$$
\begin{aligned}
&\textit{maximize } \textstyle\sum_{j=1}^{L}\sum_{k=1}^{K} n_{kj} \\
&\textit{subject to } \textstyle\sum_{k\in K_j} b_{kj}n_{kj} && \leq C_j \\
&\qquad \textstyle\sum_{j\in J_k}\left(\frac{\sum_{i\in K_j} n_{ij}}{C_j}\right) && \leq d_{k,UB}, \qquad \forall k \\
&\qquad \textstyle\prod_{j\in J_k} p_j && \geq p_{k,LB}, \qquad \forall k \\
&\qquad b_{kj}n_{kj} && \geq R_k, \qquad \forall k \\
&\qquad n_{kj} && \geq N_{kj}, \qquad \forall k \\
&\qquad p_j && \leq p_{j,UB} \\
&\qquad b_{kj}, C_j, p_j, d_{k,UB}, p_{k,LB} && \geq 0
\end{aligned}
\tag{7.11}
$$

Similar to Lemma 7.2.1, the constraints here too are posynomials. Hence, this also is a geometric program [31, 240, 163].

In the next section we shall consider some simulations for the optimization problems mentioned in this section.

7.3 Simulation

Example 7.3.1 (Single Link Formulation)

A simple four node multi-hop network is considered in this example. As shown in Figure 7.4, the network consists of four nodes A, B, C, and D, and three links 1, 2, and 3. There are two traffic classes - audio and video - that have to be transported across the network. The route for audio traffic is BCD and that for video traffic is ABC. The node D has to be moved to a new location D' such that the pre-existing QoS is not disrupted. The new location D' is 15m from D and $\angle CDD' = 120^o$. Let each link length be 10m i. e. $AB = BC = CD = 10$m. The data rate for audio and video traffic is 32kbps and 200kbps respectively. The

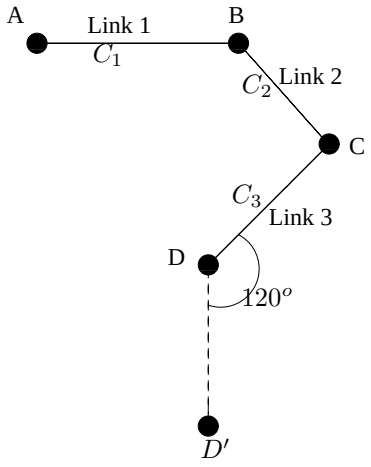


Figure 7.4: Network Topology for Example 1

affordable delay for audio traffic is 75ms and that for video traffic is 100ms. Let the bandwidth occupied by audio and video traffic on any link be 50KHz and 500KHz respectively. The link capacities and the number of packets dynamically admitted on each link will constitute the optimization variables. The objective is to calculate the distance that node D can move along the path DD'. On solving the optimization problem 7.9, we get the minimum capacity on link 3 as 150000.03bps. If the total channel width on link 3 is 1MHz, transmission power as 1W and the noise level as 40dB below then, the feasible distance that node D can be moved along DD' is 10.07m.

Example 7.3.2 (Multiple Link Formulation)

In this example we consider the displacement for a node having multiple links (see Figure 7.5). The data for this problem remains the same as in example 7.3.1 except that the node that needs to be displaced is C. On solving the optimization problem, the link capacities that we obtain are $C_2 = 650000.01$ bps and $C_3 = 150000.03$ bps. For these link capacities, the maximum displacement that node C can afford is 2.015m.

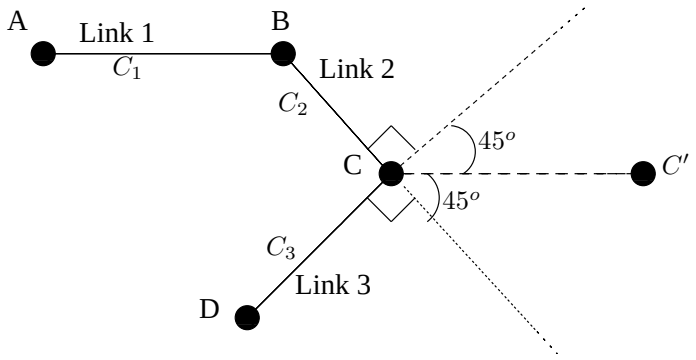


Figure 7.5: Network Topology for Example 2

Example 7.3.3 (Throughput Maximization)

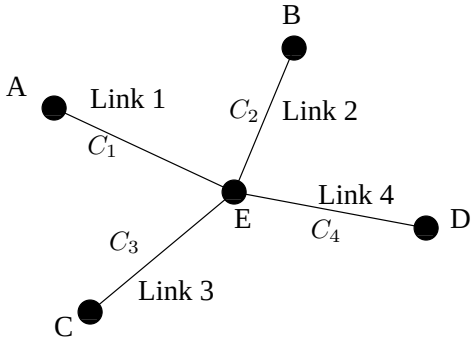


Figure 7.6: Network Topology for Example 3

Here we consider an example of maximizing throughput of the newly formed network at the node's new location. The network for this example is depicted in Figure 7.6. The audio traffic is considered over the links 1, 2 and 4, and the video traffic over the links 2, 3 and 4. The value for the data rate, bandwidth and the delay for the traffic class remain the same as in example 7.3.1. The link capacities are given by $C_1 = 500000$bps, $C_2 = 6000000$bps, $C_3 = 5000000$bps and $C_4 = 6000000$bps. The optimization variables are the number of packets dynamically admitted. Denote n_{ai} as the number of audio packets dynamically admitted on i^{th} link and n_{vi} as the number of video packets on i^{th} link. On solving

the optimization problem, $n_{a1} = 10$, $n_{a2} = 28$, $n_{v2} = 9$, $n_{v3} = 10$, $n_{a4} = 28$ and $n_{v4} = 9$.

7.4 Heuristic Approach

The algorithm in Lemma 7.2.1 and Lemma 7.2.2 requires global information of the network. Hence, it may not be readily implemented. This demands an heuristic approach so that the algorithm can be implemented in real time. The approach employs the same idea as the theoretical case but instead of executing the algorithm (geometric program) with global information, the algorithm is executed only with the local information served by one-hop neighbors of the node that has to be displaced. The other details remain the same. Hence, the requirement of having the global information for decision-making is avoided. For example, on solving the example 7.3.1 with only the local information and having a minimum of three dynamically admitted audio packets on link 3, we obtain the feasible distance that node D can be moved along DD' as 10.07m. For this case the heuristic approaches the theoretical limits. But, this certainly will not be true for a more complex network. Because of the lack of global information, the delay associated with the delivery of traffic across the network is not well defined. Various other heuristics for the geometric program are also available as stated in [119, 217].

7.5 Conclusions

For mobile ad-hoc networks, nodes must balance a variety of tasks including sensing and communications relays. Mobile nodes might thus change location or trajectory for sensing purposes, subject to constraints on disruption of network QoS. The maximization of non-communication application QoS (e. g. node

motion to facilitate sensing) with communication QoS (e. g. packet delay, etc.) constraints and the throughput maximization is considered in this chapter. This is done for mobile ad-hoc networks with multi-hop transmission. Although these formulations are non-linear, it can be readily solved by posing it as a geometric program. The heuristic approach to implement these algorithms in real time has also been discussed.

CHAPTER 8

Concluding Remarks and Future Research

If we knew what it was we were doing, it would not be called research, would it? - A. Einstein.

8.1 Concluding Remarks

Many applications of sensor networks and ad hoc networks require the large scale deployment of the nodes. However, the recent results in [98] were discouraging for such deployments indicating non-scalable behavior. Wireless ad hoc networks were shown to be scalable with the adjustment in source-destination pair distribution. In other words, if the communications in the network is made local then the scalability results could be obtained. Another approach to scalable networks is to provide extra resources such as bandwidth and to have a hierarchial architecture. If mobility is allowed in these networks then scenario changes further. For example, a network with random mobility is scalable [96]. However, there was no constraint on delay and hence the scalability could only be achieved at infinite delay. We considered the network with the combination of static nodes and controllably mobile nodes. For this network model, the worst case delay could be guaranteed and with mobile nodes being only a fraction of static nodes, both per-node throughput and delay could be made finite.

The scalability question in sensor networks transforms into the problem of

information extraction at the desired fidelity. The highest efficiency solution requires joint source-channel coding. However, if the sensors and communication relays densities are made sufficiently large, the decoupling of source and channel coding is possible, yielding scalability with suboptimal resource usage. To completely characterize the sensor network in practice, the notion of spatial fidelity must be considered. This enables the analysis of sensor density and locally cooperating sensors. The spatial fidelity for point sources is equivalent to spatial separation between them whereas for a distributed phenomenon it maps to the cut-off frequency. Based on this, the heuristics for the determination of undersampling, critical-sampling, or over sampling could be proposed.

Continuing further on rate-distortion problems in sensor networks, a generalized solution to the rate-distortion problem with side information and CEO system were derived. These results are an essential building block in the development of a complete information theory for the sensing coverage and communication performance of sensor networks. Engineers could utilize these results to formulate various resource constrained optimization problems to calculate the required sensor density. However, the sources were assumed to be Gaussian for analytical simplicity and an extension to non-Gaussian sources remain an interesting challenge.

One problem in sensor or ad hoc networks is to overcome the gap between two clusters (or groups) of nodes, that is, to communicate the data to a distant cluster. The cooperation between two transmitters and two receivers is considered to solve this problem. From the information theoretic analysis, the transmitter cooperation is found to be more significant than the receiver cooperation. Hence, the practical algorithms could be developed with the transmitter cooperation only. However, it may be noted that due to the high cost of implementation

involved in cooperative coding, it should be used only when required. Ideally, the use of any available relay channel is suggested.

At a higher layer, practical algorithms for wireless ad hoc as well as sensor networks could be formulated that guarantees the QoS constraints such as delay, bandwidth, link capacity, and transmission rate. This also takes care of an unexpected topological changes. The heuristics proposed for guaranteeing QoS in networks closely follows the theoretical optimization formulations. In particular, the maximization of non-communication application QoS (e. g. node motion to facilitate sensing) with communication QoS (e. g. packet delay, etc.) constraints and the throughput maximization could be efficiently solved.

8.2 Information Theory and Statistics

Information theory has the distinction of having an identifiable beginning. However, it also has the large set of open problems. The purpose of this section is to exploit the close relation between information theory and statistics for the possible solutions of the unknown information theoretic problems.

Information theory is the study of data compression and transmission. This field relies on characterizations of data (structures, regularities, long-run behavior) and noise (broadly, the transformations applied during transmission), as well as the costs associated with preparing and communicating the data. Not surprisingly, information theory draws on probability theory and shares many fundamental ideas with statistics. For example, the likelihood ratio test in statistics can be expressed in terms of information theory as: $D(P_{X^n} \parallel P_2) - D(P_{X^n} \parallel P_1) \geq \frac{1}{n} \log T$. Another common example is Fisher Information which has a close relation with the entropy. The relation between information theory and statis-

tics is exploited by many researchers and comprehensive summary of it could be found in [54, 135, 242].

The duality among these fields is clearly observable in detection theory. The objective of the decision problem to minimize the information loss between the source and sink terminals is equivalent to the maximization of the mutual information. The Gaussian CEO problem provides strong connections between information theory and statistics. Such interesting connections have also appeared in the investigation of the multi-terminal estimation problem introduced in [21] and studied in [262, 101, 7, 11]. The close correlation between information theory and statistics for Gaussian CEO system is described in [25, 243]. The calculation of achievable distortion is one of the most challenging part in $R(D)$ bound derivation. The usual information theoretical analysis is limited to the Gaussian distribution. But with the use of Cramer-Rao Bound for random parameter estimation, it is possible to extend to non-Gaussian distributions. For instance,

$$\begin{aligned} D &= \frac{1}{n}\sum_{t=1}^{n} E\left(X(t)-\hat{X}_t\right)^2 \\ &\geq \frac{1}{n}\sum_{t=1}^{n} \frac{1}{E\left[-\frac{\partial^2}{\partial^2 X(t)}\log \Pr\left(X(t),\hat{X}_t,C_1,C_2,\ldots,C_L\right)\right]} \end{aligned}$$

In the above, depending on the distribution, the achievable lower bound on the distribution could be calculated. However, it should be noted that this still remains an open problem.

8.3 Future Research

Shannon's separation theorem [214] states that source coding (compression) and channel coding (error protection) can be performed separately and sequentially,

while maintaining optimality. However, this is true only in the case of asymptotically long block lengths of data. Shannon also stated that if the rate-distortion function of the encoded source is smaller than the channel capacity, theoretically achievable performance is limited solely by source coding errors. However, this assumes that there are no constraints on tolerable channel encoding/decoding complexity which is never the case in real-world systems. Thus, in many practical applications, the conditions of the Shannon's separation theorem neither hold, nor can be used as a good approximation. This resulted to considerable interest in joint source-channel coding (JSCC) to optimize overall performance at reasonable complexity levels.

Consider a sensor network where sensors are observing a physical phenomenon and they communicate to a fusion center over a common wireless channel with a power constraint. This is clearly a JSCC problem. The trade-off is between the total power of the sensors and the fidelity. The optimal trade-off between them is an open problem. One special case of this has been handled in [90]. Extensive research on JSCC has been done in image processing [193, 35, 156, 157, 194, 189]. In communications, the research on JSCC is mainly focussed on the development joint source channel codes for a wireless single link [233, 95, 94]. A general approach for joint source-channel matching based on a parametric distortion model for single transmission link is addressed in [12]. The other approach for a single link is based on measure matching as in [90, 85].

As in the point-to-point case, the network source-channel communication problem is a matter of achieving the right marginal distributions. But the key problem is to identify the set of achievable marginals. The solution obviously lies in formulating the optimization problem similar to that formulated for single sender-receiver pair [214]. However, it may or may not be convex. In fact, the

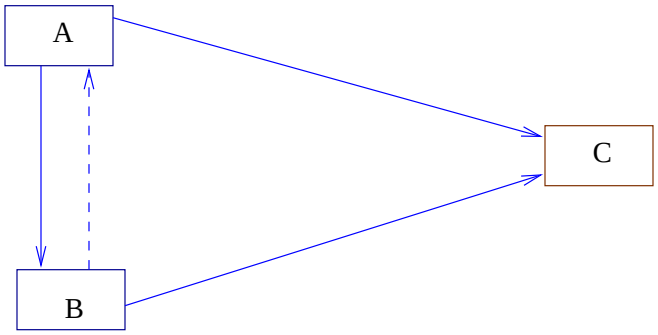


Figure 8.1: Joint Source Channel Coding for a network with three nodes.

formulation of the optimization problem itself is difficult. For example, consider a simplest possible network of 3 sources as in Figure 8.1. To further simplify things, suppose SNR at source C is zero, that is, source C will act only as the receiver. Hence, we have 3 sources (A,B,C) and 4 channels (AB, AC, BA, BC).

For such a scenario, many transmission combinations are possible depending on the Euclidean distances. For instance, SNR for all the channels could be same, or SNR for the channel between A and B maybe higher than that between A and C or B and C. Also, there are constraints on every channel and source. The constraints include distortion, rate, power, and bandwidth. Note, this problem is certainly not similar to that discussed in [176, 209, 117] for the obvious reason that they discussed only channel coding. Here, the objective is to find the optimal trade-off between power and distortion.

There can be various questions asked based on the requirements such as what should be the optimal length of codes. Some of the feasible and interesting questions would be when to use JSCC, what propagation loss law should be followed or rather the range of exponent. In [90, 85] it was argued that for the large (tending to infinity) number of sensors the JSCC becomes inevitable. But based on our results for m-helper and CEO systems in Chapter 4 and 5

respectively, one can argue that the number of sensors required is finite and hence it seems JSCC may not be required. However, joint source channel coding is a rich regime for research.

Along with joint source-channel coding, it should be challenging and interesting to extend the information theoretical analysis to non-Gaussian sources. The code design for multiterminal systems also poses a challenge. In addition, the resource constrained optimization problems to provide the number of required sensors, and communication relays is also a good direction to look. The most significant feature in sensor networks deployed to observe a distributed phenomenon is the sampling frequency. Generally, it is assumed that the sampling is done at super Nyquist rate to avoid any aliasing effect. However, it remains difficult to have a general theory that can recognize between under-sampled, critically-sampled, or over-sampled field. This also could be a possible future direction.

Thus, while this thesis has made a number of contributions, the formulation of a general information theory for sensor networks remains a fascinating and rich area for future research.

References

[1] Multimodal networks of in-situ sensors. http://mantis.cs.colorado.edu/.

[2] What you will want next and the really smart house. Newsweek, May 1999.

[3] Mobile metropolitan ad-hoc networks MobileMAN. Information Society Technologies Proposal. FET Open Initiative 2001, IST-2001-37018, February 2002. http://www.cl.cam.ac.uk/Research/SRG/netos/sla/mobileman/mobileman.pdf.

[4] J Ahlberg. *The Theory of Splines and Their Applications*. Elsevier Science & Technology Books, 1967.

[5] R Ahlswede. Multi-way communication channels. In *Proceedings 2nd. Int. Symp. Information Theory (Tsahkadsor, Armenian S.S.R.)*, pages 23–52. Publishing House of the Hungarian Academy of Sciences, 1971.

[6] R Ahlswede. The capacity region of a channel with two senders and two receivers. *Ann. Prob.*, 2:805–814, 1974.

[7] R Ahlswede and M Burnashev. On minimax estimation in the presence of side information about remote data. *Ann. Statistics*, 18(1):141–171, 1990.

[8] M Ahmed. *Decentralized Information Processing in Wireless Peer-to-Peer Networks*. PhD thesis, University of California, Los Angeles, 2003.

[9] M Ahmed and G Pottie. Information theory of wireless sensor networks: the n-helper gaussian case. In *Proceedings of the IEEE International Symposium on Information Theory*, page 436, 2000.

[10] I Akyildiz, W Su, Y Sankarasubramaniam, and E Cayirci. A survey on sensor networks. *IEEE Communications Magazine*, pages 102–114, August 2002.

[11] S Amari. Fisher information under restriction of shannon information in multiterminal situations. *Ann. Inst. Statist. Math.*, 41(4):623–648, 1989.

[12] S Appadwedula, D Jones, K Ramchandran, and I Kozintsev. Joint source channel matching for a wireless communications link. In *Proceedings of Data Compression Conference (DCC)*, 1998.

[13] S Baatz, M Frank, R Gopffarth, D Kassatkine, P Martini, M Scetelig, and A Vilavaara. Handoff support for mobility with ip over bluetooth. In

Proceedings of the 25th Annual Conference on Local Computer Networks (LCN00), Tampa, FL, November 2000.

[14] N Bansal and Z Liu. Capacity, Delay and Mobility in Wireless Ad-Hoc Networks. In *IEEE INFOCOM 2003*, San Francisco, CA, April 1–3 2003.

[15] S Basagni. Distributed clustering for ad hoc networks. In *Proceedings of IEEE International Symposium on Parallel Architectures, Algorithms, and Networks (I-SPAN)*, pages 310–315, Perth, Western Australia, June.

[16] C Bazlamacci and K Hindi. Minimum-weight spanning tree algorithms: a survey and empirical study. *Computers & Operations Research*, 28(8):767–785, 2001.

[17] R Benzel. The capacity region of a class of discrete additive degraded interference channels. *IEEE Transactions on Information Theory*, IT-25:228–231, 1979.

[18] P Bergamo, A Giovanardi, A Travasoni, D Maniezzo, G Mazzini, and M Zorzi. Distributed power control for energy efficient routing in ad hoc networks. *Wireless Networks*, 10(1):29–42, 2004.

[19] T Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Information and System Sciences Series. Prentice-Hall, Englewood Cliffs, NJ, USA, 1971.

[20] T Berger. *The Information Theory Approach to Communications*, chapter Multiterminal source coding, pages 171–231. Number 229 in CISM Courses and Lectures. Vienna/New York: Springer-Verlag, 1978.

[21] T Berger. Decentralized estimation and decision theory. In *IEEE 7th Springs Workshop in Information Theory*, Mt. Kisco, NY, September 1979.

[22] T Berger. Lossy source coding. *IEEE Transactions on Information Theory*, 44(6):2693–2723, October 1998.

[23] T Berger, K Houswright, J Omura, S Tung, and J Wolfowitz. An upper bound on the rate distortion function for source coding with partial side information at the decoder. *IEEE Transactions on Information Theory*, IT-25:664–666, November 1979.

[24] T Berger and R Yeung. Multiterminal source encoding with one distortion criterion. *IEEE Transactions on Information Theory*, IT-35:228–236, March 1989.

[25] T Berger, Z Zhang, and H Vishwanathan. The CEO problem. *IEEE Transactions on Information Theory*, 42(3):887–902, May 1996.

[26] P Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Transactions on Information Theory*, IT-19:197–207, 1973.

[27] D Bertsekas and R Gallager. *Data networks (2nd edition)*. Prentice-Hall, Inc., 1992.

[28] N Blachman. The convolution inequality for entropy powers. *IEEE Transactions on Information Theory*, IT-11:267–271, April 1965.

[29] R Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18:14–20, 1972.

[30] A Boukerche. A performance comparison of routing protocols for ad hoc networks. In *Proceedings of IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 1940–1946, San Fransisco, CA, April 2001.

[31] S Boyd and L Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[32] J Broch, D Johnson, and D Maltz. The dynamic source routing protocol for mobile ad hoc networks. Internet Draft, MANET Working group, October 1999. draft-ietf-manet-dsr-03.txt.

[33] J Broch, D Maltz, D Johnson, Y Hu, and J Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*, pages 85–97, Dallas, TX, 1998.

[34] J Burke, E Mendelowitz, J Kim, and R Lorenzo. Networking with knobs and knats: Towards ubiquitous computing for artists. In *Ubiquitous Computing, Concepts and Models Workshop*, Gothenburg, Sweden, 2002.

[35] M Bystrom and J Modestino. Combined source-channel coding schemes for video transmission over an additive white gaussian noise channel. *IEEE Journal on Selected Areas in Communications (JSAC)*, 18(6), June 2000.

[36] A Campbell, J Gomez, C Wan, and S Kim. Cellular IP. http://comet.ctr.columbia.edu/cellularip/pub/draft-ietf-mobileip-cellularip-00.txt, January 2000. Internet draft.

[37] A Carleial. Outer bounds on the capacity of the interference channel. *IEEE Transactions on Information Theory*, IT-29:602–606, 1983.

[38] K Castelino, R DSouza, and P Wright. Tool-path optimization for minimizing airtime during machining. *Accepted for publication in Journal of Manufacturing Systems*.

[39] A Cerpa, J Elson, D Estrin, L Girod, M Hamilton, and J Zhao. Habitat monitoring: Application driver for wireless communications technology. In *ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean*, 2001.

[40] A Chakrabarty, A Sabharwal, and B Aazhang. Using predictable observer mobility for power efficient design of a sensor network. In *Second International Workshop on Information Processing in Sensor Networks(IPSN)*, pages 129–145, Palo Alto, CA, USA, April 2003.

[41] J Chen, X Zhang, T Berger, and S Wicker. An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem. *IEEE Journal on Selected Areas in Communications (Special Issue on Fundamental Performance limits of wireless Sensor networks)*. to appear.

[42] J Chen, X Zhang, T Berger, and S Wicker. Rate allocation in distributed sensor network. In *Proceedings of the Annual Allerton Conference on Communication, Control, and Computing*, October 2003.

[43] M Chiang and G Carlsson. Power control, admission control, and qos analysis for ad hoc wireless networks. In *Proceedings of IEEE International conference on communications (ICC)*, pages 245–249, Helsinki, Finland, June 2001.

[44] M Chiang, D O'Neill, D Julian, and S Boyd. Resource allocation for QoS provisioning in wireless ad hoc networks. In *Proceedings of IEEE Globecom*, November 2001.

[45] C Chien, I Elgorriaga, and C McConaghy. Low-power direct-sequence spread-spectrum modem architecture for distributed wireless sensor networks. In *Proceedings of the 2001 international symposium on Low power electronics and design (ISLPED)*, pages 251–254, Huntington Beach, CA, 2001.

[46] J Chou, D Petrovic, and K Ramchandran. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks. In *In Proceeding of IEEE INFOCOM*, San Fransisco, CA, 2003.

[47] C Comaniciu and H Poor. On the capacity of mobile ad hoc networks with delay constraints. In *Proceedings of the IEEE CAS Workshop on Wireless Communications and Networking*, September 2002.

[48] M Conti. Body, personal, and local ad hoc wireless networks. *The handbook of ad hoc wireless networks*, pages 3–24, 2003.

[49] T Cover. Broadcast channels. *IEEE Transactions on Information Theory*, IT-18:2–14, 1972.

[50] T Cover. An achievable rate region for the broadcast channel. *IEEE Transactions on Information Theory*, IT-21:399–404, 1975.

[51] T Cover. A proof of the data compression theorem of slepian and wolf for ergodic sources. *IEEE Transactions on Information Theory*, 21:226–228, March 1975.

[52] T Cover and A Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, IT-25:572–584, 1979.

[53] T Cover and C Leung. An achievable rate region for the multiple access channel with feedback. *IEEE Transactions on Information Theory*, IT-27:292–298, 1981.

[54] T Cover and J Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc, 1991.

[55] I Cox and G Wilfong. *Autonomous Robot Vehicles*. Springer, 1990.

[56] I Csiszar and J Korner. Towards a general theory of source networks. *IEEE Transactions on Information Theory*, IT-26:155–165, 1980.

[57] F Cullen, J Jarvis, and H Ratliff. Set partitioning based heuristics for interactive routing networks. *Networks*, (11):125–143, 1981.

[58] S Das, C Perkins, and E Royer. Performance comparison of two on-demand routing protocols for ad hoc networks. In *Proceedings of IEEE INFOCOM*, pages 3–12, 2000.

[59] E Van der Meulen. A survey of multi-way channels in information theory. *IEEE Transactions on Information Theory*, IT-23:1–37, 1977.

[60] J Desrosiers, G Laporte, M Sauve, F Soumis, and S Taillefer. Vehicle routing with full loads. *Computer Operations Research*, 15(3):219–226, 1988.

[61] S Diggavi, M Grossglauser, and D Tse. Even one-dimensional mobility increases ad hoc wireless capacity. Lausanne, Switzerland, June 2002.

[62] M Dohler, A Gkelias, and H Aghvami. A Resource Allocation Strategy for Distributed MIMO Multi-Hop Communication Systems. *IEEE Communications Letter*, 8(2), February 2004.

[63] O Dousse and P Thiran. Connectivity vs capacity in dense ad hoc networks. In *Proceedings of IEEE INFOCOM*, March 2004.

[64] S Draper and G Wornell. Successively structured CEO problems. In *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, 2002.

[65] Y Dumas, J Desrosiers, and F Soumis. The pickup and delivery problem with time windows. *European Journal of Operational Research*, 54:7–22.

[66] A Elson and K Römer. Wireless sensor networks: A new regime for time synchronization. In *Proceedings of the First Workshop on Hot Topics In Networks (HotNets - I)*, Princeton, NJ, 2002.

[67] J Elson, L Girod, and D Estrin. Fine-grained network time synchronization using reference broadcasts. *SIGOPS Operating Systems Review*, 36(SI):147–163, 2002.

[68] J Elson, L Girod, and D Estrin. Fine-grained network time synchronization using reference broadcasts. February 2002.

[69] D Estrin, L Girod, G Pottie, and M Srivastava. Instrumenting the world with wireless sensor networks. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, UT, June 2001.

[70] D Estrin, R Govindan, and J Heidemann. Embedding the internet: introduction. *Communications of the ACM*, 43(5):38–41, 2000.

[71] A Fanimokun and J Frolik. Effects of natural propagation environments on wireless sensor network coverage area. In *Southeastern Symposium on System Theory (SSST)*, March 2003.

[72] T Flynn and R Gray. Encoding of correlated observations. *IEEE Transactions on Information Theory*, 33:773–787, November 1987.

[73] T Gaarder and J Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *IEEE Transactions on Information Theory*, IT-21:100–102, 1975.

[74] R Gallager. *Low-Density Parity-Check Codes*. PhD thesis, Massachusetts Institute of Technology (MIT), Cambridge, MA, 1962.

[75] R Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., 1968.

[76] R Gallager. Capacity and coding for degraded broadcast channels. *Problemy Peredaci Informaccii*, 10(3):3–14, 1974.

[77] R Gallager, P Humblet, and P Spira. A distributed algorithm for minimum-weight spanning trees. *ACM Transactions on Programming Languages and Systems*, 5(1):66–77, 1983.

[78] G Gallo, F Malucelli, and M Marre. Hamiltonian paths algorithms for disk scheduling. Technical Report 20/94, HP Labs , Dipartimento di Informatica, Universita di Pisa, 1994. http://www.hpl.hp.com/techreports/95/HPL-95-71.html.

[79] A Gamal. The feedback capacity of degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-24:379–381, 1978.

[80] A Gamal and T Cover. Multiple user information theory. *Proceedings IEEE*, 68:1466–1483, 1980.

[81] A Gamal, J Mammen, B Prabhakar, and D Shah. Throughput-Delay Trade-off in Wireless Networks. In *IEEE INFOCOM 2004*, Hong Kong, March 7–11 2004.

[82] H Gamal. On the transport capacity of the many-to-one dense wireless network. In *Proceedings of the 58th IEEE Vehicular Technology Conference (VTC)*, Orlando, FL, 2003.

[83] D Ganesan, S Ratnasamy, H Wang, and D Estrin. Coping with irregular spatio-temporal sampling in sensor networks. *SIGCOMM Computer Communication Review*, 34(1):125–130, 2004.

[84] J Gao. *Energy Efficient Routing for Wireless Sensor Networks*. PhD thesis, University of California, Los Angeles (UCLA), 2000.

[85] M Gastpar. *To Code Or Not To Code*. PhD thesis, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, November 2002.

[86] M Gastpar. *To Code Or Not To Code*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, November 2002.

[87] M Gastpar. The wyner-ziv problem with multiple sources. Technical Report I&C Technical Report IC/2002/36, EPFL, Lausanne, Switzerland, June 2002. submitted to IEEE Transactions on Information Theory.

[88] M Gastpar. On wyner-ziv networks. In *Proceedings of 37th Asilomar Conference on Signals, Systems, and Computers*, pages 9–12, November 2003.

[89] M Gastpar and M Vetterli. On the capacity of wireless networks: the relay case. In *Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 3, pages 1577 –1586, June 2002.

[90] M Gastpar and M Vetterli. Source channel communication in sensor networks. In *Second International Workshop on Information Processing in sensor Networks*, pages 162–177, Palo Alto, CA, April 2003.

[91] M Gerla and J Tsai. Multicluster, mobile, multimedia radio network. *ACM/Baltzer Journal of Wireless Networks*, 1(3):255–265, 1995.

[92] A Giridhar and P Kumar. Computing and communicating functions over sensor networks. *Journal on Selected Areas in Communications (JSAC)*. under submission. Revised July 30, 2004. http://black.csl.uiuc.edu/~prkumar/ps_files/04_07_30_data_fusion.pdf.

[93] A Glavieux, C Berrou, and P Thitimajshima. Near shannon limit error-correcting coding and decoding: Turbo-codes. In *Proceedings of IEEE International Conference on Communications (ICC)*, pages 1064–1070, 1993.

[94] A Goldsmith and M Effros. Iterative joint design of multiresolution source and channel codes. In *Proceedings of the IEEE International Conference on Communications (ICC)*, 1997.

[95] A Goldsmith and M Effros. Joint design of fixed-rate source codes and multiresolution channel codes. *IEEE Transactions on Communications*, 46(10):1301–1312, October 1998.

[96] M Grossglauser and D Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking (TON)*, 10(4):477–486, 2002.

[97] P Gupta and P Kumar. Critical power for asymptotic connectivity. In *Proceedings of the 37th IEEE Conference on Decision and Control*, Tampa, FL, December 1998.

[98] P Gupta and P Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, IT-46(2):388–404, March 2000.

[99] P Gupta and P Kumar. Internets in the sky: The capacity of three dimensional wireless networks. 1(1):33–50, January 2001.

[100] P Gupta and P Kumar. Towards and information theory of large networks: An achievable rate region. 49:1877–1894, 2003.

[101] T Han and S Amari. Parameter estimation with multiterminal data compression. *IEEE Transactions on Information Theory*, 41, pages =, 1995.

[102] T Han and K Kobayashi. A unified achievable rate region for a general class of multi-terminal source coding systems. *IEEE Transactions on Information Theory*, IT-26:277–288, May 1980.

[103] T Han and K Kobayashi. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory*, IT-27:49–60, 1981.

[104] Z Hass and M Pearlman. The zone routing protocol (ZRP) for ad hoc networks. Internet Draft, MANET Working group, March 2000.

[105] W Heinzelman, A Chandrakasan, and H Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, volume 8, pages 1–10. IEEE Computer Society, January 2000.

[106] W Heinzelman, J Kulik, and H Balakrishnan. Adaptive protocols for information dissemination in wireless sensor networks. In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, pages 174–185, Seattle, WA, 1999.

[107] L Herzberger and F Groen. *Intelligent Autonomous Systems*. Elsevier Science Publishers, 1987.

[108] E Hossain, R Palit, and P Thulasiraman. Clustering in mobile wireless ad hoc networks: issues and approaches. *Wireless communications systems and networks*, pages 383–424, 2004.

[109] A Host-Madsen. A new achievable rate for cooperative diversity based on generalized writing on dirty paper. In *IEEE International Symposium on Information Theory*, page 317, June 2003.

[110] K Housewright. *Source Coding Studies for Multiterminal Systems*. PhD thesis, University of California, Los Angeles (UCLA), 1977.

[111] A Hu and S Servetto. Asymptotically optimal time synchronization in dense sensor networks. In *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications*, pages 1–10, San Diego, CA, 2003. ACM Press.

[112] C Huang and Y Tseng. The coverage problem in a wireless sensor network. In *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications*, pages 115–121. ACM Press, 2003.

[113] P Ishwar, A Kumar, and K Ramchandran. Distributed sampling for dense sensor networks: A bit-conservation principle. In *Proceedings of the International Workshop on Information Processing in Sensor Networks*, April 2003.

[114] A Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, Inc., 1989.

[115] S Jain, K Fall, and R Patra. Routing in a delay tolerant network. In *SIGCOMM*, 2004.

[116] S Jain, R Shah, G Borriello, W Brunette, and S Roy. Exploiting mobility for energy efficient data collection in sensor networks. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2004.

[117] N Jindal, U Mitra, and A Goldsmith. Capacity of ad-hoc networks with node cooperation. In *IEEE International Symposium on Information Theory*, June 2004.

[118] D Johnson and D Maltz. Dynamic source routing in ad hoc wireless networks. In Imielinski and Korth, editors, *Mobile Computing*, volume 353. Kluwer Academic Publishers, 1996.

[119] D Julian, M Chiang, D O'Neill, and S Boyd. QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks. In *Proceedings of IEEE INFOCOM*, 2002.

[120] D Julian, D O'Neill, and M Chiang. Robust and QoS constrained optimization of power control in wireless cellular networks. In *Proceedings of IEEE Vehicular Technology Conference*, October 2001.

[121] E Jung and N Vaidya. A power control mac protocol for ad hoc networks. In *Proceedings of the 8th annual international conference on Mobile computing and networking*, pages 36–47, Atlanta, GA, 2002.

[122] W Kaiser, G Pottie, M Srivastava, G Sukhatme, J Villasenor, and D Estrin. Networked infomechanical systems (NIMS) for ambient intelligience. Technical Report 31, Center For Embedded Networked Sensing, UCLA, 2003. http://www.cens.ucla.edu.

[123] W Kaiser, G Pottie, M Srivastava, G Sukhatme, J Villasenor, and D Estrin. *Ambient Intelligence*, chapter Networked Infomechanical Systems (NIMS) for Ambient Intelligence. Springer-Verlag, 2004.

[124] A Kansal, A Pandya, M Srivastava, and G Pottie. Throughput and delay in networks with controlled mobility. Technical report, Center For Embedded Networked Sensing, UCLA, August 2004. submitted to IEEE International Conference on Communications (ICC), 2005. http://www.cens.ucla.edu.

[125] A Kansal, A Somasundara, D Jea, M Srivastava, and D. Estrin. Intelligent fluid infrastructure for embedded networks. In *The Second International Conference on Mobile Systems, Applications, and Services (MobiSys)*, June 2004.

[126] A Kaspi and T Berger. Rate-distortion for correlated sources with partially separated encoders. *IEEE Transactions on Information Theory*, IT-28:828–840, November 1982.

[127] J Kempf and P Yegani. Openran: A new architecture for mobile wireless internet radio access networks. *IEEE Communications Magazine*, 40:118–123, May 2002.

[128] P Keskinocak and S Tayur. Scheduling of time-shared jet aircraft. *Transportation Science*, 3:277–294.

[129] S Khuller, B Raghavachari, and N Young. Balancing minimum spanning and shortest path trees. In *Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms*, pages 243–250. Society for Industrial and Applied Mathematics, 1993.

[130] L Klein. *Sensor and Data Fusion Concepts and Applications*. SPIE Press, 1993.

[131] A Knoll and J Meinkoehn. Data fusion using large multi-agent networks: an analysis of network structure and performance. In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 113–120, Las Vegas, NV, October 1994.

[132] J Körner. chapter Some methods in multi-terminal communication: A Tutorial Survey, pages 173–224. CISM Courses and Lectures, no. 219. Springer-Verlag, Wien and New York, 1975.

[133] J Körner and K Marton. Images of a swt via two channels and their role in multi-user communication. *IEEE Transcations on Information Theory*, 23:751–761, 1977.

[134] U Kozat and L Tassiulas. Throughput capacity of random ad hoc networks with infrastructure support. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 55–65. ACM Press, 2003.

[135] S Kullback. *Information theory and statistics*. John Wiley and Sons, New York, 1959.

[136] A Kumar, P Ishwar, and K Ramchandran. On distributed sampling of smooth non-bandlimited fields. In *Proceedings of the third international symposium on Information processing in sensor networks*, pages 89–98, 2004.

[137] A LaMarca, W Brunette, D Koizumi, M Lease, S Sigurdsson, K Sikorski, D Fox, and G Borriello. Plantcare: An investigation in practical ubiquitous systems. In *Ubicomp*, September 2002.

[138] J Laneman, D Tse, and G Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*. Accepted for publication. http://www.nd.edu/ jnl/pubs/.

[139] J Li, C Blake, D De Couto, H Lee, and R Morris. Capacity of ad hoc wireless networks. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 61–69. ACM Press, 2001.

[140] X Li, P Wan, and O Frieder. Coverage in wireless ad hoc sensor networks. *IEEE Transactions on Computers*, 52(6), June 2003.

[141] H Liao. *Multiple access channels*. PhD thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[142] C Lin and M Gerla. Adaptive clustering for mobile wireless networks. *IEEE Journal of Selected Areas in Communications*, 15(7):1265–1275, 1997.

[143] C Lin and M Gerla. Adaptive clustering for mobile wireless networks. *IEEE Journal of Selected Areas in Communications*, 15(7):1265–1275, 1997.

[144] T Linder and R Zamir. High-resolution source coding for non-difference distortion measures: the rate distortion function. *IEEE Transactions on Information Theory, under revision*, 1998.

[145] S Lindsey, C Raghavendra, and K M Sivalingam. Data gathering algorithms in sensor networks using energy metrics. *IEEE Transactions on Parallel and Distributed Systems, special issue on Mobile Computing*, April 2002.

[146] H Luo, A Pandya, and G Pottie. Detection fidelity in distributed wireless sensor networks. Technical Report 20, Center For Embedded Networked Sensing, UCLA, 2003. http://www.cens.ucla.edu.

[147] A Mainwaring, J Polastre, R Szewczyk, D Culler, and J Anderson. Wireless sensor networks for habitat monitoring. In *First ACM Workshop on Wireless Sensor Networks and Applications*, Atlanta, GA,, September 2002.

[148] J Manyika and H Durran-Whyte. On sensor management in decentralized data fusion. In *Proceedings of Conference on Decision and Control*, pages 3506–3507, 1992.

[149] J Manyika and H Durrant-Whyte. *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*. Prentice Hall PTR, 1995.

[150] D Marco, E Duarte-Melo, M Liu, and D Neuhoff. On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data. In *Proceedings of the International Workshop on Information Processing in Sensor Networks (IPSN)*, April 2003.

[151] S Meguerdichian, F Koushanfar, M Potkonjak, and M Srivastava. Coverage problems in wireless ad-hoc sensor networks. In *Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 3, pages 1380–1387, April 2001.

[152] L Miller. Distribution of link distances in a wireless network. *Journal of Research of the National Institute of Standards and Technology*, 106, March – April 2001.

[153] L Miller. Probability of a two-hop connection in a random mobile network. In *Conference on Information Sciences and Systems*, John Hopkins University, March 2001.

[154] E Mingozzi. Scheduling algorithms for guaranteeing QoS in wireless networks. In *Proceedings of MTM Workshop*, February 2000.

[155] U Mitra and A Sabharwal. Complexity constrained sensor networks: achievable rates for two relay networks and generalizations. In *Proceedings of the third international symposium on Information processing in sensor networks*, pages 301–310. ACM Press, 2004.

[156] J Modestino and D Daut. Combined source-channel coding of images. *IEEE Transactions on Communications*, 27:1644–1659, 1979.

[157] J Modestino, D Daut, and A Vickers. Combined source-channel coding of images using the block cosine transform. *IEEE Transactions on Communications*, 29:1261–1274, 1981.

[158] G. Montorsi, F Pollara, S Benedetto, and D Divsalar. Serial concatenation of interleaved codes: Performance analysis, design und iterative decoding. *IEEE Transcations on Information Theory*, 44:909–920, 1998.

[159] Mica wireless measurement system. Datasheet, Crossbow Technology Inc. http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA.pdf.

[160] J Muramatsu, T Uyematsu, and T Wadayama. LDPC matrices for correlated source/channel coding. In *Proceedings of* 25^{th} *Symposium on Information Theory and Its Applications*, Ikaho, 2002.

[161] J Muramatsu, T Uyematsu, and T Wadayama. Low density parity check matrices for coding of correlated sources. In *IEEE International Symposium on Information Theory*, Yokohama, 2003.

[162] R Negi and A Rajeswaran. Capacity of power constrained ad-hoc networks. In *Proceedings of IEEE INFOCOM*, March 2004.

[163] Y Nesterov and A Nemirovsky. *Interior Point Polynomial Methods in Convex Programming*. SIAM, 1994.

[164] R Nowak and U Mitra. Boundary estimation in sensor networks: Theory and methods. In *2nd International Workshop on Information Processing in Sensor Networks*, volume 20, Palo Alto, CA, 2003.

[165] T Oates, M Prasad, and V Lesser. Cooperative information-gathering: a distributed problemsolving approach. In *IEE Proceedings of Software Engineering*, volume 144, pages 72–88, February 1997.

[166] J Omura and K Housewright. Source coding studies for information networks. In *Proceedings IEEE International Conference on Communications*, pages 237–240, 1977.

[167] Y Oohama. Gaussian multiterminal source coding. *IEEE Transactions on Information Theory*, 43(6):1912–1923, November 1997.

[168] Y Oohama. The rate-distortion function for the quadratic gauasian CEO problem. *IEEE Transactions on Information Theory*, 44(3):1057–1070, May 1998.

[169] Y Oohama. Multiterminal source coding for correlated memoryless gaussian sources with several side information at the decoder. In *Proceedings of IEEE Information Theory and Communications Workshop*, page 100, June 1999.

[170] A Pandya, A Kansal, G Pottie, and M Srivastava. Bounds on the rate-distortion of cooperative multiterminal gaussian sources. Technical report, Center For Embedded Networked Sensing, UCLA, 2003. http://www.cens.ucla.edu.

[171] A Pandya, A Kansal, G Pottie, and M Srivastava. Fidelity and resource sensitive data gathering. In *Proceedings of the* 42^{nd} *Annual Allerton Conference on Communication, Control, and Computing*, October 2004.

[172] A Pandya, A Kansal, G Pottie, and M Srivastava. Lossy source coding of multiple gaussian sources: m-helper problem. In *Proceedings of IEEE Information Theory Workshop*, October 2004.

[173] A Pandya, H Luo, and G Pottie. Spatial fidelity and estimation in sensor networks. In *Proceedings of The* 38^{th} *Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2004.

[174] A Pandya and G Pottie. On scalability and source/channel coding decoupling in large scale sensor networks. Technical Report 17, Center For Embedded Networked Sensing, UCLA, 2003. http://www.cens.ucla.edu.

[175] A Pandya and G Pottie. QoS in ad hoc networks. In *Proceedings of IEEE Vehicular Technology Conference*, October 2003.

[176] A Pandya and G Pottie. Bounds on achievable rates for cooperative channel coding. In *Proceedings of The* 38^{th} *Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2004.

[177] A Pentland, R Fletcher, and A Hasson. Daknet: rethinking connectivity in developing nations. *IEEE Computer*, 37(1):78–83, January 2004.

[178] C Perkins. Ad hoc networking: an introduction. pages 1–28, 2001.

[179] C Perkins and P Bhagwat. Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. *ACM SIGCOMM Computer Communications Review*, 24(4):234–244, 1994.

[180] C Perkins and E Royer. Ad hoc on-demand distance vector routing. In *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, pages 90–100, New Orleans, LA, February 1999.

[181] G Pottie and W Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43(5):51–58, 2000.

[182] G Pottie, H Luo, and A Pandya. *Encyclopedia of Sensors*, chapter Sensor Network Information Theory. American Scientific Publishers, 2004. To appear in 2005.

[183] V Prabhakaran, D Tse, and K Ramchandran. Rate region of the quadratic gaussian CEO problem. In *IEEE International Symposium on Information Theory*, June '2004.

[184] S Pradhan and K Ramchandran. Distributed source coding using syndromes (DISCUS): Design and construction. In *Proceedings of the Conference on Data Compression*, page 158. IEEE Computer Society, 1999.

[185] S Pradhan and K Ramchandran. Distributed source coding using syndromes (DISCUS): Design and construction. *IEEE Transactions on Information Theory*, IT-49(3):626–643, March 2003.

[186] J Proakis. *Digital Communications*. McGraw-Hill, New York, 4^{th} edition, 2001.

[187] H Psaraftis. *Vehicle Routing: methods and Studies*, chapter Dynamic vehicle routing problems, pages 223–248. 1988.

[188] W Pugh. Skip lists: A probabilistic alternative to balanced trees. In *Workshop on Algorithms and Data Structures*, pages 437–449, 1989.

[189] L Qian, D Jones, K Ramchandran, and S Appadwedula. A general joint source-channel matching method for wireless video transmission. In *Proceedings of Data Compression Conference (DCC)*, pages 414–423, 1999.

[190] J Rabaey, J Ammer, J da Silva Jr., and D Patel. Picoradio: Ad-hoc wireless networking of ubiquitous low-energy sensor/monitor nodes. pages 9–12, Orlando, Fl, USA, April 2000.

[191] M Rahimi, H Shah, G Sukhatme, J Heidemann, and D Estrin. Studying the feasibility of energy harvesting in a mobile sensor network. In *IEEE International Conference on Robotics and Automation*, 2003.

[192] B Rezaei, A Pandya, V Roychowdhury, and G Pottie. A graph theoretical approach to decentralize data fusion. Technical report, Center for Embedded Networked Sensing, UCLA, 2004. Preprint. http://www.cens.ucla.edu.

[193] M Ruf and J Modestino. Rate-distortion performance for joint source and channel coding of images. In *Proceedings of IEEE International Conference on Image Processing*, pages 77–80, 1995.

[194] M Ruf and J Modestino. Operational rate-distortion performance for joint source and channel coding of images. *IEEE Transactions on Image Processing*, 8(3):305–320, 1999.

[195] D Sakrison. A geometric treatment of the source encoding of a gaussian random variable. *IEEE Transactions on Information Theory*, 14:481–486, 1968.

[196] D Sakrison. The rate distortion function of a gaussian process with a weighted sqare error criterion. *IEEE Transactions on Information Theory*, 14:506–508, 1968. Addendum 610–611.

[197] D Sakrison. Source encoding in the presence of a random disturbance. *IEEE Transactions on Information Theory*, 14:165–167, 1968.

[198] D Sakrison. The rate of a class of random processes. *IEEE Transactions on Information Theory*, 16:10–16, 1970.

[199] D Sakrison and V Algazi. Comparison of line-by-line and two-dimensional encoding of random images. *IEEE Transactions on Information Theory*, 17(4):386–398, July 1971.

[200] N Sarshar, B Rezaei, A Pandya, V Roychowdhury, and G Pottie. Scalability of wireless capacity: A renormalization approach. Preprint.

[201] H Sato. The capacity of the Gaussian interference channel under strong interference. *IEEE Transactions on Information Theory*, IT-27:786–788, 1981.

[202] M Savelsbergh and M Sol. The general pickup and delivery problem. *Transportation Science*, 29:17–29.

[203] A Savvides and M Srivastava. A distributed computation platform for wireless embedded sensing. In *IEEE International Conference on Computer Design: VLSI in Computers and Processors*, pages 220–225, September 2002.

[204] A Savvides and M B Srivastava. A distributed computation platform for wireless embedded sensing. In *IEEE International Conference on Computer Design: VLSI in Computers and Processors*, pages 220 – 225, September 2002.

[205] A Sayed. *Fundamentals of Adaptive Filtering*. Wiley-IEEE Computer Society Press, 2003.

[206] A Scaglione and S Servetto. On the interdependence of routing and data compression in multi-hop sensor networks. In *Proceedings of the 8th annual international conference on Mobile computing and networking*, pages 140–147. ACM Press, 2002.

[207] D Schoenwald. AUVs: In space, air, water, and on the ground. *IEEE Control Systems Magazine*, 20(6):15–18, December 2000.

[208] J Scott and M Hazas. User-friendly surveying techniques for location-aware systems. In *Ubicomp*, October 2003.

[209] A Sendonaris, E Erkip, and B Aazhang. User Cooperation Diversity–Part I: System Description. *IEEE Transactions on Communications*, 51(11), November 2003.

[210] A Sendonaris, E Erkip, and B Aazhang. User Cooperation Diversity–Part II: Implementation Aspects and Performance Analysis. *IEEE Transactions on Communications*, 51(11), November 2003.

[211] S Servetto. On the feasibility of large-scale wireless sensor networks. In *Proceedings of the 40th Annual Allerton Conference on Communication, Control, and Computing*, Urbana, IL, October 2002.

[212] A Sgarro. Source coding with side information at several decoders. *IEEE Transcations on Information Theory*, 23:179–182, 1977.

[213] R Shah, S Roy, S Jain, and W Brunette. Datamules: Modelling a three tiered architecture for sparse sensor networks. In *First IEEE International Workshop on Sensor Network Protocols and Applications (SNPA)*, May 2003.

[214] C Shannon. A mathematical theory of communication. Technical Report 27, Bell System Technical Journal, 1948.

[215] C Shannon. Coding theorems for a discrete source with a fidelity criterion. In *Institute of Radio Engineers, International Convention Record*, volume 7, pages 142–163, 1959.

[216] C Shannon. Two-way communication channels. In *Proceedings 4th Berkeley Symp. Math. Stat. Prob.*, volume 1, pages 611–644. Univ. California Press, 1961.

[217] M Shyu, A Sangiovanni-Vincentelli, J Fishburn, and A Dunlop. Optimization based transistor sizing. *IEEE Journal of Solid-State Circuits*, 23(2):400–409, April 1988.

[218] S Singh and C Raghavendra. PAMAS - power aware multi-access protocol with signalling for ad hoc networks. *SIGCOMM Computer Communications Review*, 28(3):5–26, 1998.

[219] S Singh, M Woo, and C Raghavendra. Power-aware routing in mobile ad hoc networks. In *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*, pages 181–190, Dallas, TX, 1998.

[220] A Sinha and A Chandrakasan. Dynamic power management in wireless sensor networks. *IEEE Design and Test*, 18(2):62–74, 2001.

[221] D Slepian and J K Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, July 1973.

[222] K Sohrabi. *On Low Power Self Organizing Sensor Networks*. PhD thesis, Electrical Engineering Department, University of California, Los Angeles (UCLA), 2000.

[223] Sony aibo ERS–7. http://www.sony.net/Products/aibo/.

[224] M Srivastava, R Muntz, and M Potkonjak. Smart kindergarten: sensor-based wireless networks for smart developmental problem-solving enviroments. In *Proc. ACM MobiCom*, pages 132 – 138, July 2001.

[225] Stargate: Xscale processor platform. http://www.xbow.com/ Products/ Product_pdf_files/Wireless_pdf/6020-0049-01_C_Stargate.pdf. Data Sheet.

[226] P Steurer and M Srivastava. System design of smart table. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom 2003)*, pages 473 –480, March 2003.

[227] W Su and I Akyildiz. A stream enabled routing (SER) protocol for sensor networks. In *Proceedings of Med-hoc-Net*, Sardegna, Italy, September 2002.

[228] Y Szu-Tu. *Cooperative Communications among Wireless Sensor Networks*. PhD thesis, University of California, Los Angeles(UCLA), 2003.

[229] Y Szu-Tu and G Pottie. Achievable rate regions in the three-node wireless network. Preprint, http://www.ee.ucla.edu/˜pottie.

[230] H Takahashi and A Matsuyama. An approximate solution for the steiner problem in graphs. *MathematicaJaponica*, 24:573–577, 1980.

[231] A Tanenbaum. *Computer Networks (3rd edition)*. Prentice-Hall, Inc., 1996.

[232] E Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6), November 1999.

[233] H Tie, A Goldsmith, and M Effros. Joint design of fixed-rate source codes and uep channel codes for fading channels. In *Proceedings of the Thirty-Second Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA.

[234] TinyOS: a component based OS for the networked sensor regime. http://webs.cs.berkeley.edu/tos/.

[235] C Toh. *Ad Hoc Wireless Networks: Protocols and Systems*. Prentice Hall PTR, 2001.

[236] C Toh, H Cobb, and D Scott. Performance evaluation of battery-life-aware routing schemes for wireless ad hoc networks. In *Proceedings of IEEE International conference on communications (ICC)*, Helsinki, Finland, June 2001.

[237] S Toumpis and A Goldsmith. Capacity bounds for large wireless networks under fading and node mobiltiy. In *Proceedings of the Allerton Conference on Communications, Control, and Computing*, October 2003.

[238] S Toumpis and A Goldsmith. Capacity regions for wireless ad hoc networks. *IEEE Transactions on Wireless Communications*, 2(4):736–748, July 2003.

[239] H Van Tress. *Detection, Estimation and Modulation Theory - Part I*. John Wiley and Sons, New York, 1968.

[240] C Tunasar and J Rajgopal. User's manual for GPGLP - a posynomial geometric programming solver department of industrial engineering. Technical Report TR95-11, University of Pittsburgh, 1995.

[241] S Tung. *Multiterminal source coding*. PhD thesis, Cornell University, Ithaca, NY, May 1978.

[242] P Varshney. *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

[243] H Viswanathan and T Berger. The quadratic gaussian CEO problem. *IEEE Transactions on Information Theory*, 43(5):1549–1559, September 1997.

[244] T Wadayama. A lossy compression algorithm for discrete memoryless sources based on ldpc codes. In *Asia-Europe Workshop on Information Theory*, Kamogata, June 2003.

[245] W Walter. *The Living Brain*. Duckworth, London, England, 1953.

[246] X Wang, G Xing, Y Zhang, C Lu, R Pless, and C Gill. Integrated coverage and connectivity configuration in wireless sensor networks. In *ACM SenSys 2003*, Los Angeles, USA, November 2003.

[247] B Warneke, M Last, B Liebowitz, and K Pister. Smart dust: Communicating with a cubic-millimeter computer. *Computer*, 34(1):44–51, 2001.

[248] F Willems and E Van der Meulen. The discrete memoryless multiple-access channel with cribbing encoders. *IEEE Transactions on Information Theory*, IT–31:313–327, May 1985.

[249] R Willett, A Martin, and R Nowak. Backcasting: adaptive sampling for sensor networks. In *Proceedings of the third international symposium on Information processing in sensor networks*, pages 124–133, Berkeley, CA, 2004. ACM Press.

[250] P Winter. Steiner problem in networks: a survey. *Networks*, 17(2):129–167, 1987.

[251] A Woo and D Culler. A transmission control scheme for media access in sensor networks. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 221–235, Rome, Italy, 2001.

[252] A Wyner. The rate-distortion function for source coding with side information at the decoder-II: General sources. *Inform. Contr.*, 38:60–80, July 1978.

[253] A Wyner and J Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-22:1–10, January 1976.

[254] L Xie and P Kumar. A network information theory for wireless communication: Scaling laws and optimal operation. *IEEE Transactions on Information Theory*, 50(5):748–767, May 2004.

[255] E Yang and J Kieffer. Simple universal lossy data compression schemes derived from the lempel-ziv algorithm. *IEEE Transactions on Information Theory*, 42:239–245, 1996.

[256] K Yazdi, H Gamal, and P Schniter. On the design of cooperative transmission schemes. In *Allerton Communications, Computing, and Control*, October 2003.

[257] B Yener, M Inanc, and M Magdon-Ismail. Power optimal connectivity and coverage in wireless sensor networks. Technical Report TR-03-06, Department of Computer Science, Rensselaer Polytechnic Institute, July 2003.

[258] S Yi, Y Pei, and S Kalyanaraman. On the capacity improvement of ad hoc wireless networks using directional antennas. In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, pages 108–116. ACM Press, 2003.

[259] R Zamir and T Berger. Multiterminal source coding with high resolution. *IEEE Transactions on Information Theory*, 45(1):106–117, 1999.

[260] C Zeng, F Kuhlmann, and A Buzo. Achievability proof of some multiuser channel coding theorems using backward decoding. *IEEE Transactions on Information Theory*, 35:1160–1165, November 1989.

[261] H Zhang and J Hou. Maintaining sensing coverage and connectivity in large sensor networks. Technical Report UIUCDCS-R-2003-2351, University of Illinois, Urbana Champagne, June 2003.

[262] Z Zhang and T Berger. Estimation via compressed information. *IEEE Transactions on Information Theory*, 34:198–211, March 1988.

[263] W Zhao and M Ammar. Message ferrying: Proactive routing in highly-partitioned wireless ad hoc networks. In *The Ninth IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS)*, 2003.

[264] W Zhao, M Ammar, and E Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *The fifth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2004.

[265] J Ziv and A Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23:337–343, 1977.