

© 2012 Adnan Raja

TOPICS IN MULTI-TERMINAL WIRELESS NETWORKS

BY

ADNAN RAJA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Pramod Viswanath, Chair
Associate Professor Chandra Chekuri
Professor Bruce Hajek
Adjunct Professor P. R. Kumar
Professor Venugopal Veeravalli

ABSTRACT

In this dissertation, tools from information theory are used to study multiterminal wireless networks. A compress-and-forward scheme with layered decoding is presented for the unicast and multi-source wireless network and shown to be approximately optimal. This scheme is shown to allow better decoding complexity compared to previously known approximately optimal schemes. Characterizing the layered decoding scheme is shown to be equivalent to characterizing an information flow for the wireless network. A node-flow for a graph with bisubmodular capacity constraints is presented and a max-flow min-cut theorem is presented. This generalizes many well-known results of flows over capacity constrained graphs studied in computer science literature. In the final part of the dissertation, the intuitions from the reciprocal nature of networks are used to present an approximately optimal communication scheme for broadcast networks, which are the reciprocal of the multi-source wireless networks.

To the questions I failed to ask and the answers I failed to find

ACKNOWLEDGMENTS

First and foremost, I would like to thank that butterfly, the flapping of whose wings led me to find Pramod Viswanath as my adviser. Pramod has been a terrific mentor to me throughout my graduate life and has made this whole journey an extremely pleasant and memorable one. He has been extremely patient and has been kind enough to give me enough space through my graduate life to learn, explore and discover. I have learned much, going beyond the realms of information theory and wireless communications, through our long, and often amusing, discussions. Our love for the open air has seen us share many experiences outside research - the most memorable being the 100 mile bike ride from Champaign to Decatur and back, the hike to Kumara Parvata and the bike ride to the highest peak in Karnataka. Thank you Pramod for everything.

I would also like to express my gratitude to Chandra Chekuri, Venu Veeravalli, Bruce Hajek, and P. R. Kumar for agreeing to be on my dissertation committee and giving me useful advice. In particular, I would like to thank Chandra for his many useful discussions and insights on polymatroidal networks and flows.

I am extremely grateful to my colleague and collaborator, the Zen master Sreeram Kannan. We shared a very noiseless communication channel which helped us in our research in answering many interesting questions about noisy communication networks. He has brought much inspiration and clear thought in my life when I have sought them. I would also like to thank Vinod Prabhakaran, with whom I had the pleasure to work during my initial years as a graduate student. His calm and relaxed style of working and deep thinking offered a lot to learn from.

In my first couple of years at CSL, I was fortunate to have a few senior people around - in particular, Aleksander Jovicic, Saurabha Tavildar, Siddhartha Mallik and Srinivas Shakkottai - who were wonderful mentors and

were always willing to offer free and useful advice. I am also grateful to my friends and colleagues at CSL - Sreekanth Reddy, Jaykrishnan Unnikrishnan, Neelesh Khude, Quan Geng, Juan Jose Jaramillo and many others. The daily interactions I had with them were to my grad-life what salt is to my food. They have helped me mature as an individual and a researcher. Many thanks are also due to my friends in the small and fun town of Urbana-Champaign who added the pepper to my life - Pritam, Rahul, Murthy, Shilpa, Nachiket, Neha, Ravali, Chandu, Sucheta, Zeba, Nihal, Neha, Saumil, Ankur, Sarath, Tejas and many many more.

Thanks are also due to the funding agencies that supported my research - the National Science Foundation and Office of Naval Research - and to Vodafone for their generous one-year fellowship. I am also grateful to Qualcomm, where I spent almost a year in all, spanning across three separate internships at three different locations. I am grateful to the mentors I found there - Saurabha Tavildar, Amir Farajidana and Subramanya Parvathanathan - and for the lessons I learned from them. Thanks are also due to the Indian Institute of Science and Prof. Vinod Sharma for graciously hosting me there for a very memorable stay on the beautiful I.I.Sc. campus for almost six months.

I would also like to thank the excellent staff at the ECE Publications Office - Jamie, Janet and Matt - for helping me put the dots on the i's and the crosses on the t's, the result of which is this beautiful looking dissertation.

I would like to thank my family for all their support throughout: my Mom and Dad, for I owe everything that I am to them; my brothers, who I have always looked up to and on whose shoulders I have tried to climb; my sister-in-law, Tasneem and my adorable niece, Batul. Finally, I would like to thank Sarrah, for standing by me through the ups and downs that are gone and promising to stand by through those that are to come.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION	1
CHAPTER 2	MODELS	4
2.1	Network and channel model	4
2.2	Communication model	8
2.3	Miscellaneous notions	10
CHAPTER 3	COMPRESS-AND-FORWARD SCHEME WITH LAYERED DECODING	14
3.1	Compress-and-forward scheme for unicast network	15
3.2	Generalizations to multi-source networks	21
3.3	Special cases	23
CHAPTER 4	FLows WITH BISUBMODULAR CAPACITY CONSTRAINTS	25
4.1	Introduction	25
4.2	A max-flow min-cut theorem	26
4.3	A compress-and-forward scheme from flows	28
CHAPTER 5	RECIPROCITY	30
5.1	Introduction	30
5.2	Examples of reciprocity in wireless networks	31
5.3	Reciprocity in linear deterministic network under linear coding	33
CHAPTER 6	BROADCASTING IN WIRELESS RELAY NETWORKS	38
6.1	Introduction	38
6.2	Deterministic broadcast networks	41
6.3	Gaussian broadcast network	46
APPENDIX A	APPENDIX FOR CHAPTER 3	50
A.1	Proof of Theorem 1	50

APPENDIX B	APPENDIX FOR CHAPTER 4	52
B.1	Proof of Theorem 5	52
B.2	Proof of Lemma 9	54
B.3	Proof of Proposition 1	55
APPENDIX C	APPENDIX FOR CHAPTER 5	57
C.1	Layering for linear deterministic network	57
APPENDIX D	APPENDIX FOR CHAPTER 6	59
D.1	Proof of Lemma 4	59
D.2	Proof of Lemma 5	59
REFERENCES		61

CHAPTER 1

INTRODUCTION

Wireless networks are pervasive today. A casual look around and one can easily spot a number of devices that communicate with each other wirelessly. Further, this trend is only predicted to increase. It is, for this reason, important to understand wireless networks from a theoretical point of view, to understand the fundamental limits of communication in wireless networks and to invent efficient and optimal communication architecture for wireless networks. Over the last few decades, many insights from the field of information theory have spurred innovations and developments in the wireless industry.

Information theory tries to establish fundamental limits on the rates of communication with reliability guarantees. The insight also leads to designing communication architecture and schemes, which come close to meeting the fundamental limits.

The fundamental limits arise from both the physics and the engineering limitations of the real world communication systems. For instance, the physical limitations could be the thermal noise that arises from the Brownian motion of electrons in the electronic receiver systems, or the superposition and broadcast nature of signals that arise from a shared medium in wireless systems. The engineering limits could be the power limitations on the transmission, due to the limited power handling capabilities of the amplifiers or the battery life of devices, or it could be the complexity of the schemes allowed, due to the limitations on the circuits on which the schemes are implemented. These limits are captured by a mathematical model, which is the first and most important step in any theoretic analysis of a system. The model often needs to be simplified to make the analysis tractable and to give fundamental insights. The insights gained from the analysis are as useful to the real world problem as the mathematical model is to the real world setup.

Over the last 60 years or so, the simple point-to-point Gaussian channel

model first studied by Shannon has been extended to multi-terminal wireless networks. The multiple-access channel, where many source nodes transmit to a single destination node, and the broadcast channel, where a single source node transmits to many destination nodes, have been fully understood. However, going beyond these simple cases has been much more difficult. In fact, characterizing the fundamental limits on rates of reliable communication of even seemingly small and simple wireless networks, like the three-node relay channel or the two-user interference channel, remain open problems.

Many recent works have focused on characterizing communication architecture for larger networks, which can be shown to be *approximately* optimal. Two approaches among these have been noteworthy. One approach, beginning with the pioneering work of [1, 2], has been to characterize schemes which are asymptotically optimal in the size of the network, known as the scaling laws.

The second approach has been to characterize an optimal scheme when the signal power is much larger than the noise (called the high SNR regime). The approximate optimality is in the sense that the schemes are shown to achieve rates a constant gap away from the fundamental limits. Such an approach was used for the two-user interference channel [3] and for the unicast wireless relay network in [4]. A useful technique that has been employed in understanding wireless networks in the high SNR regime is the use of deterministic models, where the channel model is assumed to be noiseless, but captures the signal interaction arising due to the shared nature of the wireless medium.

In this thesis, the second approach is used to further our understanding of the wireless networks. Chapter 2 lays the basic groundwork by setting up the communication problem in multi-terminal wireless networks. It also discusses the various models and establishes some of the tools that will be used in the rest of the thesis.

In Chapter 3, the unicast relay network, where a single source node is communicating with a single destination node in the presence of multiple relay nodes, is studied. A compress-and-forward scheme is proposed and the scheme is shown to be approximately optimal. The scheme allows possibilities of a simplified decoding architecture compared to previously known schemes. The simplification comes from the characterization of information flows in wireless networks.

Flows in graphs have been used to study routing of commodities in transportation and communication networks. Information flow in a wired network like the Internet consists of information packets being routed across the nodes in the network from the source to the respective destination. The flow is through the wired links which form the edges in the network. A wireless network, on the contrary, has no edges. The notion of node-flows for a graph is proposed in Chapter 4. While the motivation here to develop the concept of node-flows comes from the study of the compress-and-forward architecture for the wireless network, it is interesting in its own right, generalizing many well known max-flow min-cut results in the computer science literature.

In Chapter 5, the notion of reciprocity in wireless networks is presented. It is suggested, by pointing to known examples in the literature, that a communication network and its reciprocal obtained by reversing the communication links and reversing the information flow are closely connected. This intuition is then used in Chapter 6 in designing communication schemes for a broadcast network, where a single source is communicating independent messages to many destination nodes through a network of relay nodes. The intuition for the approximately optimal communication schemes for such a network comes from relating it to its reciprocal network for which schemes are known and comparatively more intuitive.

CHAPTER 2

MODELS

The first task in analytically understanding the fundamental limits of a wireless communication network is to develop and describe a mathematical model. The goal of this chapter is to describe a mathematical model for the communication network and to formulate an objective function that needs to be optimized in the design of the communication architecture we seek. The main criterion that will be considered is the *rate of communication* which is a measure of the amount of information that can be exchanged between the communicating nodes and is usually measured in bits per second per hertz.

2.1 Network and channel model

The term *communication node* is used to describe any device that has a radio (or transceiver) embedded in it, which can be used to exchange signals with other compatible nodes. The network model describes the geometry of the nodes and the channel model describes the signal interaction between these nodes.

2.1.1 General discrete-time memoryless network model

Consider a communication network denoted by \mathcal{N} . The communication network consists of the following components.

1. A collection of *communication nodes* (or radios) denoted by \mathcal{V} . The nodes can be thought of as points in a 3-dimensional space. For simplicity, and as it often suffices, they can even be considered to be points on a plane. The communication channel, described next, describes relationship between pairs of nodes and can be used to define edges between the nodes. Thus, the set \mathcal{V} forms the vertex set of a graph.

2. The *communication channel* between these nodes is assumed to be a discrete-time memoryless communication channel. More precisely, time is assumed to be discrete and synchronized among all nodes. The transmit symbol at any time at a node $v \in \mathcal{V}$ is given by $x_v \in \mathcal{X}_v$ and the receive symbol is given by $y_v \in \mathcal{Y}_v$. Unless otherwise mentioned, the input and output alphabet sets, \mathcal{X}_v 's and \mathcal{Y}_v 's respectively, are assumed to be finite sets. A *memoryless* network implies that the received symbol at any node at any given time depends (stochastically when the channel is noisy) only on the current transmitted symbols at other nodes. This dependency is represented by directed edges between the nodes. An edge (u, v) indicates that the transmit symbol of node u influences the received symbol of node v . Often, by a principle of reciprocity, if a node u influences a node v , then the transmitted symbol of node v also influences the received symbol of node u and more so in a commensurate manner. Therefore, the dependency can be represented by an undirected edge uv . The set of all edges is denoted by the \mathcal{E} . The precise relationship between the transmitted and received symbols can, in general, be modeled by a conditional probability function $p(Y_{\mathcal{V}}|X_{\mathcal{V}})$. Here, X_A for any set A is used to denote the collection of random variables $\{X_v|v \in A\}$. By default, a full duplex mode of operation is assumed at each node, so that a node can transmit and receive simultaneously.

2.1.2 Gaussian network model

The term *wireless network* commonly refers to the Gaussian network model. The communication network is called a Gaussian network when the canonical Gaussian channel model describes the relationship between the transmitted and received symbols of the various nodes in the network. Denoting the baseband transmit symbol (a complex number) of node k at time t by $x_k[t]$, the average transmit power constraint at each node implies that

$$\sum_{t=1}^T |x_k[t]|^2 \leq TP_k, \quad (2.1)$$

where T is the time period over which the communication occurs. At each time t , the received signal at any node ℓ

$$y_\ell[t] = \sum_{k \neq \ell} h_{k\ell}[t] x_k[t] + z_\ell[t]. \quad (2.2)$$

Here $\{z_\ell[t]\}_t$ is i.i.d. Gaussian noise and independent across the different nodes ℓ .

For the most part here, it will be assumed that the channel is non-fading. This means that the channel coefficients $h_{k\ell}$ do not depend on time t . By scaling the channel coefficients appropriately, it can be assumed that the $z_\ell[t]$ are unit variance and the power constraint at each node is unity. The Gaussian network can equivalently be represented by the channel transition probability $p(Y_\mathcal{V}|X_\mathcal{V})$ given by

$$Y_\ell = \sum_{k \neq \ell} h_{k\ell} X_k + Z_\ell, \quad (2.3)$$

where $Z_\ell \sim \mathcal{CN}(0, 1)$.

Here a single antenna is assumed at each node. With multiple antennas, the symbols are assumed to be complex vectors and the channel coefficient is replaced by the matrix $H_{k\ell}$.

2.1.3 Deterministic network

To simplify analysis and to distill insights, a simplified *noise-free* or *deterministic* model is often used. The communication network is called a deterministic network when the received symbols are a deterministic function of the transmitted symbols from the other nodes. More precisely,

$$y_\ell[t] = g_\ell \left(\{x_k[t]\}_{k \neq \ell} \right). \quad (2.4)$$

The input and output alphabet sets, \mathcal{X}_k 's and \mathcal{Y}_ℓ 's respectively, are assumed to be finite sets. Two special cases of the deterministic network are of particular importance as they capture the most important features of a wireless network - superposition and broadcast nature of signals.

1. *Linear deterministic network:*

The linear deterministic network assumes that transmit and receive alphabet sets are vector finite field \mathbb{F}_p^q , where p is the order of the finite field and q is the length of the vector. The channel model is linear and is given by

$$y_\ell[t] = \sum_{k \neq \ell} G_{\ell k} x_k[t]. \quad (2.5)$$

Here $y_\ell[t], x_k[t] \in \mathbb{F}_p^q$ and $G \in \mathbb{F}_p^{q \times q}$. The linear deterministic network was first introduced in [4] as a simple model to capture wireless signal interaction. In particular it considered the binary field ($p = 2$). The channel was modeled by shift matrices S^{q-k} , where

$$S = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}_{q \times q} \quad (2.6)$$

and $k \in \{0, 1, \dots, q\}$ is a measure of the relative channel strength. This philosophy of the deterministic model to capture channel strengths was first introduced in the context of a compound point-to-point channel where it was successfully used to construct codes that universally achieve the diversity-multiplexing tradeoff over any fading channel [5] (see also Chapter 9 of [6]).

In [4], the capacity of the linear deterministic network with unicast and multicast traffic was determined. The insights were then used to give a coding theorem for the Gaussian network. The coding theorem was used to establish the approximate capacity of the Gaussian network, by showing that rates within a constant gap of an outer bound given by the cutset bound can be achieved.

The linear deterministic, however, does not approximate the capacity of the Gaussian network. In particular, it fails to capture the complex phase in the channel model of the Gaussian network and the power constraints. To overcome this limitation, the discrete superposition network was introduced in [7]. The model is closely related to a similar

model introduced for the two-user interference channel in [8] and the truncated deterministic model introduced in [4] .

2. Discrete superposition network:

In the Gaussian network, the channel model is given by

$$y_\ell[t] = \sum_{k \neq \ell} h_{k\ell} x_k[t] + z_\ell[t]. \quad (2.7)$$

By scaling the channel coefficient $h_{k\ell}$'s appropriately, it can be assumed that each node has unit power constraint and the noise has unit variance. For every Gaussian network model, the corresponding DSN channel model is given by

$$y_\ell^{(\text{DSN})}[t] = \left[\sum_{k \neq \ell} [h_{k\ell}] x_k^{(\text{DSN})}[t] \right], \quad (2.8)$$

where $[\cdot]$ lies in $\mathbb{Z} + i\mathbb{Z}$ and corresponds to quantizing the real and imaginary parts of the complex number by neglecting the fractional part. Further, the transmit alphabet in the DSN is restricted to a finite set, such that both the real and imaginary parts belong to the finite set with equally spaced points given by

$$\mathcal{X}_v^{(\text{DSN})} = \frac{1}{\sqrt{2}} \{0, 2^{-n}, \dots, 1 - 2^{-n}\}, \quad (2.9)$$

where $n \triangleq \max_{(i,j)} \lfloor \max \{\log_2 \text{Re}(h_{ij}), \log_2 \text{Im}(h_{ij})\} \rfloor$.

2.2 Communication model

The *information traffic* in the network refers to the message streams that need to be communicated in the network. The messages are assumed to be independent. Each message is associated with one source node and at the least one, and possibly many, destination nodes. Some of the common traffic patterns we consider are:

- *Unicast* - single message stream with one source and one destination node.

- *Broadcast* - many independent message streams from a single source node S but to multiple destination nodes - D_1, \dots, D_J .
- *Multicast* - single message stream from a single source node S but to multiple destination nodes - D_1, \dots, D_J .
- *Multiple unicast* - many independent message streams with different source nodes and different destination nodes.

Note that in general a communication network could have any or a combination of the above traffic patterns.

The communication problem can be described mathematically as follows. The communication is done over a block of T time symbols over which J independent messages (information streams) are communicated in the network.

1. Each independent *message* in the network is associated with an *independent random variables* W_i which is distributed uniformly on $[2^{TR_i}]$ for $i \in [J]$ respectively. The message W_i is assumed to be known at precisely one node in the network called the source node for that message. At the end of the communication period, the message needs to be determined at some nodes (could be one for unicast or many for multicast) in the network. These nodes are called the destination nodes for the message.
2. The *encoding* at any node $v \in \mathcal{V}$ and time t is given by

$$f_{v,t} : (\mathcal{W}_v, \mathcal{Y}_v^{t-1}) \rightarrow \mathcal{X}_v, \quad (2.10)$$

where \mathcal{W}_v represents all the messages for which node v is the source node. If a node is purely a source node, then the encoding corresponds to a source mapping and is given by

$$f_v : \mathcal{W}_v \rightarrow \mathcal{X}_v^T. \quad (2.11)$$

If a node is purely a relay node, then the encoding corresponds to a relay mapping and is given by

$$f_{v,t} : \mathcal{Y}_v^{t-1} \rightarrow \mathcal{X}_v. \quad (2.12)$$

3. The *decoding* map for a message W_i at destination node D ,

$$g_{D,W_i} : \mathcal{Y}_{D_i}^T \rightarrow \hat{W}_i. \quad (2.13)$$

The probability of error for the above decoding is given by

$$P_e(D, W_i) \triangleq \Pr\{g_{D,W_i} \neq W_i\}. \quad (2.14)$$

A rate tuple (R_1, R_2, \dots, R_J) , where R_i is the rate of communication in bits per unit time for message W_i , is said to be achievable if for any $\epsilon > 0$, there exists an encoding and decoding scheme that achieves a probability of error less than ϵ for all messages and for all corresponding destination nodes, i.e., $\max_i P_e(D, W_i) \leq \epsilon$. The capacity region \mathcal{C} is the set of all achievable rates.

2.3 Miscellaneous notions

2.3.1 Cutset bound

The following is the well known cutset outer bound to the rate tuples of reliable communication [9, 10].

A cut (or more precisely vertex-cut) in the network is represented by a set of nodes $\Omega \subset \mathcal{V}$. The cut partitions \mathcal{V} into two sets Ω and Ω^c . Let Λ denote the set of all subsets $\Omega \subset \mathcal{V}$. We say that the cut separates the message stream W_i if the source node corresponding to the message is in Ω and at the least one destination node is in Ω^c . Let $\delta(\Omega)$ denote the set of all indices of messages that are separated by the cut Ω .

The cutset bound states that if (R_1, \dots, R_J) is achievable then there is a joint distribution $p(\{X_v | v \in \mathcal{V}\})$ (denoted by Q) such that

$$R_{\delta(\Omega)} \leq I(Y_{\Omega^c}; X_{\Omega} | X_{\Omega^c}), \quad (2.15)$$

where $R_{\mathcal{A}} \triangleq \sum_{j \in \mathcal{A}} R_j$.

Let $\bar{\mathcal{C}}(Q)$ denote the set of all rate tuples that satisfy the cutset outer

bound for a given joint distribution Q

$$\bar{\mathcal{C}}(Q) \triangleq \{(R_1, \dots, R_J) : R_{\delta(\Omega)} \leq I(Y_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) \quad \forall \Omega \in \Lambda\}. \quad (2.16)$$

$$\bar{\mathcal{C}} \triangleq \bigcup_Q \bar{\mathcal{C}}(Q). \quad (2.17)$$

$\bar{\mathcal{C}}$ denotes the cutset bound.

Cutset bound for the Gaussian network: For the Gaussian network, due to the power constraint the cutset bound is evaluated by considering all $p(X_{\mathcal{V}})$ with $E[|X_v|^2] = 1$. Further, it can be shown that the cutset region is always maximized when $X_{\mathcal{V}}$ are jointly Gaussian. Therefore for the Gaussian network the cutset bound is given by restricting $X_{\mathcal{V}} \sim \mathcal{CN}(0, K_X)$,

$$\begin{aligned} \bar{\mathcal{C}}^g(K_X) \triangleq \{(R_1, \dots, R_J) : R_{\delta(\Omega)} &\leq \log(|I + H_{\Omega\Omega^c} K_X H_{\Omega\Omega^c}^*|), \\ &\forall \Omega \in \Lambda\}, \end{aligned} \quad (2.18)$$

and

$$\bar{\mathcal{C}}^g \triangleq \bigcup_{K_X, K_X(ii)=1} \bar{\mathcal{C}}^g(K_X). \quad (2.19)$$

Here,

$$Y_{\Omega^c} = H_{\Omega\Omega^c} X_{\Omega} + Z_{\Omega^c} \quad (2.20)$$

is the MIMO channel formed by the cut Ω with the nodes on Ω forming the source and the nodes in Ω^c forming the destination node.

The following lemma due to [4] characterizes the gap from cutset when the distribution is further restricted to i.i.d. Gaussian. This will come in useful in characterizing approximate optimality of schemes.

Lemma 1. (Lemma 6.6, [4]) *If the rate vector $\vec{R} \in \bar{\mathcal{C}}^g$, then $(\vec{R} - 2|\mathcal{V}|\vec{1}) \in \bar{\mathcal{C}}^g(I)$.*

2.3.2 Layered network

A network is a layered network if the underlying graph of the network $(\mathcal{V}, \mathcal{E})$, which determines the connectivity of the graph, has a layered topology as described below. A network is called an *L-layered network* if the set of vertices \mathcal{V} can be partitioned into L disjoint sets, such that the source nodes are in the 1st layer and the J destination nodes are in the L -th layer. The nodes in

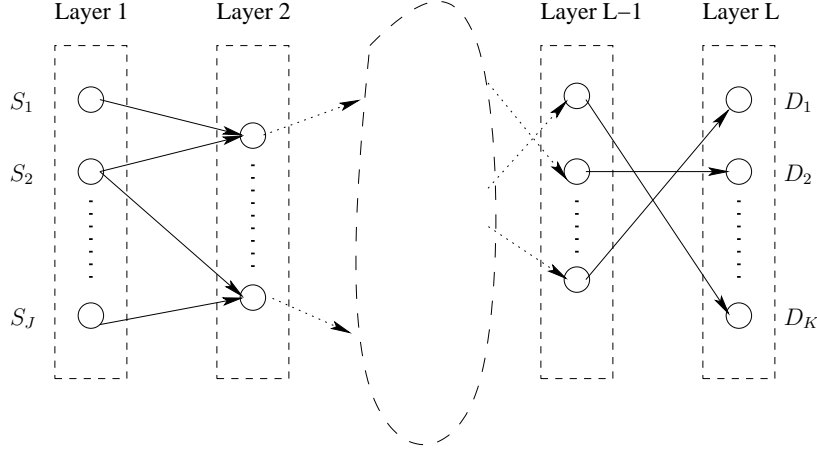


Figure 2.1: Layered network.

the intermediate layers are relaying nodes. The received signal at the nodes in the $l + 1$ -th layer only depend on the transmitted signals at the nodes in the l -th layer. This dependency is often represented by edges connecting the nodes from the l -th layer to the $(l + 1)$ -th layer. An example of a layered network is shown in Figure 2.1.

The advantage of working with a layered network is that we can consider layered schemes for such a network. The layered scheme is such that nodes operate over blocks of T symbols. The source node sends independent message in each block. The independent messages can be seen as propagating from one layer to the next without getting intertwined.

It is shown in [4] that any arbitrary network \mathcal{N} can be dealt with by considering a corresponding unfolded L -layered network \mathcal{N}^{unf} , which is constructed as follows.

- The node set for the unfolded network is as follows. The first layer has only the source nodes and the last layer has only the destination nodes. The remaining layers each have a replication of all the nodes in the original network.
- Next we describe the edge set and the channel model in the unfolded network. The edge set consists of $L - 1$ subsets (or stages), where the i -th stage gives the connection between the i -th layer of nodes and the $(i + 1)$ -th layer. A node in a layer is connected to its own replicate in the subsequent layer by an orthogonal link of infinite capacity. This

represents the memory in the network. For the first and the last stage, these are the only connections. For the other stages, any node in a layer is connected to another node in the following layer, if there was a connection between the two nodes in the original network. The channel model for every stage is identical to the channel model in the original network.

The following two lemmas prove the relationship between a network and its corresponding unfolded L -layered network

Lemma 2. *If a rate tuple $\vec{R} = (R_1, \dots, R_J)$ is achievable for the unfolded L layered network, then $\vec{R}/(L - 2)$ is achievable for the original network.*

The lemma follows from the observation that any scheme for the unfolded network over B symbols can be emulated on the original network in $B(L - 2)$ symbols.

Lemma 3. *If $\bar{\mathcal{C}}^{\mathcal{N}}$ and $\bar{\mathcal{C}}^{\mathcal{N}^{unf}}$ denote the cutset bound of the original network and the unfolded network, then*

$$\bar{\mathcal{C}}^{\mathcal{N}} = \lim_{L \rightarrow \infty} \frac{1}{L - 2} \bar{\mathcal{C}}^{\mathcal{N}^{unf}}. \quad (2.21)$$

This lemma was proved in [4] by a careful analysis of the cutsets of the unfolded network.

CHAPTER 3

COMPRESS-AND-FORWARD SCHEME WITH LAYERED DECODING

In [4] a quantize-map-forward scheme was presented for the unicast and multicast Gaussian relay network. It was shown that this scheme is approximately optimal, i.e. it gives a reliability criterion for rates within a constant gap of the cutset bound, where the constant gap depends only on the size of the network and not on the channel parameters. In this scheme, each node quantizes the received signal, symbol by symbol, at the noise level. The quantized symbols accumulated together in a block are then mapped to a transmit codeword at that node. These transmission codebooks at every node are generated independently of each other.

In [11], a related scheme was presented for the unicast Gaussian network. Here, the coding and quantization is done in a structured manner using lattices. The scheme was shown to achieve performance similar to the quantize-map-forward scheme of [4] in terms of the reliable rates.

In [12], a *noisy network coding* scheme in the more general setting of the discrete memoryless network was presented for the unicast relay network and also generalized to the case of multicast and multiple sources with single destination. In this scheme, the relay quantizes the received signal in blocks using vector-quantization, subsequently mapping each quantized codeword to a unique codeword, which is re-transmitted by the relay. Specialized to the Gaussian network, the noisy network coding can be thought of as a vector version of the quantize-map-forward scheme, where each relay does a vector quantization rather than the scalar quantization proposed in [4].

In [7], an alternate approach was provided, wherein the discrete superposition network was used as a digital interface for the Gaussian network and the scheme was constructed by lifting the scheme for the discrete superposition network. The discrete superposition network provided the quantization interface for this scheme.

In this chapter, a compress-and-forward scheme is presented for a relay

network in the general setting of the discrete memoryless network. This scheme is similar to the noisy network scheme, but generalizes it in the following sense. In the compress-and-forward scheme, the relay node *bins* the quantized received signal and subsequently maps the bin number to a unique codeword, which is then retransmitted by the relay. The feasible region of communication rate and compression rates at each relay node is characterized under the optimal maximum likelihood decoding rule and a reduced complexity layered decoding scheme.

The compress-and-forward scheme with the layered decoding scheme presents an efficient architecture for the relay network, wherein the encoding and decoding operation is done over smaller sized local sub-networks. Further, this architecture too is approximately optimal.

In the next section, the compress-and-forward scheme is presented in the context of the unicast network. The extension to multiple-source and multicast is presented in Section 3.2.

3.1 Compress-and-forward scheme for unicast network

Recall that a unicast network has a single source node, denoted by S , with the message, denoted by W , which is required at a single destination node, denoted by D . Further, only a layered network is considered as shown in Figure 3.1, so that

$$\mathcal{V} = \bigcup_{l=1}^L \mathcal{O}_l, \quad (3.1)$$

where \mathcal{O}_l denotes the m_l nodes in the l -th layer. The k -th node in the l -th layer will be denoted by $v_{(l,k)}$. The first layer has only one node which is the source node and is denoted by $v_{(1,1)}$ or S . The last layer has only the destination node and is denoted by $v_{(L,1)}$ or D . The nodes other than the source and the destination node will be referred to as the relay nodes and are denoted by \mathcal{V}_r .

In the layered network, the received symbol for a node in the $l + 1$ -th layer depends only on the transmit symbol from the nodes in the l -th layer. The probability transition function describing the general discrete-time memoryless channel model can be decomposed into a product form as

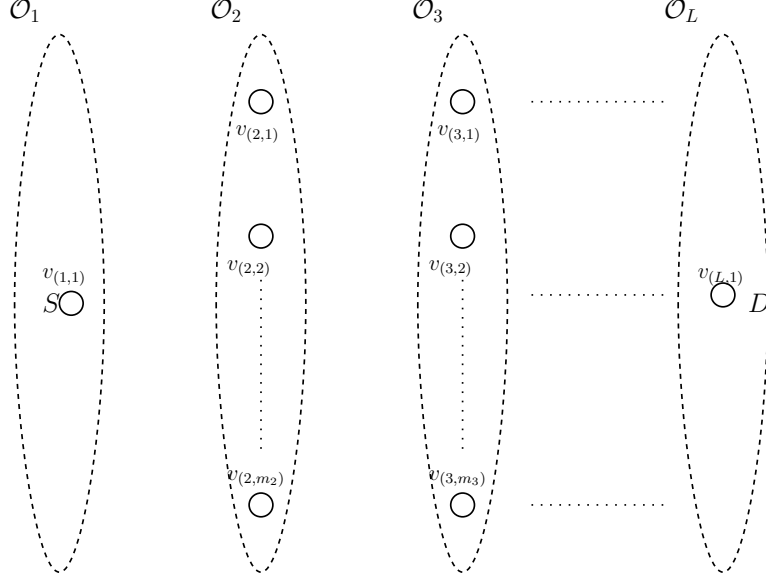


Figure 3.1: Layered unicast network.

follows.

$$p(y_{\mathcal{V}}|x_{\mathcal{V}}) = \prod_{l=1}^{L-1} p(y_{\mathcal{O}_{l+1}}|x_{\mathcal{O}_l}). \quad (3.2)$$

Here $x_{\mathcal{O}_l}$ is used to denote $\{x_v : v \in \mathcal{O}_l\}$ and $y_{\mathcal{O}_l}$'s are similarly defined.

Further, the additive noise at each node is assumed to be independent of each other. This implies that the channel model at each layer can be further decomposed as follows:

$$p(y_{\mathcal{O}_{l+1}}|x_{\mathcal{O}_l}) = \prod_{k=1}^{m_{l+1}} p(y_{v_{(l+1,k)}}|x_{\mathcal{O}_l}). \quad (3.3)$$

A block-encoded layered scheme is considered where each node performs its operation over blocks of time symbols. The relay node quantizes (or compresses) the symbols it receives over a block of time to finite bits. These bits are then transmitted in the next block. The compression rate at a relay node is defined to be the rate of transmission of the compressed bits.

Assuming that uniformly sized blocks of T symbols are used by each node for this operation, a compress-and-forward scheme is parametrized by $(T, R, \{r_v\}_{v \in \mathcal{V}_r})$, where R is the overall rate of communication and r_v 's are the compression rates at the relay nodes. A rate vector $(R, \{r_v\}_{v \in \mathcal{V}_r})$ is said to be feasible w.r.t. the compress-and-forward scheme, if for any arbitrary

$\epsilon > 0$, there exists a compress-and-forward scheme $(T, R, \{r_v\}_{v \in \mathcal{V}_r})$ which achieves a probability of error less than ϵ .

The following theorem characterizes the feasible region of $(R, \{r_v\}_{v \in \mathcal{V}_r})$ for the compress-and-forward scheme.

Theorem 1. *A rate vector $(R, \{r_v\}_{v \in \mathcal{V}_r})$ is feasible if for some collection of random variables $\{X_{\mathcal{V}}, \hat{Y}_{\mathcal{V}}\}$, henceforth denoted by Q_p , which is distributed as*

$$p(X_{\mathcal{V}}, \hat{Y}_{\mathcal{V}}, Y_{\mathcal{V}}) = \left(\prod_{v \in \mathcal{V}} p(X_v) \right) p(Y_{\mathcal{V}} | X_{\mathcal{V}}) \left(\prod_{v \in \mathcal{V}} p(\hat{Y}_v | Y_v) \right), \quad (3.4)$$

the vector $(R, \{r_v\}_{v \in \mathcal{V}_r})$ satisfies

$$R < r(\Omega^c \setminus \Phi) + I(\hat{Y}_{\Phi}; X_{\Omega} | X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}}), \quad (3.5)$$

$\forall \Omega, \Phi$, s.t., $S \in \Omega \subseteq \mathcal{V}$, $D \in \Phi \subseteq \Omega^c$, where $r(\mathcal{A}) \triangleq \sum_{v \in \mathcal{A}} r_v$.

(Note: The choice $\hat{Y}_D = Y_D$ is always optimal for (3.5)).

Proof. The proof is by random coding technique. A random ensemble of coding scheme is defined using the collection of random variables Q_p distributed as given by (3.4). A scheme in the ensemble is generated as follows.

1. *Source codebook and encoding:* For each message $w \in [2^{TR}]$, the source generates a T -length sequence $x_s^T(w)$ using i.i.d. $p(X_S)$.
2. *Relay codebooks and mappings:* For every relay node $v \in \mathcal{V}_r$ a binned quantization codebook is generated with 2^{Tr_v} bins. The binned quantization codebook is given by $\hat{y}_v^T(w_v, \bar{w}_v)$, where $w_v \in [2^{Tr_v}]$ and $\bar{w}_v \in [2^{T\bar{r}_v}]$. And it is generated using i.i.d. $p(\hat{Y}_v)$.

Every relay node also generates a transmission codebook of size 2^{Tr_v} , which consists of $x_v^T(w_v)$ sequences generated using i.i.d. $p(X_v)$.

On receiving y_v^T , the relay node finds a vector $\hat{y}_v^T(w_v, \bar{w}_v)$ in the quantization codebook and transmits $x_v^T(w_v)$ corresponding to the bin number of the quantization vector.

If the relay cannot find any quantization vector, it transmits a sequence corresponding to any bin uniformly at random. The probability that this latter event is arbitrarily small is ensured by letting

$$\bar{r}_v = I(Y_v, \hat{Y}_v) - r_v + \epsilon_1, \quad (3.6)$$

for an arbitrarily small $\epsilon_1 > 0$. This ensures that the total size of the quantization codebook is of the order $2^{TI(Y_v, \hat{Y}_v)}$.

3. *Decoding:* On receiving y_D^T , the destination node finds a *unique* \hat{w} , and *any* $\{(\hat{w}_v, \hat{\hat{w}}_v)\}_{v \in \mathcal{V}_r}$, such that

$$\left(x_S^T(\hat{w}), \left\{ \hat{Y}_v^T(\hat{w}_v, \hat{\hat{w}}_v), x_v^T(\hat{w}_v) \right\}_{v \in \mathcal{V}_r}, y_D^T \right) \in \mathcal{T}_\epsilon^T. \quad (3.7)$$

If it is successful, the destination declares \hat{w} as the decoded message; if not, the destination declares an error.

The theorem follows by the standard argument of showing that the average probability of error, averaged over the ensemble of codes and over all messages, goes to 0 as T tends to infinity. The details of the error probability analysis are in Appendix A.1.

□

In the usual communication problem setup, one is interested in only maximizing the overall communication rate R . The following corollary of the above theorem establishes the achievable rate by the compress-and-forward scheme.

Corollary 1. *The communication rate R is achievable by the compress-and-forward scheme if*

$$R < \min_{\Omega \subseteq \mathcal{V}, S \in \Omega} I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) - I(\hat{Y}_{\Omega}; Y_{\Omega} | X_{\mathcal{V}}), \quad (3.8)$$

for some collection of random variables Q_p .

Proof. The compress-and-forward scheme with $R_v = I(Y_v, \hat{Y}_v) + \epsilon_1$ achieves this rate. □

It should be noted that the achievable rate in (3.8) is the same as the one obtained in noisy network coding scheme in [12]. This is not surprising as by allowing the compression rates to be large enough, the scheme essentially reduces to the noisy network coding scheme, where every quantized codeword is uniquely mapped to a re-transmission codeword at the relay node.

3.1.1 A low-complexity layered decoding scheme

A maximum likelihood decoder maximizes the probability of the received vector conditioned on the transmitted codeword at the source. (Note that the jointly-typical-set decoding is a proof technique for the random coding argument and it upper-bounds the error probability that can be achieved by the maximum likelihood (ML) decoder.

$$\text{ML decoder: } \hat{w} = \operatorname{argmax}_w p(y_D^T | x_S^T(w)). \quad (3.9)$$

The conditional probability depends on the channel model and the operations (quantization, compression and mapping) at each node. Therefore implementing a ML decoder has very high complexity. In [13], the ML decoder is implemented for a simple one-relay network with binary LDPC codes and a reduced quantizer operation for which the decoding reduces to belief-propagation over a large Tanner graph, which comprises the Tanner graphs of the LDPC codes for each node, the quantization and mapping operation, and the network itself. Even when this simplified encoding scheme is extended to a network with multiple layers of relay nodes, the decoding complexity would be large. In this section, a simplified decoding architecture is presented for the compress-and-forward scheme which operates layer-by-layer and decodes the compressed bits transmitted by each relay node.

Layered decoding scheme: The decoder at the destination node operates backwards layer-by-layer. First, it decodes the messages (or compressed bits) transmitted by the nodes in the layer \mathcal{O}_{L-1} . Then using these decoded messages, it decodes the messages in the layer \mathcal{O}_{L-2} . This process continues till the destination node eventually decodes the source message.

The following theorem characterizes the feasible region of $(R, \{r_v\}_{v \in \mathcal{V}_r})$.

Theorem 2. *A rate vector $(R, \{r_v\}_{v \in \mathcal{V}_r})$ is feasible for the compress-and-forward scheme, under the layered decoding scheme, if for some Q_p the vector $(R, \{r_v\}_{v \in \mathcal{V}_r})$ satisfies*

$$r(U) \leq I(X_U; Y_D | X_{\mathcal{O}_{L-1} \setminus U}), \quad \forall U \subseteq \mathcal{O}_{L-1}, \quad (3.10)$$

$$\begin{aligned} r(U) - r(\mathcal{O}_{l+1} \setminus V) &\leq I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U}) - I(\hat{Y}_{\mathcal{O}_{l+1} \setminus V}; Y_{\mathcal{O}_{l+1} \setminus V} | X_{\mathcal{O}_l}), \\ &\quad \forall U \subseteq \mathcal{O}_l, V \subseteq \mathcal{O}_{l+1}, 2 \leq l \leq L-2, \end{aligned} \quad (3.11)$$

$$R - r(\mathcal{O}_2 \setminus V) \leq I(X_S; \hat{Y}_V) - I(\hat{Y}_{\mathcal{O}_{l+1} \setminus V}; Y_{\mathcal{O}_{l+1} \setminus V} | X_S), \quad \forall V \subseteq \mathcal{O}_2. \quad (3.12)$$

Proof. The proof is by backward induction. Assuming that the destination has decoded the messages transmitted by the relay nodes in layer \mathcal{O}_{l+1} , the probability of error for decoding the messages from the layer \mathcal{O}_l is considered. To do so, a hypothetical layered network as shown in Figure 3.2 is considered. This network consists of the layers \mathcal{O}_l and \mathcal{O}_{l+1} and in addition a layer with an aggregator node A . A node $v_{(l+1,j)}$ in layer \mathcal{O}_{l+1} is connected to the aggregator node with wired link of capacity $r_{v_{(l+1,j)}}$ bits per symbol. This layer represents the forward part of the network beyond layer \mathcal{O}_{l+1} .

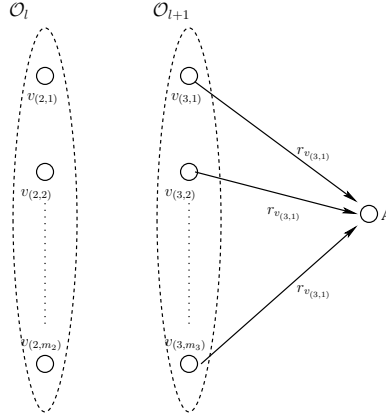


Figure 3.2: A hypothetical network.

This network is now a multiple-source single-destination relay network, with all the nodes in layer \mathcal{O}_l being source nodes and the aggregator node as the destination node. The node $v_{(l,j)}$ has a message for the aggregator node with rate $r_{v_{(l,j)}}$. The noisy network coding scheme [12] assures that the messages can be decoded with arbitrarily small probability of error, if

$$r(U) - r(\mathcal{O}_{l+1} \setminus V) \leq I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U}) - I(\hat{Y}_{V^c}; Y_{V^c} | X_{\mathcal{O}_l}), \quad (3.13)$$

$\forall U \subseteq \mathcal{O}_l, V \subseteq \mathcal{O}_{l+1}$, where the above inequality corresponds to the cut $\Omega = U \cup V^c$.

□

Note that the layered decoding scheme is weaker than the ML decoding scheme. Therefore the feasible region under the layered decoding scheme should be a strict subset of the feasible region under the ML decoding scheme.

However, the following theorem that will be proved in the next chapter in Section 4.3 shows that the compress-and-forward scheme with layered

decoding achieves similar communication rate as the noisy network coding scheme.

Theorem 3. *The communication rate R is achievable by the compress-and-forward scheme with layered decoding if for some collection of random variables Q_p ,*

$$R < \min_{\Omega \subseteq \mathcal{V}, S \in \Omega} I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) - \kappa_1, \quad (3.14)$$

where the constant κ_1 is given by the recursive relation,

$$\kappa_l = I(\hat{Y}_{\mathcal{O}_{l+1}}; Y_{\mathcal{O}_{l+1}} | X_{\mathcal{O}_l}) + \kappa_{l+1} |\mathcal{O}_{l+1}|, \quad (3.15)$$

and $\kappa_{L-1} = 0$.

(Note: It is conjectured that the constant κ_1 can be further tightened to make the achievable rate region comparable to the region in Theorem 2.)

The above theorem will be proved by characterizing an *information flow* for the network. Note that the conditions of Theorem 2 can be interpreted as a flow decomposition for the layered network. If R is the information that flows from the source to the destination, then the flow decomposition gives the effective amount of information that flows through each node. If the compression rate at each relay node is made approximately equal to the information flowing through that node, then the layered decoding where the destination ends up decoding the effective information at each node has a chance to work. Thus, in order to choose the right compression rates at each node, a flow decomposition for the network must be obtained. These notions are made more precise in the next chapter.

3.2 Generalizations to multi-source networks

Consider the communication network with multiple source nodes $\{S_i | i \in [J]\}$. The source node S_i has independent message W_i at rate R_i . There is a common destination node D . The network is illustrated in Figure 3.3. The noisy network coding scheme of [12] extends to this case as well. In fact this result was used for each layer to analyze the layered decoding scheme in the proof of Theorem 2.

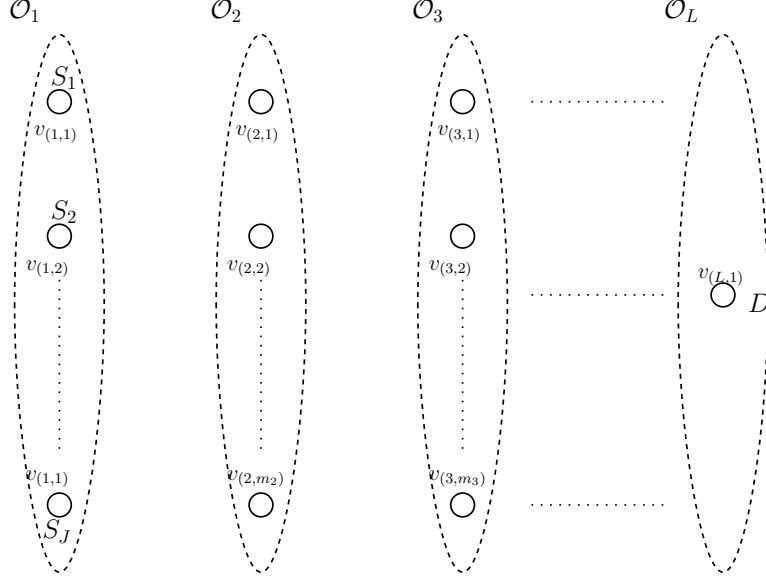


Figure 3.3: A layered multi-source network.

The results of the compress-and-forward scheme and the layered decoding scheme can be generalized to the communication network with multiple source nodes and a common destination node.

The following corollary extends the results of the compress-and-forward scheme for the unicast network to the multi-source relay network.

Theorem 4. *The communication rates $\vec{R} = (R_1, \dots, R_J)$ are achievable by the compress-and-forward scheme (with joint decoding) for the multi-source single destination network if, for some collection of random variables Q_p which is distributed as (3.4), the rates satisfy*

$$R(\Omega_1) < I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) - I(\hat{Y}_{\Omega}; Y_{\Omega} | X_{\mathcal{V}}), \quad \forall \Omega, \text{ s.t., } \Omega \subseteq \mathcal{V}, D \in \Omega^c, \quad (3.16)$$

where $\Omega_1 \triangleq \Omega \cap \mathcal{O}_1$.

Further, with the layered decoding scheme, the rates $\vec{R} = (R_1, \dots, R_J)$ are achievable if

$$R(\Omega_1) < I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) - |\Omega_1| \kappa_1, \quad (3.17)$$

where κ_1 is given by (3.15).

The results can be proved by adding a hypothetical supernode in layer

0, which is connected to the source nodes with orthogonal wired links such that the wired link to node S_i is of rate R_i .

3.3 Special cases

3.3.1 Gaussian network

For the special case of the Gaussian network described in Section 2.1.2, the achievable rates can be compared to the cutset bound.

As noted in [12], a good choice for \hat{Y}_v for the Gaussian network is given by

$$\hat{Y}_v = Y_v + \hat{Z}_v, \quad (3.18)$$

where $\hat{Z}_v \sim \mathcal{CN}(0, 1)$ is independent across nodes.

The particular choice of \hat{Y}_v implies that the quantization is done at the noise level. This also agrees with the philosophy in [4, 7], where the quantization was done at the noise level to show approximate optimality; in [4], scalar quantization was done at the noise level, and in [7], quantization was done using the discrete superposition network, which was a model obtained from the Gaussian network by clipping the signal at the noise level.

As shown in [12], with this choice of \hat{Y}_v and with $X_{\mathcal{V}} \sim \mathcal{CN}(0, I)$,

$$I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) = \log \left| I + \frac{H_{\Omega\Omega^c} H_{\Omega\Omega^c}^*}{2} \right| \quad (3.19)$$

$$\geq \log |I + H_{\Omega\Omega^c} H_{\Omega\Omega^c}^*| - \frac{|\Omega^c|}{2}. \quad (3.20)$$

And further,

$$I(\hat{Y}_v; Y_v | X_{\mathcal{V}}) \leq 1. \quad (3.21)$$

Using (3.20), (3.21) and Lemma 1, the following corollary of Theorem 4 follows.

Corollary 2. *If $\vec{R} = (R_1, \dots, R_J)$ is in \bar{C}^g , then rates $\vec{R} - 3|\mathcal{V}|\vec{1}$ are achievable by the compress-and-forward scheme (with joint decoding) for the multi-source single destination Gaussian network. Further, with the layered decod-*

ing scheme, the rates $\vec{R} - (2|\mathcal{V}| + \kappa_1^g)\vec{1}$ are achievable, where

$$\kappa_l^g = 1 + \kappa_{l+1}^g |\mathcal{O}_{l+1}|, \quad (3.22)$$

and $\kappa_{L-1}^g = 0$.

3.3.2 Deterministic network

For the special case of the deterministic network described in Section 2.1.3, the optimal choice of \hat{Y}_v is Y_v and with this choice

$$I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) = H(\hat{Y}_{\Omega^c} | X_{\Omega^c}). \quad (3.23)$$

And further,

$$I(\hat{Y}_v; Y_v | X_{\mathcal{V}}) = 0. \quad (3.24)$$

Therefore, specializing the results of Theorem 4 leads to the following corollary.

Corollary 3. *For the multi-source single-destination deterministic network, $\vec{R} = (R_1, \dots, R_J)$ is achievable by the compress-and-forward scheme with the layered decoding scheme if for some collection of random variables Q_p which is distributed as (3.4),*

$$\vec{R} \in \bar{\mathcal{C}}(Q_p). \quad (3.25)$$

Specializing further to the linear deterministic region, it can be shown that the product distribution (with uniformly distributed X_v over all input alphabets) maximizes the cutset bound, thereby showing that all rates in the cutset bound are achievable.

CHAPTER 4

FLOWS WITH BISUBMODULAR CAPACITY CONSTRAINTS

4.1 Introduction

Maximum flow problems are extensively studied in graph theory and combinatorial optimization [14]. The problems are most often motivated from the study of transportation and communication networks. A directed graph $(\mathcal{V}, \mathcal{E})$ consists of the set of vertices or nodes \mathcal{V} and the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Traditionally, flow is defined to be a non-negative function over the set of all edges which satisfy the *flow-conservation law* at each vertex other than the source and the destination node. Further, the flow over any edge is less than the capacity of that the edge. The classic max-flow min-cut result of [15] characterizes the maximum flow from the source to destination node and shows it to be equal to the min-cut of the graph. In order to distinguish from the concept of the node-flow that will be introduced here, such a flow is called an edge flow over an edge-capacitated graph. Beginning from the single commodity result of [15], various extensions of these problems have been considered. In particular, the edge-capacitated graph was extended to a polymatroidal network [16], where the flow is constrained not only by the edge-capacities but by joint capacities on sets of incoming and outgoing edges at every vertex. A special case is the node-capacitated graph [17], where the constraints on the flow are on the sum-total of the incoming and outgoing flow at each node.

In this chapter, the concept of a node-flow in the context of a layered graph with bisubmodular constraints on the flows is introduced. The node-flows can be related to the edge-flows with flow-conservation at the node. Note that the conservation law for edge-flow enforces that the net incoming flow at any node is equal to the net outgoing flow at the node and this quantity can be viewed as the node-flow for a node. The bisubmodular

constraints can be viewed as generalizations of the polymatroidal constraints of [16]. The definitions here are motivated by the layered coding scheme for the wireless network, which was presented in the previous chapter. The main result is a max-flow min-cut theorem for the single-commodity node-flow for a graph with bisubmodular capacity constraints. The result is closely related to, and can be viewed as a generalization of, the flow introduced in the context of the linear deterministic networks and polylinking systems in [18, 19].

4.2 A max-flow min-cut theorem

In this section, the max-flow min-cut theorem is proved for *single-commodity node-flow* on a *layered graph* with *bisubmodular capacity* constraints.

Layered graph: A layered graph is considered, which is represented by a set of nodes \mathcal{V} , which can be decomposed into subsets $\mathcal{O}_l, 1 \leq l \leq L$ as shown in Figure 3.1. The layering is ensured by the edges of the graph, which connect nodes in any layer l to nodes in the subsequent layer $l + 1$. Since the edges do not play any role in the problem here, beyond ensuring the layering, they will henceforth be neglected. The first layer \mathcal{O}_1 has a single node, which is the source node and the last layer \mathcal{O}_L has a single node, which is the destination node.

Bisubmodular capacity functions: The bisubmodular capacity functions are defined for the layered graph using a family of $L - 1$ functions $\{\rho_l : 1 \leq l \leq L - 1\}$, $\rho_l : 2^{\mathcal{O}_l} \times 2^{\mathcal{O}_{l+1}} \rightarrow \mathbb{R}^+$, which satisfy the following properties:

1. ρ_l is bisubmodular, i.e., $\forall U_1, U_2 \subseteq \mathcal{O}_l, V_1, V_2 \subseteq \mathcal{O}_{l+1}$,

$$\rho_l(U_1 \cup U_2, V_1 \cap V_2) + \rho_l(U_1 \cap U_2, V_1 \cup V_2) \leq \rho_l(U_1, V_1) + \rho_l(U_2, V_2). \quad (4.1)$$

2. ρ_l is non-decreasing, i.e.

$$\rho_l(U, V) \leq \rho_l(U_1, V_1), \text{ for } U \cup V \subseteq U_1 \cup V_1. \quad (4.2)$$

3. If $U = \emptyset$ or $V = \emptyset$, then

$$\rho_l(U, V) = 0. \quad (4.3)$$

Node-flow: The node-flow for the layered graph is defined as a function $f : \mathcal{V} \rightarrow \mathbb{R}^+$ which satisfies the capacity constraints, i.e.,

$$f(V) - f(\mathcal{O}_l \setminus U) \leq \rho_l(U, V), \quad \forall U \subseteq \mathcal{O}_l, V \subseteq \mathcal{O}_{l+1}, \forall l \in [L-1], \quad (4.4)$$

where $f(A)$ is an over-loaded notation, such that when $A \subseteq \mathcal{V}$ then $f(A) \triangleq \sum_{v \in A} f(v)$. Further, the destination node must sink the flow from the source. Therefore $f(D) = f(S)$.

The max-flow problem is to find the maximum $f(S)$ that can be supported given the capacity constraints on the graph. An efficient algorithm to compute the flow at each node given any $f(S)$ that can be supported is also sought.

An upper bound on the max-flow is given by the cut function.

Cut function: The cut function $C : 2^{\mathcal{V}} \rightarrow \mathbb{R}_+$ is defined as

$$C(\Omega) \triangleq \sum_{l=1}^{L-1} \rho_l(\Omega_l, \mathcal{O}_{l+1} \setminus \Omega_{l+1}), \quad (4.5)$$

where $\Omega_l \triangleq \Omega \cap \mathcal{O}_l$.

Clearly,

$$\max f(S) \leq \min_{\Omega \subseteq \mathcal{V}} C(\Omega). \quad (4.6)$$

The next theorem shows that the min-cut is achievable. The proof is constructive and gives an efficient method of computing the flow.

Theorem 5.

$$\max f(S) = \min_{\Omega \subseteq \mathcal{V}} C(\Omega). \quad (4.7)$$

Proof. The proof is based on the polymatroid intersection theorem. The details are in Appendix B.1. \square

The max-flow min-cut theorem for node-flows with bisubmodular constraints presented here is closely related to the max-flow min-cut results of [18, 19]. [18] considered linear deterministic networks, which led to bisubmodular capacity functions arising from the rank of a matrix. [19] considered polylinking systems, where the bisubmodular capacity functions are given by the polylinking function. The results of [19] generalized the results of [18] by showing that a linear deterministic network is a special case of polylinking system.

The max-flow min-cut theorem can be easily generalized to the following two cases:

- Consider a layered graph with J source nodes in \mathcal{O}_1 and a single destination node in \mathcal{O}_L , such that $f(\mathcal{O}_1) = f(D)$. For this case, the following corollary generalizes Theorem 5.

Corollary 4. $\{f(v)|v \in \mathcal{O}_1\}$ is a feasible flow iff,

$$f(\Omega_1) \leq C(\Omega), \quad \forall \Omega \subseteq \mathcal{V}, \quad (4.8)$$

where $\Omega_1 \triangleq \Omega \cap \mathcal{O}_1$.

- Consider a layered graph with a single source node in \mathcal{O}_1 and J destination nodes in \mathcal{O}_L , such that $f(S) = f(\mathcal{O}_L)$. For this case, the following corollary generalizes Theorem 5.

Corollary 5. $\{f(v)|v \in \mathcal{O}_L\}$ is a feasible flow iff,

$$f(\Omega_L) \leq C(\Omega), \quad \forall \Omega \subseteq \mathcal{V}, \quad (4.9)$$

where $\Omega_L \triangleq \Omega \cap \mathcal{O}_L$.

Note that the proof for the multiple sources (or destinations) case follows by adding a hypothetical supernode A in layer 0 (or $L + 1$) with capacity functions ρ_0 (or ρ_L) given by $\rho_0(A, V) = \sum f(v)$, $\forall V \subseteq \mathcal{O}_1$ (or $\rho_L(V, A) = \sum f(v)$, $\forall V \subseteq \mathcal{O}_L$).

4.3 A compress-and-forward scheme from flows

In this section, Theorem 3 is proved by establishing a connection between the compression rates of the compress-and-forward scheme with the layered decoding and the node-flows with bisubmodularity constraints. Recall that the achievable rates for the compress-and-forward with the layered decoding scheme are given by (3.10)-(3.12), which appear very much like the bisubmodular capacity constraints.

To make this connection more precise, first observe the following proposition.

Proposition 1. *Given the collection of random variables Q_p distributed as given by (3.4), the family of $L-1$ functions $\rho_l : \mathcal{O}_l \times \mathcal{O}_{l+1} \rightarrow \mathbb{R}^+$, $\forall l \in [L-1]$ defined by*

$$\rho_l(U, V) \triangleq I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U}) \quad (4.10)$$

forms a family of bisubmodular capacity functions.

Proof. Appendix B.3. □

For any $\Omega \subseteq \mathcal{V}$, the corresponding cut value $C(\Omega)$ is now given by

$$C(\Omega) = \sum_{l=1}^{L-1} I(X_{\Omega_l}; \hat{Y}_{\mathcal{O}_{l+1} \setminus \Omega_{l+1}} | X_{\mathcal{O}_l \setminus \Omega_l}) \quad (4.11)$$

$$= I(\hat{Y}_{\Omega_c}; X_{\Omega} | X_{\Omega^c}). \quad (4.12)$$

Theorem 5 is then used to construct a flow $f(v)$ for this network, such that

$$f(S) \leq \min_{\Omega} I(\hat{Y}_{\Omega_c}; X_{\Omega} | X_{\Omega^c}), \quad S \in \Omega, D \in \Omega^c, \quad (4.13)$$

and

$$f(V) - f(\mathcal{O}_l \setminus U) \leq \rho_l(U, V), \quad \forall U \subseteq \mathcal{O}_l, V \subseteq \mathcal{O}_{l+1}, \forall l \in [L-1]. \quad (4.14)$$

For any $v \in \mathcal{O}_l, l \in [L-1]$, let

$$r_v = f(v) - \kappa_l, \quad (4.15)$$

and $R = f(S) - \kappa_1$, where κ_l is given by (3.15).

Then $\forall U \neq \emptyset \subseteq \mathcal{O}_l, V \subseteq \mathcal{O}_{l+1}$,

$$r(U) - r(\mathcal{O}_{l+1} \setminus V) = f(U) - f(\mathcal{O}_{l+1} \setminus V) - |U|\kappa_l + |\mathcal{O}_{l+1} \setminus V|\kappa_{l+1} \quad (4.16)$$

$$\leq \rho_l(U, V) - \kappa_l + |\mathcal{O}_{l+1} \setminus V|\kappa_{l+1} \quad (4.17)$$

$$= \rho_l(U, V) - I(\hat{Y}_{\mathcal{O}_{l+1}}; Y_{\mathcal{O}_{l+1}} | X_{\mathcal{O}_l}) \quad (4.18)$$

$$\leq I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U}) - I(\hat{Y}_{\mathcal{O}_{l+1} \setminus V}; Y_{\mathcal{O}_{l+1} \setminus V} | X_{\mathcal{O}_l}). \quad (4.19)$$

Therefore $(R, \{r_v\}_{v \in \mathcal{V}_r})$ satisfies (3.10)-(3.12). This proves Theorem 3.

CHAPTER 5

RECIPROCITY

5.1 Introduction

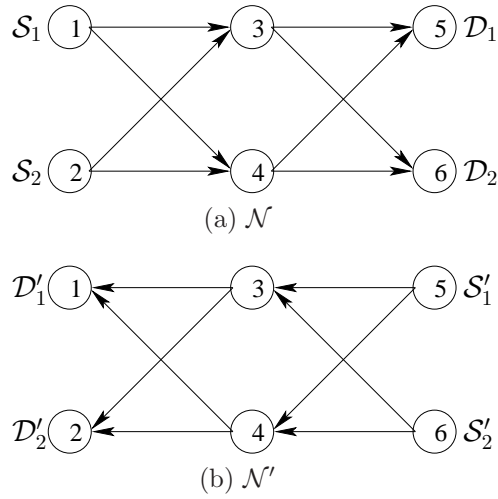


Figure 5.1: An example of a layered network with two-unicast (a), and its reciprocal (b).

Consider a network \mathcal{N} with multiple unicast traffic. The reciprocal of such a network \mathcal{N}' is illustrated with an example in Figure 5.1 and is defined as follows.

- The set of nodes in \mathcal{N}' is the same set of nodes \mathcal{V} in \mathcal{N} .
- The direction of the edges is reversed. This implies a reciprocal relation in the sense that if the transmit symbol of node v influenced the received symbols of a set of nodes in \mathcal{N} , then in the reciprocal network the received symbol of the node v is influenced by exactly the same set of nodes. To complete the picture, the exact channel model for the reciprocal network needs to be defined. It is not clear what is the most

appropriate way to do so for a general network, but we look at this in the context of two particular network models of interest.

1. *Gaussian network*: The channel attenuation between any pair of nodes in both networks, the original and its reciprocal, is assumed to be the same. This is in agreement with the physical nature of electromagnetic propagation. With multiple antennas, the channel matrix for a pair of nodes in the reciprocal network is given by the transpose of the channel matrix in the opposite direction for the original network. The subtle part is with respect to the power constraint and the additive noise. A simplification could be that all nodes have unit power constraint and the additive noise is unit variance.
 2. *Linear deterministic network*: The linear deterministic network is noiseless and does not have a power constraint. Therefore, the difficulties that arise for the case of a Gaussian network are not present. The channel matrix between any two nodes in the reciprocal network is the transpose of the channel matrix of the original network
- The reciprocal network has the same set of messages as the original network, but with the roles of the source and destination nodes swapped.

5.2 Examples of reciprocity in wireless networks

While it is unresolved whether a given network and its reciprocal (when defined appropriately) have the same capacity region, many interesting examples are known for which this is true. For some cases, this reciprocity is applicable even at the scheme level. We discuss some of the interesting examples below:

- In [20, 21], reciprocity (or duality) was shown between the *Gaussian multiple access channel (MAC)* and the *Gaussian broadcast channel (BC)*. It was shown that the capacity region of the MAC is equal to the capacity region of the BC under the same sum power constraint. This duality was also shown, interestingly, at the scheme level between

the dirty-paper pre-coding for the BC and the successive cancellation for the MAC.

- *Two-user interference channel:* The exact capacity of the linear deterministic two-user interference channel was characterized in [3]. It was shown that the capacity of this channel is the same as that of its reciprocal, which is a different two-user interference channel. For the Gaussian version of the problem, the reciprocity was shown with respect to the generalized degrees of freedom.

Two modifications of the two-user interference channel, with source cooperation and with destination cooperation, were considered in [22, 23]. The two networks are reciprocal of each other. It was shown that the capacity regions for the linear deterministic version and the generalized degrees of freedom region of the Gaussian version of the two networks are the same.

- *One-to-many and Many-to-one interference channels* are reciprocal networks and were studied in [24]. The capacity regions for the linear deterministic channel model and the generalized degrees of freedom region of the Gaussian channel model of the two networks were found to be the same.
- A *Wireline network* is a noiseless network and a special case of the deterministic network described in Chapter 2. It is a network of nodes with noiseless links between pairs of connected nodes with a certain capacity. It was been shown in [25] that wireline networks are reciprocal (also called reversible in the literature) under linear coding.

In the next section, the result of [25] for wireline networks is generalized to linear deterministic networks. It is shown that the linear deterministic network is reciprocal when the operations at each node are restricted to be linear.

5.3 Reciprocity in linear deterministic network under linear coding

A linear deterministic network with n unicast messages flowing in the network is considered. Consider a communication scheme over T transmission times (symbols). Every message W_k , for $1 \leq k \leq n$, is a vector of independent symbols of length $w_k T$, i.e. $W_k \in \mathbb{F}_p^{w_k T}$. The message W_k is available at the source node \mathcal{S}_k and is demanded by the destination node \mathcal{D}_k . The corresponding rate associated with the message W_k is $R_k = w_k \log p$ bits. Thus the network is associated with a rate requirement (R_1, \dots, R_n) .

At any time instant t , $0 \leq t \leq T - 1$, a node j transmits a signal $x_j[t]$, which is determined by the encoding function $f_j^{(t)}$. The destination node for the message with index k reconstructs an estimate \hat{W}_k using the decoding function g_k . A communication scheme is said to be *linear coding*, if the functions $f_j^{(m)}$ and g_k are linear.

A *linear deterministic network* is *solvable* if there exists a coding scheme such that $\hat{W}_k = W_k$. Further, if there exists a linear coding scheme, the network is *linearly solvable*. This follows the standard definitions in the (wireline) network coding literature [26].

Recall that for the reciprocal of the linear deterministic network, the channel gain matrix associated with the an edge (j, i) $\mathcal{N}' G_{ji}$ is given by the the reciprocal of the channel gain matrix associated with the edge (i, j) in the original network $\mathcal{N} G_{ij}^T$. Note that for the shift deterministic network, the reciprocal is no longer a shift deterministic network, but rather a *flipped* shift deterministic network, where the vectors are shifted upwards rather than downwards. This is discussed separately in Section 5.3.2.

A network is *reversible* if the reciprocal of the network is solvable. Further, a network is *linearly reversible* if the reciprocal of the network is linearly solvable. The main result of this section is the following theorem.

Theorem 6. *Any linear deterministic network which is linearly solvable is linearly reversible.*

Note that the above theorem is equivalent to the statement that *the linear deterministic network and its reciprocal have the same achievable rate region under the class of linear coding scheme.*

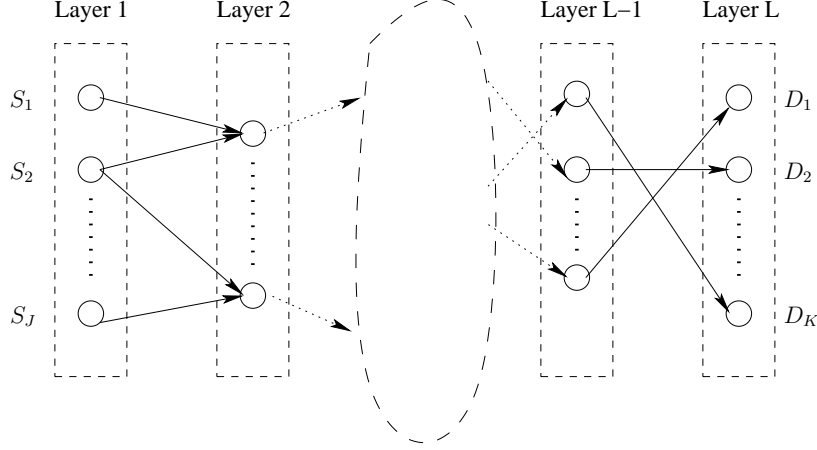


Figure 5.2: Layered linear deterministic network.

5.3.1 Layered linear deterministic network.

Consider the L -layered linear deterministic network illustrated in Figure 5.2. A *layered* transmission scheme over such a network is considered, such that each node only transmits once and it does so after it has received signals from the nodes in the previous layer. The concepts of solvable, linearly solvable, reversible and linearly reversible hold for the layered network with the layered transmission scheme too.

In Appendix C.1, it is shown that any linear deterministic network with a coding scheme over a block of time $L - 1$ can be unfolded over time to create a layered linear deterministic network. Further, it is straightforward to see that the reciprocal of the original network corresponds to the reciprocal of the layered network. Thus it suffices to prove Theorem 6 for the layered network only.

Proof of Theorem 6 (Layered networks): Consider a linear layered network \mathcal{N} , which is solvable by a linear coding scheme. The linear coding scheme is specified by a set of linear matrices. The coding matrices at the source nodes are denoted by $C_k \in \mathbb{F}_p^{q \times w_k}$, for $k = 1, \dots, n$. If j is a source node for a subset of the messages, $\Omega_j \subseteq \{1, \dots, n\}$, then the signal transmitted by node j is given by

$$x_j = \sum_{k \in \Omega_j} C_k W_k. \quad (5.1)$$

The decoding matrices are denoted by $D_k \in \mathbb{F}_p^{w_k \times q}$, for $k = 1, \dots, J$. If j

is a destination node for a subset of the messages, $\Omega_j \subseteq \{1, \dots, J\}$, then the node j reconstructs the messages from the received signal y_j using the decoding matrix,

$$\hat{W}_k = D_k y_j, \quad \forall k \in \Omega_j. \quad (5.2)$$

Any intermediate relay node is associated with relay coding matrix F_j such that

$$x_j = F_j y_j. \quad (5.3)$$

Note that since transmission happens at each node in the layered network only once, we can conveniently drop the time index.

The original messages at the source nodes, $\{W_k\}_1^J$, and the reconstructed messages at the destination nodes, $\{\hat{W}_k\}_1^J$, can be related linearly as

$$\hat{W}_k = \sum_l \Gamma_{lk} W_l. \quad (5.4)$$

Γ_{lk} is the transfer coefficient matrix between the source node \mathcal{S}_l and the destination node \mathcal{D}_k . It can be obtained by considering any path from the source node to the destination node, taking the product of the encoding and the channel matrices as you move along that path and then summing up this product for all such paths. In the example of Figure 5.1, the transfer coefficient matrices for the network \mathcal{N} are

$$\begin{aligned} \Gamma_{11} &= D_1 G_{35} F_3 G_{13} C_1 + D_1 G_{45} F_4 G_{14} C_1 \\ \Gamma_{12} &= D_2 G_{36} F_3 G_{13} C_1 + D_2 G_{46} F_4 G_{14} C_1 \\ \Gamma_{22} &= D_2 G_{36} F_3 G_{23} C_2 + D_2 G_{46} F_4 G_{24} C_2 \\ \Gamma_{21} &= D_1 G_{35} F_3 G_{23} C_2 + D_1 G_{45} F_4 G_{24} C_2. \end{aligned}$$

Since the network is solvable by this linear scheme, we must have $\hat{W}_k = W_k$, $\forall 1 \leq k \leq J$. Therefore, the transfer coefficient matrices Γ_{lk} must satisfy the following condition:

$$\Gamma_{lk} = \delta_{lk} I. \quad (5.5)$$

Similarly, consider the reciprocal network with the linear coding scheme parameterized by some coding matrices C'_k , decoding matrices D'_k and relay

coding matrices F'_j . The transmitted messages W_k and the reconstructed version \hat{W}_k can again be linearly related by the matrices Γ'_{lk} . In the example of Figure 5.1, the transfer coefficient matrices for the reciprocal network \mathcal{N}' are

$$\begin{aligned}\Gamma'_{11} &= D'_1 G_{13}^T F'_3 G_{35}^T C'_1 + D'_1 G_{14}^T F'_4 G_{45}^T C'_1 \\ \Gamma'_{12} &= D'_2 G_{23}^T F'_3 G_{35}^T C'_1 + D'_2 G_{24}^T F'_4 G_{45}^T C'_1 \\ \Gamma'_{22} &= D'_2 G_{23}^T F'_3 G_{36}^T C'_2 + D'_2 G_{24}^T F'_4 G_{46}^T C'_2 \\ \Gamma'_{21} &= D'_1 G_{13}^T F'_3 G_{36}^T C'_2 + D'_1 G_{14}^T F'_4 G_{46}^T C'_2.\end{aligned}$$

Note that this follows from the fact that $G'_{ji} = G_{ij}^T$. Letting

$$\begin{aligned}C'_k &= D_k^T, \quad \forall 1 \leq k \leq n, \\ D'_k &= C_k^T, \quad \forall 1 \leq k \leq n, \\ \text{and } F'_j &= F_j^T,\end{aligned}$$

it can be seen that $\Gamma'_{lk} = \Gamma_{kl}^T$. Finally, from (5.5), it follows that

$$\Gamma'_{lk} = \delta_{lk} I. \quad (5.6)$$

Therefore, the reciprocal network \mathcal{N}' is linearly solvable and hence the network \mathcal{N} is linearly reversible. \square

5.3.2 Some special cases

The linear deterministic model captures two important special cases: the noiseless wireline network and the shift linear deterministic network, which models the wireless network.

Noiseless wireline networks: Noiseless wireline networks have been extensively studied in network coding literature [27], [28]. They are characterized by orthogonal communication links, so that they are free of both features, interference and broadcast, present in wireless networks. The linear deterministic network model captures the wireline network as a special case obtained by choosing the channel gain matrices G_{ij} such that the incoming and outgoing links become orthogonal. The linear reversibility result of theorem 6, specialized to the wireline networks, gives the known result in [25].

Shift linear deterministic networks: Another special case is the linear deterministic network with the channel gain matrices being shift matrices $S_{ij} = S^{(q-g_{ij})}$. The shift matrix shifts a vector of length q downwards by $q-g_{ij}$ levels and hence g_{ij} represents the strength of the channel. Basic electromagnetic principles suggest that physical media are reversible, i.e. the communication link (channel) behaves the same way in both the forward and the reverse direction. Therefore in the reciprocal network, we should expect the channel gain matrix to be the same. The reciprocal of the linear deterministic network is obtained by taking the transpose of the channel gain matrices. The transpose of the shift matrix S_{ij} shifts the vector of length q by $q - g_{ij}$, but instead of downwards this *shift is upwards*. An important observation here is the following: the transpose of the shift matrix can be interpreted as a *flipping* of all the signal vectors.

More formally: consider the physical reciprocal network with the same channel gain matrix S_{ij} on each edge as the original network. Given any coding scheme for this reciprocal network, we modify it as follows. Every node *flips* its vector before transmitting and *flips* the vector it receives before coding. The *flipping* operation is denoted by left-multiplying the vector with the matrix J , where

$$J = \begin{bmatrix} 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 1 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}_{q \times q} . \quad (5.7)$$

These matrices can then be “absorbed” into the channel matrix S_{ij} to give the effective channel matrix $JS_{ij}J$. We readily see that $JS_{ij}J$ is the same as S_{ij}^T . In the context of the reciprocal network, the actually encoding matrices for the *physical* reciprocal network should be $JC'_k J$, $JD'_k J$ and $JF_j J$.

CHAPTER 6

BROADCASTING IN WIRELESS RELAY NETWORKS

6.1 Introduction

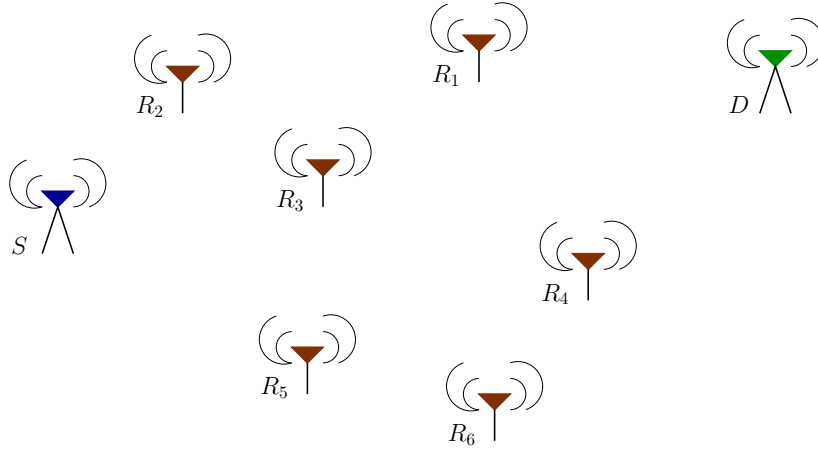


Figure 6.1: A wireless broadcast network.

Consider a communication network with broadcast traffic as illustrated using Figure 6.1. A single source node S is reliably communicating J independent messages, W_1, \dots, W_J , to multiple destination nodes, D_1, \dots, D_J , respectively, at rate R_1, \dots, R_J respectively.

In the example of a cellular system, the setting represents downlink communication where the base-station is transmitting to multiple terminals with the potential help of relay stations. Note that some of the terminals can themselves act as relays.

The broadcasting setup we present here captures two important special cases, which have been extensively studied before - *unicast relay network* and *broadcast channel*.

The unicast relay network is a special case when there is only one destina-

tion node. The unicast relay network was discussed in Chapter 3. It has been studied extensively. In particular [4, 7, 12] presented schemes for the relay network, which all had the underlying philosophy of a quantize-and-forward operation at the relay nodes. In Chapter 3 a compress-and-forward scheme was presented with a layered decoding architecture. All these schemes when specialized to the Gaussian network model were shown to be approximately optimal in the sense that it achieves rates, within a constant gap of the well known cutset outer bound, where the constant gap does not depend on the power and the channel parameters but only on the size of the network.

The broadcast channel is a special case with only a source node and multiple destinations (i.e., no relays). The capacity of the Gaussian broadcast channel with multiple antennas (MIMO broadcast channel) was characterized in [29]. The capacity achieving scheme is based on the Marton's coding scheme [30], which is the best known achievable scheme for the general broadcast channel.

The coding scheme presented here for broadcasting in wireless relay networks is based on combining the schemes for the unicast relay network and Marton's coding scheme for the broadcast channel.

Broadcasting in wireless networks has been studied previously for a simple scenario in [31]. A simple two-user broadcast channel was considered where the destination nodes could also transmit, thereby also acting as relay nodes. Decode-and-forward schemes were considered and specialized outer bounds were given for this network, which were shown to be better than cut-set bounds. More recently, [32] considered broadcasting over two classes of information networks: (a) a network composed of multiple-access channels alone and (b) a network composed of deterministic broadcast channels alone. For such networks, it was shown there that cut-set bound can be achieved. The scheme was a separation based scheme - a local physical layered scheme over the constituent networks to create a wired overlay network and a global routing scheme over the overlay network. Here, generalization to any arbitrary network is considered. The coding scheme presented here is shown to be approximately optimal by comparing it to the cut-set outer bound.

The main result of interest is the following.

Theorem 7. *For the Gaussian broadcast network, a rate vector $\vec{R} = (R_1, \dots, R_J)$*

is achievable if $\forall \mathcal{J}$,

$$\vec{R} + k\vec{1} \in \bar{\mathcal{C}} \quad (6.1)$$

for some constant k , which depends only on the number of nodes, and not on the channel coefficients, and $k = O(|\mathcal{V}| \log |\mathcal{V}|)$.

The approach to prove the above theorem is the following. A coding theorem is first proved for the deterministic network. The proposed scheme operates in two steps. In the *inner* code, the relays essentially perform a quantize-map-forward operation. This induces a vector broadcast channel end-to-end between the source node and the destination nodes. The *outer* code is essentially a Marton code ([30,33]) for the broadcast channel induced by the relaying scheme.

The coding scheme for the discrete superposition network is then *lifted* to the Gaussian network by using the discrete superposition network as a digital interface for the Gaussian network.

6.1.1 A lesson from reciprocity

The key intuition motivating the scheme presented here is the lesson learned from reciprocity. Note that the reciprocal of the broadcast network is the network with the multiple-source and single-destination. It was shown in Section 3.2 that the schemes for the unicast network naturally extend to the multi-source single-destination case. These schemes were further shown to be approximately optimal by comparing them to the cutset bound. Reciprocity would suggest the existence of similar schemes for the broadcast network.

In going from the multi-source to the reciprocal broadcast case, certain difficulties naturally arise. These difficulties were seen even in the simple case of the multiple-access and the broadcast channel. While the capacity schemes for the multiple access channel generalized simply from the point-to-point case, schemes for the broadcast channel involved clever coding at the source node. The difference can be attributed to where (at the transmitter or the receiver) the complexity of the scheme lies. For the multiple-access channel the complexity lies at the decoder. It is easy to show the existence of good coding schemes for the multiple access channel using simple random codebooks and analyzing performance of the complex joint decoder. For the broadcast channel (reciprocally!), the complexity is at the encoder. The

complex encoding is often elusive. In fact, for a general broadcast channel the optimal scheme and the capacity are still unknown. The best known scheme is a family of schemes due to Marton [30] which is optimal only for certain special cases. For the Gaussian channel with multiple antennas it is known that a particular Marton coding scheme (called Costa's dirty paper coding) is optimal [20, 21, 29]. For the degraded broadcast channel, superposition coding is optimal. For the deterministic channel, a simple Marton coding scheme achieves the cutset bound [34].

Carrying this intuition forward to networks, the scheme for the multiple-source network suggests the following scheme for the broadcast network. The relays and the destination node perform a quantize-map-forward operation. The source takes into account the effect of the channel and operations performed by the relay node and needs to do a well-designed scheme for the effective broadcast channel. The limitations on the understanding of the broadcast channel restrict us to design such a scheme only when the channel model is deterministic. Even when the channel model is Gaussian, taking into account the relay operation leads to an end-to-end non-Gaussian broadcast channel. However, for the Gaussian network, the discrete superposition network (DSN) is used as a quantization interface for the Gaussian network.

The rest of the chapter makes these ideas more precise. In Section 6.2, a coding scheme is given for the deterministic broadcast networks. In Section 6.3, we prove Theorem 7 by giving a coding scheme for the Gaussian network.

6.2 Deterministic broadcast networks

For the broadcast problem, the cutset outer bound that was described in Section 2.3.1 can be simplified as follows. If (R_1, \dots, R_J) is achievable, then $\forall \mathcal{J} \subseteq [J]$, and there is a joint distribution $p(\{X_v | v \in \mathcal{V}\})$ (denoted by Q) such that

$$R_{\mathcal{J}} \leq \bar{C}_{\mathcal{J}}(Q) \triangleq \min_{\Omega \in \Lambda_{\mathcal{J}}} I(Y_{\Omega^c}; X_{\Omega} | X_{\Omega^c}). \quad (6.2)$$

Here $\Lambda_{\mathcal{J}}$ is the set of all cuts Ω , such that $S \in \Omega$ and $D_{\mathcal{J}} \triangleq \{D_i | i \in \mathcal{J}\} \subseteq \Omega^c$.

The following theorem characterizes an achievable region for the deterministic broadcast network.

Theorem 8. *For the deterministic broadcast network, a rate vector (R_1, \dots, R_J)*

is achievable if $\forall \mathcal{J}$, and there is a product distribution $\prod_{v \in \mathcal{V}} p(X_v)$ (denoted by Q_p) such that

$$R_{\mathcal{J}} \leq \bar{C}_{\mathcal{J}}(Q_p). \quad (6.3)$$

Remark 1. For many special classes of deterministic networks such as the linear deterministic network and the network composed of deterministic broadcast channels [32], it can be shown that the cutset bound is also maximized by the product distribution, thereby characterizing the capacity of such networks completely.

Proof. Theorem 8 is proved next for the layered network alone. Layering a non-layered network was described in Section 2.3.2.

6.2.1 Coding scheme: Outline

The basic idea of the coding scheme is as follows:

- The broadcast network is converted into a unicast network by adding a super-destination D which has links from each of the destinations D_i by a *wired* link of capacity r_i . If (r_1, \dots, r_J) is in $\bar{C}_{\mathcal{J}}(Q_p)$ (the cutset region of the broadcast network evaluated under product-distributions), then $\bar{C}_{\mathcal{J}}^{\text{uc}}(Q_p)$ (the min-cut of the unicast network evaluated under product-distributions) is equal to $r \triangleq \sum_j r_j$.
- For the relay network, a zero-error coding scheme is employed that operates over T_1 time instants, which achieves the rate r .
- The relaying scheme creates an end-to-end deterministic vector broadcast channel between the source and the destinations over vectors of T_1 symbols.
- A Marton code is used over T_2 vectors to achieve the cutset of the induced deterministic broadcast channel.

6.2.2 Coding scheme in detail

The random ensemble of coding operations is described for a fixed product distribution Q_p . Further, the random coding is described to achieve an ar-

bitrary rate tuple $(r_1, \dots, r_J) \in \bar{C}_{\mathcal{J}}(Q_p)$. The coding is done over a period of $T_1 T_2$ time instants.

Creating the unicast network:

We add a super-sink D to the deterministic broadcast network to obtain a deterministic unicast network. The unicast network is obtained by adding *wired* links of capacity r_i from destination D_i to super-sink D .

Lemma 4. *If $(r_1, \dots, r_J) \in \bar{C}_{\mathcal{J}}(Q_p)$, where $\bar{C}_{\mathcal{J}}(Q_p)$ is the cut-set of the broadcast network evaluated under a distribution Q , then the cut-set of the unicast network with wired links r_1, \dots, r_J evaluated under Q_p is equal to r .*

Proof. See Appendix D.1. □

For a deterministic unicast network, Theorem 4.1 in [4] shows that the cut-set under product form distributions is achievable using an ϵ -error scheme (this is proved by a random coding argument). Since the channel is deterministic, this also implies that there is a zero error scheme which can achieve arbitrarily close to the cut-set bound under product form distributions. Thus the rate r is achievable using such a scheme. Suppose this relaying scheme operates over a block length of T_1 . Let $\vec{x}_v \triangleq x_v^{T_1}$ and $\vec{y}_v \triangleq y_v^{T_1}$ denote the transmit and receive block at any node $v \in \mathcal{V}$. Thus, we have a source codebook for the unicast network given by \mathcal{C}_S , a collection of 2^{rT_1} vectors of length T_1 each. And the relay mappings,

$$f_v : \mathcal{Y}_v \rightarrow \mathcal{X}_v, \quad (6.4)$$

for the relay node $v \in \mathcal{V}_R$.

Relay mappings:

The scheme for the broadcast network operates over $T_1 T_2$ time interval and this entire time duration is divided into T_2 blocks, each composed of T_1 time interval. Each set of T_1 time instants is treated as a block and the vector $\vec{x}(t_2)$ denotes x over the T_1 time instants corresponding to the t_2 -th block: $\vec{x}(t_2) = (x((t_2 - 1)t_1 + 1), x((t_2 - 1)t_1 + 2), \dots, x(t_2 t_1))$. Furthermore \vec{x}^T denotes $(\vec{x}(1), \dots, \vec{x}(T))$.

The relaying operation for the broadcast network is performed in blocks using the relaying scheme for the unicast network as follows. Each relay transmits a T_1 block using only the information from the previous received

T_1 block. Thus

$$\vec{x}_v(t_2) = f_v(\vec{y}_v(t_2 - 1)), \quad \forall t_2 = 1, 2, \dots, T_2. \quad (6.5)$$

Source mappings:

With the fixed relaying operations for the relaying nodes, as defined above, an end-to-end deterministic channel results, as shown in Figure 6.2, between the source and the destination nodes. Note that the input alphabet set at the source node is given by the source codebook of the unicast network \mathcal{C}_S . The deterministic broadcast channel is time-invariant since the same relay mappings are used for all t_2 and is characterized by the functions

$$\vec{y}_{D_i} = \psi_i\{\vec{x}_S\} \quad \forall i = 1, 2, \dots, J. \quad (6.6)$$

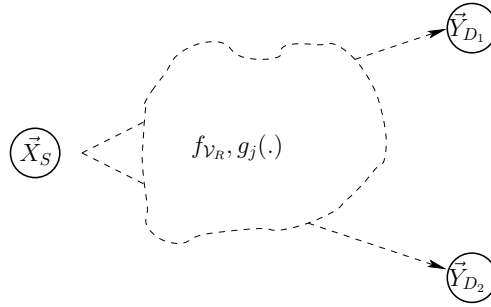


Figure 6.2: Effective end-to-end deterministic broadcast channel created by an inner code.

The capacity of the deterministic broadcast channel is well known (see [30, 35]). In particular, the coding scheme described for the deterministic broadcast channel in [33], commonly referred to as the “Marton code,” can be used and is described below succinctly.

A description of the Marton code is given here for completeness: the reader is referred to [33] for further details. The random code ensemble is constructed as follows. Consider a uniform distribution over \mathcal{C}_S , which is a collection of \vec{X}_S . The channel and the relay mapping ψ_i induce the joint distribution over the random variables (\vec{X}_S, \vec{Y}_V) . Create auxiliary random variables \vec{U}_{D_i} such that $p_{\vec{X}, \vec{U}_{D_1}, \vec{U}_{D_2}, \dots, \vec{U}_{D_J}}$ is the same as $p_{\vec{X}, \vec{Y}_{D_1}, \vec{Y}_{D_2}, \dots, \vec{Y}_{D_J}}$.

The set $\mathcal{T}_\delta^{T_2}(\vec{U}_{D_i})$ of all typical $\vec{u}_{D_i}^{T_2}$ are binned into $2^{T_1 T_2 R_i}$ bins, where each bin index corresponds to a message, for $i = 1, 2, \dots, J$. For each vector

$(\vec{u}_{D_1}^{T_2}, \dots, \vec{u}_{D_J}^{T_2}) \in \mathcal{T}_\delta^{T_2}(\vec{Y}_{D_1}, \dots, \vec{Y}_{D_J})$, there exists a sequence $\vec{x}_S^{T_2}(\vec{u}_{D_1}^{T_2}, \dots, \vec{u}_{D_J}^{T_2})$, since the channel is deterministic, such that

$$(\vec{x}_S^{T_2}, \vec{u}_{D_1}^{T_2}, \dots, \vec{u}_{D_J}^{T_2}) \in \mathcal{T}_\delta^{T_2}(\vec{X}_S, \vec{Y}_{D_1}, \dots, \vec{Y}_{D_J}). \quad (6.7)$$

Encoding: To transmit the message (W_1, \dots, W_J) , the source tries to find a vector $(\vec{u}_1^{T_2}, \dots, \vec{u}_J^{T_2}) \in \mathcal{T}_\delta^{T_2}(\vec{U}_1, \dots, \vec{U}_J)$ such that $\vec{u}_i^{T_2}$ is also in the bin with index W_i . If the source can find such a vector, it transmits $\vec{x}_S^{T_2}(\vec{u}_1^{T_2}, \dots, \vec{u}_J^{T_2})$. If the source cannot find such a sequence it transmits a random sequence.

Decoding: The destination D_i finds the bin in which the received vector $\vec{y}_{D_i}^{T_2}$ falls and decodes that bin index as the transmitted message.

6.2.3 Performance analysis

First, the rate constraints for the Marton code are identified under which arbitrarily low probability of error is guaranteed provided a large enough T_2 is chosen. It is shown in [33] that this is guaranteed, provided the rate tuple (R_1, \dots, R_J) satisfies

$$\sum_{j \in \mathcal{J}} R_j < \frac{1}{T_1} H(\vec{Y}_{D_{\mathcal{J}}}) \quad \forall \mathcal{J} \subseteq \{1, \dots, J\}, \quad (6.8)$$

where $D_{\mathcal{J}} = \{D_j\}_{j \in \mathcal{J}}$.

Next, $H(\{\vec{Y}_{D_i}\}_{D_i \in L})$ is evaluated with the relaying operations that was chosen using the following lemma:

Lemma 5. *Given arbitrary $\epsilon > 0$, $\exists T_1$ s.t.,*

$$H(\vec{Y}_{D_{\mathcal{J}}}) \geq T_1 \left(\sum_{j \in \mathcal{J}} r_j - \epsilon \right), \quad \forall \mathcal{J} \subseteq \{1, \dots, J\}. \quad (6.9)$$

Proof. See Appendix D.2 for the proof. \square

Using (6.8) and Lemma 5, it can be concluded that the rate tuple (r_1, \dots, r_J) is achievable.

Since (r_1, \dots, r_J) was chosen to be any point in $\bar{C}(Q_p)$, the region $\bar{C}(Q_p)$ is achievable. This proves the theorem. \square

6.3 Gaussian broadcast network

As for the deterministic network, only a layered network is considered here. For the Gaussian network, while it is possible to do the inner code as done in the deterministic network and induce an end-to-end broadcast channel; the induced broadcast channel would be a vector non-linear non-Gaussian broadcast channel due to the complicated nature of the relay mappings. For general broadcast channels, it is unknown whether a Marton coding scheme or any other scheme achieves rates within a constant gap of the cut-set bound.

Therefore, a different approach along the lines of [7], with the discrete superposition network (DSN) approximation for the Gaussian network as a digital interface is used. The DSN is a deterministic network. The germane code for this deterministic network is constructed and then appropriately “lifted” to construct the code for the Gaussian network.

6.3.1 Unicast network: Connection between Gaussian and DSN

First, the connection between Gaussian and DSN unicast networks, which was established in [7], is revisited.

The following lemma establishes a crucial relationship between the two networks by relating the cutset bounds in the two networks.

Lemma 6. *(Theorem 3.2 in [7]) There exists a constant $k_1 = O(|\mathcal{V}| \log |\mathcal{V}|)$, such that if R is the min-cut of a Gaussian unicast network, then $R - k_1$ is the min-cut for the corresponding DSN unicast network evaluated under product form distributions.*

In [7], a coding scheme for the Gaussian network was presented, which used the corresponding DSN as a digital interface. A coding scheme for the DSN was first constructed and the coding scheme for the Gaussian network was constructed by defining an emulation function that operated on top of the DSN scheme. This strategy is revisited next.

Emulation scheme for the Gaussian unicast network:

Consider a unicast Gaussian network and its corresponding DSN unicast network. The transmitted and received symbols at node v in the DSN are denoted by \hat{x}_v and \hat{y}_v , and the transmitted and received symbols at node v in

the Gaussian network are denoted by x_v and y_v respectively. Let $\hat{\mathcal{X}}_v, \hat{\mathcal{Y}}_v$ be the output and input alphabets at node v of the DSN and $\mathcal{X}_v, \mathcal{Y}_v$ the output and input alphabets for the Gaussian network.

The coding scheme for the DSN network is comprised of the following:

1. A source codebook $f_S : [2^{TR}] \rightarrow \hat{\mathcal{X}}_S^T$, i.e., $\hat{x}_S^T = f_S(W)$
2. The relay mappings $f_v : \hat{\mathcal{Y}}_v^T \rightarrow \hat{\mathcal{X}}_v^T$, i.e., $\hat{x}_v^T = f_v(\hat{y}_v^T)$
3. The destination decoder $g_D : \hat{\mathcal{Y}}_D^T \rightarrow [2^{TR}]$, i.e., $\hat{W} = g_D(\hat{y}_D^T)$

For the Gaussian network, the DSN coding scheme can be emulated on the Gaussian network using “emulation mappings” e_v that convert the received vector in the Gaussian network to the received vector in the DSN given by

$$e_v : \mathcal{Y}_v^T \rightarrow \hat{\mathcal{Y}}_v^T. \quad (6.10)$$

The emulation mapping along with the coding scheme of the DSN comprises the coding scheme for the Gaussian network.

The probability of error for emulation is defined as the probability that the emulated vector is different from the vector in the DSN and is given by

$$\mathbb{P}\{\exists v : e_v(y_v^T(W)) \neq \hat{y}_v^T(W)\}. \quad (6.11)$$

In [7], it has been shown that there exists an emulation mapping such that the probability of error for emulation can be made arbitrarily small for rate within a constant of the cutset bound. This is stated more precisely in the following lemma.

Lemma 7. *[7] Given a zero-error coding scheme for the DSN unicast network of rate R , a pruned coding scheme of rate $R - \kappa$ (with $\kappa = \log(6|\mathcal{V}| - 1) + 11$) can be created for the DSN unicast network and an emulation scheme can be created for the Gaussian network with probability of emulation error less than ϵ , for any arbitrary $\epsilon > 0$.*

Proof. For proof, refer Theorem 3.4 in [7]. □

6.3.2 Coding scheme for the Gaussian broadcast network

Let us consider a specific rate vector (r_1, \dots, r_J) in the interior of \bar{C}^g , the cutset region for the Gaussian network. A coding scheme with rate vector a

constant away from the rate vector (r_1, \dots, r_J) is constructed as follows.

1. Consider the corresponding DSN network to the Gaussian network. Next, construct the unicast network by adding a super-destination D to both the Gaussian and the DSN network. This unicast network is constructed by adding incoming edges from each of the destinations D_i by a rate-limited *wired* link of capacity r_i .

The cutset bound of the Gaussian unicast network is equal to $r \triangleq \sum_i r_i$. The cutset bound of the DSN unicast network (under product form distributions) is given by $\tilde{r} \geq \sum_i r_i - k_1$ where $k_1 = O(|\mathcal{V}| \log(|\mathcal{V}|))$ by Lemma 6. Theorem 8 then implies that there exists a zero-error coding scheme for the DSN unicast network at rate $r - k_1$.

2. Construct a $(2^{(r-k)T_1}, T_1)$ -pruned coding scheme for this DSN unicast network at rate $r - k$, with $k = k_1 + \kappa = O(|\mathcal{V}| \log(|\mathcal{V}|))$ and $\kappa = \log(6|\mathcal{V}| - 1) + 11$, as given by Lemma 7. This scheme can be emulated on the Gaussian unicast network with an arbitrarily small error probability.
3. The relay mappings from the DSN unicast network can then be used to create a coding scheme for the DSN broadcast network as described in Section 6.2.2. This is done using the relay mapping to construct a deterministic end-to-end broadcast channel and then using the Marton code.

Recall that the coding scheme is over $T_1 T_2$ time instants, where each set of T_1 time instants is treated as a block and the vector $\vec{x}(t_2)$ denotes x over the T_1 time instants corresponding to the t_2 -th block and is denoted by $\vec{x}(t_2)$. The relay mappings are given by \hat{f}_v and operate over the blocks of T_1 time instants.

4. For the Gaussian broadcast network, the emulation mapping is then used to emulate the received vectors on the DSN and hence convert the scheme for the DSN broadcast network to a scheme for the Gaussian broadcast network.

6.3.3 Performance analysis

First, the rates that can be achieved for the DSN broadcast network are characterized. As seen in Section 6.2.3, this is given by

$$\sum_{j \in \mathcal{J}} R_j < \frac{1}{T_1} H(\vec{Y}_{D_{\mathcal{J}}}) \quad \forall \mathcal{J} \subseteq \{1, \dots, J\}. \quad (6.12)$$

Note that $\vec{Y}_{D_{\mathcal{J}}}$ is obtained by assuming a uniform distribution over the pruned codebook for the DSN unicast network. The following lemma analogous to Lemma 5 characterizes $H(\vec{Y}_{D_{\mathcal{J}}})$.

Lemma 8. *Given arbitrary $\epsilon > 0$, $\exists T_1$ s.t.,*

$$H(\vec{Y}_{D_{\mathcal{J}}}) \geq T_1 \left(\sum_{j \in \mathcal{J}} r_j - k - \epsilon \right) \quad \forall \mathcal{J} \subseteq \{1, \dots, J\}. \quad (6.13)$$

Proof. The proof of this lemma is the same as the proof of Lemma 5 with $r - k$ replacing r as the rate of the DSN unicast scheme. \square

Therefore, the rates $R_j = r_j - k$ can be achieved for the DSN broadcast network.

Lemma 7 ensures an emulation mapping with arbitrarily small emulation error probability and thus the rate vector $(r_1 - k, \dots, r_J - k)$ can be achieved for the Gaussian broadcast network. This completes the proof of Theorem 7.

APPENDIX A

APPENDIX FOR CHAPTER 3

A.1 Proof of Theorem 1

Without loss of generality it is assumed that the message with index 1 is transmitted at the source and the index corresponding to the quantized vectors at each node is $(1, 1)$. Next, the probability of error that this message is wrongly decoded at the destination is found.

Let $\mathcal{E}_{w, \{(w_v, \bar{w}_v)\}_{v \in \mathcal{V}_r}}$ denote the event that

$$\left(x_S^T(w), \{ \hat{y}_v^T(w_v, \bar{w}_v), x_v^T(w_v) \}_{v \in \mathcal{V}_r}, y_D^T \right) \in \mathcal{T}_\epsilon^T, \quad (\text{A.1})$$

where \mathcal{T}_ϵ^T is the set of all jointly typical sequences.

The required probability of error is then given by

$$\mathbb{P}(\text{error}) = \mathbb{P}(\mathcal{E}_{1, \{(1,1)\}_{v \in \mathcal{V}_r}}^c) + \mathbb{P} \left(\bigcup_{w \neq 1} \mathcal{E}_{w, \{(w_v, \bar{w}_v)\}_{v \in \mathcal{V}_r}} \right). \quad (\text{A.2})$$

From the properties of joint typicality, it can be shown that the former term goes to 0 as $T \rightarrow \infty$. The latter term can be simplified by decomposing the corresponding union of events into disjoint events using cut-set partitions. Consider any $\Omega \subset \mathcal{V}_r$, and $\Phi \subseteq \mathcal{V}_r \setminus \Omega$; then

$$\mathbb{P} \left(\bigcup_{w \neq 1} \mathcal{E}_{w, \{(w_v, \bar{w}_v)\}_{v \in \mathcal{V}_r}} \middle| \mathcal{E}_{1, \{(1,1)\}_{\mathcal{V}_r}} \right) = \sum_{\Omega, \Phi} P_{(\Omega, \Phi)}, \quad (\text{A.3})$$

where $P_{(\Omega, \Phi)}$ is the probability corresponding to the typical event $\mathcal{E}_{w, \{(w_v, \bar{w}_v)\}_{v \in \mathcal{V}_r}}$ with $w \neq 1$, $w_v \neq 1$ for only $v \in \Omega$ and $\bar{w}_v = 1$ for only $v \in \Phi$. It can be

shown that

$$\begin{aligned}
P_{(\Omega, \Phi)} &= 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} \\
&\quad \times 2^{T(H(Y_D, \hat{Y}_\Phi, \hat{Y}_{\Phi^c}, X_\Omega, X_{\Omega^c}, X_S) - H(X_\Omega, X_S) - H(Y_D, \hat{Y}_\Phi, X_{\Omega^c}) - \sum_{v \in \Phi^c} H(\hat{Y}_v))} \\
&= 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} \\
&\quad \times 2^{T(H(Y_D, \hat{Y}_\Phi, \hat{Y}_{\Phi^c} | X_\Omega, X_{\Omega^c}, X_S) - H(Y_D, \hat{Y}_\Phi | X_{\Omega^c}) - \sum_{v \in \Phi^c} H(\hat{Y}_v))} \\
&= 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} \\
&\quad \times 2^{-T(H(Y_D, \hat{Y}_\Phi | X_{\Omega^c}) - H(Y_D, \hat{Y}_\Phi | X_\Omega, X_{\Omega^c}, X_S) + \sum_{v \in \Phi^c} H(\hat{Y}_v) - H(\hat{Y}_{\Phi^c} | X_\Omega, X_{\Omega^c}, X_S))} \\
&= 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} \times 2^{-T(I(Y_D, \hat{Y}_\Phi; X_\Omega, X_S | X_{\Omega^c}) + \sum_{v \in \Phi^c} I(\hat{Y}_v; X_{\mathcal{V}_r}, X_S))}
\end{aligned}$$

The Markovian property of the random variables implies that

$$I(\hat{Y}_v; X_{\mathcal{V}_r}, X_S) = H(\hat{Y}_v) - H(\hat{Y}_v | X_{\mathcal{V}_r}, X_S) \quad (\text{A.4})$$

$$= H(\hat{Y}_v) - H(\hat{Y}_v | Y_v) + H(\hat{Y}_v | Y_v) - H(\hat{Y}_v | X_{\mathcal{V}_r}, X_S) \quad (\text{A.5})$$

$$= H(\hat{Y}_v) - H(\hat{Y}_v | Y_v) + H(\hat{Y}_v | Y_v, X_{\mathcal{V}_r}, X_S) - H(\hat{Y}_v | X_{\mathcal{V}_r}, X_S) \quad (\text{A.6})$$

$$= I(\hat{Y}_v; Y_v) - I(\hat{Y}_v; Y_v | X_{\mathcal{V}_r}, X_S). \quad (\text{A.7})$$

Using the above and using (3.6) leads to

$$P_{(\Omega, \Phi)} = 2^{T(R-r(\Omega^c \setminus \Phi) - I(Y_D, \hat{Y}_\Phi; X_\Omega, X_S | X_{\Omega^c}) + I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}_r}, X_S))}. \quad (\text{A.8})$$

Therefore $P_{(\Omega, \Phi)} \rightarrow 0$, if

$$R < r(\Omega^c \setminus \Phi) + I(Y_D, \hat{Y}_\Phi; X_\Omega, X_S | X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}_r}, X_S). \quad (\text{A.9})$$

APPENDIX B

APPENDIX FOR CHAPTER 4

B.1 Proof of Theorem 5

The theorem will be proved in a slightly general setting, allowing multiple nodes in layer \mathcal{O}_1 and layer \mathcal{O}_L . Assuming that the flow values for these layers \mathcal{O}_1 and \mathcal{O}_L are given and satisfy

$$f(\mathcal{O}_1) = f(\mathcal{O}_L), \quad (\text{B.1})$$

$$f(\Omega_1) - f(\Omega_L) \leq C(\Omega), \quad \forall \Omega \subseteq \mathcal{V}, \quad (\text{B.2})$$

the flow for all intermediate layers will be constructed.

The proof is by inductive construction.

For $L=2$, there are no intermediate layers and the theorem holds by definition. Consider $L > 2$. The induction hypothesis assumes that the flow can be constructed with fewer than L layers and the flow for the boundary layers are specified with the constraints given by (B.2)

Consider any $L_0 \in \{2, \dots, L-1\}$. Define networks \mathcal{N}_A and \mathcal{N}_B to be the sub-networks of \mathcal{N} with the set of vertices $\mathcal{V}_A = \cup_{l=1}^{L_0} \mathcal{O}_l$ and $\mathcal{V}_B = \cup_{l=L_0}^L \mathcal{O}_l$ respectively. Similarly, denote the cut for the two networks by C_A and C_B respectively.

Next, a flow for the layer \mathcal{O}_{L_0} will be constructed which satisfies the following conditions.

$$f(\mathcal{O}_{L_0}) = f(\mathcal{O}_1), \quad (\text{B.3})$$

$$f(\Omega_A \cap \mathcal{O}_1) - f(\Omega_A \cap \mathcal{O}_{L_0}) \leq C_A(\Omega_A), \quad \forall \Omega_A \subseteq \mathcal{V}_A, \text{ and} \quad (\text{B.4})$$

$$f(\Omega_B \cap \mathcal{O}_{L_0}) - f(\Omega_B \cap \mathcal{O}_L) \leq C_B(\Omega_B), \quad \forall \Omega_B \subseteq \mathcal{V}_B. \quad (\text{B.5})$$

The induction hypothesis would then guarantee that the flows for the inter-

mediate layers in the sub-networks \mathcal{N}_A and \mathcal{N}_B can be constructed.

The set of linear inequalities given by (B.4) and (B.5) can be rewritten as

$$f(T) \leq r_A(T) \triangleq \min \{C_A(\Omega_A) + f(\Omega_A^c \cap \mathcal{O}_1) : \Omega_A^c \cap \mathcal{O}_{L_0} = T\}, \quad (\text{B.6})$$

$$f(T) \leq r_B(T) \triangleq \min \{C_B(\Omega_B) + f(\Omega_B \cap \mathcal{O}_L) : \Omega_B \cap \mathcal{O}_{L_0} = T\}, \quad (\text{B.7})$$

$\forall T \subseteq \mathcal{O}_{L_0}$.

The following properties for the functions $r_A(T)$ and $r_B(T)$ can be established.

Lemma 9. *The functions $r_A(T)$ and $r_B(T)$ are*

- *submodular,*
- *non-decreasing, and*
- *satisfy $r_A(\emptyset) = 0$ and $r_B(\emptyset) = 0$.*

Proof. Appendix B.2. □

Define the following polymatroids with the functions r_A and r_B .

$$P_A = \{\mathbf{x} \in \mathbb{R}_+^{m_{L_0}} : x(U) \leq r_A(U), \forall U \in \mathcal{O}_{L_0}\} \quad (\text{B.8})$$

$$P_B = \{\mathbf{x} \in \mathbb{R}_+^{m_{L_0}} : x(U) \leq r_B(U), \forall U \in \mathcal{O}_{L_0}\}, \quad (\text{B.9})$$

where $\mathbf{x} = [x(1) \dots x(m_{L_0})]$ and $x(U) \triangleq \sum_{u \in U} x(u)$. The conditions (B.3)-(B.5) are now equivalent to finding

$$[f(L_0, 1) \dots f(L_0, m_{L_0})] \in P_A \cap P_B, \quad (\text{B.10})$$

such that $f(\mathcal{O}_{L_0}) = f(\mathcal{O}_1)$. It then follows from Edmond's polymatroid intersection ([14], Corollary 46.1c) that:

$$\max \{x(\mathcal{O}_{L_0}) : \mathbf{x} \in P_A \cap P_B\} = \min_{T \subseteq \mathcal{O}_{L_0}} \{r_A(\mathcal{O}_{L_0} \setminus T) + r_B(T)\}. \quad (\text{B.11})$$

Therefore the required flow exists since

$$f(\mathcal{O}_1) \leq \min_{T \subseteq \mathcal{O}_{L_0}} \{r_A(\mathcal{O}_{L_0} \setminus T) + r_B(T)\} \quad (\text{B.12})$$

$$= \min_{\Omega \in \mathcal{V}} \{C(\Omega) + f(\mathcal{O}_1 \setminus \Omega_1) + f(\Omega_L)\}. \quad (\text{B.13})$$

Further, in Theorem 47.1 of [14] it is shown that the maximizing \mathbf{x} in (B.11) can be computed in *polynomial* time in the dimension of \mathbf{x} . Hence, the flow can also be computed in polynomial time in the number of nodes.

B.2 Proof of Lemma 9

We will prove the lemma for $r_B(T)$. The proof for $r_A(T)$ is similar.

1. Submodularity:

Let

$$\begin{aligned} r_B(T^{(1)}) &= C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L), \\ &\quad \forall \Omega_B^{(1)} \cap \mathcal{O}_{L_0} = T^{(1)} \end{aligned} \tag{B.14}$$

$$\begin{aligned} r_B(T^{(2)}) &= C_B(\Omega_B^{(2)}) + d(\Omega_B^{(2)} \cap \mathcal{O}_L), \\ &\quad \forall \Omega_B^{(2)} \cap \mathcal{O}_{L_0} = T^{(1)}. \end{aligned} \tag{B.15}$$

Since

$$(\Omega_B^{(1)} \cup \Omega_B^{(2)}) \cap \mathcal{O}_{L_0} = T^{(1)} \cup T^{(2)}, \tag{B.16}$$

$$(\Omega_B^{(1)} \cap \Omega_B^{(2)}) \cap \mathcal{O}_{L_0} = T^{(1)} \cap T^{(2)}, \tag{B.17}$$

it follows that

$$\begin{aligned} r_B(T^{(1)} \cup T^{(2)}) &\leq C_B(\Omega_B^{(1)} \cup \Omega_B^{(2)}) \\ &\quad + d((\Omega_B^{(1)} \cup \Omega_B^{(2)}) \cap \mathcal{O}_L), \end{aligned} \tag{B.18}$$

$$\begin{aligned} r_B(T^{(1)} \cap T^{(2)}) &\leq C_B(\Omega_B^{(1)} \cap \Omega_B^{(2)}) \\ &\quad + d((\Omega_B^{(1)} \cap \Omega_B^{(2)}) \cap \mathcal{O}_L). \end{aligned} \tag{B.19}$$

By definition of cut and the bi-submodularity of ρ_l , it is easy to verify that $C_B(\Omega_B)$ is submodular. And since d is an additive function, it then follows that $r_B(T)$ is submodular.

2. Non-decreasing:

Consider $T^{(1)} \subseteq T^{(2)}$. Let

$$\begin{aligned} r_B(T^{(1)}) &= C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L), \\ &\forall \Omega_B^{(1)} \cap \mathcal{O}_{L_0} = T^{(1)}. \end{aligned} \quad (\text{B.20})$$

Let $\Omega_B = \Omega_B^{(1)} \cup T^{(2)} \setminus T^{(1)} \supseteq \Omega_B^{(1)}$, so that $\Omega_B \cap \mathcal{O}_{L_0} = T^{(2)}$. By the definition of cut and the non-decreasing property of ρ_l , it follows that $C_B(\Omega_B^{(1)}) \leq C_B(\Omega_B)$. Also $d(\Omega_B^{(1)} \cap \mathcal{O}_L) \leq d(\Omega_B \cap \mathcal{O}_L)$. Therefore

$$r_B(T^{(2)}) = C_B(\Omega_B) + d(\Omega_B \cap \mathcal{O}_L) \quad (\text{B.21})$$

$$\geq C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L) \quad (\text{B.22})$$

$$= r_B(T^{(1)}). \quad (\text{B.23})$$

3. $r_B(\emptyset) = 0$:

When $T = \emptyset$, by letting $\Omega_B = \emptyset$, it follows that $r_B(\emptyset) = 0$.

B.3 Proof of Proposition 1

To prove the lemma, it needs to be shown that $I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U})$ satisfies the three properties of bisubmodular capacity functions.

- $I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U})$ is bi-submodular.

$$I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U}) = H(\hat{Y}_V | X_{\mathcal{O}_l \setminus U}) - H(\hat{Y}_V | X_{\mathcal{O}_l}) \quad (\text{B.24})$$

$$\begin{aligned} &= H(\hat{Y}_V, X_{\mathcal{O}_l \setminus U}) - H(X_{\mathcal{O}_l \setminus U}) \\ &\quad - H(\hat{Y}_V | X_{\mathcal{O}_l}). \end{aligned} \quad (\text{B.25})$$

The submodularity of entropy [36] implies that $H(\hat{Y}_V, X_{\mathcal{O}_l \setminus U})$ is bi-submodular. The submodularity of entropy follows from the fact that given collection of random variables Υ_1 and Υ_2 , we have

$$\begin{aligned} &H(\Upsilon_1) + H(\Upsilon_2) - H(\Upsilon_1 \cup \Upsilon_2) - H(\Upsilon_1 \cap \Upsilon_2) \\ &= I(\Upsilon_1 \setminus \Upsilon_2; \Upsilon_2 \setminus \Upsilon_1 | \Upsilon_1 \cap \Upsilon_2) \\ &\geq 0. \end{aligned}$$

The product form of the random variables implies that $H(X_{\mathcal{O}_l \setminus U})$ and $H(\hat{Y}_V | X_{\mathcal{O}_l})$ are modular or additive. Therefore, $I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U})$ is bi-submodular.

- $I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U})$ is non-decreasing. Given $U_1 \subseteq U \subseteq \mathcal{O}_l$ and $V_1 \subseteq V \subseteq \mathcal{O}_{l+1}$,

$$I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \setminus U}) = H(X_U | X_{\mathcal{O}_l \setminus U}) - H(X_U | X_{\mathcal{O}_l \setminus U} \hat{Y}_V) \quad (\text{B.26})$$

$$\geq H(X_U | X_{\mathcal{O}_l \setminus U}) - H(X_U | X_{\mathcal{O}_l \setminus U} \hat{Y}_{V_1}) \quad (\text{B.27})$$

$$= I(X_U; \hat{Y}_{V_1} | X_{\mathcal{O}_l \setminus U}) \quad (\text{B.28})$$

$$= H(\hat{Y}_{V_1} | X_{\mathcal{O}_l \setminus U}) - H(\hat{Y}_{V_1} | X_{\mathcal{O}_l}) \quad (\text{B.29})$$

$$\geq H(\hat{Y}_{V_1} | X_{\mathcal{O}_l \setminus U_1}) - H(\hat{Y}_{V_1} | X_{\mathcal{O}_l}) \quad (\text{B.30})$$

$$= I(X_{U_1}; \hat{Y}_{V_1} | X_{\mathcal{O}_l \setminus U_1}), \quad (\text{B.31})$$

where both the inequalities follow from the fact that conditioning reduces entropy.

- $I(\emptyset; \hat{Y}_V | X_{\mathcal{O}_l}) = I(X_U; \emptyset | X_{\mathcal{O}_l \setminus U}) = 0$ follows trivially.

APPENDIX C

APPENDIX FOR CHAPTER 5

C.1 Layering for linear deterministic network

We show that a coding scheme (not necessarily linear) over T time instants for any linear deterministic network can be equivalently represented as a layered coding scheme over an unfolded $T + 1$ layered network. We unfold the network to $T + 1$ stages such that the m^{th} -stage is representing what happens in the network during the m^{th} time duration. Every node ν in the unlayered network appears at stage $1 \leq m \leq T + 1$ in the unfolded network as $\nu[m]$. If there is an edge connecting node ν_i to node ν_j with channel gain matrix $G_{ij} \in \mathbb{F}_p^{q \times q}$ in the unlayered network, then there is an edge connecting the nodes $\nu_i[m]$ to node $\nu_j[m + 1]$ in the layered network with channel gain matrix given by $\hat{G}_{ij} \in \mathbb{F}_p^{q(T+2) \times q(T+2)}$, where

$$\hat{G}_{ij} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ G_{ij} & 0 & 0 \end{bmatrix}. \quad (\text{C.1})$$

Note that if G_{ij} is a shift matrix, then so is \hat{G}_{ij} . Further, every node $\nu[m]$ is connected to its next instance $\nu[m + 1]$ by a link with channel gain $I_{q(T+2) \times q(T+2)}$. Figure C.1 illustrates a simple example of a network and the corresponding layered network.

We first show that any coding scheme for the unlayered network can be used to construct a layered coding scheme for the layered network. If $x_{\nu_j}[m] \in \mathbb{F}_p^q$ is the vector transmitted by the node ν_j in the unlayered network, then the vector transmitted by the node $\nu_j[m]$ is $\hat{x}_{\nu_j[m]} \in \mathbb{F}_p^{q(T+2)}$ and is given

by

$$\hat{x}_{\nu_j[m]} = \begin{pmatrix} x_{\nu_j}[m] \\ s_{\nu_j}[m-1] \\ 0_{q \times 1} \end{pmatrix}, \quad (\text{C.2})$$

where $s_{\nu_j}[m-1]$ represents the state of the node ν_j at time instant $m-1$ in the unlayered network and is the stack of all received vectors at node ν_j until that time instant. Note that the received vector at a node $\nu_j[m+1]$ in the layered network is

$$\hat{y}_{\nu_j[m+1]} = \begin{pmatrix} x_{\nu_j}[m] \\ s_{\nu_j}[m-1] \\ y_{\nu_j}[m] \end{pmatrix}. \quad (\text{C.3})$$

It is essentially all the information at node ν_j till time instant m . Thus any coding function for the node ν_j at time instant $m+1$ can be converted to a coding function for node $\nu_j[m+1]$.

Conversely, for any scheme on the layered network, the corresponding scheme on the unlayered network is given by

$$x_{\nu_j}[m] = (\hat{x}_{\nu_j[m]}[1] \dots \hat{x}_{\nu_j[m]}[q])^T. \quad (\text{C.4})$$

It is easy to see that $x_{\nu_j}[m]$ can be written as a function of the previous received vectors and the source messages, if any, at that node.

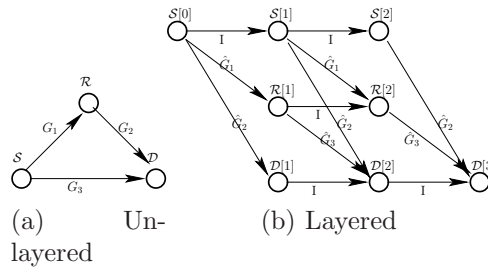


Figure C.1: Layering a network by unfolding over time.

APPENDIX D

APPENDIX FOR CHAPTER 6

D.1 Proof of Lemma 4

Consider any cut Ω such that $S \in \Omega$ and $D \in \Omega^c$. There are two components that contribute to the value of the cut: one part c_1 comes from the added rate limited links and the other part c_2 comes from the original network. Let \mathcal{J} be such that $D_{\mathcal{J}^c} \subseteq \Omega$ and $D_{\mathcal{J}} \subseteq \Omega^c$; this implies that the rate limited links of capacity $c_1 = \sum_{j \in \mathcal{J}^c} r_j$ are included in the cuts. Recall that $\bar{C}_{\mathcal{J}}(Q)$ denotes the value of the cutset bound evaluated under the distribution Q for separating the source from the set $D_{\mathcal{J}}$. As the cut Ω separates S from $D_{\mathcal{J}}$, the value of cut gained from the original network is bigger than $\bar{C}_{\mathcal{J}}(Q)$. Furthermore, since $(r_1, \dots, r_J) \in \bar{C}(Q)$, $\sum_{j \in \mathcal{J}} r_j \leq \bar{C}_{\mathcal{J}}(Q)$. This implies that the value of cut gained from the original network is bigger than this value: $c_2 \geq \sum_{j \in \mathcal{J}} r_j$. Thus the total value of the cut is $c = c_1 + c_2 \geq \sum_j r_j$. The min-cut value is actually equal to $\sum_j r_j$ since the cut that separates D_1, \dots, D_J from D has value $\sum_j r_j$.

D.2 Proof of Lemma 5

Let the min-cut between the source and the destination be r . Since the relaying scheme can achieve any rate close to the cut-set bound for large enough T_1 , the information transmitted by all the sinks should be greater

than rate r ; therefore for any subset $\mathcal{J} \subseteq \{1, 2, \dots, J\}$,

$$T_1 r \leq H(\vec{X}_{D_1}, \dots, \vec{X}_{D_J}) \quad (\text{D.1})$$

$$\leq H(\vec{X}_{D_{\mathcal{J}}}) + H(\vec{X}_{D_{\mathcal{J}^c}}) \quad (\text{D.2})$$

$$\leq H(\vec{X}_{D_{\mathcal{J}}}) + \sum_{j \in \mathcal{J}^c} H(\vec{X}_{D_j}) \quad (\text{D.3})$$

$$= H(\vec{X}_{D_{\mathcal{J}}}) + \sum_{j \in \mathcal{J}^c} H(X_{D_j}^{T_1}) \quad (\text{D.4})$$

$$\leq H(\vec{Y}_{D_{\mathcal{J}}}) + \sum_{j \in \mathcal{J}^c} \sum_{t=1}^{T_1} H(X_{D_j}(t)) \quad (\text{D.5})$$

$$\leq H(\vec{Y}_{D_{\mathcal{J}}}) + \sum_{j \in \mathcal{J}^c} T_1 r_j \quad (\text{D.6})$$

$$\Rightarrow \frac{1}{T_1} H(\vec{Y}_{D_{\mathcal{J}}}) \geq r - \sum_{j \in \mathcal{J}^c} r_j. \quad (\text{D.7})$$

Note that (D.6) follows due to the rate-limited links. Furthermore, the min-cut (under product distributions) is $r = \sum_i r_i$ by Lemma 4, and this gives

$$\frac{1}{T_1} H(\vec{Y}_{D_{\mathcal{J}}}) \geq \sum_{j \in \mathcal{J}} r_j \quad \forall \mathcal{J} \subseteq \{1, 2, \dots, J\}. \quad (\text{D.8})$$

REFERENCES

- [1] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Transactions on Information Theory*, vol. IT-46, no. 2, pp. 388–404, March 2000.
- [2] P. Gupta and P. R. Kumar, “A network information theory for wireless communication: Scaling laws and optimal operation,” *IEEE Transactions on Information Theory*, vol. IT-50, no. 5, pp. 748–767, May 2004.
- [3] R. H. Etkin, D. N. C. Tse, and H. Wang, “Gaussian interference channel capacity to within one bit,” *IEEE Transactions on Information Theory*, vol. IT-54, no. 12, pp. 5534–5562, December 2008.
- [4] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: A deterministic approach,” *IEEE Transactions on Information Theory*, vol. IT-57, no. 4, pp. 1872 – 1905, April 2011.
- [5] S. Tavildar and P. Viswanath, “Approximately universal codes over slow-fading channels,” *IEEE Transactions on Information Theory*, vol. IT-52, no. 7, pp. 3233–3258, July 2006.
- [6] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005.
- [7] M. Anand and P. R. Kumar, “A digital interface for Gaussian relay and interference networks: Lifting codes from the discrete superposition model,” *IEEE Transactions on Information Theory*, vol. IT-57, no. 5, pp. 2548 – 2564, May 2011.
- [8] G. Bresler and D. Tse, “The two-user Gaussian interference channel: a deterministic view,” *European Transactions on Telecommunications*, vol. 19, pp. 333–354, June 2008.
- [9] A. E. Gamal, “On information flow in relay networks,” in *IEEE National Telecommunications Conference*, November 1981, pp. D4.1.1–D4.1.4.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing, 2006.

- [11] A. Ozgur and S. N. Diggavi, “Approximately achieving Gaussian relay network capacity with lattice codes,” in *IEEE International Symposium on Information Theory*, Austin, TX, USA, June 2010, pp. 669 – 673.
- [12] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, “Noisy Network Coding,” 2010, submitted to *IEEE Transactions on Information Theory*. [Online]. Available: <http://arxiv.org/abs/1002.3188>
- [13] V. Nagpal, I. Wang, M. Joragovanovic, D. N. C. Tse, and B. Nikolić, “Quantize-Map-and-Forward Relaying: Coding and System Design,” in *Allerton Conference on Communications, Control, and Computing*, September 2010, pp. 443 – 450.
- [14] A. Schrijver, *Combinatorial Optimization*. Springer, 2003.
- [15] L. R. Ford and D. R. Fulkerson, “Maximal flow through a network,” *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.
- [16] E. L. Lawler and C. U. Martel, “Computing maximal polymatroidal network flows,” *Mathematics of Operations Research*, vol. 7, August 1982.
- [17] U. Feige, M. Hajiaghayi, and J. Lee, “Improved approximation algorithms for minimum-weight vertex separators,” *SIAM Journal on Computing*, vol. 38, pp. 334–347, 2008.
- [18] S. M. S. T. Yazdi and S. A. Savari, “A max-flow/min-cut algorithm for linear deterministic relay networks,” *IEEE Transactions on Information Theory*, vol. IT-57, no. 5, pp. 3005 – 3015, May 2011.
- [19] M. X. Goemans, S. Iwata, and R. Zenklusen, “A flow model based on polylinking system,” *Mathematical Programming, Series A*, pp. 1–23, March 2011.
- [20] S. Vishwanath, N. Jindal, and A. J. Goldsmith, “Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channel,” *IEEE Transactions on Information Theory*, vol. IT-49, no. 10, pp. 2658–2668, 2003.
- [21] P. Viswanath and D. Tse, “Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality,” *IEEE Transactions on Information Theory*, vol. IT-49, no. 8, pp. 1912 – 1921, August 2003.
- [22] V. Prabhakaran and P. Viswanath, “Interference channel with source cooperation,” *IEEE Transactions on Information Theory*, vol. IT-57, no. 1, pp. 156–186, January 2011.

- [23] V. Prabhakaran and P. Viswanath, “Interference channel with destination cooperation,” *IEEE Transactions on Information Theory*, vol. IT-57, no. 1, pp. 187–209, January 2011.
- [24] G. Bresler, A. Parekh, and D. N. C. Tse, “The approximate capacity of the many-to-one and one-to-many Gaussian interference channels,” *IEEE Transactions on Information Theory*, vol. IT-56, no. 9, pp. 4566–4592, September 2010.
- [25] S. Riis, “Reversible and irreversible information networks,” *IEEE Transactions on Information Theory*, vol. IT-53, no. 11, pp. 4339–4349, November 2007.
- [26] R. Dougherty, K. Zeger, and C. Freiling, “Nonreversibility and equivalent constructions of multiple-unicast networks,” *IEEE Transactions on Information Theory*, vol. IT-52, no. 11, pp. 5067–5077, November 2008.
- [27] R. Koetter and M. Médard, “An algebraic approach to network coding,” *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, October 2003.
- [28] S. R. Li, R. W. Yeung, and N. Cai, “Linear network coding,” *IEEE Transactions on Information Theory*, vol. IT-49, no. 2, pp. 371–381, February 2003.
- [29] H. Weingarten, Y. Steinberg, and S. Shamai, “The capacity region of the Gaussian multiple-input multiple-output broadcast channel,” *IEEE Transactions on Information Theory*, vol. IT-52, no. 9, pp. 3936–3964, September 2006.
- [30] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” *IEEE Transactions on Information Theory*, vol. IT-25, no. 3, pp. 306–311, 1979.
- [31] Y. Liang and V. V. Veeravalli, “Cooperative relay broadcast channels,” *IEEE Transactions on Information Theory*, vol. IT-53, no. 3, pp. 900–928, March 2007.
- [32] D. Vasudevan and S. B. Korada, “Polymatroidal flows on two classes of information networks,” *IEEE Transactions on Information Theory*, vol. IT-57, no. 1, pp. 227–233, January 2011.
- [33] A. E. Gamal and E. V. der Meulen, “A proof of Marton’s coding theorem for the discrete memoryless broadcast channel,” *IEEE Transactions on Information Theory*, vol. IT-27, no. 1, pp. 120–122, January 1981.
- [34] M. S. Pinsker, “Capacity of noiseless broadcast channels,” *Problems of Information Transmission*, vol. 14, no. 2, pp. 97–102, 1978.

- [35] M. S. Pinsker, “Capacity of noiseless broadcast channels,” *Probl. Pered. Inform.*, vol. 14, no. 2, pp. 28–34, April-June 1978.
- [36] A. K. Kelmans and B. N. Kimelfeld, “Multiplicative submodularity of a matrix’s principal minor as a function of the set of its rows and some combinatorial applications,” *Discrete Mathematics*, vol. 44, no. 1, pp. 113–116, 1980.