A study of some problems in network information theory

Sudeep Kamath

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2013-148
http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-148.html

August 15, 2013

A study of some problems in network information theory

by

Sudeep Uday Kamath

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

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Tse, Co-chair
Professor Venkat Anantharam, Co-chair
Professor David Aldous

Fall 2013

A study of some problems in network information theory

Abstract

A study of some problems in network information theory

by

Sudeep Uday Kamath

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor David Tse, Co-chair

Professor Venkat Anantharam, Co-chair

Shannon theory has been very successful in studying fundamental limits of communication in the classical setting, where one sender wishes to communicate a message to one receiver over an unreliable medium. The theory has also been successful in studying networks of small to moderate sizes, with multiple senders and multiple receivers. However, it has become well-known recently that understanding the fundamental limits of communication in a general network is a hard problem on numerous accounts.

In this dissertation, we suggest that a significant aspect of the difficulty in studying limits of communication over networks lies in the unidirectional aspect of the problem. Under different assumptions that rid the problem of this particular aspect by introducing a suitable symmetry, either in the underlying network or in the traffic model, we find that simple schemes are approximately optimal in achieving these fundamental limits. We demonstrate this as a meta-theorem in the class of wireline networks and Gaussian networks. The key contribution driving these results is a new outer bound that we call the Generalized Network Sharing bound.

We also study a problem of simulation of joint distributions in a non-interactive setup. Two agents observe correlated random variables and wish to simulate a certain joint distribution. We consider a non-asymptotic formulation of this problem and study tools that help prove impossibility results. We also study connections of this problem to existing problems in the literature.

To my family for all their love and support

Contents

Contents **ii**

List of Figures **iv**

1 Introduction **1**

2 The Generalized Network Sharing bound **4**
2.1 Preliminaries . . . 5
2.2 Equivalence to the functional dependence bound . . . 10
2.3 Edge-cut bounds based only on source-destination connectivity . . . 11
2.4 NP-completeness of minimum GNS-cut . . . 15
2.5 Discussion . . . 16

3 An information-theoretic meta-theorem for wireline networks **18**
3.1 Preliminaries . . . 19
3.2 k-unicast undirected networks . . . 20
3.3 k-pair unicast directed symmetric-demand networks . . . 21
3.4 k-groupcast directed networks: Sum-rate . . . 27
3.5 Discussion . . . 28

4 Wireless Networks: Network Capacity under Traffic Symmetry **29**
4.1 Gaussian networks composed of Broadcast and Multiple Access Channels . . 30
4.2 General Gaussian Networks . . . 41

5 Non-Interactive Simulation of Joint Distributions **48**
5.1 Problem formulation . . . 51
5.2 Maximal Correlation and the Hypercontractivity Ribbon . . . 53
5.3 Main Results . . . 58
5.4 Proofs . . . 59
5.5 Non-interactive simulation of a three random variable joint distribution . . . 63

6 Discussion and concluding remarks **65**

Bibliography **66**

List of Figures

2.1 A butterfly network 8
2.2 Profiles of edge-cuts for 2-unicast networks 12
2.3 Two examples of profiles and their corresponding canonical networks 13
2.4 Edge to Gadget 16
2.5 Bat network 17

3.1 Connectivity graph $\mathcal{G}_c$ 24
3.2 Reduced connectivity graph $\mathcal{G}_r$ 24
3.3 Connectivity pattern 26
3.4 Example of a groupcast network 28

4.1 Example of a MAC+BC Gaussian network 31
4.2 GNS bound for MAC+BC Gaussian networks 34

5.1 Gohari-Anantharam formulation 49
5.2 A generalization of the Gács-Körner and Wyner formulations 50
5.3 Non-Interactive Simulation of Joint Distributions 50
5.4 Hypercontractivity ribbon 55
5.5 Three random variable simulation example 63

Acknowledgments

It has been an incredible experience being a student of David Tse and Venkat Anantharam for the last five years. There was never a time that I did not feel cared for, thanks to their constant support and guidance.

Working with David has helped me appreciate the necessity of being a good scholar, the importance of strategy while devising research goals, and the need to constantly ask oneself bold and important questions. Having served as a teaching assistant in a course taught by him, I have been greatly inspired by his teaching philosophy and lucid presentation style. His emphasis on simplicity and intuition, in teaching as well as in research, will continue to influence me in the years ahead.

Working with Venkat has helped me understand the importance of rigor and clarity in understanding. His deep knowledge of many disciplines and crystal-clear intuitions for complex ideas are things I am only capable of aspiring for, but which I hope to inculcate into my research style to the extent I can. I am inspired by his tenacity to focus on fundamental problems, and hope to have a similar theme in my future research plans.

I am very grateful to Satish Rao for many helpful discussions that helped shape some of the results in this dissertation. I am also grateful to David Aldous for being present on my dissertation committee. Searching for answers to questions posed by him has been a very helpful endeavour.

I would like to thank Pramod Viswanath and Sreeram Kannan for fruitful collaboration over the past few years. A large section of this dissertation is comprised of results from this collaboration. I would also like to thank Pramod for useful advice and guidance throughout graduate school.

I would like to thank Alcatel-Lucent Bell Labs for hosting me as an intern during the summer of 2010, and to Urs Niesen and Piyush Gupta for their mentorship during my stay.

I would like to thank Amin Gohari and Chandra Nair for some exciting research directions that we have been exploring recently. Their brilliant ideas and unrelenting passion have been very inspiring.

Thanks to the administrative staff at EECS Berkeley, especially to Ruth Gjerde, Kim Kail, and Shirley Salanio, for their patience with my countless requests and administrative questions.

Wireless Foundations has been an awesome place. Special thanks to Pulkit Grover for all the help and advice throughout graduate life, and to Sreeram Kannan for his infectious enthusiasm in our numerous research collaborations. I would like to thank all my WiFo mates for making my time here very enjoyable: Guy Bresler, Venkatesan Ekambaram, Naveen Goela, Amin Gohari, Kate Harrison, Varun Jog, Rashmi K.V., Po-Ling Loh, Nebojsa Milosavljevic, Nima Noorshams, Barlas Oğuz, Se Yong Park, Sameer Pawar, Gireeja Ranade, Nihar Shah, Changho Suh, I-Hsiang Wang, Kristen Woyach, and Baosen Zhang.

Life at Berkeley has been always fun thanks to the wonderful friends around here: Mangesh Bangar, Onkar Dalal, Rutooj Deshpande, Sukhada Fadnavis, Kartik Ganapati, Pavan Hosur, Kaushik Jayaram, Sudeep Juvekar, Yashodhan Kanoria, Subhransu Maji, Aditya Medury, Dhawal Mujumdar, Debanjan Mukherjee, Sharanya Prasad, Arunanshu Roy, Pranav Shah, Shaama Mallikarjun Sharada, and Anuj Tewari.

Thanks to my long-distance friends who have been close even when they were far: Shantanu Gangal, Sonal Jain, Vignesh Kasinath, Robin Kothari, Ritu Priyanka Marpu, Kartik Mohta, Vamsi Pingali, Sangram Raje, Sushant Sachdeva, Karthik Shekhar, Shriharsh Tendulkar, Srikant Veeraraghavan, and Anisha Zaveri.

Lastly, I would like to thank my family - my late father Uday Kamath (who remains alive in our memories), my mother Urmila Kamath, my brother Pritish Kamath, my aunt Naina Kamath (whom we all call Atee), and my grandmom Mukta Kamath (lovingly called Ammu). Their endless love and support throughout the years has meant more to me than words can express. I would like to dedicate my dissertation to them.

Chapter 1

Introduction

Information theory concerns the study of the problem of reliable communication between nodes under the presence of uncertainty. The classical setting with one sender and one receiver is called the point-to-point setup, and fundamental limits of communication for this setting were obtained by Shannon in his seminal work [55].

Network information theory considers the problem of communication between multiple senders and receivers. Characterizing the fundamental limits of communication for a general network, called the capacity region, has come to be seen as a very hard problem. One way to simplify the problem in order to obtain insight into general networks is to study the capacity regions of the simplest class of networks - wireline networks - where links between nodes of the network are noise-free and orthogonal. It was observed first in [3] that if nodes of the network perform coding operations rather than simply route information treating it as a commodity, then communication rates can be significantly improved. This setting with nodes performing coding operations has been termed *network coding* and shows promise in implementation of the networks of tomorrow.

However, recent results suggest that characterizing the capacity region of a general network is a hard problem in multiple ways:

- Certain so-called Non-Shannon inequalities (which are themselves not completely understood [63, 47]) are important for characterizing capacity [20], [10].
- A simple class of schemes called linear coding schemes does not achieve capacity in general [18].

One of the central contributions of this dissertation is the following:

Routing is approximately optimal for achieving capacity in wireline networks, provided that the network or traffic has a suitable symmetry.

This statement highlights that the difficulty of studying capacity regions of networks lies, generally speaking, in the lack of symmetry in the problem. The above theme turns out to have more generality. A similar statement can be worked out in the case of wireless networks as well. Wireless networks are more complicated than wireline networks due to the presence of noise as well as the aspects of broadcast, superposition and interference. Yet, the general theme holds and we have:

> *A simple separation based scheme is approximately optimal for achieving capacity in Gaussian networks, provided that the network or traffic has a suitable symmetry.*

The key tool that we use to obtain our results is a simple but novel outer bound on the capacity regions of networks called the Generalized Network Sharing (GNS) bound that we developed in an earlier work [32, 28].

In the latter part of the dissertation, we study a problem of simulation of joint distributions in a non-interactive setup. We use the tools of maximal correlation and hypercontractivity for studying this problem and our key contribution is a connection between the two tools.

A rough outline of this dissertation is as follows:

- Chapter 2 introduces the Generalized Network Sharing bound and discusses some of its properties.

- Chapter 3 obtains the approximation results for capacity regions of wireline networks under various symmetry assumptions.

  Here, we consider the problem of communication between k sources with their respective destinations in a wireline network. As we mentioned earlier, network coding is known to have significant advantages for a general network. However, when the network or traffic has some kind of symmetry, we show that the potential advantages of network coding are small. We demonstrate this in the case of undirected networks, networks with bidirectional traffic and symmetric demands (for every source communicating to a destination at a certain rate, the destination communicates an independent message back to the source at the same rate), and *groupcast* networks (networks with a special group of nodes, each of which has an independent message for every other node in the group).

- Chapter 4 obtains these approximation results for wireless networks under one specific symmetry assumption on the traffic: bidirectional traffic with symmetric demands.

  It has been shown recently that a simple layering principle - local physical-layer schemes combined with global routing - can achieve approximately optimal performance in wireless networks [35]. However, this result depends heavily on the assumption of reciprocity of wireless networks, which may be violated in general networks, due to asymmetric power constraints, directional antennas, or frequency-duplexing. We show that the approximate optimality continues to hold even for wireless networks modeled as directed graphs as long as there is a symmetric demand constraint: for every demand from a source to a destination at a particular rate, there runs a counterpart demand from the destination to the source at the same rate. This models several practical scenarios including voice calls, video calls, and interactive gaming. We prove this result in the context of several channel models for which good local schemes exist.

- Chapter 5 takes a look at a problem of simulation of joint distributions.

  We consider the following problem: Agents Alice and Bob observe different random variables occuring in nature which are correlated. They are required to output one random variable each, based on their observations, such that the pair of output random variables has a specified joint distribution. This problem has important applications to existing problems in simulation of joint distributions. The emphasis here is on the non-asymptotic nature of the problem: The number of observations of the agents is allowed to be arbitrarily large even though the agents are required to output just one random variable each. We use the tools of maximal correlation and hypercontractivity to obtain negative results for this problem. We hope that these tools will be useful more generally to attack other problems in network information theory.

- Chapter 6 contains a discussion and some concluding remarks.

Publications in which some of this work has appeared

The results in this dissertation have appeared in various conferences. The main results of Chapter 2 can be found in [31]. The results of Chapters 3 and 4 can be found in [33] and [30] respectively. Most of the results of Chapter 5 can be found in [29], while a few others have not appeared in print, but were developed in the course of finalizing this dissertation [5].

Chapter 2

The Generalized Network Sharing bound

The central problem of network information theory is to characterize the capacity region of a general network. Wireline networks are a special class of such networks where the edges between vertices are *unidirectional*, *orthogonal* and *noise-free*. In this class of networks, network coding has the potential to provide significant advantages in comparison to flow (i.e. routing strategies) for multicast problems [3] as well as for multiple unicast problems [26]. Recent results suggest that characterizing the capacity region of a multiple unicast network is a hard problem [19], [18], [10]. In particular, even coding strategies such as linear codes do not achieve capacity in general [18].

Nonetheless, it is useful to develop outer bounds on the capacity region as they can then provide useful guarantees on the performance of any suggested coding schemes. The simplest and oldest such bound is the cutset bound [21], [14]. This bound however, is often quite loose. The tightest known explicitly computable bound is the so-called LP bound [62] that harnesses the full power of the basic information inequalities or so-called Shannon inequalities. It is however, computationally intractable since the linear program has size exponential in that of the network. The Generalized Network Sharing (GNS) bound was introduced in our previous work [32, 28] as a new bound that is an improvement over the cutset bound. In this chapter, we study more properties of this bound for wireline networks.

The literature has numerous outer bounds derived from the graph-theoretic structure of the network. These *edge-cut* bounds have conventionally served as outer bounds to commodity flow problems. Indeed, such commodity flow bounds derived from edge-cuts are not in general, *fundamental*, i.e. they are not bounds on the capacity region and can potentially be beaten by network coding [39]. It is of interest to study these edge-cut bounds because they tend to be simpler and more intuitive than the LP bound, while also being tighter than the cutset bound. Different works have studied what makes edge-cut bounds fundamental. [60] proposed the Network Sharing bound which was subsequently improved to the Generalized Network Sharing (GNS) bound in [32]. [39], [56] study bounds derived from functional dependence graphs and [26] studies bounds derived from information dominance. Recently, there has also been some progress in studying *weighted* sum-rate edge-cut bounds. [40] exhibits such bounds for multimessage multicast problems while [59] produces bounds for multiple unicast problems based on a class of Shannon inequalities and a knowledge of the network graph structure.

In this chapter, we focus on sum-rate edge-cut bounds for multiple-unicast networks. We show that for multiple unicast networks, the GNS bound is equivalent to the more sophisticated functional dependence bound derived from functional dependence graphs [56]. Next, we study edge-cut bounds on network coding capacity based purely on what we define as the 'profile' of the edge-cut, which is simply the knowledge of the residual connectivity between sources and destinations after a set of edges has been removed from the graph. We show that the only edge-cut profiles for which every edge-cut with the said profile always leads to a fundamental bound on network coding rates are the profiles of GNS-cuts. Furthermore, we provide the tight constant associated with every edge-cut profile up to which network coding may potentially outperform the associated edge-cut. Finally, we consider the problem of computation of the GNS bound. We show that this problem is NP-complete, even for two-unicast networks, and discuss the implications of this result.

The rest of the chapter is organized as follows. We describe notation and preliminaries in Section 2.1. We show the equivalence between the GNS bound and the functional dependence bound of [56] in Section 2.2. We study bounds from edge-cuts based purely on source-destination connectivity in Section 2.3. We prove NP-completeness of the GNS-cut in Section 2.4. Finally, we conclude with a discussion in Section 2.5.

In Chapter 3 and Chapter 4, we will use the GNS bound to obtain approximate capacity characterizations for wireline and wireless networks under suitable symmetry assumptions on the network or traffic.

2.1 Preliminaries

We briefly describe some notation that will be used throughout this chapter and the next.

Definition: A k-unicast wireline network $\mathcal{N}$ for source-destination pairs $\{(s_i; d_i)\}_{i\in\mathcal{I}}$ with $|\mathcal{I}| = k$ is a tuple $(\mathcal{G}, \underline{\mathrm{C}})$ where

- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the underlying directed or undirected graph with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, with $s_i, d_i \in \mathcal{V}(\mathcal{G})$ for $i \in \mathcal{I}$,
- $\underline{\mathrm{C}} = (C_e : e \in \mathcal{E}(\mathcal{G}))$ is the edge-capacity vector, with $C_e \in \mathbb{R}_{\geq 0} \cup \{\infty\}\ \forall e \in \mathcal{E}(\mathcal{G})$.

For each $i \in \mathcal{I}$, s_i has independent information to be communicated to d_i at rate R_i.

Unless otherwise stated, it will be assumed that $\mathcal{I} = \{1, 2, \ldots, k\}$.

Notation: For directed graphs with $v \in \mathcal{V}(\mathcal{G})$, let $\mathsf{In}(v)$ and $\mathsf{Out}(v)$ denote the edges entering into and leaving v respectively. For undirected graphs with $v \in \mathcal{V}(\mathcal{G})$, we denote the set of edges incident onto v as $\mathsf{Inc}(v)$.

Definition: Given a k-unicast network $\mathcal{N} = (\mathcal{G}, \underline{C})$ for source-destination pairs $\{(s_i; d_i)\}_{i\in\mathcal{I}}$, we say that the non-negative rate tuple $(R_i : i \in \mathcal{I})$ is *achievable* if for any $\epsilon > 0$, there exist positive integers N and T (called block length and number of epochs respectively), a finite alphabet $\mathcal{A}$ with $|\mathcal{A}| \geq 2$ and using notation $H_v := \Pi_{i\in\mathcal{I}:v=s_i}\mathcal{A}^{\lceil NTR_i \rceil}$ (with an empty product being the singleton set),

- for the case of directed graphs,
  - encoding functions for $1 \leq t \leq T, e = (u, v) \in \mathcal{E}$,
    $f_{e,t} : H_u \times \Pi_{e'\in\mathsf{In}(u)} \left(\mathcal{A}^{\lfloor NC_{e'} \rfloor}\right)^{(t-1)} \mapsto \mathcal{A}^{\lfloor NC_e \rfloor}$,
  - decoding functions at destinations d_i for $i \in \mathcal{I}$,
    $f_{d_i} : H_{d_i} \times \Pi_{e'\in\mathsf{In}(d_i)} \left(\mathcal{A}^{\lfloor NC_{e'} \rfloor}\right)^{T} \mapsto \mathcal{A}^{\lceil NTR_i \rceil}$
- and for the case of undirected graphs,
  - a *subdivision of capacity* for each edge $e = (u, v) \in \mathcal{E}$ and each epoch t, $1 \leq t \leq T$, $C^u_{e,t}, C^v_{e,t} \geq 0$ such that $C^u_{e,t} + C^v_{e,t} \leq C_e$, where $C^u_{e,t}$ is the capacity for outgoing data from u on edge e,
  - for $1 \leq t \leq T, e = (u, v) \in \mathcal{E}$, two encoding functions $f^u_{e,t}, f^v_{e,t}$ as
    $f^u_{e,t} : H_u \times \Pi_{e'=(u,w)\in\mathsf{Inc}(u)}\Pi_{l=1}^{t-1}\mathcal{A}^{\lfloor NC^w_{e',l} \rfloor} \mapsto \mathcal{A}^{\lfloor NC^u_{e,t} \rfloor}$,
    $f^v_{e,t} : H_v \times \Pi_{e'=(v,w)\in\mathsf{Inc}(v)}\Pi_{l=1}^{t-1}\mathcal{A}^{\lfloor NC^w_{e',l} \rfloor} \mapsto \mathcal{A}^{\lfloor NC^v_{e,t} \rfloor}$,
  - decoding functions at destinations d_i for $i \in \mathcal{I}$,
    $f_{d_i} : H_{d_i} \times \Pi_{e'=(d_i,w)\in\mathsf{Inc}(d_i)}\Pi_{l=1}^{T}\mathcal{A}^{\lfloor NC^w_{e',l} \rfloor} \mapsto \mathcal{A}^{\lceil NTR_i \rceil}$

with the property that under the uniform probability distribution on $\Pi_{i\in\mathcal{I}}\mathcal{A}^{\lceil NTR_i \rceil}$,

$$\mathsf{Pr}\left(g(m_1, m_2, \ldots, m_k) \neq (m_1, m_2, \ldots, m_k)\right) \leq \epsilon,$$

where $g : \Pi_{i\in\mathcal{I}}\mathcal{A}^{\lceil NTR_i \rceil} \mapsto \Pi_{i\in\mathcal{I}}\mathcal{A}^{\lceil NTR_i \rceil}$ is the global decoding function induced inductively by

- $\{f_{e,t} : e \in \mathcal{E}(\mathcal{G}), 1 \leq t \leq T\}$ and $\{f_{d_i} : i \in \mathcal{I}\}$ in the directed graph case and
- $\{f^u_{e,t}, f^v_{e,t} : e = (u, v) \in \mathcal{E}(\mathcal{G}), 1 \leq t \leq T\}$ and $\{f_{d_i} : i \in \mathcal{I}\}$ in the undirected graph case.

The closure of the set of achievable rate tuples is called the *capacity region* and is denoted by $\mathcal{C}$. Define the *sum-rate-capacity* by $C_{\text{sum-rate}} := \sup_{(R_i:i\in\mathcal{I})\in\mathcal{C}} \sum_{i\in\mathcal{I}} R_i$.

Definition: Given a k-unicast network $\mathcal{N} = (\mathcal{G}, \underline{C})$ for source-destination pairs $\{(s_i; d_i)\}_{i\in\mathcal{I}}$, we say that the non-negative rate tuple $(R_i : i \in \mathcal{I})$ is *achievable by routing flow* if there exist for each $i \in \mathcal{I}$ and each $e = (u, v) \in \mathcal{E}(\mathcal{G})$, real numbers $f_{i,e} \geq 0$ in the directed graph case and $f^u_{i,e}, f^v_{i,e} \geq 0$ in the undirected graph case such that

- $\sum_{i \in \mathcal{I}} f_{i,e} \leq C_e \; \forall \; e \in \mathcal{E}(\mathcal{G})$, and for each $i \in \mathcal{I}$ and each $v \in \mathcal{V}(\mathcal{G})$,

$$\sum_{e \in \mathsf{Out}(v)} f_{i,e} - \sum_{e \in \mathsf{In}(v)} f_{i,e} = \begin{cases} 0 & \text{if } v \neq s_i, d_i, \\ R_i & \text{if } v = s_i, \\ -R_i & \text{if } v = d_i. \end{cases}$$

  in the directed graph case and

- $\sum_{i \in \mathcal{I}} f^u_{i,e} + f^v_{i,e} \leq C_e \; \forall \; e = (u,v) \in \mathcal{E}(\mathcal{G})$, and for each $i \in \mathcal{I}$ and each $v \in \mathcal{V}(\mathcal{G})$,

$$\sum_{e=(v,w) \in \mathsf{Inc}(v)} f^v_{i,e} - f^w_{i,e} = \begin{cases} 0 & \text{if } v \neq s_i, d_i, \\ R_i & \text{if } v = s_i, \\ -R_i & \text{if } v = d_i. \end{cases}$$

  in the undirected graph case.

The closure of the set of rate tuples achievable by routing flow is called the *flow region* and is denoted by $\mathcal{F}$. Define the *sum-rate-max-flow* by $F_{\text{sum-rate}} := \sup_{(R_i : i \in \mathcal{I}) \in \mathcal{F}} \sum_{i \in \mathcal{I}} R_i$.

Definition: For a k-unicast network $\mathcal{N} = (\mathcal{G}, \underline{\mathrm{C}})$, we call $\mathcal{G}$ an *uncapacitated* k-unicast network. This uncapacitated network is converted to a capacitated network by assigning non-negative capacities $\underline{\mathrm{C}} := (C_e : e \in \mathcal{E}(\mathcal{G})) \in \mathbb{R}_{\geq 0}^{|\mathcal{E}(\mathcal{G})|}$ to the edges of $\mathcal{G}$.

Edge-cut bounds have traditionally been studied in the context of commodity flow problems since they are simple outer bounds on the commodity flow region. Fix a k-unicast uncapacitated network $\mathcal{G}$, and an edge set $E \subseteq \mathcal{E}(\mathcal{G})$. Define the *edge-cut-disconnected-indices* derived from E, denoted $\mathcal{D}_{\mathcal{G},E}$, to be the subset of $\{1, 2, \ldots, k\}$ where index $j \in D_{\mathcal{G},E}$ if and only if there is no path from s_j to d_j in $\mathcal{G} \setminus E$. Consider Statements 1 and 2 below.

Statement 1 *: For any assignment of capacities* $\underline{C}_{\mathcal{G}} \in \mathbb{R}_{\geq 0}^{|\mathcal{E}(\mathcal{G})|}$*, and any rate tuple* $(R_1, R_2, \ldots, R_k) \in \mathcal{F}(\mathcal{G}, \underline{C}_{\mathcal{G}})$,

$$\sum_{j \in \mathcal{D}_{\mathcal{G},E}} R_j \leq \sum_{e \in E} C_e. \tag{2.1}$$

Statement 2 *: For any assignment of capacities* $\underline{C}_{\mathcal{G}} \in \mathbb{R}_{\geq 0}^{|\mathcal{E}(\mathcal{G})|}$*, and any rate tuple* $(R_1, R_2, \ldots, R_k) \in \mathcal{C}(\mathcal{G}, \underline{C}_{\mathcal{G}})$,

$$\sum_{j \in \mathcal{D}_{\mathcal{G},E}} R_j \leq \sum_{e \in E} C_e. \tag{2.2}$$

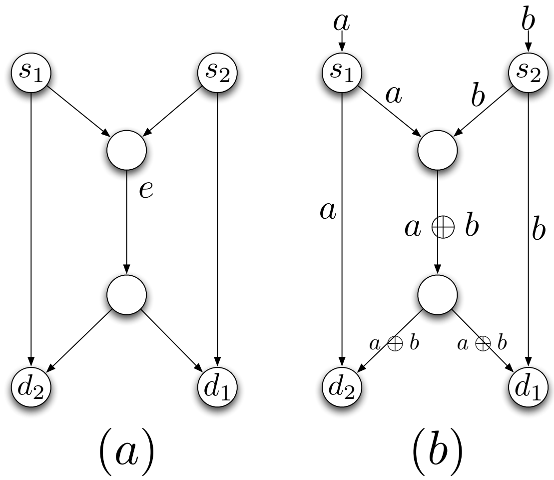


Figure 2.1: A butterfly network

In this example, all edges having unit capacity. $E = \{e\}$ yields the edge-cut inequality $R_1 + R_2 \leq 1$. The inequality is violated by $R_1 = R_2 = 1$ achieved by coding scheme shown.

Inequality (2.1) is the edge-cut bound derived from E. Statement 1 is obviously true since any commodity flow must use up capacity on one of the edges of E. However, Statement 2 is not true in general, as evidenced by the butterfly counterexample in Fig. 2.1.

We say that the edge-cut derived from E is *fundamental* if Statement 2 holds, i.e. if Inequality (2.2) holds for all capacity assignments and all rate tuples in the corresponding capacity region. We shall be interested in what kinds of edge-cuts are fundamental. We define the Generalized Network Sharing (GNS) bound via GNS-cuts as follows.

Definition: Given a k-unicast uncapacitated network $\mathcal{G}$, an edge-cut derived from $E \subseteq \mathcal{E}(\mathcal{G})$ is called a *GNS-cut* for the network if there exists a permutation $\pi : \{1, 2, \ldots, k\} \mapsto \{1, 2, \ldots, k\}$ such that for any i, j there are no paths from s_i to d_j in $\mathcal{G} \setminus E$ whenever $\pi(i) \geq \pi(j)$. If there exists a subset $D \subseteq \{1, 2, \ldots, k\}$, and a bijection $\pi : D \to \{1, 2, \ldots, |D|\}$ such that for any $i, j \in D$, there are no paths from s_i to d_j in $\mathcal{G} \setminus E$ whenever $\pi(i) \geq \pi(j)$, then we say that E is a GNS-cut for the network for the sources indexed by D.

Note that for $k = 1$, GNS-cuts are just vertex bipartition cuts that feature in the cutset bound. The usefulness of GNS-cuts arises from the following theorem that describes the GNS bound.

Theorem 1 *(from [32]) GNS-cuts are fundamental.*

We first present the proof of Theorem 1. The essential idea is contained in [32, 28] but we provide a proof here for completeness.

Proof: Consider a communication scheme over alphabet $\mathcal{A}$ with block length N and number of epochs T that achieves for $1 \leq i \leq r$, rate R_i for the message from s_i to d_i with error probability at most ϵ. Let E be a GNS-cut for $\{s_1, s_2, \ldots, s_r; d_1, d_2, \ldots, d_r\}$ with the identity permutation without loss of generality. Thus, there are no paths in $\mathcal{G} \setminus E$ from node s_i to node d_j whenever $i \geq j$. For $1 \leq i \leq r$, let W_i be the source message at s_i that is required to be delivered to d_i, for $1 \leq i \leq r$. Let W_0 denote the vector of all other source messages in the network. $W_0, W_1, \ldots, W_r$ are mutually independent and each $W_i, 0 \leq i \leq r$ has the uniform distribution. Let X_E denote the vector of all symbols transmitted on the edges of E over the duration of the complete scheme. For $1 \leq i \leq r$, let $\hat{W}_i$ denote the estimate at d_i of the source message W_i upon completion of the coding scheme. Assume $|\mathcal{A}| = 2$; the proof for larger alphabet size is identical. Note that

$$
\begin{aligned}
& H(W_1, W_2, \ldots, W_r | X_E, W_0) && (2.3) \\
& = \sum_{i=1}^{r} H\left(W_i | X_E, W_0, \{W_j : 1 \leq j < i\}\right) && (2.4) \\
& = \sum_{i=1}^{r} H\left(W_i | X_E, W_0, \{W_j : 1 \leq j < i\}, \hat{W}_i\right) && (2.5) \\
& \quad \text{[since } \hat{W}_i \text{ is a function of } X_E, W_0, \{W_j : 1 \leq j < i\} \\
& \quad\quad \text{from the connectivity properties of } \mathcal{G} \setminus E\text{]} \\
& \leq \sum_{i=1}^{r} H\left(W_i | \hat{W}_i\right) && (2.6) \\
& \leq \sum_{i=1}^{r} h(\epsilon) + \epsilon \lceil NTR_i \rceil = rh(\epsilon) + \epsilon \sum_{i=1}^{r} \lceil NTR_i \rceil, && (2.7)
\end{aligned}
$$

where $h(\cdot)$ is the binary entropy function. The last inequality follows from Fano's inequality. Thus, we have

$$
\begin{aligned}
& \sum_{i=1}^{r} \lceil NTR_i \rceil = H(W_1, W_2, \ldots, W_r) && (2.8) \\
& = I(W_1, W_2, \ldots, W_r; X_E, W_0) \\
& \quad\quad + H(W_1, W_2, \ldots, W_r | X_E, W_0) && (2.9) \\
& \leq I(W_1, W_2, \ldots, W_r; X_E | W_0) + rh(\epsilon) + \epsilon \sum_{i=1}^{r} \lceil NTR_i \rceil && (2.10) \\
& \leq H(X_E) + rh(\epsilon) + \epsilon \sum_{i=1}^{r} \lceil NTR_i \rceil && (2.11) \\
& \leq \sum_{e \in E} T \lfloor NC_e \rfloor + rh(\epsilon) + \epsilon \sum_{i=1}^{r} \lceil NTR_i \rceil && (2.12)
\end{aligned}
$$

This establishes that $\sum_{i=1}^{r} R_i \leq \sum_{e\in E} C_e$.

Given a k-unicast uncapacitated network $\mathcal{G}$, one can set out to build a collection of fundamental edge-cut bounds as follows:

- Fix a non-empty subset $D \subseteq \{1, 2, \ldots, k\}$ and consider the $|D|$-unicast network where for each $j \in D$, s_j communicates to d_j at rate R_j.
- For every edge cut derived from a set of edges $E \subseteq \mathcal{E}(\mathcal{G})$ that forms a GNS-cut for the $|D|$-unicast problem, we include the edge-cut bound $\sum_{j\in D} R_j \leq \sum_{e\in E} C_e$.
- Repeat for all choices of non-empty subsets D.

This collection of fundamental edge-cut bounds will be called the *GNS edge-cut bound collection* for $\mathcal{G}$.

Through most of this chapter, we shall only consider *complete* edge-cuts, namely edge-cuts which disconnect all sources from their respective destinations so that $\mathcal{D}_{\mathcal{G},E} = \{1, 2, \ldots, k\}$. This is without loss of generality, since if we have a non-complete but fundamental edge-cut, one only needs to prove the necessary bound by considering a complete edge-cut for a suitable $|D|$-unicast problem with $D = \mathcal{D}_{\mathcal{G},E} \subset \{1, 2, \ldots, k\}$.

2.2 Equivalence to the functional dependence bound

The problem of identifying edge-cut bounds that are fundamental has been approached using different techniques. These include the following:

- PdE bound [39]
- Information Dominance bound [26]
- Functional Dependence bound [56]

The aforementioned works provide algorithms to show that certain edge-cut bounds are fundamental based on properties of the underlying graph. We show that the GNS bound is equivalent to the functional dependence bound [56]. Connection of the GNS bound to the PdE bound [39] and the information dominance bound [26] is discussed in Section 2.5.

Theorem 2 *For multiple-unicast networks, the GNS edge-cut bound collection is equivalent to the functional dependence bound [56].*

Proof: It is easy to check that the GNS bound is a special case of the functional dependence bound [56]. Now, given a k-unicast uncapacitated network $\mathcal{G}$, the functional dependence bound [56] says that the inequality

$$\sum_{i=1}^{k} R_i \leq \sum_{e\in E} C_e \tag{2.13}$$

holds for all capacity assignments $\underline{C}_{\mathcal{G}}$ and all $(R_1, R_2, \ldots, R_k) \in \mathcal{C}(\mathcal{G}, \underline{C}_{\mathcal{G}})$, for a set of edges E that correspond to a so-called *maximal irreducible set* (defined below). We will show that such a set of edges always yields a GNS-cut and that will complete the proof.

We describe the construction from [56] of the functional dependence graph (FDG) denoted by $\mathcal{Z}$. Corresponding to the information message of source s_i, introduce a (so-called) pseudo-variable Y_i and corresponding to each edge e, introduce a pseudo-variable U_e. For each e, draw incoming edges into U_e from each of the pseudo-variables associated with all incoming sources and edges incident on the tail of e. For each destination d_i, draw incoming edges into Y_i from each of the pseudo-variables associated with all incoming edges and sources incident on d_i. This completes the construction of $\mathcal{Z}$. In the network $\mathcal{G}$, each source s_i must have at least one path to its own destination d_i. So, we have that the FDG $\mathcal{Z}$ is *cyclic* (in the notation of [56]). A maximal irreducible set is a subset of vertices $\mathcal{A}$ of the FDG $\mathcal{Z}$ with the property that after one removes all edges outgoing from vertices in $\mathcal{A}$ and successively removes all vertices and edges with no incoming edges and vertices respectively, then no vertex in $\mathcal{Z}$ remains. It must also be that no proper subset of $\mathcal{A}$ has the same property but we will not need this latter condition.

We start with a maximal irreducible set $\mathcal{A}$ that has none of the source variables Y_i, say $\mathcal{A} = \{U_e : e \in E\}$. Consider vertices and edges being removed *from the graph* by this procedure one at a time. Since the process ends with all the Y_i's removed from the graph, let the order in which they get removed be given by a permutation π, i.e. let the order be $Y_{\pi(1)}, Y_{\pi(2)}, \ldots, Y_{\pi(k)}$. Then, none of the sources $s_1, s_2, \ldots, s_k$ have a path to $d_{\pi(1)}$ in $\mathcal{G} \setminus E$. Further, none of the sources with the possible exception of $s_{\pi(1)}$ can have a path to $d_{\pi(2)}$ in $\mathcal{G} \setminus E$. Continuing this chain of reasoning, we find that the edge-cut derived from E is a GNS-cut for the network $\mathcal{G}$ with permutation π. This completes the proof.

2.3 Edge-cut bounds based only on source-destination connectivity

Fundamentality of an edge-cut bound is a purely graph-theoretic property. A simple way to classify different edge-cuts is to look at connectivity from all sources to all destinations. For a k-unicast uncapacitated network $\mathcal{G}$, and a subset of edges $E \subseteq \mathcal{E}(\mathcal{G})$ which yield a complete edge-cut, we define the *profile of the edge-cut* derived from E, denoted $\mathcal{P}_{\mathcal{G},E}$, to be a directed graph with nodes having labels $s_1, s_2, \ldots, s_k, d_1, d_2, \ldots, d_k$ with s_i's having only outgoing edges, d_i's having only incoming edges and an edge from s_i to d_j if and only if there is a path from s_i to d_j in $\mathcal{G} \setminus E$. If the edge-cut derived from E is a GNS-cut, then we call the corresponding profile a *GNS profile*. Fig. 2.2 shows all possible profiles of complete edge-cuts for a 2-unicast network.

From Theorem 1, all edge-cuts with a GNS profile result in fundamental bounds. A natural question to ask is whether there are other edge-cut profiles for which it is also true that all edge-cuts with that profile result in fundamental bounds. Furthermore, it is of

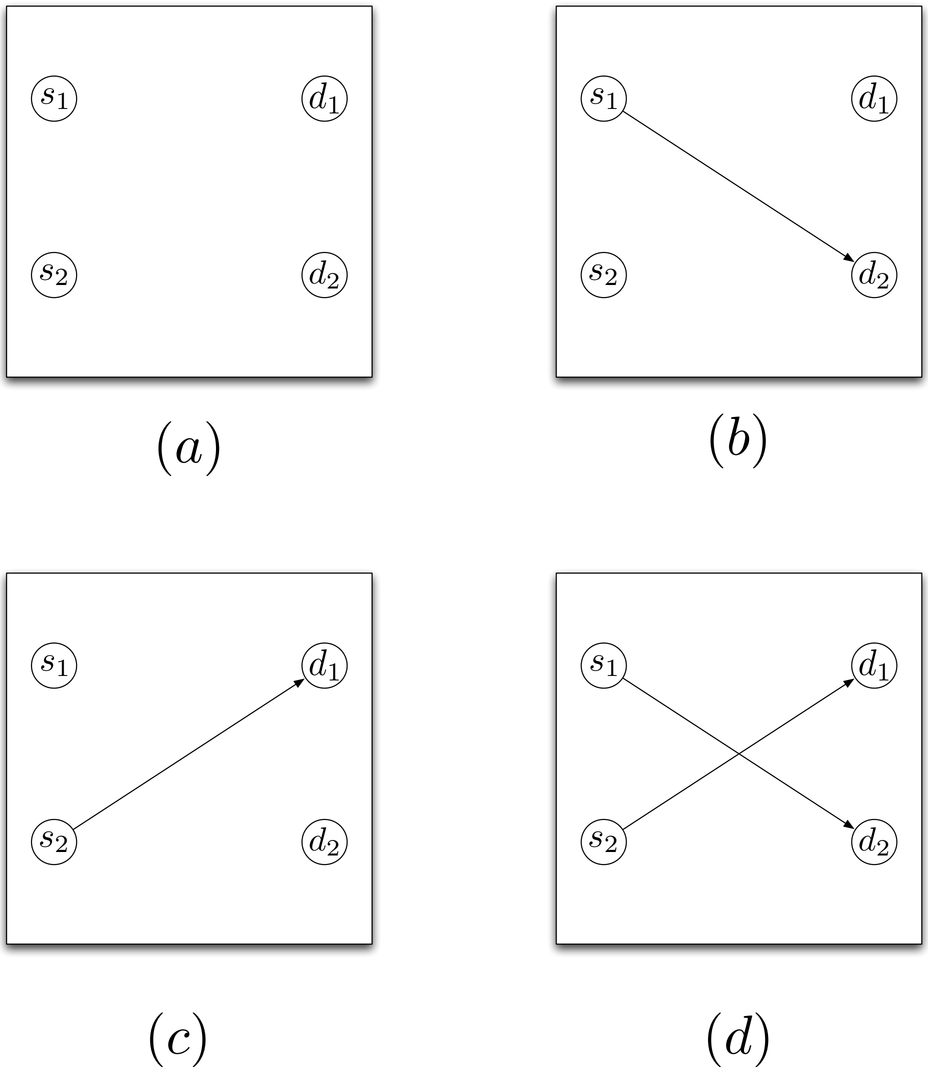
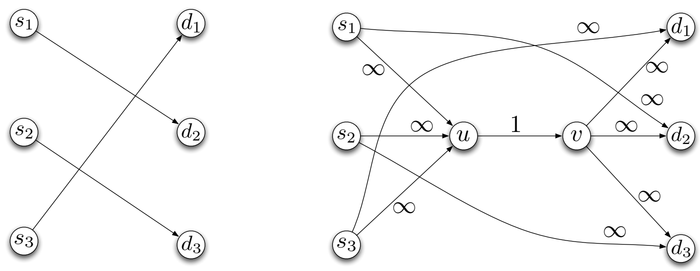


Figure 2.2: Profiles of edge-cuts for 2-unicast networks

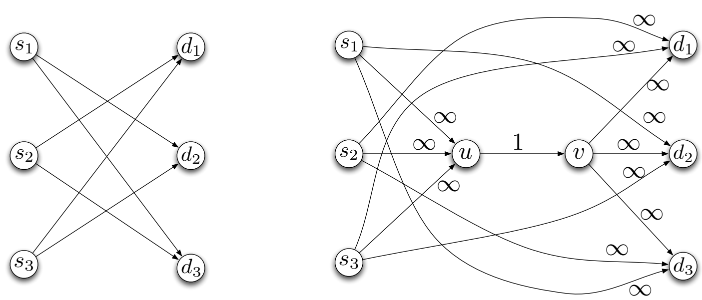
(a), (b), (c) are GNS profiles while (d) is not.

interest to provide some bounds in the case of an edge-cut profile that does not necessarily give fundamental bounds for all networks. Both these issues are addressed by Theorem 3. As an example, the profile in Fig. 2.2(d) is a non-GNS profile and this profile happens to not give fundamental bounds in all networks as seen by the example in Fig. 2.1.

To state the main result of this section, we need one more definition. Given a profile $\mathcal{P}$ of a complete edge-cut for a k-unicast network, we define a specific capacitated network - its canonical network $\mathcal{N}(\mathcal{P})$ - an index coding [6] instance, as follows. Take the directed graph represented by the profile $\mathcal{P}$ and add two nodes u and v. Add edges from all the s_i's to u, from v to all the d_i's and from u to v. All edges have infinite capacity except the edge from u to v which has capacity 1 unit. For each i, s_i has independent information to be communicated to d_i. Let the sum-capacity of this network be denoted by $\rho(\mathcal{P})$. Fig. 2.3 shows two examples of profiles of edge-cuts and their corresponding canonical networks.



(a) Profile $\mathcal{P}_1$ (b) Canonical network $\mathcal{N}(\mathcal{P}_1)$



(c) Profile $\mathcal{P}_2$ (d) Canonical network $\mathcal{N}(\mathcal{P}_2)$

Figure 2.3: Two examples of profiles and their corresponding canonical networks

It can be shown that $\rho(\mathcal{P}_1) = \frac{3}{2}$ and $\rho(\mathcal{P}_2) = 3$.

Theorem 3 *Fix an edge-cut profile $\mathcal{P}$. For any k-unicast uncapacitated network $\mathcal{G}$, and any complete edge-cut derived from edge set $E \subseteq \mathcal{E}(\mathcal{G})$ with $\mathcal{P}_{\mathcal{G},E} = \mathcal{P}$, we have the inequality*

$$\sum_{j=1}^{k} R_j \leq \rho(\mathcal{P}) \sum_{e \in E} C_e, \tag{2.14}$$

for any assignment of capacities $\underline{C}_{\mathcal{G}} \in \mathbb{R}_{\geq 0}^{|\mathcal{E}(\mathcal{G})|}$, and any rate tuple within capacity, $(R_1, R_2, \ldots, R_k) \in \mathcal{C}(\mathcal{G}, \underline{C}_{\mathcal{G}})$. Moreover, the constant $\rho(\mathcal{P})$ in Inequality (2.14) *cannot be improved upon and satisfies $\rho(\mathcal{P}) \geq 1$, with equality if and only if $\mathcal{P}$ is a GNS-profile.*

Proof: Suppose we have a k-unicast capacitated network $(\mathcal{G}, \underline{C}_{\mathcal{G}})$, and a complete edge-cut derived from edge set E, whose profile is $\mathcal{P}$. We will perform modifications to the network and its capacities which can only enhance its capacity region.

- For each directed edge (x, y) in E, add an edge from each of the sources to x and from y to each of the destinations.
- Now, assign infinite capacities to all edges of this network that do not belong to E.

Now all source messages can be assumed to be present in their entirety at the tails of each edge in E and all destinations are connected with an infinite capacity path to the heads of each edge in E, and therefore, any coding scheme operating on this network can be translated to a coding scheme on a $\left(\sum_{e\in E} C_e\right)$-scaled copy of $\mathcal{N}(\mathcal{P})$ and vice versa. Therefore, the sum capacity of this enhanced network is $\rho(\mathcal{P})\sum_{e\in E} C_e$ and so, for any rate tuple $(R_1, R_2, \ldots, R_k) \in \mathcal{C}(\mathcal{G}, \underline{C}_{\mathcal{G}})$, we have the desired Inequality (2.14). The constant $\rho(\mathcal{P})$ cannot be improved upon since $\mathcal{N}(\mathcal{P})$ is an example of a network with edge-cut derived from $E = \{(u, v)\}$ having the desired profile $\mathcal{P}$ and for which Inequality (2.14) is tight by the definition of $\rho(\mathcal{P})$.

$\rho(\mathcal{P}) \geq 1$ is obvious from the definition since commodity flow can achieve a sum-rate of 1 in $\mathcal{N}(\mathcal{P})$. For a GNS profile $\mathcal{P}$, Theorem 1 gives $\rho(\mathcal{P}) \leq 1$. We only need to show that $\rho(\mathcal{P}) > 1$ for any non-GNS profile $\mathcal{P}$. It is easy to show that for any non-GNS profile $\mathcal{P}$, one can find a sequence of $t \geq 2$ distinct indices $i_1, i_2, \ldots, i_t \in \{1, 2, \ldots, k\}$ such that in the directed graph represented by $\mathcal{P}$, we have that s_{i_r} has an edge to $d_{i_{r+1}}$ for $r = 1, 2, \ldots, t-1$ and s_{i_t} has an edge to d_{i_1}. (For example, Fig. 2.3(a), (b) are both 3-unicast non-GNS profiles for which one can set $i_1 = 1, i_2 = 2, i_3 = 3$.) We now propose a coding scheme for $\mathcal{N}(\mathcal{P})$ which achieves a sum-rate of $\frac{t}{t-1}$ thus showing $\rho(\mathcal{P}) \geq \frac{t}{t-1} > 1$.

Assume that only the sources $s_{i_\alpha}, \alpha = 1, 2, \ldots, t$ wish to deliver one message symbol from a finite field $\mathbb{F}$ to their respective destinations and that the edge (u, v) can carry one finite symbol per time slot. We will accomplish this task in $t-1$ time slots. Node u receives all the t finite field message symbols, say $X_1, X_2, \ldots, X_t$. It sends out $t-1$ random linear combinations of these symbols on the edge (u, v), where the co-efficients are uniformly chosen from $\mathbb{F}$ independent across the different symbols and across time. Each destination $d_{i_\alpha}, \alpha = 1, 2, \ldots, t$ receives all of these $t-1$ symbols and also has one message symbol of side information from the source directly connected to it. Standard calculations similar to those in [38] can then be used to show that each destination can recover its intended message with high probability as the size of the finite field $\mathbb{F}$ goes to infinity. Thus, there exists some coding scheme that delivers the desired performance.

Remark: We note that $\rho(\mathcal{P})$ may be quite hard to compute, especially for large k. However, once computed for a profile for a specific k, it gives useful bounds for all k-unicast networks with no restrictions on the size of such networks. Recent work in [6] provides inner bounds on the entire capacity region for the index coding problem. In particular, their bounds are tight for upto five-node networks which would allow us to evaluate the sum-capacity $\rho(\mathcal{P})$ exactly for all canonical networks $\mathcal{N}(\mathcal{P})$ with $k \leq 5$.

2.4 NP-completeness of minimum GNS-cut

The works of [26], [39], [56] provide algorithms to check if their approach can deduce the fundamentality of a given edge-cut. However, the number of edge-cuts is exponential in the size of the network and so listing all of them and checking if they provide fundamental bounds is computationally intractable. For a single-unicast problem, we know that the algorithm of Ford and Fulkerson reveals the mincut efficiently in spite of there being exponentially many edge-cuts. Given a capacitated k-unicast network, can we have any algorithm that efficiently finds, among all complete edge-cuts E that are GNS-cuts, the one that has the smallest value of $\sum_{e\in E} C_e$? Theorem 4 will show unfortunately that we cannot, even for $k = 2$, unless P=NP. Let us define the following decision problem.

MIN 2-GNS-CUT

Instance: A two-unicast uncapacitated network $\mathcal{G}$ and an assignment of non-negative capacities $\underline{\mathrm{C}}_{\mathcal{G}}$ to the edges.

Question: Is there a set of edges $E \subseteq \mathcal{E}(\mathcal{G})$ with $\sum_{e\in E} C_e \leq K$ such that the edge-cut derived from E is a GNS-cut?

Theorem 4 *MIN 2-GNS-CUT is NP-complete.*

Proof: It is clear that MIN 2-GNS-CUT is in NP. We give a polynomial transformation from the multiterminal cut problem for three terminals which is known to be NP-complete [17]. In the multiterminal cut problem, we are given a number K and an unweighted undirected graph $\mathcal{H}$ with three special vertices or "terminals" x, y, z. We are asked whether there is a subset of edges F of the graph $\mathcal{H}$ with $|F| \leq K$ such that $\mathcal{H} \setminus F$ has no paths between any two of x, y, z. Given $(\mathcal{H}, K)$, we construct a corresponding instance of MIN 2-GNS-CUT as follows. Let the number of edges of $\mathcal{H}$ be N with $K \leq N$.

The two-unicast capacitated network $\mathcal{G}$ is obtained by replacing each undirected edge (u, v) of $\mathcal{H}$ with a gadget as shown in Fig. 2.4. The gadget introduces two new vertices w, w' and constitutes five edges, the one *central* edge having unit capacity and four *flank* edges each having capacity $N + 1$ units. Finally, s_1 is identified with terminal x, d_2 with terminal y and both s_2 and d_1 with terminal z.

We show that $\mathcal{G}$ has a GNS-cut derived from a set of edges E with $\sum_{e\in E} C_e \leq K$ if and only if $\mathcal{H}$ has a set of edges F forming a multiterminal cut with $|F| \leq K$.

Suppose that in the undirected graph $\mathcal{H}$, there is a multiterminal cut F with at most K edges. Then, picking the central edge of the gadgets corresponding to the edges in F gives a GNS-cut in $\mathcal{G}$ with edge set E such that $\sum_{e\in E} C_e = |E| \leq K$.

Conversely, suppose there is a GNS-cut in $\mathcal{G}$ derived from an edge set E which satisfies $\sum_{e\in E} C_e \leq K$. As s_2 and d_1 are identified, the GNS-cut must have the profile shown in Fig. 2.2 (c). Moreover, as $K \leq N$, the edge set E cannot contain any flank edge and must consist exclusively of central edges of gadgets. Choosing the undirected edges of $\mathcal{H}$

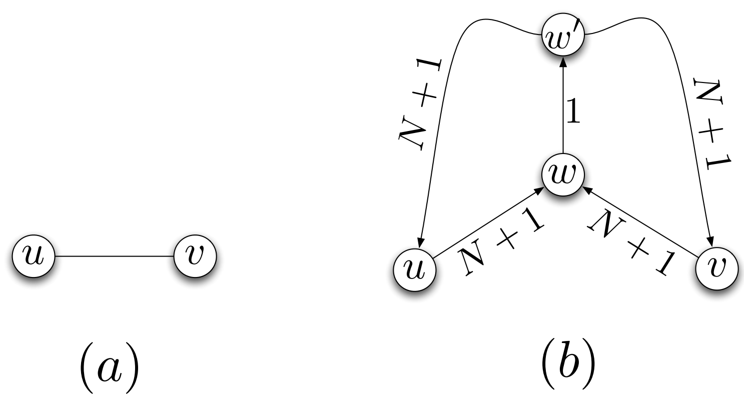


Figure 2.4: Edge to Gadget

(a) shows an undirected edge and (b) the corresponding gadget

corresponding to the gadgets whose central edges lie in E gives an edge set F of $\mathcal{H}$ that has at most K edges and is a multiterminal cut in $\mathcal{H}$.

The significance of Theorem 4 can be appreciated using Theorem 5 below.

Theorem 5 *(Theorem 5 of [32]) For 2-unicast networks, the GNS edge-cut bound collection is the tightest edge-cut bound collection possible. In particular, therefore, it is equivalent to the PdE bound [39], Information Dominance bound [26] and Functional Dependence bound [56] for 2-unicast networks.*

Thus, the hardness result in Theorem 4 and the fact that the difficulty of computing the tightest edge-cut bound in any class of bounds for the k-unicast problem can only increase with k, we obtain a hardness result for computation of the tightest edge-cut bounds for the k-unicast problem that can be dereived from the PdE, information dominance and functional dependence bounds.

2.5 Discussion

Theorem 3 does not say that an edge-cut with a non-GNS profile must necessarily not be fundamental. Consider the bat network example in Fig. 2.5(a). The edge-cut derived from $E = \{e_1, e_2\}$ is not a GNS-cut for the two-unicast network shown, yet $R_1 + R_2 \leq C_{e_1} + C_{e_2}$ is a fundamental edge-cut bound. The reason of course, is that $R_1 \leq C_{e_2}$ and $R_2 \leq C_{e_1}$ follow from the cutset bound. Fig. 2.5(b) shows each edge assigned unit capacity and a specific coding scheme. This coding scheme makes it clear why functional dependence or information dominance do not capture this bound for the 2-unicast problem: The information flowing on $\{e_1, e_2\}$ does not dominate all the source messages.

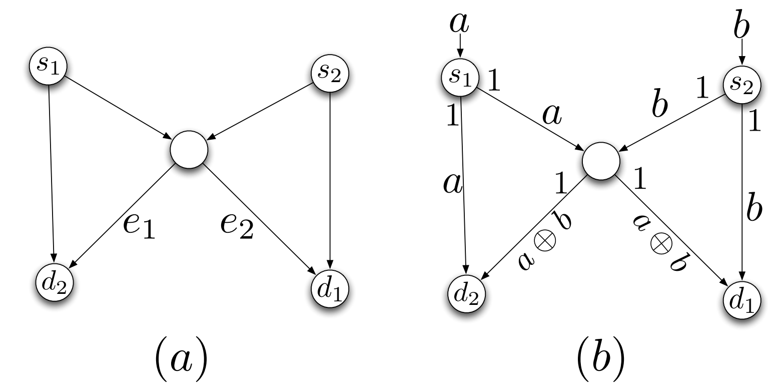


Figure 2.5: Bat network

(a) shows a two-unicast uncapacitated network. (b) shows a specific coding scheme on the capacitated network with all edges having unit capacity.

Although this example is somewhat daft, the general question is not. Is there a fundamental sum-rate edge-cut bound for a k-unicast network that is not implied by the GNS edge-cut bound collection? The answer is No for $k = 1$ (by the Max-Flow-Min-Cut theorem) and also No for $k = 2$ by Theorem 5. The question is open for $k \geq 3$. In particular, it is of interest to know whether or not the PdE bound [39] and Information Dominance bound [26] are strictly stronger than the GNS bound collection for k-unicast networks with $k \geq 3$.

It is also of interest to explore weighted sum-rate edge-cut bounds as studied in [59]. It would be useful to determine other classes of Shannon (or non-Shannon) inequalities and other information about the graph structure of the network that can help in deriving such bounds.

Chapter 3

An information-theoretic meta-theorem for wireline networks

In the previous chapter, we studied edge-cut bounds as conventional outer bounds on commodity flow or routing, and distinguished them from fundamental edge-cut bounds that are outer bounds on the network capacity region. However, the literature on hardness of cut problems typically deals with edge-cut bounds - the outer bounds on commodity flow. Although these edge-cut bounds in directed networks are not fundamental, they are combinatorially well-represented. They are however, hard to approximate in general [13], [1].

One class of networks for which edge-cut bounds can be approximated well is the class of undirected networks. A series of works [44], [46] has shown that for the problem of k-unicast in undirected networks, flow solutions approach the edge-cut bounds up to a factor of $\kappa' \log(k+1)$, for some universal constant κ'. There has also been discovered a semi-definite programming relaxation approach that allows an approximation of edge-cut bounds up to a factor of $\Theta(\sqrt{\log k})$ [7]. Interestingly, for undirected networks, edge-cut bounds can be derived from the vertex bipartition cutset bound and are hence fundamental outer bounds on the capacity region. Thus, [44], [46] also characterize up to a factor of $\kappa' \log(k+1)$, (with κ' being a universal constant), the capacity region of k-unicast in undirected networks. It has been conjectured that flow solutions in fact achieve capacity [45], [25].

Another setting in which edge-cut bounds can be approximated well is the problem of multiple unicast in directed wireline networks with *symmetric demands*, i.e. for each source communicating to its destination at a certain rate there is an independent message to be communicated from the destination back to the source at the same rate. Klein, Plotkin, Rao, and Tardos [37] show under this model that flow solutions achieve within $\kappa \log^3(k+1)$ of the edge-cut bounds for some universal constant κ. We ask the question: "Are these edge-cut bounds fundamental outer bounds on the capacity region?" Surprisingly, the answer turns out to be yes and the proof of this result is one of the main contributions of this chapter. The key tool we use in the proof is the Generalized Network Sharing (GNS) bound that we described in Chapter 2. This completes an approximate characterization of the capacity region for this class of problems.

Yet another setting which allows a flow-cut closeness result is the groupcast problem in directed wireline networks. In this setting, there is a group of nodes and each node in the group has an independent message to be relayed to every other node in the group. [51] shows that the maximum sum-rate achievable by routing flow for groupcast is at least half the so-called multicut, a simple edge-cut based outer bound. We ask the question: "Is the multicut a fundamental outer bound on the sum-rate?". We show that the answer is no but that twice the multicut is indeed a fundamental outer bound. This shows that routing flow is approximately optimal for maximizing sum-rate in groupcast.

When there is some kind of symmetry in the network, either in the underlying graph (undirected or bidirected networks) or in the traffic (directed network with symmetric demands, sum-rate in groupcast), the following picture seems to emerge.

- (Achievability) *Algorithmic Meta-Theorem*: Edge-cut bounds can be well-approximated either by flows [44], [46], [37], [12], [51] or by other means [7].
- (Converse) *Information-Theoretic Meta-Theorem*: Edge-cut bounds are fundamental or close to fundamental outer bounds on the capacity region.
- *Combined Meta-Theorem*: Flows approximately achieve capacity.

In Chapter 4, we use achievability results similar to [46], [37], and [51] obtained for polymatroidal networks (a generalization of wireline networks), and an extension of the GNS bound to Gaussian networks to study the capacity regions of wireless networks. The rest of this chapter is organized as follows. We define the edge-cut outer bound in Section 3.1. We then discuss

- k-unicast undirected networks in Section 3.2,
- k-unicast directed symmetric-demand networks in Section 3.3,
- k-groupcast directed networks in Section 3.4.

Finally, we end with a discussion in Section 3.5.

3.1 Preliminaries

We add the following definition to the objects defined in Chapter 2.

Definition: Given a k-unicast network $\mathcal{N} = (\mathcal{G}, \underline{C})$ for source-destination pairs $\{(s_i; d_i)\}_{i\in\mathcal{I}}$, we define the *edge-cut outer bound* denoted by $\mathcal{R}_{\text{edge-cut}}$, to be the set of all non-negative tuples $(R_i : i \in \mathcal{I})$ that satisfy for every $E \subseteq \mathcal{E}(\mathcal{G})$, the inequality $\sum_{i\in J} R_i \leq \sum_{e\in E} C_e$ where index $i \in J \subseteq \mathcal{I}$ if and only if $\mathcal{G} \setminus E$ has no paths from s_i to d_i. (The paths are directed paths in the directed graph case and undirected paths in the undirected graph case.) We define the *multicut* denoted by R_{multicut}, to be the minimum value of $\sum_{e\in E} C_e$ over all $E \subseteq \mathcal{E}(\mathcal{G})$ with the property that $\mathcal{G} \setminus E$ has no paths from s_i to d_i for each $i \in \mathcal{I}$.

While it is clear that $\mathcal{F} \subseteq \mathcal{R}_{\text{edge-cut}}$ and $\mathcal{F} \subseteq \mathcal{C}$, the connection between $\mathcal{C}$ and $\mathcal{R}_{\text{edge-cut}}$ is unclear. The example in Fig. 2.1 in Chapter 2 shows that $\mathcal{C} \nsubseteq \mathcal{R}_{\text{edge-cut}}$ in general. Thus, simple edge-cut based outer bounds are not *fundamental* fundamental, i.e. they are not outer bounds on the capacity region. Likewise it is clear that $F_{\text{sum-rate}} \leq R_{\text{multicut}}$ and $F_{\text{sum-rate}} \leq C_{\text{sum-rate}}$ but $C_{\text{sum-rate}}$ and R_{multicut} have no apparent connection. Indeed, [26] provides a series of k-unicast networks, one for each k with $k = 2^n$, with $F_{\text{sum-rate}} = R_{\text{multicut}} = \frac{1}{k} C_{\text{sum-rate}}$ and $\mathcal{C} \nsubseteq (k-\epsilon)\mathcal{R}_{\text{edge-cut}}$ for any $\epsilon > 0$.

Remark: Note that R_{multicut} may in general be strictly larger than the tighter bound on $F_{\text{sum-rate}}$ given by $\sup_{(R_i : i \in \mathcal{I}) \in \mathcal{R}_{\text{edge-cut}}} \sum_{i \in \mathcal{I}} R_i$.

3.2 k-unicast undirected networks

Theorems 6 and 7, and Corollary 1 will refer to k-unicast undirected networks.

Theorem 6 *(Leighton-Rao [44], Linial-London-Rabinovich [46])*

$$\frac{\mathcal{R}_{\text{edge-cut}}}{\kappa' \log(k+1)} \subseteq \mathcal{F} \subseteq \mathcal{R}_{\text{edge-cut}}, \tag{3.1}$$

for a universal constant κ'.

Theorem 7

$$\mathcal{C} \subseteq \mathcal{R}_{\text{edge-cut}} \tag{3.2}$$

Theorems 6 and 7 together imply that routing flow is approximately capacity-achieving:

Corollary 1

$$\frac{\mathcal{R}_{\text{edge-cut}}}{\kappa' \log(k+1)} \subseteq \mathcal{F} \subseteq \mathcal{C} \subseteq \mathcal{R}_{\text{edge-cut}}, \tag{3.3}$$

for a universal constant κ'.

Although Theorem 7 is easy to show, it is not very well-known. So, we provide a proof here.

Proof: Consider a k-unicast undirected network $\mathcal{N}$ for source-destination pairs $\{(s_i; d_i)\}_{i \in \mathcal{I}}$. Consider a coding scheme that achieves rates $(R_i : i \in \mathcal{I})$ in T epochs of block length N with overall error probability at most ϵ. Let $E \subseteq \mathcal{E}(\mathcal{G})$ be any subset of edges and let $J \subseteq \mathcal{I}$ denote the set of all indices $i \in \mathcal{I}$ which have s_i and d_i disconnected from each other in $\mathcal{G} \setminus E$. Let $(\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_r)$ denote the partition of $\mathcal{V}$ obtained as the connected components of $\mathcal{G} \setminus E$ and let $E' \subseteq E$ denote the set of edges that connect a vertex in $\mathcal{V}_s$ to a vertex in $\mathcal{V}_{s'}$, for $1 \leq s, s' \leq r, s \neq s'$.

For $1 \leq q \leq r$, let J_q denote the set of all indices $i \in \mathcal{I}$ which have $s_i \in \mathcal{V}_q$ and $d_i \notin \mathcal{V}_q$. $(J_1, J_2, \ldots, J_r)$ is a partition of J. For $1 \leq q \leq r$, let X_q denote the vector of all symbols transmitted on edges going out of $\mathcal{V}_q$ and into $\mathcal{V}_s$ for some $s \neq q$. For $i \in \mathcal{I}$, let W_i be the source message at s_i that is required to be decoded at d_i. By the cutset bound written for vertex bipartition $(\mathcal{V}_q, \mathcal{V} \setminus \mathcal{V}_q)$, we have

$$\sum_{i \in J_q} H(W_i) \leq H(X_q) \tag{3.4}$$

Adding up these inequalities over $1 \leq q \leq r$, we get $\sum_{i \in J} H(W_i) \leq \sum_{q=1}^{r} H(X_q)$, which yields $\sum_{i \in J} R_i \leq \sum_{e \in E'} C_e \leq \sum_{e \in E} C_e$. This establishes $\mathcal{C} \subseteq \mathcal{R}_{\text{edge-cut}}$.

Remark: Theorems analogous to 6 and 7 can be similarly proved for k-unicast in bidirected networks, i.e. directed networks in which for every edge from node u to node v there is another edge from node v to node u with the same capacity.

It has been conjectured that a much stronger result than Corollary 1 holds:

Conjecture 1 *(Li and Li conjecture [45], [25]) For k-pair unicast undirected networks,*

$$\mathcal{F} = \mathcal{C}.$$

3.3 k-pair unicast directed symmetric-demand networks

Definition: A *k-pair unicast directed symmetric-demand network* is a $2k$-unicast directed network $\mathcal{N}$ with $2k$ distinct distinguished nodes (source-destination nodes) $u_1, u_2, \ldots, u_k$, $v_1, v_2, \ldots, v_k$ with source-destination pairs $\{s_i; d_i\}_{i \in \mathcal{I}}$ where $\mathcal{I} = \{1, 2, \ldots, k\} \cup \{-1, -2, \ldots, -k\}$ and for $i > 0$, $s_i = u_i, d_i = v_i$, while for $i < 0, s_i = v_{-i}, d_i = u_{-i}$. The rate tuple $(R_i : 1 \leq i \leq k)$ is defined to be in the capacity region $\mathcal{C}$, flow region $\mathcal{F}$, or edge-cut outer bound region $\mathcal{R}_{\text{edge-cut}}$ for the k-pair unicast directed symmetric-demand network if the rate tuple $(R'_i : i \in \mathcal{I})$, given by $R'_i = R_{|i|}$ for $i \in \mathcal{I}$, lies in the capacity region, flow region, or edge-cut outer bound region respectively of the $2k$-unicast directed network.

Remark: There is no loss of generality in assuming that $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_k$ are distinct since if they aren't we can add more nodes and infinite capacity edges to make them distinct while obtaining a network with an identical capacity region.

Definition: Given a k-pair unicast directed symmetric-demand network $\mathcal{N} = (\mathcal{G}, \underline{C})$ with source-destination nodes $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_k$, we define the *GNS-cut outer bound* denoted by $\mathcal{R}_{\text{GNS-cut}}$, to be the set of all non-negative tuples $(R_i : 1 \leq i \leq k)$ that satisfy for every $E \subseteq \mathcal{E}(\mathcal{G})$, the inequality $\sum_{i \in J} R_i \leq \sum_{e \in E} C_e$ whenever E is a GNS-cut for $\{w_1, w_2, \ldots, w_r; w'_1, w'_2, \ldots, w'_r\}$ with some permutation π where

- $J \subseteq \{1, 2, \ldots, k\}, |J| = r$,
- $w_1, w_2, \ldots, w_r, w'_1, w'_2, \ldots, w'_r$ are distinct,
- for $1 \leq j \leq r$, $(w_j, w'_j) = (u_i, v_i)$ or (v_i, u_i) for some $i \in J$.

We define a *weak edge-cut* outer bound for this class of networks.

Definition: Given a k-pair unicast directed symmetric-demand network $\mathcal{N} = (\mathcal{G}, \underline{C})$ with source-destination nodes $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_k$, we define the *weak edge-cut outer bound* denoted by $\mathcal{R}_{\text{w.e.c.}}$, to be the set of all non-negative tuples $(R_i : 1 \leq i \leq k)$ that satisfy for every $E \subseteq \mathcal{E}(\mathcal{G})$, the inequality $\sum_{i \in J} R_i \leq \sum_{e \in E} C_e$ where index $i \in J \subseteq \{1, 2, \ldots, k\}$ if and only if $\mathcal{G} \setminus E$ has no directed paths from either u_i to v_i or v_i to u_i or both.

Remark: For $E \subseteq \mathcal{E}(\mathcal{G})$, if J_1 is the set of indices $i, 1 \leq i \leq k$ for which $\mathcal{G} \setminus E$ has no directed paths from either u_i to v_i or from v_i to u_i but not both and J_2 is the set of indices $i, 1 \leq i \leq k$ for which $\mathcal{G} \setminus E$ has no directed paths from u_i to v_i and from v_i to u_i, then the edge-cut outer bound has the inequality $\sum_{i \in J_1} R_i + 2 \sum_{j \in J_2} R_j \leq \sum_{e \in E} C_e$ while the weak edge-cut outer bound has the inequality $\sum_{i \in J_1} R_i + \sum_{j \in J_2} R_j \leq \sum_{e \in E} C_e$. It is therefore, clear that

$$\mathcal{R}_{\text{edge-cut}} \subseteq \mathcal{R}_{\text{w.e.c.}} \subseteq 2\mathcal{R}_{\text{edge-cut}}.$$

Theorems 8, 9 and 10 and Corollary 2 will refer to k-pair unicast directed symmetric-demand networks.

Theorem 8 *(Klein-Plotkin-Rao-Tardos [37])*

$$\frac{\mathcal{R}_{\text{w.e.c.}}}{\kappa \log^3(k+1)} \subseteq \mathcal{F} \subseteq \mathcal{R}_{\text{w.e.c.}}, \tag{3.5}$$

for a universal constant κ.

Theorem 9 *(stated as Theorem 1 in Chapter 2)*

$$\mathcal{C} \subseteq \mathcal{R}_{\text{GNS-cut}} \tag{3.6}$$

Theorem 10

$$\mathcal{R}_{\text{w.e.c.}} = \mathcal{R}_{\text{GNS-cut}} \tag{3.7}$$

Theorems 8, 9 and 10 together imply that routing flow is approximately capacity-achieving:

Corollary 2

$$\frac{\mathcal{R}_{\text{w.e.c.}}}{\kappa \log^3(k+1)} \subseteq \mathcal{F} \subseteq \mathcal{C} \subseteq \mathcal{R}_{\text{w.e.c.}} = \mathcal{R}_{\text{GNS-cut}}, \tag{3.8}$$

for a universal constant κ.

Remark: The GNS bound is to the capacity region what the edge-cut bound is to the flow region, namely an intuitive outer bound that arises from simple connectivity properties of the underlying graph of the network. While more sophisticated bounds [39], [26], and [56] include the GNS bound as a special case, it is the simplicity of the GNS bound that becomes useful for Theorem 10, which shows that weak edge-cuts and GNS-cuts are identical for k-pair directed symmetric-demand networks. We also note that the outer bound $\mathcal{R}_{\text{GNS}-\text{cut}}$ is in general strictly tighter than the cutset outer bound and that, in general, the capacity region $\mathcal{C}$ is not contained in $\mathcal{R}_{\text{edge}-\text{cut}}$ although it is always contained in $\text{R}_{\text{w.e.c.}}$.

We conjecture that a much stronger result than Corollary 2 holds:

Conjecture 2 *(Analog of Li and Li conjecture [45], [25]) For k-pair unicast directed symmetric-demand networks,*

$$\mathcal{F} \subseteq \mathcal{C} \subseteq 2\mathcal{F},$$

i.e. network coding can improve rates beyond routing flow by at most a factor 2.

We prove the equivalence between weak edge-cuts and GNS-cuts for k-pair unicast directed symmetric-demand networks, thus proving Theorem 10.

Proof: It is easy to see that the inequality obtained from a GNS-cut can always be obtained from a weak edge-cut since a GNS-cut requires stronger disconnections as compared to a weak edge-cut. This gives $\mathcal{R}_{\text{w.e.c.}} \subseteq \mathcal{R}_{\text{GNS}-\text{cut}}$. To show $\mathcal{R}_{\text{GNS}-\text{cut}} \subseteq \mathcal{R}_{\text{w.e.c.}}$, we now consider $E \subseteq \mathcal{E}(\mathcal{G})$, and say $i \in J \subseteq \{1, 2, \ldots, k\}$ if and only if $\mathcal{G} \setminus E$ has no directed paths from either u_i to v_i or from v_i to u_i or both. We show that E is a GNS-cut for $\{w_1, w_2, \ldots, w_r; w_1', w_2', \ldots, w_r'\}$ with some permutation π where the $2r$ vertices $w_1, w_2, \ldots, w_r, w_1', w_2', \ldots, w_r'$ are all distinct and for $1 \leq j \leq r$, $(w_j, w_j') = (u_i, v_i)$ or (v_i, u_i) for some $i \in J$ with $|J| = r$. We will prove this for the case $J = \{1, 2, \ldots, k\}$. The proof for other choices of J is similar.

Define the *connectivity graph* $\mathcal{G}_c$ as a directed graph over $2k$ vertices $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_k$ as follows. For every pair of distinct vertices w and z, there is a directed edge from w to z in $\mathcal{G}_c$ if and only if w has a directed path to z in $\mathcal{G} \setminus E$. See Fig. 3.1 for an example. $\mathcal{G}_c$ is transitively closed, i.e. for three distinct vertices w, z, x, if w has an edge to z and z has an edge to x, then w has an edge to x. Define two distinct vertices u and v in $\mathcal{G}_c$ as *associated*, if u has an edge to v and v has an edge to u. If we define every vertex to be associated with itself, this relation is reflexive and symmetric. As $\mathcal{G}_c$ is transitively closed, this relation is also transitive and so association is an equivalence relation. Further, for each $i = 1, 2, \ldots, k$, we have that u_i and v_i are not associated since there are no paths in $\mathcal{G} \setminus S$ from either u_i to v_i or from v_i to u_i.

Now, define the *reduced connectivity graph* $\mathcal{G}_r$ as a directed graph with vertices represented by the equivalence classes defined from being associated in $\mathcal{G}_c$. See Fig. 3.2 for an example. There is a directed edge from equivalence class $\mathcal{E}_1$ to $\mathcal{E}_2$ in $\mathcal{G}_r$ if there is a directed edge in $\mathcal{G}_c$ from each vertex in $\mathcal{E}_1$ to each vertex in $\mathcal{E}_2$. By transitive closure of $\mathcal{G}_c$, this happens if and

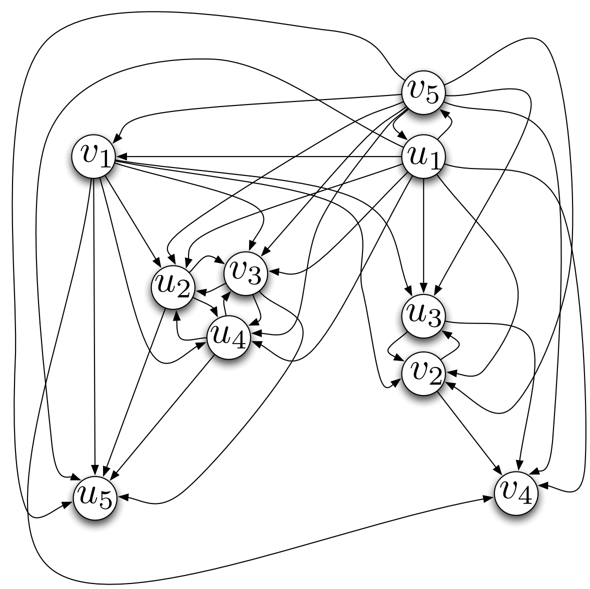


Figure 3.1: Connectivity graph $\mathcal{G}_c$

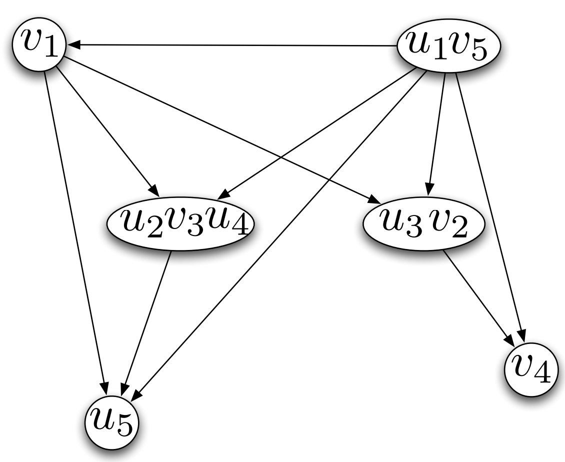


Figure 3.2: Reduced connectivity graph $\mathcal{G}_r$

only if there is a directed edge in $\mathcal{G}_c$ from some vertex in $\mathcal{E}_1$ to some vertex in $\mathcal{E}_2$. $\mathcal{G}_r$ has at least two vertices since u_1 and v_1 cannot belong to the same equivalence class.

Now, note that $\mathcal{G}_r$ is a directed acyclic graph. Suppose not, i.e. suppose the equivalence classes $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots, \mathcal{E}_r, \mathcal{E}_1$ in that order describe a directed cycle. Then, in the graph $\mathcal{G}_c$, for vertex w_j chosen from equivalence class $\mathcal{E}_j$ for $j = 1, 2, \ldots, r$, we have w_j has a directed edge to w_{j+1} for $j = 1, 2, \ldots, r-1$ and w_r has a directed edge to w_1. Transitive closure of $\mathcal{G}_c$ implies that there must be a directed edge from w_j to w_k for $j, k = 1, 2, \ldots, r, j \neq k$, leading to a collapse of the $r \geq 2$ equivalence classes into one equivalence class, a contradiction.

We now describe an algorithm $\mathcal{P}$ that fills the cells of a $k \times 2$ table with vertex names from $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_k$ such that the following properties hold:

(α) Each vertex shows up exactly once in the table.

(β) Each row of the table is made up of vertices u_i and v_i for some i.

(γ) In graph $\mathcal{G}_c$, vertex u obtained from the first column of row i does not have an edge to vertex v obtained from the second column of row j whenever $i \geq j$.

A directed acyclic graph has at least one sink vertex, i.e. a vertex with no outgoing edges. This is the proposed algorithm $\mathcal{P}$.

(1) Pick any sink vertex in directed acyclic graph $\mathcal{G}_r$.

(2) List the vertices of $\mathcal{G}_c$ in the equivalence class represented by the chosen sink vertex.

  (a) Pick a vertex w from the list.

  (b) If vertex w has been entered previously in the table, do nothing. Else, add vertex w in the first column of the lowest row in the table not yet filled. Add the destination of vertex w in the second column of the same row, i.e. if v_3 was entered in the first column of the lowest available row, then fill u_3 in the second column.

  (c) Remove w from the list and go back to (a) if the list is still non-empty, else proceed to (3)

(3) Modify graph $\mathcal{G}_r$ by deleting the chosen sink vertex. The modified graph continues to be a directed acyclic graph. If this graph has non-zero number of vertices, go to step (1), else quit.

Let us verify the claimed properties. By step (b), it is clear that each non-empty row of the table is filled with a vertex and its destination, i.e. the vertices u_i and v_i for some i. As the algorithm terminates only when all vertices have been listed and checked for their presence in the table, and as the vertices are added only when they have not been added previously, it follows that each vertex shows up exactly once and the table is completely filled upon termination of the algorithm. This verifies claimed properties (α) and (β). Now, we verify property (γ).

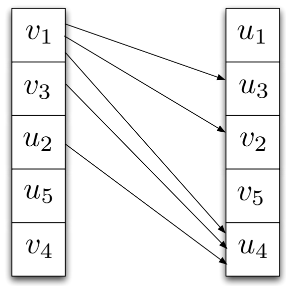


Figure 3.3: Connectivity pattern

One of the several 5×2 tables generated by running algorithm $\mathcal{P}$ on the $\mathcal{G}_c, \mathcal{G}_r$ shown in Fig. 3.1, Fig. 3.2 respectively. The order of choosing sinks was v_4, u_5, $u_2v_3u_4$, u_3v_2, v_1, u_1v_5. The arrows show connectivity from the vertices on the left to the vertices on the right in $\mathcal{G}_c$. Note that no arrows go 'horizontally' or go 'upward'. They always go 'downward' which is the desired GNS-cut property.

- Consider vertices in row j of the table, say w and w' with w in the first column. These are source-destination pairs u_i, v_i for some i. We claim that w has no edge to w' in $\mathcal{G}_c$. Suppose it did. Then, there would be an edge in $\mathcal{G}_r$ from the equivalence class $\mathcal{E}$ containing w to the equivalence class $\mathcal{E}'$ containing w'. These equivalence classes must be distinct as w and w' are source-destination pairs. This means that the algorithm $\mathcal{P}$ must pick the equivalence class containing w' before picking the equivalence class containing w. When w' is probed in the list of vertices, w must not have been entered into the table as yet, and thus w' would then be entered in the first column of some row and w in the second column of the same row. This contradicts the assumed structure of the table. Thus, we have no edge from w to w' in $\mathcal{G}_c$.

- Now, consider rows i and j with $i > j$. Let the vertices in row i be w and w' with w in the first column and the vertices in row j be z and z' with z in the first column. We claim that there is no edge from w to z' in $\mathcal{G}_c$. Suppose there is. Then, either w and z' are in the same equivalence class in $\mathcal{G}_c$ or they are not. If they are not, then the equivalence class containing z' has an incoming edge from the equivalence class containing w and thus, the former ought to have been picked by the algorithm before the latter. This is inconsistent with the table which was filled with w in the first column of a row while z' had not yet been filled. Now, if w and z' are in the same equivalence class, then clearly z does not fall in that equivalence class. Moreover, the equivalence class containing z is picked after the equivalence class containing w and z'. The algorithm $\mathcal{P}$, when exhausting the list of vertices in the equivalence class containing w and z' is supposed to have accepted w and added it to the first column of a row and rejected z'. But when z' was probed, we are still in the same equivalence class as w, so z had not been probed yet. Then z' must have been added to the first column of some row, which contradicts the structure of the table. Thus, there is no edge from w to z' in $\mathcal{G}_c$.

Now, if the j^{th} row of the table consists of u_i, v_i, we set $\pi(j) = i$ and $(w_i, w_i') = (u_i, v_i)$ or (v_i, u_i) depending on whether the first entry in the row is u_i or v_i. This shows that S is a GNS-cut for $\{w_1, w_2, \ldots, w_k; w_1', w_2', \ldots, w_k'\}$ with permutation π. This gives $\mathcal{R}_{\text{w.e.c.}} \supseteq \mathcal{R}_{\text{GNS-cut}}$ and completes the proof.

3.4 k-groupcast directed networks: Sum-rate

Definition: A *k-groupcast directed network* is a $k(k-1)$-unicast directed network $\mathcal{N}$ with k distinct distinguished nodes (group-nodes) $v_1, v_2, \ldots, v_k$ with source-destination pairs $\{s_{(i,j)}; d_{(i,j)}\}_{(i,j)\in\mathcal{I}}$ where $\mathcal{I} = \{(i,j) : 1 \leq i, j \leq k, i \neq j\}$ and $s_{(i,j)} = v_i, d_{(i,j)} = v_j$.

Theorems 11 and 12 and Corollary 3 will refer to k-groupcast directed networks.

Theorem 11 *(Naor-Zosin [51])*

$$\frac{1}{2}R_{\text{multicut}} \leq F_{\text{sum-rate}} \leq R_{\text{multicut}} \tag{3.9}$$

Theorem 12

$$C_{\text{sum-rate}} \leq 2R_{\text{multicut}} \tag{3.10}$$

Theorems 11 and 12 together imply that routing flow is approximately capacity-achieving for sum-rate:

Corollary 3

$$\frac{1}{2}R_{\text{multicut}} \leq F_{\text{sum-rate}} \leq C_{\text{sum-rate}} \leq 2R_{\text{multicut}} \tag{3.11}$$

We give the proof of Theorem 12.

Proof: Consider a k-groupcast directed network $\mathcal{N}$ with group-nodes $v_1, v_2, \ldots, v_k$. Let E be a set of edges such that $\mathcal{G} \setminus E$ has no directed paths from v_i to v_j for each $(i,j) \in \mathcal{I}$. Let $(R_{(i,j)} : (i,j) \in \mathcal{I}) \in \mathcal{C}$. Observe that E is a GNS-cut for source-destination pairs $\{s_{(i,j)}; d_{(i,j)}\}_{(i,j)\in\mathcal{I}:i>j}$. Theorem 9 gives $\sum_{(i,j)\in\mathcal{I}:i>j} R_{(i,j)} \leq \sum_{e\in E} C_e$. Similarly, we can get $\sum_{(i,j)\in\mathcal{I}:i<j} R_{(i,j)} \leq \sum_{e\in E} C_e$. Adding, we obtain $\sum_{(i,j)\in\mathcal{I}} R_{(i,j)} \leq 2\sum_{e\in E} C_e$, which completes the proof.

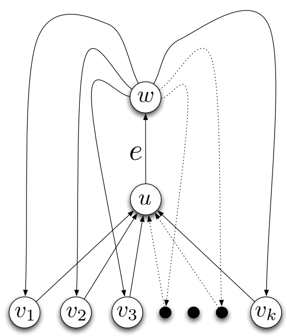


Figure 3.4: Example of a groupcast network

Remark: For the groupcast network in Fig. 3.4 with all edges having unit capacity, a simple XOR coding scheme achieves the rate tuple given by

$$R_{(i,j)} = \begin{cases} 1 & \text{if } (i,j) = (1,2) \text{ or } (2,1) \\ 0 & \text{otherwise,} \end{cases} \tag{3.12}$$

is achievable, while the multicut $R_{\text{multicut}} = 1$. This shows that the factor 2 in the inequality of Theorem 12 cannot be improved upon.

3.5 Discussion

It is intriguing that the kind of symmetry that allows results suggesting the closeness of flow and edge-cuts also leads to the near-fundamentality of such edge-cuts. It would be interesting to see whether there is a deeper explanation of this phenomenon.

Chapter 4

Wireless Networks: Network Capacity under Traffic Symmetry

The capacity region of multiple unicast in general wireless networks is an interesting open problem. Recent work [34], [35], [36] has made progress in this direction by giving an approximate characterization of this capacity region by using the reciprocity in wireless channels. It has been shown that simple layered architectures involving local physical-layer schemes combined with global routing can achieve approximately optimal performance in wireless networks.

In many practical scenarios, the reciprocity may be affected due to asymmetric power constraints, directional antennas, or frequency-duplexing. The question we address in this chapter is: "do layered architectures continue to be optimal even in this case?" We answer this question in the affirmative under the *symmetric demands* model: there are k specially-marked pairs of nodes $(s_i, d_i), i = 1, 2, ..., k$ with s_i wanting to communicate an independent message to d_i at rate R_i *and* d_i wanting to communicate an independent message to s_i at rate R_i. This traffic model is valid in several practical scenarios including voice calls, video calls, and interactive gaming.

As mentioned in Chapter 3, the symmetric demands traffic model was originally studied for *wireline networks* by Klein, Plotkin, Rao, and Tardos [37], who established that the routing rate region and *edge-cuts* are within a factor $\kappa \log^3(k + 1)$ of each other, where κ is a universal constant. This result, however, does not establish that routing is approximately optimal since edge-cuts do not, in general, bound the rate of general communication schemes. Chapter 3 showed that edge-cuts do in fact form fundamental upper bounds for the communication rates under this traffic model.

In this chapter, we prove this result for wireless networks under several channel models for which good schemes are known at a local level. Our results for wireless networks with symmetric demands include:

- Capacity approximations for networks comprised of Gaussian MAC and broadcast channels,

- Degrees-of-freedom approximation for fixed Gaussian networks, and
- Capacity approximations for fading Gaussian networks.

At the heart of our achievable scheme is a connection to "polymatroidal networks" for which the symmetric demands problem was recently addressed [11]. Our outer bound is based on the Generalized Network Sharing bound which we extend to wireless networks in this chapter.

This chapter is organized as follows. We first study a special class of Gaussian networks that we call MAC+BC networks in Section 4.1. Then, we move to general Gaussian networks in Section 4.2, with a study of fixed Gaussian networks and ergodic Gaussian networks.

4.1 Gaussian networks composed of Broadcast and Multiple Access Channels

We set up a network model for this class of networks that we call MAC+BC networks. The communication network is represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and an edge coloring $\psi : \mathcal{E} \to C$, where C is the set of colors. Each node v has a set of colors $C(v) \subseteq C$ on which it operates. Each color can be thought of as an orthogonal resource, so that the broadcast and interference constraints for the wireless channel apply only *within* a given color. The set of edges A_c corresponding to color c interact with each other and can be said to constitute a channel.

The channel model can therefore be written as,

$$y_i^c(t) = \sum_{j \in \mathsf{In}_c(i)} h_{ji}^c x_j^c(t) + z_i^c(t) \quad \forall c \in C(i), \tag{4.1}$$

where $x_i^c(t), y_i^c(t), z_i^c(t)$ are the transmitted vector, received vector, and noise vector on color c at time instant t, h_{ji}^c is the (fixed) channel coefficient between node i and node j on color c and $\mathsf{In}_c(i)$ represents the set of in-neighbors of node i who are operating on color c. We denote by y_i the vector comprised of $\{y_i^c\}$ for all $c \in C(i)$. We do not assume any symmetry in the channels, so that, in general h_{ij}^c may be different from h_{ji}^c.

We say that a given $c \in C$ corresponds to a multiple access channel (MAC) of degree d, if the set of edges A_c is of the form $A_c = \{i_1 j, i_2 j, ..., i_d j\}$, i.e., all edges are directed towards a particular node j. Similarly a channel c is said to correspond to a broadcast channel (BC) of degree d if $A_c = \{ij_1, ij_2, ..., ij_d\}$ for some node i. If A_c is a singleton set, we say that the channel c corresponds to an orthogonal link. A network is said to be a MAC+BC network if the set C can be decomposed as $C = \mathcal{M} \cup \mathcal{B}$, where $\mathcal{M}$ is the set of MAC channels and $\mathcal{B}$ is the set of broadcast channels or orthogonal links. Stated alternately, a network is composed of broadcast and multiple access channels if and only if no edge is involved simultaneously in a broadcast and interference constraint inside the same color. We will call such a network a "Gaussian MAC+BC network". An example of such a network is shown in Fig. 4.1.

Each node i has an average power constraint P to transmit for each color that it transmits in. If there are distinct power constraints for different nodes, they can be absorbed into the channel co-efficient without loss of generality. We assume that the channel h_{ij}^c is fixed (time-invariant) and is known at all the nodes.

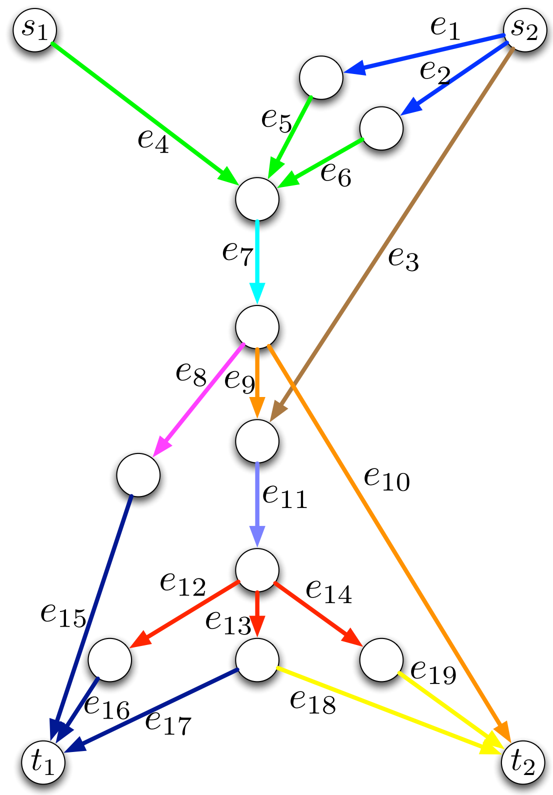


Figure 4.1: Example of a MAC+BC Gaussian network

Multiple Unicast in Gaussian MAC+BC Networks

A k-unicast Gaussian network has k pairs of nodes $s_i, d_i,\ i = 1, 2, ..., k$, where node s_i has a message to send to d_i and d_i has an independent message to send to s_i, both at rate R_i. We would like to characterize the closure of the set of all achievable rate tuples, called the capacity region $\mathcal{C}$. We will use $\mathcal{R}_{\text{ach}}$ to denote rates achievable by a simple scheme that we will propose. Also, for clarity, we will use $\mathcal{X}_v, \mathcal{Y}_v$ to denote the input and output alphabets at node v, implicitly understanding that in the Gaussian network, $\mathcal{X}_v = \mathbb{R}^a, \mathcal{Y}_v = \mathbb{R}^b$, for suitable integers a and b. Formally, a $(\lceil 2^{TR_1} \rceil, \lceil 2^{TR_2} \rceil, \ldots, \lceil 2^{TR_k} \rceil, T)$ coding scheme for this network which communicates over T time instants is comprised of the following.

1. Independent random variables W_i which are distributed uniformly on $\mathcal{W}_i := \{1, \ldots, \lceil 2^{TR_i} \rceil\}$ for $i = 1, \ldots, k$ respectively. W_i denotes the message intended from source s_i to destination d_i.

2. The source mapping at source s_i for time t,
$$f_{s_i,t} : (\mathcal{W}_i, \mathcal{Y}_{s_i}^{t-1}) \to \mathcal{X}_{s_i}^t. \tag{4.2}$$

3. The relay mappings for each $v \in \mathcal{V} \backslash \{s_1, d_1, s_2, d_2, \ldots, s_k, d_k\}$ and time t,
$$f_{v,t} : \mathcal{Y}_v^{t-1} \to \mathcal{X}_v^t. \tag{4.3}$$

4. The decoding map at destination d_i,
$$g_{d_i} : \mathcal{Y}_{d_i}^T \to \mathcal{W}_i. \tag{4.4}$$

If $\hat{W}_i$ is the decoded symbol at d_i, then the probability of error for destination d_i under this coding scheme is given by
$$P_e^i \ := \ \Pr\{\hat{W}_i \neq W_i\}. \tag{4.5}$$

A rate tuple $(R_1, R_2, \ldots, R_k)$, where R_i is the rate of communication in bits per unit time from source s_i to destination d_i, is said to be achievable if for any $\epsilon > 0$, there exists a $(\lceil 2^{TR_1} \rceil, \lceil 2^{TR_2} \rceil, \ldots, \lceil 2^{TR_k} \rceil, T)$ scheme that achieves a probability of error lesser than ϵ for all nodes, i.e., $\max_i P_e^i \leq \epsilon$. The capacity region $\mathcal{C}$ is the closure of the set of all achievable rate tuples.

The capacity region can similarly be defined for the k-unicast problem with symmetric demands: For each $i = 1, 2, \ldots, k$, s_i has a message to be communicated to d_i at rate R_i and d_i has an independent message to be communicated to s_i at the same rate R_i.

Weak edge-cut bound

Similar to the wireline network case, we define a weak edge-cut bound region for the wireless network with demand symmetry. The weak edge-cut bound region for the Gaussian MAC+BC network is defined by the following: consider any set of edges $F \subseteq \mathcal{E}$, and let $K(F)$ denote the set of $i \in \{1, 2, ..., k\}$ such that either there is no path from s_i to d_i or there is no path from d_i to s_i in $\mathcal{G} \setminus F$. The value of the cut F is defined by $\nu(F) := \sum_c \nu(F^c)$, where $F = \cup_c F^c$ with F^c being the set of edges that participate in color c and $\nu(F^c)$ is the capacity under complete coordination of source nodes in channel c. More formally, if c is a broadcast channel, $\nu(F^c)$ is equal to the sum-capacity of the broadcast channel specified only by edges in F^c, under complete coordination of destination terminals in F^c. Similarly, if c is a MAC channel, $\nu(F^c)$ is equal to the sum-capacity of the MAC channel specified by edges in F^c, under complete coordination of source terminals in F^c.

The weak edge-cut bound region is now given as

$$\mathcal{R}_{\text{w.e.c.}} = \{(R_1, ..., R_k) : \sum_{i \in K(F)} R_i \leq \nu(F) \ \forall F \subseteq \mathcal{E}\}.$$

As in the wireline network case, it is not clear if $\mathcal{R}_{\text{w.e.c.}}$ is an outer bound to the capacity region $\mathcal{C}$.

Our main result is the following:

Theorem 13 *For the k-unicast problem with symmetric demands in a Gaussian MAC+BC network, the weak edge-cut bound is a fundamental outer bound on the capacity region and a simple separation strategy can achieve $\mathcal{R}_{ach}(P)$ which satisfies,*

$$\frac{\mathcal{R}_{\text{w.e.c.}}(\frac{P}{d_{\max}})}{\kappa \log^3(k+1)} \subseteq \mathcal{R}_{ach}(P) \subseteq \mathcal{C}(P) \subseteq \mathcal{R}_{\text{w.e.c.}}(P), \tag{4.6}$$

where κ is a universal constant independent of problem parameters and $d_{\max}$ is the maximum degree of any broadcast or MAC component channel c.

Thus, the edge-cut bound is a fundamental upper bound on the capacity region and furthermore the edge-cut bound, scaled down in power by a factor $d_{\max}$ and scaled down in rate by a factor $\frac{1}{\kappa \log^3(k+1)}$, can be achieved by the proposed scheme.

Outer bound

We first establish that the weak edge-cut bound is fundamental, i.e., every communication scheme must have rate pairs that lie inside this region: $\mathcal{C} \subseteq \mathcal{R}_{\text{w.e.c.}}$. We will prove this result using a GNS bound for Gaussian networks.

Given an ℓ-unicast MAC+BC Gaussian network with source destination pairs $\{s_i, d_i\}_{i=1}^{\ell}$, we define a set of edges $F \subseteq \mathcal{E}$ to be a GNS-cut if there exists a permutation $\pi : \{1, 2, \ldots, \ell\} \to \{1, 2, \ldots, \ell\}$ such that there are no paths from s_i to d_i in $\mathcal{G} \setminus F$, whenever $\pi(i) \geq \pi(j)$.

Lemma 1 *(GNS bound for MAC+BC Gaussian networks) For an ℓ-unicast Gaussian MAC+BC network, every GNS cut F is fundamental, i.e., $\sum_{i \in K(F)} R_i \leq \nu(F)$ for any communication scheme achieving $(R_1, ..., R_\ell)$. Alternately $\mathcal{C} \subseteq \mathcal{R}_{\text{GNS}-\text{cut}}$, where*

$$\mathcal{R}_{\text{GNS}-\text{cut}} = \cap_{F \subseteq \mathcal{E}} \{(R_1, .., R_\ell) : \sum_{i \in K(F)} R_i \leq \nu(F)\}. \tag{4.7}$$

An instance of Lemma 1 can be found in Fig. 4.2.

Proof: Let F be a GNS-cut disconnecting s_i from d_i for $i = 1, 2, \ldots, \ell$ with say, the identity permutation π_{id}. Thus, $K(F) = \{1, 2, \ldots, \ell\}$. A similar proof holds in the case when $K(F) \subset$

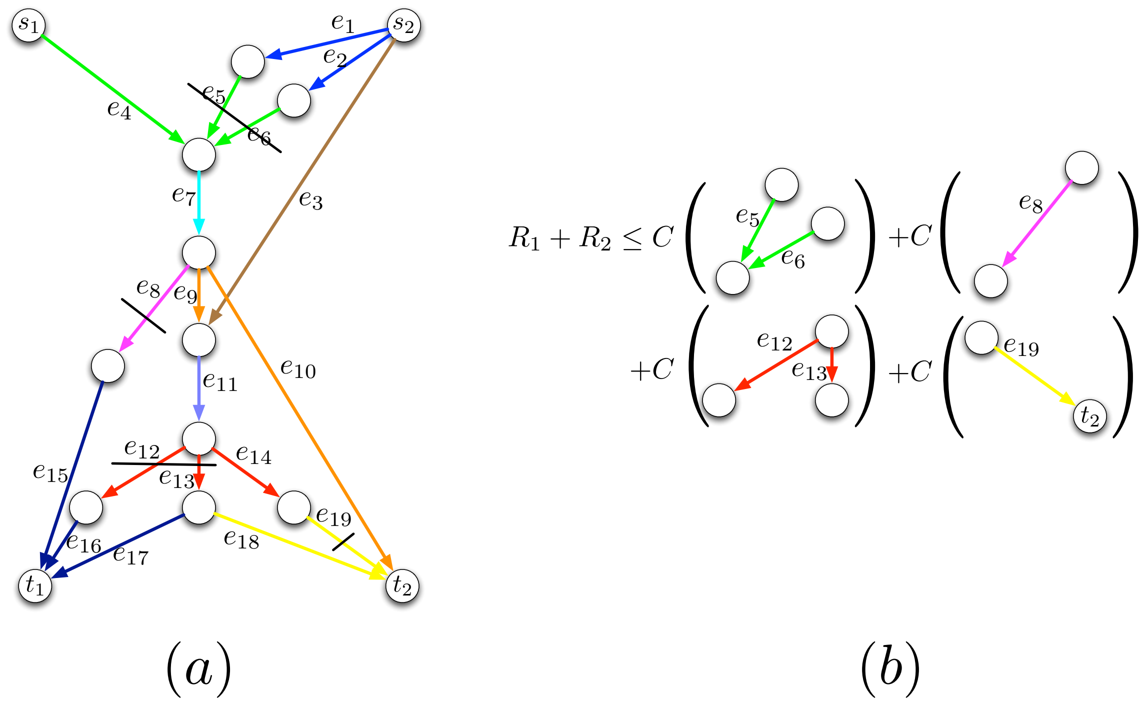


Figure 4.2: GNS bound for MAC+BC Gaussian networks

The set of edges $\{e_5, e_6, e_8, e_{12}, e_{13}, e_{19}\}$ forms a GNS-cut in the two-unicast network of (a). The outer bound on the capacity region of the two-unicast network, that can be derived from the GNS-cut is shown in (b).

$\{1, 2, \ldots, \ell\}$. We first provide a proof of the GNS bound when the network has an acyclic underlying graph $\mathcal{G}$.

Recall that the set of colors $C = \mathcal{M} \cup \mathcal{B}$ where $\mathcal{M}$ consists of the colors of edges involved in MAC components and $\mathcal{B}$ consists of colors of edges involved in broadcast components or orthogonal links. For $\mu \in C$, let A_μ denote the set of edges involved in μ. Now, construct a directed graph $\mathcal{G}'$ as follows: for each $\mu \in C$, there is a node in $\mathcal{G}'$ and add a directed edge from node μ to node ν in $\mathcal{G}'$ if there exists an edge in A_μ that is upstream to some edge in A_ν in the original DAG $\mathcal{G}$. Since the set of all edges with a given color constitute either a MAC or a BC, we have that $\mathcal{G}'$ is a directed acyclic graph. Thus, we can have a total order on the vertices of $\mathcal{G}'$ consistent with the partial order of ancestry in $\mathcal{G}'$. This gives a total order on C and therefore also a total order on the subset $\mathcal{D} := \{\mu \in C : F \cap A_\mu \neq \emptyset\}$, which we will denote by $\mu_1 < \mu_2 < \ldots < \mu_r$, where μ_1 is the most "upstream".

- For $\mu \in \mathcal{M}$, we denote transmissions along edge e in A_μ by X_e and we denote the reception by Y_μ so that $Y_\mu = \sum_{e \in A_\mu} X_e + Z_\mu$ where Z_μ is Gaussian noise. Further, define $U_\mu := \{X_e : e \in F \cap A_\mu\}$, and $V_\mu := \sum_{e \in F \cap A_\mu} X_e + Z_\mu$. Intuitively, U_μ (V_μ) is the transmission (reception) on channel μ if only edges in F were present in the channel.

- For $\mu \in \mathcal{B}$, we denote the transmission on the broadcast component or orthogonal link by X_μ and the receptions at heads of $e \in A_\mu$ by Y_e so that $Y_e = X_\mu + Z_e$ where $\{Z_e, e \in A_\mu\}$ are independent Gaussian noise random variables. Further define $U_\mu := X_\mu$, and $V_\mu := \{Y_e : e \in F \cap A_\mu\}$.

Define $\tilde{Y}_{d_i} = \{Y_\mu^n : \mathsf{head}(e) = d_i, e \in A_\mu, \mu \in \mathcal{M}\} \cup \{Y_e^n : \mathsf{head}(e) = d_i, e \in A_\mu, \mu \in \mathcal{B}\}$.

$$
\begin{aligned}
n[\sum_{i=1}^{\ell} R_i - \epsilon_n] &\leq \sum_{i=1}^{\ell} I(W_i; \tilde{Y}_{d_i}) \\
&\leq \sum_{i=1}^{\ell} I(W_i; \{V_\mu^n : \mu \in \mathcal{D}\}, \{W_j : j < i\}) \\
&\qquad [\text{since } W_i - \{V_\mu^n : \mu \in \mathcal{D}\}, \{W_j : j < i\} - \tilde{Y}_{d_i} \\
&\qquad \text{as } F \text{ is a GNS cut with identity permutation}] \\
&= \sum_{i=1}^{\ell} I(W_i; \{V_\mu^n : \mu \in \mathcal{D}\} | \{W_j : j < i\}) \\
&\qquad [\text{since } W_i \text{ is independent of } \{W_j : j < i\}] \\
&= I(\{W_i : 1 \leq i \leq \ell\}; \{V_\mu^n : \mu \in \mathcal{D}\}) \\
&= h(\{V_\mu^n : \mu \in \mathcal{D}\}) \\
&\quad - h(\{V_\mu^n : \mu \in \mathcal{D}\} | \{W_i : 1 \leq i \leq \ell\}) \\
&\leq \sum_{\mu \in \mathcal{D}} h(V_\mu^n) - h(\{V_\mu^n : \mu \in \mathcal{D}\} | \{W_i : 1 \leq i \leq \ell\}) \\
&=: \sum_{\mu \in \mathcal{D}} h(V_\mu^n) - A.
\end{aligned}
$$

Now, we consider the negative term A above.

$$
\begin{aligned}
A &= h(\{V_\mu^n : \mu \in \mathcal{D}\}|\{W_i : 1 \leq i \leq \ell\}) \\
&= h(V_{\mu_1}^n, V_{\mu_2}^n, \ldots, V_{\mu_r}^n|\{W_i : 1 \leq i \leq \ell\}) \\
&\geq h(V_{\mu_1}^n, V_{\mu_2}^n, \ldots, V_{\mu_r}^n|\{W_i : 1 \leq i \leq k\}, U_{\mu_1}^n) \\
&= h(V_{\mu_1}^n|\{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n) \\
&\quad + h(V_{\mu_2}^n, \ldots, V_{\mu_r}^n|\{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n, V_{\mu_1}^n) \\
&= h(V_{\mu_1}^n|U_{\mu_1}^n) \\
&\quad + h(V_{\mu_2}^n, \ldots, V_{\mu_r}^n|\{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n, V_{\mu_1}^n) \\
&\qquad\qquad [\text{since } \{W_i : 1 \leq i \leq \ell\} - U_{\mu_1}^n - V_{\mu_1}^n \\
&\qquad\qquad\quad \text{as } \mu_1 \text{ is the most upstream channel}] \\
&\geq h(V_{\mu_1}^n|U_{\mu_1}^n) \\
&\quad + h(V_{\mu_2}^n, \ldots, V_{\mu_r}^n|\{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n, V_{\mu_1}^n, U_{\mu_2}^n) \\
&= h(V_{\mu_1}^n|U_{\mu_1}^n) + h(V_{\mu_2}^n|\{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n, V_{\mu_1}^n, U_{\mu_2}^n) \\
&\quad + h(V_{\mu_3}^n, \ldots, V_{\mu_r}^n|\{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n, V_{\mu_1}^n, U_{\mu_2}^n, V_{\mu_2}^n) \\
&= h(V_{\mu_1}^n|U_{\mu_1}^n) + h(V_{\mu_2}^n|U_{\mu_2}^n) \\
&\quad + h(V_{\mu_3}^n, \ldots, V_{\mu_r}^n|\{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n, V_{\mu_1}^n, U_{\mu_2}^n, V_{\mu_2}^n) \\
&\qquad\qquad [\text{since } \{W_i : 1 \leq i \leq \ell\}, U_{\mu_1}^n, V_{\mu_1}^n - U_{\mu_2}^n - V_{\mu_2}^n \\
&\qquad\qquad\quad \text{as only } \mu_1 \text{ could be more upstream than } \mu_2] \\
&\geq \sum_{\mu \in \mathcal{D}} h(V_\mu^n|U_\mu^n),
\end{aligned}
$$

from repeating these steps. Thus, we obtain

$$
\begin{aligned}
n[\sum_{i=1}^{\ell} R_i - \epsilon_n] &\leq \sum_{\mu \in \mathcal{D}} I(U_\mu^n; V_\mu^n) && (4.8) \\
&\leq n \sum_{\mu \in \mathcal{D}} \nu(F_\mu) && (4.9) \\
&= n\nu(F), && (4.10)
\end{aligned}
$$

and therefore it follows that the GNS bound is a fundamental upper bound on the capacity region for acyclic networks.

For a general cyclic network, we can employ a standard time-layering argument in order to complete the proof. While the details of our method and the use of time-layering to deal with cyclic networks are fairly standard, see [3, 8], one key difference is that here we are using time-layering in order to prove an outer-bound, whereas the earlier works utilized time layering to show achievability. We will provide a brief sketch of the method here. Given a cyclic network $\mathcal{G}$ and a coding scheme over n time instants, we construct a time-layered graph $\mathcal{G}^n$ as follows. The nodes in the graph $\mathcal{G}^n$ are arranged in $n+1$ layers $0, 1, ..., n$. For each i, layer i has a copy of all the nodes V in the original graph, we call this $V[i]$ and the copy of node v in layer i is called $v[i]$. Add directed edges in the graph in the following manner.

- For each $(u, v) \in \mathcal{E}$ in the original graph with channel coefficient h_{vu}^c on color c, we add edges $(u[i], v[i+1])$ for $i = 0, 1, ..., n-1$ with channel coefficient $h_{vu}^{c_i}$ on color c_i.
- Create an edge from $v[i]$ to $v[i+1]$ for each v of infinite capacity in an independent channel (in order to model memory of the link).

Thus the time-layered graph $\mathcal{G}^n$ is created. This graph defines an instance of a Gaussian MAC+BC network, which is acyclic. For this new graph, we define a communication problem by specifying that sources $s_1[0], ..., s_\ell[0]$ wish to communicate independent information to destinations $d_1[n], d_2[n], ..., d_\ell[n]$. Observe that any scheme on the original network utilizing n time instants gives a valid scheme on this graph $\mathcal{G}^n$. Thus upper bounds on the communication rates in this graph serve as upper bounds to $n(R_1, R_2, ..., R_\ell)$ whenever $(R_1, R_2, \ldots, R_\ell)$ lies in the capacity region of the original network. Now given a GNS cut on the original graph with the identity permutation, defined by a set of edges F, we can define a cut on this graph $\mathcal{G}^n$ by $F^n := \cup_{i \in [n]} F[i]$, where $F[i] = \{(u[i-1], v(i), \ \forall (u, v) \in F\}$. If F disconnected source s_a from destination d_b in the original graph, this cut F^n disconnects $s_a[0]$ from $d_b[n]$ in the time-layered graph because any remaining path from $s_a[0]$ to $d_b[n]$ would imply a path in the original graph from s_a to d_b. This implies that any GNS cut on the original graph can produce a GNS cut on $\mathcal{G}^n$ with n times the value, as each edge occurs n times in F^n. Since the rate is also scaled by n times in this time layered graph, this proves that the GNS bound is a valid upper bound on the rate of an arbitrary (cyclic) graph.

Now that we know that GNS-cut is a valid upper bound on the Gaussian network for any ℓ unicast problem, we will define the GNS-cut for a symmetric demands problem (in exactly the same way as it was defined for wireline networks in Chapter 3).

Definition: Given a k-pair unicast directed symmetric-demand MAC + BC Gaussian network with source-destination nodes $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_k$, we define the *GNS-cut outer bound* denoted by $\mathcal{R}_{\mathrm{GNS-cut}}$, to be the set of all non-negative tuples $(R_i : 1 \leq i \leq k)$ that satisfy for every $F \subseteq \mathcal{E}(\mathcal{G})$, the inequality $\sum_{i \in J} R_i \leq \nu(F)$ whenever F is a GNS-cut for $\{w_1, w_2, \ldots, w_r; w_1', w_2', \ldots, w_r'\}$ with some permutation π where

- $J \subseteq \{1, 2, \ldots, k\}, |J| = r$,
- $w_1, w_2, \ldots, w_r, w_1', w_2', \ldots, w_r'$ are distinct,
- for $1 \leq j \leq r$, $(w_j, w_j') = (u_i, v_i)$ or (v_i, u_i) for some $i \in J$.

By Theorem 10 of Chapter 3, we have that $\mathcal{R}_{\text{w.e.c.}} = \mathcal{R}_{\text{GNS-cut}}$. Using this result in conjunction with Lemma 1, gives us the desired result,

$$\mathcal{C} \subseteq \mathcal{R}_{\text{w.e.c.}}. \tag{4.11}$$

We parametrize both $\mathcal{C}$ and $\mathcal{R}_{\text{w.e.c.}}$ by power constraint P in order to emphasize its dependence.

Coding Scheme

The coding scheme we propose is a separation-based strategy: each component broadcast or multiple access channel is coded for independently creating bit-pipes on which information is routed globally. In order to evaluate the rate region of this scheme, we use polymatroidal networks as an interface for which we can show that the flow region corresponding to routing and the bounding region defined by edge-cuts are close to each other. For simplicity of notation, we will assume that all MAC and broadcast channels have degree $d = d_{\max}$. It will be clear from the details that this assumption is not necessary.

For a finite set V, a set function $f : 2^V \mapsto \mathbb{R}$ is said to satisfy the *polymatroidal axioms* if

- $f(\emptyset) = 0$,
- $A \subseteq B \implies f(A) \leq f(B)$,
- The function f is submodular, i.e. for any two sets $A, B \subseteq V$, $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$.

A bounding region $\mathcal{B}$ defined over $\mathbb{R}^{|\mathbb{V}|}$ is said to be a *polymatroidal region* if it is of the form $\mathcal{B} = \{(R_v : v \in V) : R_v \geq 0 \text{ and for any } A \subseteq V, \sum_{v \in A} R_v \leq f(A)\}$ for some functin f that satisfies the polymatroidal axioms.

Let us first consider the coding for the multiple access channel with channel coefficients $h_1, ..., h_d$ and power constraint P at each of the d nodes. Let the rate region achievable on this multiple access channel be denoted by

$$\mathcal{R}_{\text{ach}}^{\text{MAC}}(P) = \{\overline{R} : \sum_{i \in A} R_i \leq \log\left(1 + \sum_{i \in A} |h_i|^2 P\right) \ \forall A\}. \tag{4.12}$$

This region is known to be polymatroidal. The outer bound for the MAC under arbitrary source cooperation is given by

$$\mathcal{R}_{\text{cut}}^{\text{MAC}}(P) = \{\overline{R} : \sum_{i \in A} R_i \leq \log\left(1 + \left(\sum_{i \in A} |h_i|\right)^2 P\right) \ \forall A\}. \tag{4.13}$$

The capacity region of a broadcast channel with channel coefficients $h_1, ..., h_k$ and power constraint P is not a polymatroidal region. However, it can be approximated by a polymatroidal region [34]. In particular, the achievable region includes the following polymatroidal region, and we will restrict our broadcast channel scheme to operate inside the following polymatroidal region, as we will show that this is not too far from the cutset outer bound:

$$\mathcal{R}_{\text{ach}}^{\text{BC}}(P) = \{\overline{R} : \sum_{i\in A} R_i \leq \log\left(1 + \sum_{i\in A} |h_i|^2 \frac{P}{d}\right)\ \forall A\}, \tag{4.14}$$

The cutset bound on the broadcast channel under arbitrary destination cooperation is

$$\mathcal{R}_{\text{cut}}^{\text{BC}}(P) = \{\overline{R} : \sum_{i\in A} R_i \leq \log\left(1 + \sum_{i\in A} |h_i|^2 P\right)\ \forall A\}. \tag{4.15}$$

It can be easily verified that $\mathcal{R}_{\text{cut}}^{\text{MAC}}(P) \subseteq \mathcal{R}_{\text{ach}}^{\text{MAC}}(dP)$ and $\mathcal{R}_{\text{cut}}^{\text{BC}}(P) \subseteq \mathcal{R}_{\text{ach}}^{\text{BC}}(dP)$. Now, each multiple access or broadcast channel can be replaced by a set of d bit-pipes whose rates are jointly constrained by the corresponding capacity constraints. It turns out that this falls inside a class of networks called polymatroidal networks, that have been already studied [12]. We will now give a short description of polymatroidal networks and some results for these networks.

Polymatroidal Networks

Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We have considered networks in the previous chapter with capacity constraints on the edges. A polymatroidal network has more general capacity constraints coupling edges that meet at a node. In the polymatroidal network, for each node v there are two associated submodular functions: ρ_v^{In} and ρ_v^{Out} which impose joint capacity constraints on the edges in $\mathsf{In}(v)$ and $\mathsf{Out}(v)$ respectively. That is, for any set of edges $S \subseteq \mathsf{In}(v)$, the total capacity available on the edges in S is constrained to be at most $\rho_v^{\mathsf{In}}(S)$. Similarly, ρ_v^{Out} constrains edges in $\mathsf{Out}(v)$.

For any subset of edges $F \subseteq \mathcal{E}$, we define the disconnection set $K(F)$ as the set of indices i for which source s_i has no paths to destination d_i in $\mathcal{G} \setminus F$. In standard networks, the value of the cut F is simply $\sum_{e\in F} c(e)$ where $c(e)$ is the capacity constraint on edge e. The value of a cut F in polymatroidal networks is defined as follows: each edge (u, v) in F is first assigned to either u or v; we say that an assignment of edges to nodes $g : F \to V$ is *valid* if it satisfies this restriction. Once this assignment is made, we can compute the value of the cut according to this assignment by evaluating the submodular functions corresponding to the set of edges grouped together. The value of the cut $\nu(F)$ is the minimum over all assignments, that is,

$$\nu(F) := \min_{g:F\to V, g \text{ valid}} \sum_v \left\{\rho_v^{\mathsf{In}}(\mathsf{In}(v) \cap g^{-1}(v)) + \rho_v^{\mathsf{Out}}(\mathsf{Out}(v) \cap g^{-1}(v))\right\}. \tag{4.16}$$

A max-flow min-cut theorem for the single unicast problem in directed polymatroidal networks is known in the literature [43, 27]. For the k-unicast problem in a directed polymatroidal network with symmetric demands, the following theorem is proved in [11], which generalizes the result of [37] from edge capacity constraints to polymatroidal capacity constraints. The weak edge-cut bound for symmetric-demand polymatroidal networks is defined similarly as for standard networks.

Theorem 14 *(from [11]) For a symmetric-demand directed polymatroidal network with k source-destination pairs, any rate tuple in the weak edge-cut rate region divided by a factor* $\kappa(\log^3(k+1))$ *is achievable by routing, where* κ *is a universal constant.*

If $\mathcal{R}^{\text{poly}}_{\text{w.e.c.}}$ stands for the weak edge-cut region in the polymatroidal network and $\mathcal{R}^{\text{poly}}_{\text{ach}}$ stands for the region achievable by flow in the polymatroidal network, then Theorem 14 can be rewritten as:

$$\frac{\mathcal{R}^{\text{poly}}_{\text{w.e.c.}}}{\kappa \log^3(k+1)} \subseteq \mathcal{R}^{\text{poly}}_{\text{ach}}. \tag{4.17}$$

Analysis of Achievable Rates in Gaussian MAC+BC Network

The proposed separation strategy converts the Gaussian MAC+BC network into a directed polymatroidal network with symmetric demands. Using the achievable rates for the corresponding MAC and BC channels from (4.12) and (4.14), we can see that this polymatroidal network has the following submodular functions at any given node v,

$$\rho_v^{\mathsf{In}}(S) = \sum_c \log\left(1 + \sum_{(uv)\in S} |h^c_{vu}|^2 P\right), \tag{4.18}$$

$$\rho_v^{\mathsf{Out}}(S) = \sum_c \log\left(1 + \sum_{(vu)\in S} |h^c_{vu}|^2 \frac{P}{d}\right). \tag{4.19}$$

This fully defines the polymatroidal network. Now any rate tuple achievable on this polymatroidal network is achievable in the Gaussian MAC+BC network using the proposed separation architecture.

Now we can use Theorem 14 to show that the achieved rate tuple in the Gaussian network is within a poly-logarithmic factor of the weak edge-cut bound in the polymatroidal network, i.e.

$$\frac{\mathcal{R}^{\text{poly}}_{\text{w.e.c.}}(P)}{\kappa \log^3(k+1)} \subseteq \mathcal{R}^{\text{poly}}_{\mathsf{flow}}(P) = \mathcal{R}^{\text{g}}_{\text{ach}}(P). \tag{4.20}$$

Here we have parametrized all the rate regions by the power constraint in order to make this dependence explicit. In order to prove our main result, we still need to connect the weak edge-cut bound in the polymatroidal network to the weak edge-cut bound in the Gaussian network.

We will connect the value of the cut in the polymatroidal network to the value of the cut in the Gaussian network. Let us take a cut derived from a set of edges F in the polymatroidal network and a valid assignment g of F, which yields the minimum among all possible valid assignments. In this assignment, each edge (u, v) of F is assigned either to u or to v. Thus for any node, some incoming edges are assigned together and some outgoing edges are assigned together and the value of the cut is given by

$$\nu(F) = \sum_v \{\rho_v^{\mathsf{In}}(\mathsf{In}(v) \cap g^{-1}(v)) + \rho_v^{\mathsf{Out}}(\mathsf{Out}(v) \cap g^{-1}(v))\}.$$

Note that each of these functions ρ_v^{In} and ρ_v^{Out} corresponds to the constraints in the achievable region of the original MAC and broadcast channels. If we take the cut corresponding to F in the original network, then these functions will be replaced by the functions corresponding to the cut in the MAC and broadcast channels, whose equations are given in (4.13) and (4.15). It has been observed earlier that,

$$\mathcal{R}_{\text{cut}}^{\text{MAC}}(P) \subseteq \mathcal{R}_{\text{ach}}^{\text{MAC}}(dP) \tag{4.21}$$
$$\mathcal{R}_{\text{cut}}^{\text{BC}}(P) \subseteq \mathcal{R}_{\text{ach}}^{\text{BC}}(dP). \tag{4.22}$$

Note that the value of any cut in the polymatroidal network is a function of the power constraint implicitly, since ρ^{In} and ρ^{Out} are functions of the power constraint. Let us call this function $v(P)$. Now if we look at the same cut in the Gaussian network then the value of this cut here is at most $v(dP)$ because of (4.21) and (4.22). Thus the weak edge-cut region in the polymatroidal network and the weak edge-cut region in the Gaussian network can be related to each other as follows,

$$\mathcal{R}_{\text{w.e.c.}}^{\text{g}}(P) \subseteq \mathcal{R}_{\text{w.e.c.}}^{\text{poly}}(dP), \tag{4.23}$$

or alternately $\mathcal{R}_{\text{w.e.c.}}^{\text{g}}(\frac{P}{d}) \subseteq \mathcal{R}_{\text{w.e.c.}}^{\text{poly}}(P)$. This result, when combined with (4.20) and (4.11), yields the following relationship,

$$\frac{\mathcal{R}_{\text{w.e.c.}}^{\text{g}}(\frac{P}{d})}{\kappa \log^3(k+1)} \subseteq \mathcal{R}_{\text{ach}}^{\text{g}}(P) \subseteq \mathcal{C} \subseteq \mathcal{R}_{\text{w.e.c.}}^{\text{g}}(P), \tag{4.24}$$

which completes the proof of Theorem 13.

4.2 General Gaussian Networks

In this section, we consider general Gaussian networks, i.e., networks where broadcast and MAC can occur simultaneously. Our basic idea will remain similar to what we did in the case of the Gaussian MAC+BC network, where we employed a separation strategy. A good physical layer scheme that approximately achieved the cut-set bound converted the

MAC+BC Gaussian network into a polymatroidal network on which routing was shown to be approximately optimal.

In order to provide guarantees on the performance of our layered architecture in general Gaussian networks, we need physical layer schemes that achieve close to the cut-set bound on the component channels. This class of networks includes as a special case the interference channel and the X-channel (where every transmitter has a message to transmit to every receiver). In a general Gaussian network, it is not clear what a component channel is. In [35], it was identified that the X-channel can be viewed as a basic component channel of a general Gaussian network. Good communication schemes are known for the X- channel under the following scenarios:

1) Degrees-of-freedom in fixed Gaussian channels

2) Capacity approximation in ergodic Gaussian channels.

As such, our network-level results also apply under these two settings.

Fixed Gaussian channels

The communication network is represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The edges (j, i) that are present have fading coefficients h_{ij} on them, each of which are chosen independently from a continuous fading distribution. Note that in a general Gaussian network, the presence of edge (j, i) does not necessarily imply the presence of edge (i, j). There are k pairs of specially designated nodes s_i, d_i such that s_i has a message to send to d_i at rate R_i and d_i has an independent message to send to s_i at the same rate R_i. The channel model can be written as

$$y_i(t) = \sum_{j \in \mathsf{In}(i)} h_{ij} x_j(t) + z_i(t) \; \forall t = 1, 2, ..., T, \tag{4.25}$$

where $x_i(t), y_i(t), z_i(t)$ are the transmitted vector, received vector, and noise vector at time t, and $\mathsf{In}(i)$ represents the set of neighbors of node i who have an incoming edge to i. The noise vector is assumed to have unit variance and is independent at each node. There is an average power constraint of P per node. The degrees-of-freedom (DOF) tuple $(d_1, ..., d_k)$ is said to be achievable if, for each P, there is a scheme achieving rate tuple $(R_1(P), ..., R_k(P))$ such that $d_i = \lim_{P \to \infty} \frac{R_i}{\log P}$. The closure of the set of all achievable DOF tuples is called the DOF region $\mathcal{D}$. Let $\mathcal{D}_{\text{ach}}$ denote the DOF tuples achievable by our specific strategy and let $\mathcal{D}_{\text{w.e.c.}}$ correspond to the weak edge-cut bound (which is defined formally below).

Weak Edge-Cut Bound:

In order to define the weak edge-cut bound in the general Gaussian network case, we will utilize the weak edge-cut bound region in the Gaussian MAC+BC network case. We will observe that there are several ways to upper bound the capacity of a given general Gaussian network by a related network of Gaussian MAC+BC channels. This can be done by "deactivation" of some of the MAC and BC constraints. Formally, consider a coloring on the set of edges such that the channel corresponding to each color is a MAC or broadcast channel. This coloring produces a Gaussian MAC+BC network whose DOF region contains the DOF region of the original network, as we shall show. For a MAC+BC network, each orthogonal MAC or BC component contributes a DOF of 1. Thus, the DOF value of the cut in the colored network is the total number of MAC or BC channels involved in the cut. Now, we define $d(F)$ as the tightest bound obtainable on the DOF by bounding it based on any suitable coloring giving a MAC+BC network. Finally, define $\mathcal{D}_{\text{w.e.c.}} = \{(d_1, ..., d_k) : \sum_{i:i \in K(F)} d_i \leq d(F) \; \forall F \subseteq \mathcal{E}\}$. It is not clear, as earlier, that $\mathcal{D}_{\text{w.e.c.}}$ is a fundamental upper bound on any achievable DOF tuple under symmetric demands.

Our main result is the following.

Theorem 15 *For a directed wireless network with symmetric demands, if the fixed channel coefficients are drawn from a continuous distribution, the DOF region given by $\mathcal{D}_{ach}$ satisfying*

$$\frac{\mathcal{D}_{\text{w.e.c.}}}{\kappa \log^3(k+1)} \subseteq \mathcal{D}_{ach} \subseteq \mathcal{D} \subseteq \mathcal{D}_{\text{w.e.c.}}, \tag{4.26}$$

is achievable, with probability 1.

Coding Scheme

Our coding scheme involves a conversion of the Gaussian network into a bit-pipe network (more specifically, a polymatroidal network) and then routing on this induced polymatroidal network.

First, consider an interference channel with l sources $s_1, s_2, \ldots, s_k$, wishing to communicate to their respective destinations $d_1, d_2, \ldots, d_k$. Suppose that the connectivity between sources and destinations is described by a suitable bipartite graph and that for each i, there is an edge from s_i to d_i. Then, it is well-known that the real interference alignment scheme [50] achieves half DOF per transmitter-receiver pair simultaneously, almost surely over the randomness in the channel co-efficients.

In our Gaussian network, we choose two subsets of nodes of equal size. We choose one set to be a set of "transmitter" nodes and the other set to be a set of "receiver" nodes. This creates for us an interference channel. We choose a suitable matching in the connectivity from the transmitters to the receivers to give us the transmitter-receiver pairs. This creates bit pipes that offer half DOF for each transmitter-receiver pair. We then time-share between all possible choices of subset pairs and over all possible matchings.

It has been shown in [35] that such a choice in fact leads to a simple polymatroidal network with the same set of of edges as the original Gaussian network, and with the following sub modular functions at each node, almost surely

$$\rho_v^{\mathsf{In}}(S) = \frac{1}{2} \quad \forall S \neq \emptyset \subseteq \mathsf{In}(v),\ \forall v \in V, \tag{4.27}$$

$$\rho_v^{\mathsf{Out}}(T) = \frac{1}{2} \quad \forall T \neq \emptyset \subseteq \mathsf{Out}(v),\ \forall v \in V. \tag{4.28}$$

Now, on this polymatroidal network, information is routed from the source nodes to the destination nodes. Since the traffic demand is symmetric between s_i and d_i, the result of Theorem 14 shows that the rate achieved on the polymatroidal network is within a poly-logarithmic factor of the edge-cut bound in the polymatroidal network, i.e.,

$$\frac{\mathcal{R}^{\text{poly}}_{\text{w.e.c.}}}{\kappa \log^3(k+1)} \subseteq \mathcal{R}^{\text{poly}}_{\text{ach}}. \tag{4.29}$$

Observe that if $(R_1, ..., R_k) \in \mathcal{R}^{\text{poly}}_{\text{ach}}$ is achievable in the polymatroidal network, a degrees of freedom tuple $(D_1, ..., D_k) = (R_1, ..., R_k)$ is achievable almost surely in the Gaussian network using the strategy described above. Thus $\mathcal{D}^{\text{g}}_{\text{ach}} = \mathcal{R}^{\text{poly}}_{\text{ach}}$. So, (4.29) gives

$$\frac{\mathcal{R}^{\text{poly}}_{\text{w.e.c.}}}{\kappa \log^3(k+1)} \subseteq \mathcal{R}^{\text{poly}}_{\text{ach}} = \mathcal{D}^{\text{g}}_{\text{ach}}. \tag{4.30}$$

Now, we need to connect $\mathcal{R}^{\text{poly}}_{\text{w.e.c.}}$ to $\mathcal{D}^{\text{g}}_{\text{w.e.c.}}$ in order to obtain the desired result. For this, we want to find the DOF value of a cut in the Gaussian network. We upper bound the DOF in the original Gaussian network by creating a modified network. We will color the edges in a certain way such that broadcast and superposition constraints are active only *inside* a color (similar to the coloring defined for MAC+BC networks (see Sec. 4.1).

We need the following lemma, which relates the rates achieved in the two networks.

Lemma 2 *(from [35]) If we color the edges in the Gaussian network to obtain a "colored network", then the capacity regions of the two networks are related by the following:*

$$\mathcal{R}^g_{ach}(P) \subseteq \mathcal{R}^g(P) \subseteq \mathcal{R}^{col}(dP), \tag{4.31}$$

where the superscript ^{g} is used for the original Gaussian network and the superscript ^{col} is used to denote the colored Gaussian network and d is the maximum degree of any node in the network, that is, $\mathcal{R}^g_{ach}(P)$ is the set of rate tuples achieved by our scheme, $\mathcal{R}^g(P)$ is the capacity region and $\mathcal{R}^{col}(dP)$ is the capacity region of the colored Gaussian network at power constraint dP.

Now, if the coloring yields a MAC+BC network, Lemma 1 informs us that the GNS bound is a fundamental outer bound on the capacity region. Also, from Theorem 10 for the symmetric demands problem, the weak edge cut bound is identical to the GNS bound, that is,

$$\mathcal{R}^{\text{col}}(P) \subseteq \mathcal{R}^{\text{col}}_{\text{w.e.c.}}(P). \tag{4.32}$$

(4.31) and (4.32) imply

$$\mathcal{R}^{\mathrm{g}}_{\mathrm{ach}}(P) \subseteq \mathcal{R}^{\mathrm{g}}(P) \subseteq \mathcal{R}^{\mathrm{col}}(dP) \subseteq \mathcal{R}^{\mathrm{col}}_{\mathrm{w.e.c.}}(dP). \tag{4.33}$$

Since degrees of freedom do not change by constant factor power scaling, we get

$$\mathcal{D}^{\mathrm{g}}_{\mathrm{ach}} \subseteq \mathcal{D}^{\mathrm{g}} \subseteq \mathcal{D}^{\mathrm{col}} \subseteq \mathcal{D}^{\mathrm{col}}_{\mathrm{w.e.c.}}. \tag{4.34}$$

This justifies that $\mathcal{D}^{\mathrm{col}}_{\mathrm{w.e.c.}}$ is indeed an upper bound on the DOF of the original network. From the definition of $\mathcal{D}^{g}_{\mathrm{w.e.c.}}$, we have $\mathcal{D}^{\mathrm{g}}_{\mathrm{w.e.c.}} = \cap_{\mathrm{col}\in\mathcal{Z}}\mathcal{D}^{\mathrm{col}}_{\mathrm{w.e.c.}}$, where $\mathcal{Z}$ denotes the set of all possible colorings that yield MAC+BC Gaussian networks. Using this and (4.34) we have

$$\mathcal{D}^{\mathrm{g}}_{\mathrm{ach}} \subseteq \mathcal{D}^{\mathrm{g}} \subseteq \mathcal{D}^{\mathrm{g}}_{\mathrm{w.e.c.}}. \tag{4.35}$$

Now, consider a subset of edges $F \subseteq \mathcal{E}$ in the polymatroidal network. Let the value of this cut in the polymatroidal network be $\nu(F)$. By the definition of cuts in polymatroidal network, $\nu(F)$ is achieved by an optimal assignment g of each edge $(u, v) \in F$ to either node u or v and the summing up the sub-modular functions in the polymatroidal network. Recall that the value of the cut $\nu(F)$ is given by,

$$\nu(F) := \sum_{v}\{\rho^{\mathsf{In}}_{v}(\mathsf{In}(v) \cap g^{-1}(v)) + \rho^{\mathsf{Out}}_{v}(\mathsf{Out}(v) \cap g^{-1}(v))\}.$$

Each sub-modular function has the same value $(\frac{1}{2})$ for a non-empty subset. Hence the total value of the edge-cut in the polymatroidal network is equal to one half of the number of nodes that are assigned at least one incoming edge by the assignment plus one half of the number of nodes that are assigned at least one outgoing edges. From this optimal assignment g, we describe a specific coloring from a partition of F into equivalence classes using the following equivalence relation $\sim$, where

$$\begin{aligned} e_1 \sim e_2 \quad \Longleftrightarrow \quad & g(e_1) = g(e_2) = \mathsf{head}(e_1) = \mathsf{head}(e_2) \text{ or} \\ & g(e_1) = g(e_2) = \mathsf{tail}(e_1) = \mathsf{tail}(e_2), \end{aligned}$$

i.e., $e_1 \sim e_2$ if they are assigned by g to the same node and share the head or tail. We assign distinct color to equivalence classes and each edge in the equivalence class is assigned the same color. The edges not in F are all assigned distinct colors, which are disjoint from the colors already assigned. This defines a colored network, whose DOF upper bounds the DOF of the original network.

The DOF value of this GNS cut is simply the sum of the number of broadcast or MAC channels involved in the cut. This is exactly equal to the number of colors in the cut, which is the same as the number of distinct equivalence classes, which is equal to $2\nu(F)$. Thus the degrees of freedom of the cut F, $d(F)$ is upper bounded by $2\nu(F)$, and therefore,

$$\mathcal{D}^{\mathrm{g}}_{\mathrm{w.e.c.}} \subseteq \mathcal{D}^{\mathrm{col}}_{\mathrm{w.e.c.}} \subseteq 2\mathcal{R}^{\mathrm{poly}}_{\mathrm{w.e.c.}}. \tag{4.36}$$

From (4.30), (4.35) and (4.36), we get

$$\frac{\mathcal{D}^{\mathrm{g}}_{\mathrm{w.e.c.}}}{\kappa \log^3(k+1)} \subseteq \mathcal{D}^{\mathrm{g}}_{\mathrm{ach}} \subseteq \mathcal{D}^{\mathrm{g}} \subseteq D^{\mathrm{g}}_{\mathrm{w.e.c.}}, \tag{4.37}$$

which proves our claim.

Ergodic Gaussian Channels

In an ergodic wireless network, the channel model is similar to the fixed Gaussian network, except that instead of assuming constant channel coefficients $h_{ij}(t) = h_{ij}$, we assume that for each $(j,i) \in \mathcal{E}$, the channel coefficient $h_{ij}(t)$ is varying as a function of time. In particular, we assume that $h_{ij}(t)$ is i.i.d. according to a fading distribution $p(h)$ across i, j, t, i.e., identical and independent across all the edges and time. We make another assumption, that the fading distribution $p(h)$ is symmetric, i.e., $p(h) = p(-h)$ with $\mathbb{E}_h|h|^2 = 1$ and also that it satisfies the following weak tail assumption,

$$a := e^{-\mathbb{E}(\log|h|^2)} < \infty. \tag{4.38}$$

Many common fading distributions satisfy this assumption. One particularly useful example of such a fading distribution is the i.i.d. complex Gaussian distribution, for which $a \approx 1.78$ [53].

Our main result in this setting is the following.

Theorem 16 *For directed Gaussian network with symmetric demands, with a symmetric weak-tailed ergodic fading distribution, the rate region given by $\mathcal{R}_{ach}(P)$ satisfying*

$$\frac{\mathcal{R}_{\mathrm{w.e.c.}}(\frac{P}{ad^3})}{\kappa \log^3(k+1)} \subseteq \mathcal{R}_{ach}(P) \subseteq \mathcal{C}(P) \subseteq \mathcal{R}_{\mathrm{w.e.c.}}(P), \tag{4.39}$$

is achievable, where κ is a universal constant, d is the maximum degree of any node and $a = e^{-\mathbb{E}(\log|h|^2)}$.

The proof of this result is similar to the case of general static networks in Subsection 4.2 and has been carried out for bidirected wireless networks in [35]. We only sketch the basic differences in the proof here.

The wireless network is converted into a polymatroidal network using the following physical layer scheme: At any given instant, we choose a subset of nodes as transmitter and another subset of nodes of the same size as receiver, and pair them up as into transmitter-receiver pairs. This creates an interference channel between the set of transmitting nodes and receiving nodes, in this interference channel, ergodic interference alignment [52] can achieve the following rate,

$$r := \frac{1}{2}\mathbb{E}_h \log\left(1 + |h|^2 P\right), \tag{4.40}$$

for every transmitter receiver pair which is connected. When we time-share between all such possible transmitter subset, receiver subset and matching choices, we get a polymatroidal network specified by the same set of edges as the original wireless network with submodular functions on the nodes given by,

$$\rho_v^{\mathsf{In}}(S) = r \quad \forall S \neq \emptyset \subseteq \mathsf{In}(v)\ \forall v \in V, \tag{4.41}$$
$$\rho_v^{\mathsf{Out}}(T) = r \quad \forall T \neq \emptyset \subseteq \mathsf{Out}(v)\ \forall v \in V. \tag{4.42}$$

We define the weak edge-cut for the ergodic wireless network similar to the definition of the weak edge-cut for fixed wireless network, using the sharpest upper bound given by the coloring of the edges that induces a MAC+BC network, i.e.,

$$\mathcal{R}^{\mathrm{g}}_{\mathrm{w.e.c.}}(P) = \cap_{\mathrm{col}\in\mathcal{Z}} \mathcal{R}^{\mathrm{col}}_{\mathrm{w.e.c.}}(dP), \tag{4.43}$$

where the superscript 'col' here denotes a particular instance of a colored network and $\mathcal{Z}$ is the set of all possible colorings of the network that lead to a MAC+BC network. By Lemma 2 and Lemma 1, we know that $\mathcal{C}(P) \subseteq \mathcal{R}^{\mathrm{g}}_{\mathrm{w.e.c.}}(P)$, i.e., the region $\mathcal{R}^{\mathrm{g}}_{\mathrm{w.e.c.}}(P)$ is a fundamental upper bound for the capacity region.

We then perform routing on this polymatroidal network from the sources to the sinks in the given traffic model. By Theorem 14, routing can achieve rate tuples close to the cut-set bound in the polymatroidal network to within a poly-logarithmic factor,

$$\mathcal{R}^{\mathrm{g}}_{\mathrm{ach}}(P) = \mathcal{R}^{\mathrm{poly}}_{\mathrm{ach}}(P) \supseteq \frac{\mathcal{R}^{\mathrm{poly}}_{\mathrm{w.e.c.}}(P)}{\kappa \log^3(k+1)}. \tag{4.44}$$

In order to connect the cuts of the Gaussian network to the polymatroidal network, we take any edge-cut F and the optimizing assignment g of the cut, and color the edges of the Gaussian network using equivalence classes of F, in exactly the same way as in the fixed Gaussian network case. The cuts in the colored network can be related to the cuts in the polymatroidal network, along the lines of [35] as follows:

$$\mathcal{R}^{\mathrm{col}}_{\mathrm{w.e.c.}}(P) \subseteq \mathcal{R}^{\mathrm{poly}}_{\mathrm{w.e.c.}}(ad^2P). \tag{4.45}$$

Therefore, we get,

$$\mathcal{R}^{\mathrm{g}}_{\mathrm{w.e.c.}}(P) \subseteq \mathcal{R}^{\mathrm{poly}}_{\mathrm{w.e.c.}}(ad^3P), \tag{4.46}$$

which completes the sketch of proof of Theorem 16.

Chapter 5

Non-Interactive Simulation of Joint Distributions

In this chapter, we consider the problem of simulation of one sample of a joint distribution by physically separated non-interacting agents observing i.i.d. copies of correlated random variables. Related problems have been well-studied in the literature. Wyner [58] studied the problem of simulating a joint distribution from shared randomness while Gács and Körner [23] studied the problem of extracting common randomness from correlated observations. Cuff studied communication requirements for simulating a channel [15]. Gohari and Anantharam generalized Cuff's formulation in [24] and Yassaee, Gohari and Aref recently solved this problem in [61]. Cuff, Permuter, and Cover studied communication requirements for establishing dependence among nodes in a network setting [16].

Non-Interactive Correlation Distillation, a setup in which non-interacting agents have to each output a uniform random bit which agree with high probability, has been studied in [57, 49, 9]. In this chapter, we propose a generalization of this problem. Below, we summarize different existing formulations of the problem of simulation of joint distributions.

a) The formulation shown in Fig. 5.1 was proposed by Gohari and Anantharam [24] as a generalization of Cuff's formulation [15]. Yassaee, Gohari, and Aref [61] recently solved this problem completely. The task is for two agents to simulate i.i.d. samples of a specified joint distribution $P(x, y, u, v)$. Nature supplies i.i.d. copies of (X, Y) with the marginal distribution $P(x, y)$ as shown and the agents can use a certain rate of common randomness and certain rate-limited communication and an infinite stream of their own private randomness to accomplish the desired task.

b) In this formulation (Fig. 5.2), two agents having access to their own infinite stream of private randomness observe n i.i.d. copies of samples generated according to a specified law $P(x, y)$ as shown and are required to output nR samples drawn from a distribution that is close (in total variation) to the the distribution constructed by taking i.i.d. copies of a specified law $Q(u, v)$. Let R^* be the supremum of all achievable rates.

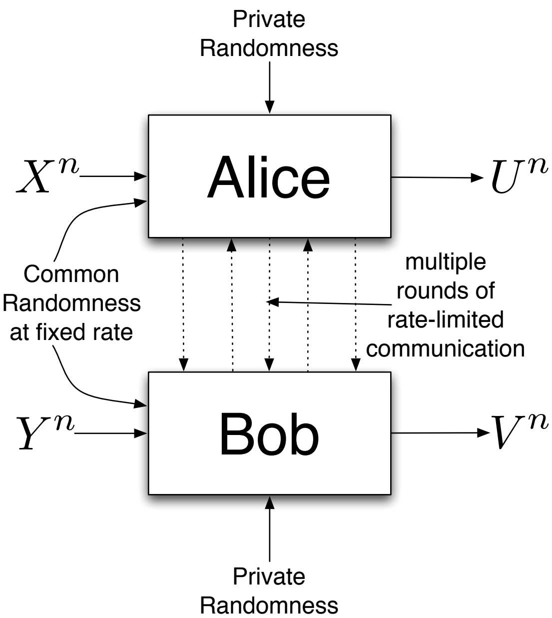


Figure 5.1: Gohari-Anantharam formulation

– When $(U,V) \sim Q(u,v)$ has $U = V \sim \text{Ber}(1/2)$, we have $R^* = K(X;Y)$, the Gács-Körner common information [23] of X and Y.
– When $(X,Y) \sim P(x,y)$ has $X = Y \sim \text{Ber}(1/2)$, we have $\frac{1}{R^*} = C(U;V)$, the Wyner common information [58] of U and V.

The problem of characterizing R^* is open for general distributions $P(x,y), Q(u,v)$ and indeed, so is the problem of characterizing when $R^* > 0$.

c) Since the problem of characterizing when $R^* > 0$ in formulation b) is also non-trivial, we propose a relaxed problem where two agents observe an arbitrary but finite number of samples drawn i.i.d. from $P(x,y)$ as shown in Fig. 5.3 and are required to output one random variable each with the requirement that the output distribution be close in total variation to a specified $Q(u,v)$. Clearly, if it is impossible to generate even a single sample, we obtain $R^* = 0$. We therefore, focus on impossibility results for this problem which will be relevant to formulation b) above. It is not clear if the converse is true, i.e. it is unclear whether the possibility of generating one sample implies that we may generate samples at a positive rate $R > 0$.

When $(U,V) \sim Q(u,v)$ has $U = V \sim \text{Ber}(1/2)$, the problem has recently come to be called Non-Interactive Correlation Distillation [49]. We therefore, call our formulation the problem of Non-Interactive Simulation of Joint Distributions. In a remarkable strengthening of the Gács-Körner result [23], Witsenhausen showed that unless the

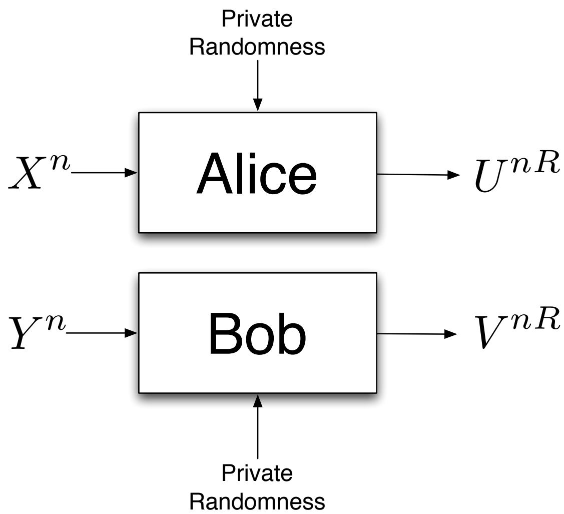


Figure 5.2: A generalization of the Gács-Körner and Wyner formulations

Gács-Körner common information $K(X;Y)$ is positive (i.e. the joint distribution of (X,Y) is decomposable), non-interactive correlation distillation is impossible to achieve [57] .

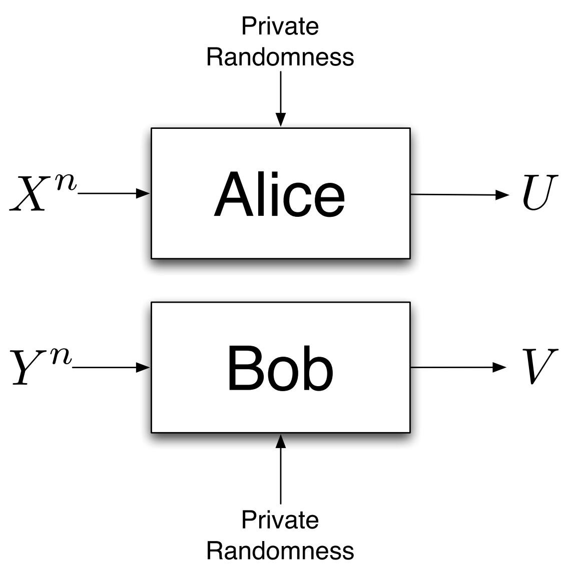


Figure 5.3: Non-Interactive Simulation of Joint Distributions

Our main contribution is a comparison between two tools - maximal correlation and hypercontractivity/reverse hypercontractivity - that help in proving impossibility results for the problem formulation c). We show that under suitable conditions on the source distribution (X, Y), hypercontractivity/reverse hypercontractivity is a stronger tool than maximal correlation.

This chapter is organized as follows. In Section 5.1, we discuss the problem formulation. In Section 5.2, we describe the key tools used: maximal correlation, hypercontractivity and reverse hypercontractivity. We describe properties of these tools that make them well-suited for use in our problem formulation. Section 5.3 contains our main results and Section 5.4 contains the proofs of the main results. We conclude with an interesting example of an extension of this problem setup to three agents in Section 5.5.

5.1 Problem formulation

Definition: Let $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{V}$ denote finite sets. Given a *source distribution* $P(x, y)$ over $\mathcal{X} \times \mathcal{Y}$ and a *target distribution* $Q(u, v)$ over $\mathcal{U} \times \mathcal{V}$, we say that *non-interactive simulation* of $Q(u, v)$ using $P(x, y)$ is possible, if for any $\epsilon > 0$, there exists a positive integer n and functions $f : \mathcal{X}^n \mapsto \mathcal{U}, g : \mathcal{Y}^n \mapsto \mathcal{V}$ such that

$$d_{\text{TV}}\left((f(X^n), g(Y^n)), (U, V)\right) \leq \epsilon$$

where $\{(X_i, Y_i)\}_{i=1}^n$ is a sequence of i.i.d. samples drawn from $P(x, y)$, (U, V) is drawn from $Q(u, v)$ and $d_{\text{TV}}(\cdot, \cdot)$ is the total variation distance.

For a fixed $P(x, y)$, the set of distributions $Q(u, v)$ on $\mathcal{U} \times \mathcal{V}$ for which non-interactive simulation is possible can be shown to be the closure of the set of marginal distributions of (U, V) satisfying $U - X^k - Y^k - V$ for some k. However, this set of distributions seems to be very hard to characterize. We focus on outer bounds on this set or in other words, impossibility results for non-interactive simulation.

Note that the simulation problem specified in the above definition does not have any more generality if we allow the agents to use their own private randomness: Agents can obtain as much private randomness as desired by using extended observations that are non-overlapping in time, i.e. the agents observe $n_1 + n_2 + n_3$ symbols, they use $X_1, \ldots, X_{n_1}$ as their correlated observations, one agent uses $X_{n_1+1}, \ldots, X_{n_2}$ as her private randomness and the other agent uses $X_{n_2+1}, \ldots, X_{n_3}$ as his private randomness.

Note also that the notion of *simulation* we consider is distinct from the notion of *exact generation*. If we have a strategic setting, such as a distributed game, in which a player, represented by a number of distributed agents, is playing against an adversary, the agents would often need to generate a joint distribution exactly [4].

We will consider two examples to motivate the focus of this study.

Example 1

Let X be a uniform Bernoulli random variable, $X \sim \text{Ber}(\frac{1}{2})$. Let Y be a noisy copy of X, i.e. $Y = X + N$ where $N \sim \text{Ber}(\alpha)$ for $0 < \alpha < \frac{1}{2}$, is independent of X. We say that (X, Y) has the Doubly Symmetric Binary Source distribution with parameter α denoted DSBS(α) following the notation of Wyner [58]. We consider $(U, V) \sim \text{DSBS}(\beta)$ for $0 \leq \beta < \frac{1}{2}$. We ask whether non-interactive simulation of DSBS(0) using DSBS(α) is possible. Witsenhausen answered this question in the negative in [57], thus significantly strengthening the result of Gács and Körner [23]. Witsenhausen established this by proving the tensorization of the Hirschfeld-Gebelein-Rényi maximal correlation, henceforth simply called the maximal correlation (both tensorization and maximal correlation are defined and discussed in Section 5.2). Witsenhausen's approach easily allows us to conclude that if non-interactive simulation is possible, then the maximal correlation of the target distribution can be no more than that of the source distribution. The maximal correlation of a pair of binary random variables distributed as DSBS(α) equals $|1 - 2\alpha|$. Thus, for instance, if the non-interactive simulation of DSBS(β) using DSBS(α) is possible, with $0 \leq \alpha, \beta \leq \frac{1}{2}$, then we must have $\alpha \leq \beta$. It is easy to see in this case that if $\alpha \leq \beta$, then non-interactive simulation is indeed possible: one agent outputs the first bit of her observation while the other agent outputs a suitable noisy copy of his first bit, the noise realization created from his other $n - 1$ observations. Thus, for $0 \leq \alpha, \beta \leq \frac{1}{2}$, non-interactive simulation of DSBS(β) using DSBS(α) is possible if and only if $\alpha \leq \beta$.

Example 2

Let $(X, Y) \sim \text{DSBS}(\alpha)$ with $0 < \alpha < \frac{1}{2}$. Consider binary random variables (U, V) distributed as $Q(u, v)$ given by: $Q(0, 0) = 0, Q(0, 1) = Q(1, 0) = Q(1, 1) = \frac{1}{3}$. We ask if non-interactive simulation of $Q(u, v)$ using DSBS(α) is possible. The maximal correlation of a DSBS(α) source distribution is $1 - 2\alpha$ while that of $Q(u, v)$ is $\frac{1}{2}$. The approach of comparing maximal correlations of the source and target informs us that the inequality $1 - 2\alpha \leq \frac{1}{2}$, if violated, makes non-interactive simulation impossible. Thus, if $\frac{1}{4} < \alpha < \frac{1}{2}$, then non-interactive simulation is impossible. But what about the case when $0 < \alpha \leq \frac{1}{4}$? Can we come up with a suitable scheme to simulate $Q(u, v)$? The answer turns out to be no for each $0 < \alpha \leq \frac{1}{4}$ and can be proved using the so-called reverse hypercontractive inequalities [48]. The following inequality holds for $\{(X_i, Y_i)\}_{i=1}^{\infty}$ being i.i.d DSBS(α), and for arbitrary sets $S, T \subseteq \{0, 1\}^n$:

$$\mathsf{Pr}\left(X^n \in S, Y^n \in T\right) \geq \mathsf{Pr}\left(X^n \in S\right)^{\frac{1}{2\alpha}} \mathsf{Pr}\left(Y^n \in T\right)^{\frac{1}{2\alpha}} . \tag{5.1}$$

If non-interactive simulation of $Q(u, v)$ using DSBS(α) were possible, we should be able to find sets S, T such that $\mathsf{Pr}\left(X^n \in S\right) \approx \frac{1}{3}, \mathsf{Pr}\left(Y^n \in T\right) \approx \frac{1}{3}$ and $\mathsf{Pr}\left(X^n \in S, Y^n \in T\right) \approx 0$. Inequality (5.1) rules out this possibility. Thus, hypercontractivity or reverse hypercontractivity can provide impossibility results when the maximal correlation approach cannot. Is it true that one is always stronger than the other? We show indeed that the approach using hypercontractivity and reverse hypercontractivity subsumes the maximal correlation approach for the case when $P(x, y)$ is of the form DSBS(α). More generally, we give necessary and sufficient conditions on $P(x, y)$ for this subsumption. This arises from an inequality obtained by Ahlswede and Gács [2] in the hypercontractive case which we extend to the reverse hypercontractive case.

The rest of the chapter is organized as follows. Section 5.2 discusses preliminaries on maximal correlation, hypercontractivity and reverse hypercontractivity. We present our main results in Section 5.3. Section 5.4 contains all the proofs. Finally, Section 5.5 discusses an interesting example of non-interactive simulation of a joint distribution of three random variables.

5.2 Maximal Correlation and the Hypercontractivity Ribbon

In our notation, sets $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{V}$ are finite and all probability distributions are discrete and have finite support. For a finite set $\mathcal{X}$, let $\mathcal{F}_{\mathcal{X}}, \mathcal{F}_{\mathcal{X}}^{+}$ denote the set of all functions from $\mathcal{X}$ to $\mathbb{R}$ and to $\mathbb{R}_{\geq 0}$ respectively. We will also assume without loss of generality that the marginals of $P(x, y)$ and $Q(u, v)$ (denoted P_X, P_Y and Q_U, Q_V respectively) assign zero probability only to the null set.

Maximal Correlation and its properties

For jointly distributed random variables (X, Y), define their *maximal correlation* $\rho(X; Y) := \sup \mathbb{E} f(X) g(Y)$ where the supremum is over $f : \mathcal{X} \mapsto \mathbb{R}, g : \mathcal{Y} \mapsto \mathbb{R}$ such that $\mathbb{E} f(X) = \mathbb{E} g(Y) = 0$ and $\mathbb{E}\left(f(X)\right)^2 = \mathbb{E}\left(g(Y)\right)^2 = 1$ and with the convention that the supremum over the empty set evaluates to 0.

The following theorem was proved by Witsenhausen in [57]. Kumar has obtained simpler proofs of the same result [41].

Theorem 17 *(Witsenhausen [57]) If $(X_1, Y_1), (X_2, Y_2)$ are independent, then $\rho(X_1, X_2; Y_1, Y_2) = \max\{\rho(X_1; Y_1), \rho(X_2; Y_2)\}$. If $(X_1, Y_1), (X_2, Y_2)$ are i.i.d., then $\rho(X_1, X_2; Y_1, Y_2) = \rho(X_1; Y_1)$.*

The following monotonicity lemma is immediate.

Lemma 3 *If $\phi(X) = U, \psi(Y) = V$, then $\rho(X; Y) \geq \rho(U; V)$.*

The following properties hold for the maximal correlation of two discrete valued random variables with finite support [54].

1. If $(X, Y) \sim$ DSBS(α), then $\rho(X; Y) = |1 - 2\alpha|$.

2. $\rho(X; Y) = 0$ if and only if X is independent of Y.

3. $\rho(X; Y) = 1$ if and only if the Gács-Körner common information $K(X; Y) > 0$, i.e. if and only if (X, Y) is *decomposable*.

It is easy to show that maximal correlation of (U, V) seen as a function of its joint distribution $Q(u, v)$ is continuous at $Q(u, v)$ whenever Q_U, Q_V assign a positive probability to each $u \in \mathcal{U}$ and each $v \in \mathcal{V}$ respectively. Using this fact and putting together Theorem 17 and Lemma 3, we get the following corollary.

Corollary 4 *Non-interactive simulation of* $(U, V) \sim Q(u, v)$ *using* $(X, Y) \sim P(x, y)$ *is possible only if* $\rho(X; Y) \geq \rho(U; V)$.

Hypercontractivity ribbon and its properties

Definition: For any random variable W and real number $p \neq 0$, define $||W||_p := (\mathbb{E}|W|^p)^{1/p}$. Define $||W||_0 := \exp(\mathbb{E} \log |W|)$. For $p \leq 0$, $||W||_p = 0$ if $\mathsf{Pr}(|W| = 0) > 0$.

$||W||_p$ is continuous and non-decreasing in p. If W is non-constant, then $||W||_p$ is strictly increasing for $p \geq 0$. If in addition, $\mathsf{Pr}(|W| = 0) = 0$, then $||W||_p$ is strictly increasing for all p.

Definition: For a pair of random variables $(X, Y) \sim P(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, define the operator $T_{X;Y} : \mathcal{F}_{\mathcal{Y}} \mapsto \mathcal{F}_{\mathcal{X}}$ as

$$(T_{X;Y} f)(x) := \mathbb{E}[f(Y)|X = x]. \tag{5.2}$$

Likewise, define $T_{Y;X} : \mathcal{F}_{\mathcal{X}} \mapsto \mathcal{F}_{\mathcal{Y}}$ as

$$(T_{Y;X} g)(y) := \mathbb{E}[g(X)|Y = y]. \tag{5.3}$$

Definition: For a pair of random variables $(X, Y) \sim P(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, we define the *hypercontractivity ribbon*

$$\mathcal{R}_{X;Y} \subseteq \{(p, q) : 1 \leq q \leq p \text{ or } 1 \geq q \geq p\}$$

as follows:

- For $1 \leq q \leq p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ if

$$||T_{X;Y} f(X)||_p \leq ||f(Y)||_q \quad \forall f \in \mathcal{F}_{\mathcal{Y}}; \tag{5.4}$$

- For $1 \geq q \geq p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ if

$$||T_{X;Y} f(X)||_p \geq ||f(Y)||_q \quad \forall f \in \mathcal{F}^+_{\mathcal{Y}}. \tag{5.5}$$

Likewise, we can define $\mathcal{R}_{Y;X}$. These are both regions in $\mathbb{R}^2$ pinching to a point at $(1, 1)$ resembling a ribbon, explaining our choice of the name (see Fig. 5.4). $\mathcal{R}_{X;Y}$ and $\mathcal{R}_{Y;X}$ are intimately connected by a duality relationship which we will discuss later. $T_{X;Y}$ is contractive in the p-norm when $p \geq 1$ and inequality (5.4) is a hypercontractive inequality since $q \leq p$. $T_{X;Y}$ is reverse contractive for non-negative valued functions f under the p-pseudo-norm when $p \leq 1$, (the triangle inequality is violated) and inequality (5.5) is called a reverse hypercontractive inequality and has been studied in [48].

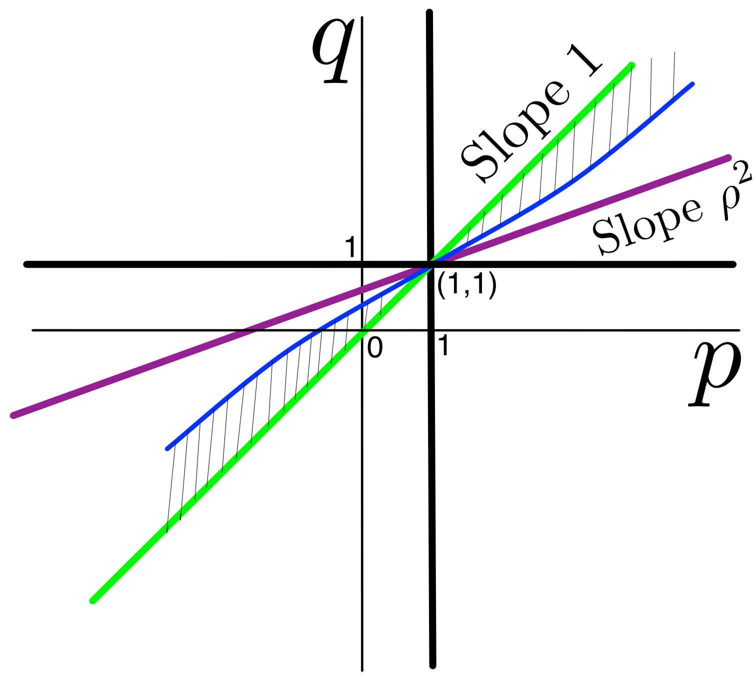


Figure 5.4: Hypercontractivity ribbon

The hypercontractivity ribbon $\mathcal{R}_{X;Y}$ is the shaded region. Also shown a straight line of slope $\rho^2 := \rho^2(X;Y)$ through $(1,1)$.

Definition: For any real $p \neq 0, 1$, define its *Hölder conjugate* p' by $\frac{1}{p} + \frac{1}{p'} = 1$. For $p = 0$, define $p' = 0$.

Remark: An equivalent definition of $\mathcal{R}_{X;Y}$ which does not use the definition of the operator $T_{X;Y}$ can be provided by observing how much the corresponding Hölder's and reverse Hölder's inequalities may be tightened.

- For $1 \leq q < p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ iff

$$\mathbb{E}f(X)g(Y) \leq ||f(X)||_{p'}||g(Y)||_q \quad \forall f \in \mathcal{F}_{\mathcal{X}}, g \in \mathcal{F}_{\mathcal{Y}}; \tag{5.6}$$

- For $1 \geq q > p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ iff

$$\mathbb{E}f(X)g(Y) \geq ||f(X)||_{p'}||g(Y)||_q \quad \forall f \in \mathcal{F}^+_{\mathcal{X}}, g \in \mathcal{F}^+_{\mathcal{Y}}; \tag{5.7}$$

- $(1, 1) \in \mathcal{R}_{X;Y}$.

To see the equivalence, observe that for $p > 1$, if (5.4) holds, then by Hölder's inequality, we get

$$\begin{aligned}
\mathbb{E}f(X)g(Y) &= \mathbb{E}f(X)\left(T_{X;Y}g\right)(X) && (5.8)\\
&\leq ||f(X)||_{p'}||\left(T_{X;Y}g\right)(X)||_p && (5.9)\\
&\leq ||f(X)||_{p'}||g(Y)||_q. && (5.10)
\end{aligned}$$

Conversely, if the inequality in (5.4) fails for some non-negative f, say $f = h$, then by choosing the function $e(X) = (T_{X;Y}h(X))^{p/p'}$, we have equality in Hölder's inequality as follows:

$$\begin{aligned}
\mathbb{E}e(X)h(Y) &= \mathbb{E}e(X)\left(T_{X;Y}h\right)(X) && (5.11)\\
&= ||e(X)||_{p'}||\left(T_{X;Y}h\right)(X)||_p && (5.12)\\
&> ||e(X)||_{p'}||h(Y)||_q, && (5.13)
\end{aligned}$$

since $||e(X)||_{p'} > 0$, thus producing the desired contradiction to (5.6). It suffices to consider non-negative f, since $-|f| \leq f \leq |f|$ holds pointwise and so $|T_{X;Y}f| \leq T_{X;Y}|f|$ holds pointwise so that if (5.4) fails for some f then it also fails for $|f|$. A similar equivalence can be observed for $p < 1$, using the reverse Hölder's inequality:

$$\mathbb{E}[WZ] \geq ||W||_{p'}||Z||_p, \quad (5.14)$$

which holds when $p < 1$ and W, Z are non-negative random variables. The contradiction is first observed for strictly positive functions with $p/p' := -1$ in the case $p = 0$ and then for non-negative functions by taking limits.

$\mathcal{R}_{X;Y}$ is closed and connected in $\mathbb{R}^2$. Moreover, $\{(p,q) : p = q\} \subseteq \mathcal{R}_{X;Y}$. So, $\mathcal{R}_{X;Y}$ is completely characterized by its other boundary, a continuous non-decreasing function $q^*_{X;Y} : \mathbb{R} \mapsto \mathbb{R}$ such that

- $q^*_{X;Y}(p) \leq p$ whenever $p \geq 1$, and $q^*_{X;Y}(p) \geq p$ whenever $p \leq 1$, so $q^*_{X;Y}(1) = 1$;
- $\mathcal{R}_{X;Y} = \{(p,q) : 1 \leq q^*_{X;Y}(p) \leq q \leq p\} \cup \{(p,q) : 1 \geq q^*_{X;Y}(p) \geq q \geq p\}$.

Hypercontractive inequalities and reverse hypercontractive inequalities tensorize [48].

Theorem 18 *Suppose $(p,q) \in \mathcal{R}_{X_1;Y_1}$ and $(p,q) \in \mathcal{R}_{X_2;Y_2}$. If $(X_1,Y_1),(X_2,Y_2)$ are independent, then $(p,q) \in \mathcal{R}_{X_1,X_2;Y_1Y_2}$, so that $\mathcal{R}_{(X_1,X_2);(Y_1,Y_2)} = \mathcal{R}_{X_1;Y_1} \cap \mathcal{R}_{X_2;Y_2}$. If $(X_1,Y_1),(X_2,Y_2)$ are i.i.d., then $\mathcal{R}_{(X_1,X_2);(Y_1,Y_2)} = \mathcal{R}_{X_1;Y_1}$.*

The following lemma provides a monotonicity property for the hypercontractivity ribbon [48].

Lemma 4 *If $\phi(X) = U, \psi(Y) = V$, then $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$.*

It is easy to show that for any fixed p, the property $q^*_{U;V}(p)$ of (U,V) seen as a function of its joint distribution $Q(u,v)$ is continuous at $Q(u,v)$ whenever Q_U, Q_V assign a positive probability to each $u \in \mathcal{U}$ and each $v \in \mathcal{V}$ respectively. Using this fact and putting together Theorem 18 and Lemma 4 , we get the following corollary.

Corollary 5 *Non-interactive simulation of $(U,V) \sim Q(u,v)$ using $(X,Y) \sim P(x,y)$ is possible only if $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$.*

The following properties hold for the hypercontractivity ribbon for two discrete valued random variables with finite support [48].

1. If $(X,Y) \sim \text{DSBS}(\alpha)$, then
   $q^*_{X;Y}(p) - 1 = (1-2\alpha)^2(p-1)$ [49].
2. $q^*_{X;Y}(p) \equiv 1$ if and only if X and Y are independent, i.e. $I(X;Y) = 0$.
3. $q^*_{X;Y}(p) \equiv p$ if and only if $P(x,y)$ is decomposable, i.e. the Gács-Körner common information $K(X;Y) > 0$.
4. If $K(X;Y) = 0$ but $I(X;Y) > 0$, then for $p > 1$, we have the strict inequalities $1 < q^*_{X;Y}(p) < p$ [2].

Proving impossibility results for non-interactive simulation using the hypercontractivity ribbon $\mathcal{R}_{X;Y}$

While Corollary 5 describes the technique for proving impossibility results, it is worthwhile noting that this is equivalent to the techniques that were originally used to produce inequalities like (5.1).

Suppose that non-interactive simulation of $Q(u,v)$ using $P(x,y)$ is possible, i.e. suppose for any $\epsilon > 0$, there exists n and functions $\phi : \mathcal{X}^n \mapsto \mathcal{U}, \psi : \mathcal{Y}^n \mapsto \mathcal{V}$ so that $\phi(X^n) = \tilde{U}, \psi(Y^n) = \tilde{V}$ produces $(\tilde{U}, \tilde{V})$ satisfying $d_{\mathsf{TV}}((\tilde{U},\tilde{V});(U,V)) \le \epsilon$ when $(U,V) \sim Q(u,v)$ and $\{(X_i,Y_i)\}_{i=1}^n$ are generated i.i.d. from $P(x,y)$. Choose

$$f(x^n) = \sum_{u \in \mathcal{U}} \lambda_u \mathbb{1}_{[\phi(x^n)=u]}, \tag{5.15}$$

$$g(y^n) = \sum_{v \in \mathcal{V}} \mu_v \mathbb{1}_{[\psi(y^n)=v]}. \tag{5.16}$$

For $(p,q) \in \mathcal{R}_{X;Y}$, with $p > 1$, using (5.6), we obtain upon taking the limit as $\epsilon \to 0$,

$$\begin{aligned} &\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \lambda_u \mu_v Q(u,v) \\ &\le \left(\sum_{u \in \mathcal{U}} \lambda_u^{p'} Q_U(u) \right)^{1/p'} \cdot \left(\sum_{v \in \mathcal{V}} \mu_v^q Q_V(v) \right)^{1/q}. \end{aligned} \tag{5.17}$$

For $(p, q) \in \mathcal{R}_{X;Y}$, with $p < 1$, using (5.7), we obtain for non-negative $\{\lambda_u\}_{u\in\mathcal{U}}, \{\mu_v\}_{v\in\mathcal{V}}$, the inequality

$$\begin{aligned}&\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}} \lambda_u \mu_v Q(u,v)\\ &\geq \left(\sum_{u\in\mathcal{U}} \lambda_u^{p'} Q_U(u)\right)^{1/p'} \cdot \left(\sum_{v\in\mathcal{V}} \mu_v^{q} Q_V(v)\right)^{1/q}.\end{aligned} \tag{5.18}$$

Indeed, (5.1) is a version of (5.18) with $(p, q) \in \mathcal{R}_{X;Y}$ for $(X, Y) \in \mathrm{DSBS}(\alpha)$ given by $p = -\frac{2\alpha}{1-2\alpha}, q = 2\alpha$.

The inclusion $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$ implies the collection of inequalities (5.17) for any choice of real $\{\lambda_u\}_{u\in\mathcal{U}}, \{\mu_v\}_{v\in\mathcal{V}}$ and the collection of inequalities (5.18) for any choice of non-negative $\{\lambda_u\}_{u\in\mathcal{U}}, \{\mu_v\}_{v\in\mathcal{V}}$. By an argument similar to the one proving equivalence of the two definitions of $\mathcal{R}_{X;Y}$, one can prove the reverse implication from the collection of inequalities (5.17), (5.18) to $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$.

5.3 Main Results

Theorem 19

$$\rho(X;Y) \leq \inf_{(p,q)\in\mathcal{R}_{X;Y}, p\neq 1} \sqrt{\frac{q-1}{p-1}} = \inf_{p\neq 1} \sqrt{\frac{q^*_{X;Y}(p)-1}{p-1}}. \tag{5.19}$$

Theorem 19 is obtained in [2] for the case of hypercontractive inequalities. We provide an alternate proof of the same result and derive it for the reverse hypercontractive inequalities. In the current form of the statement of Theorem 19, the maximal correlation is afforded a geometric meaning, namely its square is the slope of a straight line bound constraining the hypercontractivity ribbon (see Fig 5.4). Indeed, for $(X, Y) \sim \mathrm{DSBS}(\alpha)$, the hypercontractivity ribbon is precisely the wedge obtained by the straight lines $p = q$, and $q - 1 = \rho(X; Y)^2(p - 1)$ [49].

Theorem 20 *The following are equivalent:*

- *For all (U, V), we have $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V} \implies \rho(X;Y) \geq \rho(U;V)$.*
- $$\rho(X;Y) = \inf_{(p,q)\in\mathcal{R}_{X;Y}, p\neq 1} \sqrt{\frac{q-1}{p-1}}. \tag{5.20}$$

Theorem 20 states that Corollary 5 subsumes Corollary 4 for all $Q(u, v)$ if and only (5.19) holds with equality.

Ahlswede and Gács [2] show that $\lim_{p\to\infty} \frac{q^*_{X;Y}(p)}{p}$ exists and equals a quantity $s^*(X;Y)$, defined as follows: Consider finite sets $\mathcal{X}, \mathcal{Y}$ and let $P(x,y)$ be a joint distribution over the product $\mathcal{X} \times \mathcal{Y}$. Let $R(x)$ be an arbitrary probability distribution on $\mathcal{X}$. Let $\sum_X P_{Y|X} * R$ denote the probability distribution on $\mathcal{Y}$ whose probability mass at y is $\sum_{x\in\mathcal{X}} \frac{P(x,y)}{P_X(x)} R(x)$. If $(X,Y) \sim P(x,y)$, then we define $s^*(X;Y) = \sup_{R:R\neq P_X} \frac{D(\sum_X P_{Y|X} * R || P_Y)}{D(R||P_X)}$.

We prove the same result, also extending it to reverse hypercontractive inequalities by a simpler approach.

Theorem 21

$$\lim_{p\to 1} \frac{q^*_{X;Y}(p) - 1}{p-1} = s^*(Y;X). \tag{5.21}$$

Corollary 6 follows from Theorem 21 upon using a duality result connecting $\mathcal{R}_{X;Y}$ and $\mathcal{R}_{Y;X}$.

Corollary 6

$$\lim_{p\to\infty} \frac{q^*_{X;Y}(p) - 1}{p-1} = \lim_{p\to-\infty} \frac{q^*_{X;Y}(p) - 1}{p-1} = s^*(X;Y). \tag{5.22}$$

Corollary 7 provides a sufficient condition for (5.20) to hold.

Corollary 7 *If* $\rho(X;Y) = \min\{\sqrt{s^*(X;Y)}, \sqrt{s^*(Y;X)}\}$, *then*

$$\forall \quad (U,V), \quad \mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V} \implies \rho(X;Y) \geq \rho(U;V).$$

Note that from properties listed for the hypercontractivity ribbon, DSBS sources always satisfy the condition in Corollary 7.

Finally, we find the right constant for a strong data processing inequality in the literature. Erkip and Cover make the following claim in [22]:

Claim 1 $\sup_{U:U-X-Y} \frac{I(U;Y)}{I(U;X)} = \rho^2(X;Y)$.

We show in fact, that

Theorem 22 $\sup_{U:U-X-Y} \frac{I(U;Y)}{I(U;X)} = s^*(X;Y)$.

5.4 Proofs

We first present the proof of Theorem 19.

Proof: The proof proceeds from a perturbative argument. Let (X, Y) distributed as $P(x, y)$. Fix functions $\phi : \mathcal{X} \mapsto \mathbb{R}, \psi : \mathcal{Y} \mapsto \mathbb{R}$ such that

$$\mathbb{E}\phi(X) = \mathbb{E}\psi(Y) = 0, \ \mathbb{E}\phi(X)^2 = \mathbb{E}\psi(Y)^2 = 1. \tag{5.23}$$

Fix $r > 0$. Define $f : \mathcal{X} \mapsto \mathcal{R}, g : \mathcal{Y} \mapsto \mathcal{R}$ by $f(x) = 1 + \frac{\sigma}{r}\phi(x), g(y) = 1 + \sigma r\psi(y)$. Note that for sufficiently small σ, the functions f, g take only positive values. Fix $(p, q) \in \mathbb{R}_{X;Y}$ with $p > 1$. Using (5.6), we have

$$\begin{aligned}&\mathbb{E}[(1 + \frac{\sigma}{r}\phi(X))(1 + \sigma r\psi(Y))] \leq \\ &\left(\mathbb{E}[(1 + \frac{\sigma}{r}\phi(X))^{p'}]\right)^{1/p'} \cdot \left(\mathbb{E}[(1 + \sigma r\psi(Y))^q]\right)^{1/q}.\end{aligned} \tag{5.24}$$

For Z satisfying $\mathbb{E}Z = 0, \mathbb{E}Z^2 = 1$,

$$\begin{aligned}&\left(\mathbb{E}[(1 + aZ)^l]\right)^{1/l} \\ &= \left(1 + l \cdot a\mathbb{E}Z + \frac{l(l-1)}{2} \cdot a^2\mathbb{E}Z^2 + O(a^3)\right)^{1/l} \\ &= \left(1 + \frac{l-1}{2}a^2 + O(a^3)\right).\end{aligned}$$

The first two terms of the expansion on both sides of (5.24) match. Comparing the coefficient of σ^2 on both sides, we get

$$\mathbb{E}\phi(X)\psi(Y) \leq \frac{p'-1}{2r^2} + \frac{(q-1)r^2}{2}.$$

Taking the supremum over all ϕ, ψ satisfying (5.23) and the infimum over all $r > 0$, we have

$$\rho(X; Y) \leq \sqrt{\frac{q-1}{p-1}}.$$

We can similarly prove the inequality in the case when $p < 1$. We get $\mathbb{E}\phi(X)\psi(Y) \geq -\sqrt{\frac{q-1}{p-1}}$ in this case and we replace ϕ by $-\phi$ and perform similar steps to get the desired. This completes the proof.

Next, we prove Theorem 20.

Proof: The if part of the statement follows immediately from Theorem 19. For the only if part, suppose that for $(X, Y) \sim P(x, y)$, we have for some $\delta > 0$,

$$\rho(X; Y) = \inf_{(p,q)\in\mathcal{R}_{X;Y}, p\neq 1} \sqrt{\frac{q-1}{p-1}} - \delta.$$

A classical result [49] states that for $(U,V) \sim \mathrm{DSBS}(\epsilon)$,

$$\frac{q^*_{U;V}(p)-1}{p-1} = (1-2\epsilon)^2 = \rho(U;V)^2.$$

Choosing ϵ so that $\rho(U;V) = 1-2\epsilon = \inf_{(p,q)\in\mathcal{R}_{X;Y}, p\neq 1}\sqrt{\frac{q-1}{p-1}}$, we have $\rho(X;Y) < \rho(U;V)$ and $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$. This completes the proof.

A perturbative argument provides the proof of Theorem 21 as below.

Proof: As noted earlier, the inequality (5.4) holds for all functions f only if it holds for all non-negative functions f. Now, for non-negative f, we always have

$$||T_{X;Y}f(X)||_1 = ||f(Y)||_1 \quad \forall f \in \mathcal{F}^+_{\mathcal{Y}}. \tag{5.25}$$

As in [48], we define for any non-negative random variable X, the function $\mathrm{Ent}(X) := \mathbb{E}X\log X - \mathbb{E}X\cdot\log\mathbb{E}X$, where by convention $0\log 0 := 0$. By strict convexity of the function $x \mapsto x\log x$, we get using Jensen's inequality that $\mathrm{Ent}(X) \geq 0$ and equality holds if and only if X is a constant almost surely. Also, note that $\mathrm{Ent}(\cdot)$ is homogenous, that is, $\mathrm{Ent}(aX) = a\,\mathrm{Ent}(X)$ for any $a \geq 0$.

Define $s := \sup\frac{\mathrm{Ent}(T_{X;Y}f(X))}{\mathrm{Ent}(f(Y))}$, where the supremum is taken over non-constant functions $f \in \mathcal{F}^+_{\mathcal{Y}}$. As P_Y assigns a positive probability to all elements of $\mathcal{Y}$, this rules out the possibility of a non-constant function f with $f(Y)$ being a constant almost surely.

If $m < s$, then $(1+\tau, 1+m\tau) \notin \mathcal{R}_{X;Y}$ for all sufficiently small $\tau > 0$. To see this, fix f_0 to be any (non-constant) function in $\mathcal{F}^+_{\mathcal{Y}}$ that satisfies

$$\frac{\mathrm{Ent}(T_{X;Y}f_0(X))}{\mathrm{Ent}(f_0(Y))} \geq m + \frac{\delta}{2}, \tag{5.26}$$

where $\delta := s - m$. Now,

$$||f_0(Y)||_{1+m\tau} = ||f_0(Y)||_1 + m\tau\,\mathrm{Ent}(f_0(Y)) + O(\tau^2), \tag{5.27}$$

$$||T_{X;Y}f_0(X)||_{1+\tau} = ||T_{X;Y}f_0(X)||_1 + \tau\,\mathrm{Ent}(T_{X;Y}f_0(X)) + O(\tau^2). \tag{5.28}$$

Putting together (5.25), (5.26), (5.27), (5.28), we get the existence of $\tau_0 > 0$ such that

$$||T_{X;Y}f_0(X)||_{1+\tau} > ||f_0(Y)||_{1+m\tau} \quad \forall \tau : 0 < \tau \leq \tau_0. \tag{5.29}$$

If $m > s$, then consider the set $\mathcal{H}$ of all functions $f : \mathcal{Y} \mapsto \mathcal{R}^+$ that satisfy $||f||_1 = 1$ and define $\tau(f) := \max\{\tau : 0 \leq \tau \leq 1, ||T_{X;Y}f(X)||_{1+\tau} \leq ||f(Y)||_{1+m\tau}\}$. As $\tau(f)$ is continuous over the compact set $\mathcal{H}$, showing $\tau(f) > 0\ \forall f \in \mathcal{H}$ would yield $\tau_1 := \inf_{f\in\mathcal{H}}\tau(f) > 0$. But that is obvious since for f constant, $\tau(f) = 1$ and for f non-constant, $\tau(f) > 0$ from (5.25), (5.26), (5.27), (5.28).

This gives $||T_{X;Y}f(X)||_{1+\tau} \leq ||f(Y)||_{1+m\tau}$ for all $f \in \mathcal{H}, 0 < \tau \leq \tau_1$. By homogeneity of the p-norm, it follows that $||T_{X;Y}f(X)||_{1+\tau} \leq ||f(Y)||_{1+m\tau}$ $\forall f \in \mathcal{F}_{\mathcal{Y}^+}, 0 < \tau \leq \tau_1$, thus proving that

$$\lim_{p \to 1^+} \frac{q^*_{X;Y}(p) - 1}{p - 1} = s. \tag{5.30}$$

Similarly, one can prove the same limit for $p \to 1^-$. The final step is to show $s = s^*(Y;X)$. For any distribution $R(\cdot)$ on $\mathcal{Y}$, that is not equal to $P_Y(\cdot)$ consider the non-constant function f given by $f(y) := \frac{R(y)}{P_Y(y)}$. This choice yields $\text{Ent}(f(Y)) = D(R||P_Y)$ and $\text{Ent}(T_{X;Y}f(X)) = D(\sum_Y P_{X|Y} * R||P_X))$ which gives $s \geq s^*(Y;X)$. Homogeneity of $\text{Ent}(\cdot)$ then completes the proof.

Corollary 6 follows simply from Theorem 21 as below.

Proof: The existence of the limit and its value both follow from Theorem 21 and the following well-known duality result that follows from the equivalent formulations of the hypercontractivity ribbon in inequalities (5.6), (5.7): For $1 < q < p$ or $1 > q > p$,

$$(p,q) \in \mathcal{R}_{X;Y} \iff (q', p') \in \mathcal{R}_{Y;X}. \tag{5.31}$$

Corollary 7 also follows easily as an easy consequence of Theorems 20, 21 and Corollary 6. We present the proof of Theorem 22 below.

Proof: Suppose U takes values in $\mathcal{U}$ and satisfies $U - X - Y$. Let $\Pr(U = u) = p(u) > 0, \Pr(X = x|U = u) = R_X^{(u)}(x)$ so that $\sum_u p(u) R_X^{(u)}(x) = P_X(x)$. Then, $\Pr(Y = y|U = u) = \sum_X P_{Y|X} * R_X^{(u)}$. Thus,

$$\frac{I(U;Y)}{I(U;X)} = \frac{\sum_u p(u) D(\sum_X P_{Y|X} * R_X^{(u)}||P_Y)}{\sum_u p(u) D(R_X^{(u)}||P_X)} \tag{5.32}$$

$$\leq \max_{u: R_X^{(u)} \neq P_X} \frac{D(\sum_X P_{Y|X} * R_X^{(u)}||P_Y)}{D(R_X^{(u)}||P_X)} \tag{5.33}$$

$$\leq \sup_{R: R \neq P_X} \frac{D(\sum_X P_{Y|X} * R||P_Y)}{D(R||P_X)} = s^*(X;Y). \tag{5.34}$$

Thus, $\sup_{U:U-X-Y} \frac{I(U;Y)}{I(U;X)} \leq \sup_{R:R\neq P_X} \frac{D(\sum_X P_{Y|X}*R||P_Y)}{D(R||P_X)}$.

Conversely, fix a distribution $R \neq P_X$. Define U_ϵ satisfying $U_\epsilon - X - Y$ taking values in $\mathcal{U} = \{1,2\}$, with $\Pr(U_\epsilon = 1) = \epsilon, \Pr(U_\epsilon = 2) = 1 - \epsilon$. Let $\Pr(X = x|U_\epsilon = 1) = R(x), \Pr(X = x|U_\epsilon = 2) = \frac{P_X(x)}{1-\epsilon} - \frac{\epsilon R(x)}{1-\epsilon}$. For sufficiently small ϵ, this defines the probability distribution of a joint triple (U, X, Y). A simple calculation gives

$$\lim_{\epsilon \downarrow 0} \frac{I(U_\epsilon; Y)}{I(U_\epsilon; X)} = \frac{D(\sum_X P_{Y|X} * R||P_Y)}{D(R||P_X)}. \tag{5.35}$$

This gives $\sup_{U:U-X-Y} \frac{I(U;Y)}{I(U;X)} \geq \sup_{R:R \neq P_X} \frac{D(\sum_X P_{Y|X} * R||P_Y)}{D(R||P_X)}$.

Thus, we have $\sup_{U:U-X-Y} \frac{I(U;Y)}{I(U;X)} \leq \sup_{R:R \neq P_X} \frac{D(\sum_X P_{Y|X} * R||P_Y)}{D(R||P_X)} = s^*(X;Y)$, completing the proof.

5.5 Non-interactive simulation of a three random variable joint distribution

This section discusses an interesting example. Consider joint distributions $P(x,y,z), Q(u,v,w)$ with binary random variables X, Y, Z and U, V, W. Fix $0 < \epsilon < \frac{1}{2}$. Let $X \sim \text{Ber}(\frac{1}{2})$ and $Y = X + N_1, Z = Y + N_2$ where $N_1, N_2 \sim \text{Ber}(\epsilon)$ are independent of X with $P(N_1 = N_2 = 0) = 1 - \frac{3\epsilon}{2}, P(N_1 = 0, N_2 = 1) = P(N_1 = 1, N_2 = 0) = P(N_1 = N_2 = 1) = \frac{\epsilon}{2}$. Let $U \sim \text{Ber}(\frac{1}{2})$ and $V = U + N_3, W = V + N_4$ where $N_3, N_4 \sim \text{Ber}(\epsilon)$ such that U, N_3, N_4 are independent. Note that $(X,Y),(Y,Z),(X,Z),(U,V),(V,W) \sim \text{DSBS}(\epsilon)$ and $(U,W) \sim \text{DSBS}(2\epsilon(1-\epsilon))$ as shown in the Fig. 5.5

Consider the problem where three agents try to simulate a triple joint distribution as follows. Agents A, B, C observe X^n, Y^n, Z^n respectively and output $\tilde{U}, \tilde{V}, \tilde{W}$, respectively which is required to be close in total variation to the target distribution (U, V, W) as shown.

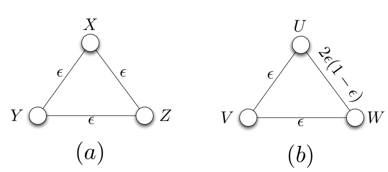


Figure 5.5: Three random variable simulation example

(a) represents the source distribution and (b) represents the target distribution.

As discussed earlier, non-interactive simulation of a DSBS target distribution with parameter $q < \frac{1}{2}$ using a DSBS source distribution with parameter $p < \frac{1}{2}$ is possible if and only if the target distribution is more noisy, i.e. $p \leq q$. Thus, for this example, each pair of agents can perform the marginal pair simulation desired of them. However, the three agents cannot simulate the desired triple joint distribution. Calculation shows

$$\rho(X, Z; Y) = \frac{1 - 2\epsilon}{\sqrt{1 - \epsilon}}, \tag{5.36}$$

$$\rho(U, W; V) = \frac{1 - 2\epsilon}{\sqrt{1 - 2\epsilon + 2\epsilon^2}}. \tag{5.37}$$

For $0 < \epsilon < \frac{1}{2}$, we have $1-2\epsilon+2\epsilon^2 < 1-\epsilon$, which gives $\rho(X, Z; Y) < \rho(U, W; V)$. This shows that even if agents A and C were merged into one agent $\tilde{A}$, then $\tilde{A}$ and B cannot achieve the desired non-interactive simulation.

Chapter 6

Discussion and concluding remarks

In this dissertation, we studied a simple outer bound for the capacity region of wireline and wireless networks - the Generalized Network Sharing bound. This bound turned out to be strong enough to show approximate optimality of routing in wireline networks and simple separation strategies in wireless networks under suitable symmetry assumptions. The following are some future directions:

Generalized Network Sharing bound for general networks

The cutset bound [21, 14] is a general outer bound on the capacity region of a general discrete memoryless network. A natural direction is to use the simple idea of the GNS bound to obtain a new bound for a general discrete memoryless network that is always at least as good as the cutset bound and gives an improvement whenever the GNS idea can kick in. This can be useful in other network problems such as broadcast packet erasure networks.

Boolean functions

The following interesting conjecture appeared recently in [42]:

Conjecture 3 *(from [42]) Let* $X \sim \mathrm{Ber}\left(\frac{1}{2}\right)$ *and let* Y *be the output of a binary symmetric channel with crossover probability* ϵ*, under input* X*. If* $\{(X_i, Y_i)\}_{i=1}^n$ *are independent and identically distributed with the same distribution as* (X, Y)*, and* $b : \{0,1\}^n \to \{0,1\}$ *is any Boolean function, then* $I(b(X^n); Y^n) \leq 1 - h(\epsilon)$*, where* $h(\epsilon) = \epsilon \log_2 \frac{1}{\epsilon} + (1-\epsilon)\log_2 \frac{1}{1-\epsilon}$.

It can be shown that using maximal correlation to constrain the set of possible joint distributions that $(b(X^n), Y^n)$ can have is not sufficient to prove this conjecture. From Theorem 19, we can show that hypercontractivity is stronger at constraining the set of possible joint distributions. This sounds like a promising approach to prove the conjecture.

Bibliography

[1] A. Agarwal, N. Alon, and M. Charikar. "Improved approximation for directed cut problems". In: *Proc. of ACM STOC*. 2007, pp. 671–680.

[2] R. Ahlswede and P. Gács. "Spreading of sets in product spaces and hypercontraction of the Markov operator". In: *Annals of Probability* 4 (1976), pp. 925–939.

[3] R. Ahlswede et al. "Network Information Flow". In: *IEEE Transactions on Information Theory* 46.4 (2000), pp. 1204–1216.

[4] V. Anantharam and V. Borkar. "Common Randomness and Distributed Control: A counterexample". In: *Systems and Control Letters* 56.7-8 (2007), pp. 568–572.

[5] V. Anantharam et al. "On Maximal Correlation, Hypercontractivity, and the Data Processing Inequality studied by Erkip and Cover". In: *arXiv:1304.6133 [cs.IT]* (Apr. 2013).

[6] F. Arbabjolfaei et al. "On the Capacity Region for Index Coding". In: *arXiv:1302.1601 [cs.IT]* (Feb. 2013).

[7] S. Arora, S. Rao, and U. Vazirani. "Expander Flows, Geometric Embeddings, and Graph Partitionings". In: *JACM* 56.2 (2009).

[8] A. Avestimehr, S. Diggavi, and D. Tse. "Wireless network information flow: A deterministic approach". In: *IEEE Transactions on Information Theory* 57.4 (2011), pp. 1872–1905.

[9] A. Bogdanov and E. Mossel. "On extracting common random bits from correlated sources". In: *preprint at arxiv: 1007.2315v2* (2010).

[10] T. Chan and A. Grant. "Mission Impossible: Computing the network coding capacity region". In: *Proc. of IEEE ISIT*. Toronto, Canada, 2008.

[11] C. Chekuri et al. "Multicommodity Flows and Cuts in Polymatroidal Networks". In: *arXiv:1110.6832 [cs.IT]* (Oct. 2011).

[12] C. Chekuri et al. *Multicommodity flows in polymatroidal capacity networks*. `http://www.ifp.illinois.edu/~pramodv/pubs/techreport.pdf`. 2011.

[13] J. Chuzhoy and S. Khanna. "Polynomial Flow-Cut Gaps and Hardness of Directed Cut problems". In: *Proc. ACM STOC*. 2007, pp. 179–188.

[14] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* 2nd. Wiley-Interscience, 2006.

[15] P. Cuff. “Communication requirements for generating correlated random variables”. In: *Proc. of IEEE ISIT.* Toronto, Canada, 2008.

[16] P. Cuff, H. Permuter, and T. Cover. “Coordination Capacity”. In: *IEEE Transactions on Information Theory* 56.9 (2010), pp. 4181–4206.

[17] E. Dahlhaus et al. “The complexity of multiterminal cuts”. In: *SIAM Journal on Computing* 23.4 (1994), pp. 864–894.

[18] R. Dougherty, C. Freiling, and K. Zeger. “Insufficiency of Linear Coding in Network Information Flow”. In: *IEEE Transactions on Information Theory* 51.8 (2005), pp. 2745–2759.

[19] R. Dougherty, C. Freiling, and K. Zeger. “Linear Network Codes and Systems of Polynomial Equations”. In: *IEEE Transactions on Information Theory* 54.5 (2008), pp. 2303–2316.

[20] R. Dougherty, C. Freiling, and K. Zeger. “Networks, Matroids and Non-Shannon Information Inequalities”. In: *IEEE Transactions on Information Theory* 53.6 (2007), pp. 1949–1969.

[21] Abbas El Gamal. “On Information Flow in Relay Networks”. In: *IEEE National Telecommunications Conference* 2 (1981), pp. D4.1.1–D4.1.4.

[22] E. Erkip and T. Cover. “The efficiency of investment information”. In: *IEEE Transactions On Information Theory* 44 (1998), pp. 1026–1040.

[23] P. Gács and J. Körner. “Common information is far less than mutual information”. In: *Problems of Control and Information Theory* 2.2 (1972), pp. 119–162.

[24] A.A. Gohari and V. Anantharam. “Generating Dependent Random Variables over Networks”. In: *Proceedings of the IEEE Information Theory Workshop.* Paraty, Brazil, 2011, pp. 698–702.

[25] N.J.A. Harvey, R.D. Kleinberg, and A.R. Lehman. “Comparing network coding with the multicommodity flow for the k-pairs communication problem”. In: *MIT LCS, Tech.Rep. 964.* 2004.

[26] N.J.A. Harvey, R.D. Kleinberg, and A.R. Lehman. “On the capacity of information networks”. In: *IEEE Transactions on Information Theory* 52.6 (2006), pp. 2345–2364.

[27] R. Hassin. “On Network Flows”. In: *Ph.D Dissertation, Yale University* (1978).

[28] S. Kamath. *Generalized Network Sharing Outer Bound and the Two-Unicast Problem.* `http://www.eecs.berkeley.edu/~sudeep/Masters_Thesis_GNS_Kamath.pdf`. 2011.

[29] S. Kamath and V. Anantharam. "Non-interactive Simulation of Joint Distributions: The Hirschfeld-Gebelein-Rényi Maximal Correlation and the Hypercontractivity Ribbon". In: *Proc. of the 50th Allerton Annual Conference on Communication, Control and Computing*. 2012.

[30] S. Kamath, S. Kannan, and P. Viswanath. "Wireless networks under symmetric demands". In: *Proc. of IEEE ISIT*. MIT, Cambridge, MA, 2012.

[31] S. Kamath and D. Tse. "On the Generalized Network Sharing bound and edge-cut bounds for network coding". In: *Proc. of IEEE ISIT*. Istanbul, Turkey, 2013.

[32] S. Kamath, D.N.C. Tse, and V. Anantharam. "Generalized Network Sharing Outer Bound and the Two-Unicast Problem". In: *Proc. of International Symposium on Network Coding*. Beijing, China, 2011.

[33] S. Kamath and P. Viswanath. "An information-theoretic meta-theorem on edge-cut bounds". In: *Proc. of IEEE ISIT*. MIT, Cambridge, MA, 2012.

[34] S. Kannan, A. Raja, and P. Viswanath. "Local PHY + Global Flow: A Layering Principle for Wireless Networks". In: *Proc. of IEEE ISIT*. Saint Petersburg, Russia, 2011.

[35] S. Kannan and P. Viswanath. "Capacity of Multiple Unicast in Wireless Networks: A Polymatroidal Approach". In: *arXiv:1111.4768 [cs.IT]* (Nov. 2011).

[36] S. Kannan and P. Viswanath. "Multiple-Unicast in Fading Wireless Networks: A Separation Scheme is Approximately Optimal". In: *Proc. of IEEE ISIT*. St. Petersburg, Russia, 2011.

[37] P.N. Klein et al. "Bounds on the Max-Flow Min-Cut Ratio for Directed Multicommodity Flows". In: *J. Algorithms* 22 (1997), pp. 241–269.

[38] R. Koetter and M. Médard. "An algebraic approach to network coding". In: *IEEE/ACM Transactions on Networking* 11.5 (2003).

[39] G. Kramer and S. Savari. "Edge-cut bounds on network coding rates". In: *Journal of Network and Systems Management* 14.1 (2006), pp. 49–67.

[40] G. Kramer, S. Yazdi, and S. Savari. "Refined edge-cut bounds for network coding". In: *Proc. of the Int. Zurich Seminar on Communications (IZS)*. 2010.

[41] G. Kumar. "On sequences of pairs of dependent random variables: A simpler proof of the main result using SVD". In: *On webpage* (2010, `http://www.stanford.edu/~gowthamr/research/Witsenhausen_simpleproof.pdf`).

[42] G. Kumar and T. Courtade. "Which Boolean Functions are Most Informative?" In: *Proc. of IEEE ISIT*. Istanbul, Turkey, 2013.

[43] E.L. Lawler and C.U. Martel. "Computing maximal polymatroidal network flows". In: *Math. Oper. Res.* 7.3 (1982).

[44] F.T. Leighton and S. Rao. "An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms". In: *Proc. of 28th Annual Symposium on Foundations of Computer Science*. Los Alamitos, California, 1988.

[45] Z. Li and B. Li. "Network coding: The case of multiple unicast sessions". In: *Proc. of the 42nd Allerton Annual Conference on Communication, Control and Computing.* 2004.

[46] N. Linial, E. London, and Y. Rabinovich. "The Geometry of Graphs and some of its Algorithmic Applications". In: *Combinatorica* 15.2 (1995), pp. 215–245.

[47] F. Matus. "Infinitely many information inequalities". In: *Proc. of IEEE ISIT*. Nice, France, 2007, pp. 24–29.

[48] E. Mossel, K. Oleszkiewicz, and A. Sen. "On Reverse Hypercontractivity". In: *preprint at arxiv: 1108.1210v1* (2011).

[49] E. Mossel et al. "Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality". In: *preprint at arxiv: 0410560v1* (2004).

[50] A. Motahari, S.O. Gharan, and A. Khandani. "Real Interference Alignment with Real Numbers". In: *arXiv:0908.1208 [cs.IT]* (Aug. 2009).

[51] J. Naor and L. Zosin. "A 2-approximation algorithm for the directed multiway cut problem." In: *SIAM Journal on Computing* 31.2 (2001), pp. 477–482.

[52] B. Nazer et al. "Ergodic Interference Alignment". In: *Proc. of IEEE ISIT*. Seoul, South Korea, 2009.

[53] O. Oyman et al. "Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels". In: *Proc. of IEEE Globecom.* 2002, pp. 1172–1176.

[54] A. Rényi. "On measures of dependence". In: *Acta. Math. Acad. Sci. Hung.* 10 (1959), pp. 441–451.

[55] C.E. Shannon. "A mathematical theory of communication". In: *Bell System Technical Journal* 27 (1948), pp. 379–423.

[56] S. Thakor, A. Grant, and T. Chan. "Network coding capacity: A functional dependence bound". In: *Proc. of IEEE ISIT*. 2009.

[57] H.S. Witsenhausen. "On sequences of pairs of dependent random variables". In: *SIAM Journal on Applied Mathematics* 28.1 (1975), pp. 100–113.

[58] A.D. Wyner. "The common information of two dependent random variables". In: *IEEE Transactions On Information Theory* 21.2 (1975), pp. 163–179.

[59] X. Xu, S. Thakor, and Y.L. Guan. "Weighted Sum-Rate Functional Dependence Bound for Network Coding Capacity". In: *Proc. of the International Symposium on Information Theory and its Applications (ISITA).* Honolulu, Hawaii, 2012.

[60] X. Yan, J. Yang, and Z. Zhang. "An outer bound for multisource multisink network coding with minimum cost consideration". In: *IEEE Transactions on Information Theory* 52.6 (2006), pp. 2373–2385.

[61] M.H. Yassaee, A.A. Gohari, and M.R. Aref. "Channel Simulation via Interactive Communications". In: *preprint at arxiv: 1203.3217* (2012).

[62] R.W. Yeung and Z. Zhang. "Distributed Source Coding for Satellite Communications". In: *IEEE Transactions on Information Theory* 45.4 (1999), pp. 1111–1120.

[63] Z. Zhang and R.W. Yeung. "On characterization of entropy function via information inequalities". In: *IEEE Transactions on Information Theory* 44.4 (1998), pp. 1440–1452.