ON THE FUNDAMENTAL LIMITS AND SYMMETRIC DESIGNS FOR DISTRIBUTED INFORMATION SYSTEMS

A Dissertation

by

AMIR SALIMI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Shuguang Cui
Co-Chair of Committee,	Tie Liu
Committee Members,	P. R. Kumar
	Catherin Yan
Head of Department,	Miroslav M. Begovic

December 2016

Major Subject: Electrical Engineering

ABSTRACT

Many multi-terminal communication networks, content delivery networks, cache networks, and distributed storage systems can be modeled as a broadcast network. An explicit characterization of the capacity region of the general network coding problem is one of the best known open problems in network information theory. A simple set of bounds that are often used in the literature to show that certain rate tuples are infeasible are based on the graph-theoretic notion of cut. The standard cut-set bounds, however, are known to be loose in general when there are multiple messages to be communicated in the network. This dissertation focuses on broadcast networks, for which the standard cut-set bounds are closely related to union as a specific set operation to combine different simple cuts of the network. A new set of explicit network coding bounds, which combine different simple cuts of the network via a variety of set operations (not just the union), are established via their connections to extremal inequalities for submodular functions. The tightness of these bounds are demonstrated via applications to combination networks.

The tightness of generalized cut-set bounds has been further explored by studying the problem of "latency capacity region" for a broadcast chanel. An *implicit* characterization of this region has been proved by Tian, where a rate splitting based scheme was shown to be optimal. However, the *explicit* characterization of this region was only available when the number of receivers are less than three. In this dissertation, a precise polyhedral description of this region for a symmetric broadcast channel with complete message set and arbitrary number of users has been established. It has been shown that a set of *generalized cut-set bounds*, characterizes the entire symmetrical multicast region. The achievability part is proved by showing that every maximum rate vector is feasible by using a successive encoding scheme. The framework for achievability strongly relies on polyhedral combi-

natorics and it can be useful in network information theory problems when a polyhedral description of a region is needed.

Moreover, It is known that there is a direct relationship between network coding solution and characterization of entropy region. This dissertation, also studies the symmetric structures in network coding problems and their relation with symmetrical projections of entropy region and introduces new aspects of entropy inequalities. First, inequalities relating average joint entropies rather than entropies over individual subsets are studied. Second, the existence of non-Shannon type inequalities under partial symmetry is studied using the concepts of Shannon and non-Shannon groups. Finally, due to the relationship between linear entropic vectors and representability of integer polymatroids, construction of such vector has been discussed. Specifically, It is shown that representability of the particularly constructed matroid is a sufficient condition for integer polymatroids to be linearly representable over real numbers. Furthermore, it has been shown that any real-valued submodular function (such as Shannon entropy) can be approximated (arbitrarily close) by an integer polymatroid.

DEDICATION

To my mom,

Nahideh Baradaran.

ACKNOWLEDGEMENTS

I would like to genuinely thank my family members who are behind me all the time. Thank my parents for their endless love and support. I would like to especially thank my advisors, Prof. Shuguang Cui and Prof. Tie Liu, for providing me this opportunity to pursue my PhD. It was a privilege working closely with them, and I learned something new anytime we had discussions on research problems.

I would like also like to thank Prof. Muriel Mèdard, for her help and thoughtful discussions, which led to the results in the last chapter of this dissertation.

Finally, I would like to thank all my committee members, Prof. P. R. Kumar, Prof. Chatherine Yan, for their time and support on my prelim exam, defense and the dissertation. Their valuable comments help me a lot in improving the quality of my work.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION*

Recent developments in the distributed storage systems and the importance of content delivery in distributed cache networks reemphasize the necessity to design reliable distributed networks. As the size of data centers increases, theoretical analysis of the robust and efficient encoding schemes over storage nodes becomes more challenging. In this dissertation, a systematic framework has been proposed to efficiently compute the fundamental limits of such systems, using mathematical tools from combinatorial optimization and polyhedral combinatorics.

In recent years, there has been a significant interest around developing computationally efficient tools to design and analyze complex information systems. One perspective to reduce the complexity is to exploit the underlying combinatorial structure in the given system. When we are dealing with a structured networks, such as distributed cloud systems, the combinatorial structure of the underlying graph and distribution of the messages, can be extremely helpful in developing such efficient tools. This perspective has been explored in this thesis, where this theory is developed using tools form information theory, network coding and combinatorial optimization. The second part of the dissertation, focuses on the projections of entropy region, which has been another well known open problem in information theory.

1.1 Generalized Cut-Set Bounds: Applications in Distributed Cloud Storage

From a theoretical point of view, many multi-terminal communication networks, such as distributed storage systems and content delivery/cache networks, can be modeled as

*Part of this chapter is reprinted, with permission, from [A. Salimi, T. Liu, and S. Cui, "Generalized cut-set bounds for broadcast networks," *IEEE Transactions on Information Theory*, vol. 61, no. 6, pp. 1–14, Jun. 20155]

broadcast networks. Nevertheless, characterizing the fundamental limits of communication for such networks is a well-known open problem. The typical approach in the literature to show the infeasibility of certain rate of data has been the simple cut-set bound, which is defined based on the graph theoretic notion of the cut. The standard cut-set bounds, however, are known to be loose. In this dissertation, this gap has been systematically improved via introducing the "generalized cut-set bounds". Specifically, a new mathematical notion of "extremal submodular inequality" has been proposed; and new set of explicit network coding bounds, which combine different simple cuts of the network via a variety of set operations, were established via their connections to extremal inequalities for submodular functions. Furthermore, the tightness of these proposed bounds were demonstrated via applications to networks arising from distributed storage systems, known as combination networks.

1.2 Latency Capacity Region

The capacity region of broadcast channel has been studied in various different settings [1]. The general broadcast channel with complete message set represents a communication scenario in which the source node transmits a different message to each distinct subset of receivers. More specifically, for the general broadcast channel with the source node s and K receivers $\{t_k : k \in [K]\}$, a *complete* message set at the source node s consists of $2^K - 1$ independent messages pertaining to each non-empty subset of receivers:

$$\mathcal{W} = \{w_{\mathcal{U}} : \emptyset \neq \mathcal{U} \subseteq [K]\}$$

where the message $w_{\mathcal{U}}$ is intended for all sink nodes from $\{t_k : k \in \mathcal{U}\}$. Thus, the set of the messages intended for the sink nodes t_k is given by:

$$\mathcal{W}_k = \{w_{\mathcal{U}} : k \in \mathcal{U} \subseteq [K]\}, \quad \forall k \in [K]. \tag{1.1}$$

With a slight abuse of notation, we shall denote the rate of the message $w_{\mathcal{U}}$ by $R_{\mathcal{U}}$ (instead of the more consistent notation $R_{w_{\mathcal{U}}}$).

In this set up, suppose that the achievability of a rate tuple $\mathbf{C} := (C_{\mathcal{U}} : \mathcal{U} \subseteq [K])$ in the $2^K - 1$ dimensional capacity region is given. The fundamental question of interest is that what are the set of all points that their achievability can be inferred just by knowing the achievability of the rate tuple $\mathbf{C}$)? The closure of all such achievable rate tuples is referred to as $\mathbf{C}$-*multicast region* for a broadcast channel. It is trivial to see that all the rate tuples that are element-wise marginalized by $\mathbf{C}$, belong to the $\mathbf{C}$-multicast region. For simplicity, we may drop the term $\mathbf{C}$ and we will refer to this region as multicast region.

It has been shown that the multicast region of the broadcast channel with complete message set, is independent of the quality of the broadcast channel [2]. Furthermore, it can be *essentially* transformed into a network coding problem over a *combination network* [3]. In this regard, the capacity region of the associated network coding problem is the multicast region of a broadcast channel. An *implicit* characterization of the "symmetrical multicast region" for a broadcast channel has been proved by Tian, where a rate splitting based scheme was shown to be optimal. However, the *explicit* characterization of this region was only available when the number of receivers are less than three. In this paper, we establish a precise polyhedral description of this region for a symmetric broadcast channel with complete message set and arbitrary number of users. We show that a set of *generalized cut-set bounds*, characterizes the entire symmetrical multicast region. The achievability part is proved by showing that every maximum rate vector is feasible by using a successive

encoding scheme. Our framework for achievability strongly relies on polyhedral combinatorics and it can be useful in network information theory problems when a polyhedral description of a region is needed.

1.3 Group-Induced Symmetrical Structures in Distributed Storage Systems

The importance of the characterization of entropy region is due to its direct relation to the capacity region of general network coding problem. From the theoretical viewpoint, the problem of characterizing the entropy region, is very challenging due to the existence of the so-called *non-Shannon type* inequalities; which essentially means that submodularity of entropy function is not sufficient for understanding the fundamental limits of such engineering problems. From the practical point of viewpoint, a main challenge for using entropy region to characterize network coding capacity regions is the huge dimension of this region. This research was motivated by the observation that for many network coding coding problems arising from the communication engineering contexts, the structure of the problem may strongly suggest that instead of considering the entire entropy region, one may only need to consider the projection of the (Shannon) entropy region. In this dissertation, it has been shown that that imposing certain symmetric structures to the design of such systems can alleviate both problems. From practical viewpoint, symmetry is ubiquitous in real world engineering models, and in fact many current distributed storage systems, such as Hadoop or Google File System, adopt symmetric architecture in their design. To make symmetric principles mathematically rigorous, we define symmetry as an invariance with respect to a permutation group. Under some subgroups of the symmetric group and cyclic group with certain orders, our results show a significant reduction in the dimension of the problem. Our results show that under some subgroups of the symmetric group and cyclic group with certain orders we can significantly reduce the dimension of the problem. Additionally, the submodularity of entropy function is sufficient to charac-

terize the entire capacity region. This result will be helpful to identify alternative flexible structures for distributed file systems such as Hadoop.

2. GENERALIZED CUT-SET BOUNDS FOR BROADCAST NETWORKS

2.1 Introduction

A *classical network* is a capaciated directed acyclic graph $((V, A), (C_a : a \in A))$, where V and A are the node and the arc sets of the graph respectively, and C_a is the link capacity for arc $a \in A$. A *broadcast network* is a classical work for which all source messages are *collocated* at a single source node.

Consider a general broadcast network with one source node s and K sink nodes t_k, $k = 1, \ldots, K$ (see Figure 2.1). The source node s has access to a collection of *independent* messages $\mathsf{W}_I = (\mathsf{W}_i : i \in I)$, where I is a finite index set. The messages intended for the sink node t_k are given by W_{I_k}, where I_k is a nonempty subset of I. When all messages from W_I are *unicast* messages, i.e., each of them is intended for *only* one of the sink nodes, it follows from the celebrated max-flow min-cut theorem [4] that routing can achieve the entire capacity region of the network. On the other hand, when some of the messages from W_I are *multicast* messages, i.e., they are intended for *multiple* sink nodes, the capacity region of the network is generally *unknown* except when there is only one multicast message at the source node [5–7] or there are only two sink nodes ($K = 2$) in the network [8–10].

In this chapter, we are interested in establishing strong network coding bounds for *general* broadcast networks with multiple (multicast) messages and more than two sink nodes ($K \geq 3$). In particular, we are interested in network coding bounds that rely only on the *cut* structure of the network. The rational behind this particular interest is two-folded. First, cut is a well-understood combinatorial structure for networks. Second, the fact that standard cut-set bounds [1, Ch. 15.10] are *tight* for the aforementioned special cases [4–10] suggests that cut as a combinatorial structure can be useful for more general

broadcast-network coding problems as well.

The starting point of this work is the following simple observation. For each $k = 1, \ldots, K$, let A_k be a "basic" cut that separates the source node s from the (single) sink node t_k. Then, for any nonempty subset $U \subseteq [K] := \{1, \ldots, K\}$ the union $\cup_{k \in U} A_k$ is also a cut that separates the source node s from the "super" sink node t_U, whose intended messages are given by $\mathsf{W}_{\cup_{k \in U} I_k}$. By the standard cut-set bound [1, Ch. 15.10], we have

$$R(\cup_{k \in U} I_k) \leq C(\cup_{k \in U} A_k) \tag{2.1}$$

for any achievable rate tuple $R_I := (R_i : i \in I)$. Here, $R : 2^I \to \mathbb{R}^+$ is the *rate* function that corresponds to the rate tuple R_I and is given by

$$R(I') := \sum_{i \in I'} R_i, \quad \forall I' \subseteq I, \tag{2.2}$$

and $C : 2^A \to \mathbb{R}^+$ is the *capacity* function of the network where

$$C(A') := \sum_{a \in A'} C_a, \quad \forall A' \subseteq A. \tag{2.3}$$

Note that the above observation depends critically on the fact that all messages W_I are *collocated* at the source node s. When the messages are *distributed* among several source nodes, it is well known that the union of several basic cuts may *no longer* be a cut that separates the super source node from the super sink node and hence may not lead to any network coding bounds [11].

Based on the above discussion, it is clear that for broadcast networks the *standard* cut-set bounds [1, Ch. 15.10] are closely related to *union* as a specific set operation to combine different basic cuts of the network. Therefore, a natural question that one may

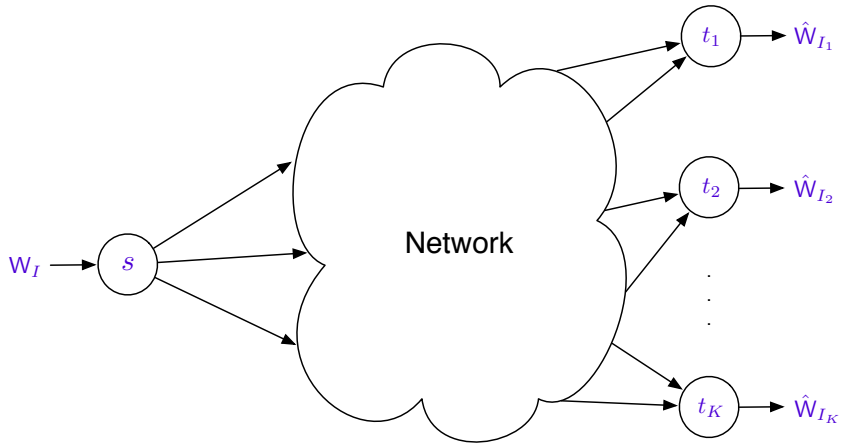

Figure 2.1: Illustration of a general broadcast network.

ask is whether there are any other set operations (besides the union) that will also lead to nontrivial network coding bounds.

In this chapter, we provide a positive answer to the above question by establishing a new set of network coding bounds for general broadcast networks. We term these bounds *generalized* cut-set bounds based on the facts that: 1) they rely only on the cut structure of the network; and 2) the set operations within the rate and the capacity functions are *identical* (but not just the union any more), both similar to the case of standard cut-set bounds as in (2.1). From the proof viewpoint, as we shall see, these bounds are established via only the *Shannon-type* inequalities. It is well known that all Shannon-type inequalities can be derived from the simple fact that Shannon entropy as a set function is *submodular* [12, Ch. 14.A]. So, at heart, the generalized cut-set bounds are reflections of several new results that we establish on submodular function optimization.

The rest of the chapter is organized as follows. In Section 2.2 we establish several new results on submodular function optimization, which we shall use to prove the generalized cut-set bounds. A new set of network coding bounds that relate *three* basic cuts of the

network is provided in Section 2.3. The proof of these bounds is rather "hands-on" and hence provides a good illustration on the essential idea on how to establish the generalized cut-set bounds. In Section 2.4, a new set of network coding bounds that relate arbitrary K basic cuts of the network is provided, generalizing the bounds provided in Section 2.3. In Section 2.5, the tightness of the generalized cut-set bounds is demonstrated via applications to *combination* networks [3]. Finally, in Section 3.3 we conclude the chapter with some remarks.

2.2 Modular and Submodular Functions

Let S be a finite ground set. A function $f : 2^S \to \mathbb{R}^+$ is said to be *submodular* if

$$f(S_1) + f(S_2) \geq f(S_1 \cup S_2) + f(S_1 \cap S_2), \quad \forall S_1, S_2 \subseteq S, \tag{2.4}$$

and is said to be *modular* if

$$f(S_1) + f(S_2) = f(S_1 \cup S_2) + f(S_1 \cap S_2), \quad \forall S_1, S_2 \subseteq S. \tag{2.5}$$

More generally, let S_k, $k = 1, \ldots, K$, be a subset of S. For any nonempty subset U of $[K]$ and any $r \in [|U|]$, let

$$S^{(r)}(U) := \cup_{\{U' \subseteq U : |U'| = r\}} \cap_{k \in U'} S_k. \tag{2.6}$$

Clearly, we have

$$\cup_{k \in U} S_k = S^{(1)}(U) \supseteq S^{(2)}(U) \supseteq \cdots \supseteq S^{(|U|)}(U) = \cap_{k \in U} S_k \tag{2.7}$$

for any nonempty $U \subseteq [K]$ and

$$S^{(r)}(U') \subseteq S^{(r)}(U) \tag{2.8}$$

for any $\emptyset \subset U' \subseteq U \subseteq [K]$ and any $r \in [|U'|]$. Furthermore, it is known that [13, Th. 2]

$$\sum_{k \in U} f(S_k) \geq \sum_{r=1}^{|U|} f(S^{(r)}(U)) \tag{2.9}$$

if f is a submodular function, and

$$\sum_{k \in U} f(S_k) = \sum_{r=1}^{|U|} f(S^{(r)}(U)) \tag{2.10}$$

if f is a modular function.

Note that the standard submodularity (2.9) relates $S^{(r)}(U)$ for different r but a *fixed* U. To establish the generalized cut-set bounds, however, we shall need the following technical results on modular and submodular functions that relate $S^{(r)}(U)$ for not only different r but also *different* U.

Lemma 1. *Let r' and J be two integers such that $0 \leq r' \leq J \leq K$. We have*

$$\sum_{r=1}^{r'} f(S_r) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r])) \geq \sum_{r=1}^{r'} f(S^{(r)}([J])) + \sum_{r=r'+1}^{J} f(S^{(r'+1)}([r])) \tag{2.11}$$

if f is a submodular function, and

$$\sum_{r=1}^{r'} f(S_r) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r])) = \sum_{r=1}^{r'} f(S^{(r)}([J])) + \sum_{r=r'+1}^{J} f(S^{(r'+1)}([r])) \tag{2.12}$$

if f is a modular function.

Note that when $r' = 0$, we have $S^{(r'+1)}([r]) = S^{(1)}([r]) = \cup_{k=1}^{r} S_k \supseteq S_r$ for any $r = 1, \ldots, J$. In this case, the inequality (2.11) reduces to the trivial equality

$$\sum_{r=1}^{J} f(S^{(1)}([r])) = \sum_{r=1}^{J} f(S^{(1)}([r])). \tag{2.13}$$

On the other hand, when $r' = J$, the inequality (2.11) reduces to the standard submodularity

$$\sum_{r=1}^{J} f(S_r) \geq \sum_{r=1}^{J} f(S^{(r)}([J])). \tag{2.14}$$

For the general case where $0 < r' < J$, a proof of the lemma is provided in Appendix A.1.

Let $S'_k := S_k \cup S_0$ for $k = 1, \ldots, K$. For any nonempty $U \subseteq [K]$ and any $r = 1, \ldots, |U|$ we have

$$\begin{aligned} S'^{(r)}(U) &= \cup_{\{U' \subseteq U : |U'| = r\}} \cap_{k \in U'} S'_k && (2.15)\\ &= \cup_{\{U' \subseteq U : |U'| = r\}} \cap_{k \in U'} (S_k \cup S_0) && (2.16)\\ &= \left(\cup_{\{U' \subseteq U : |U'| = r\}} \cap_{k \in U'} S_k\right) \cup S_0 && (2.17)\\ &= S^{(r)}(U) \cup S_0. && (2.18) \end{aligned}$$

Applying Lemma 1 for S'_k, $k = 1, \ldots, K$, and (2.18), we have the following corollary.

Corollary 2. *Let r' and J be two integers such that $0 \leq r' \leq J \leq K$, and let S_0 be a*

subset of S. We have

$$\sum_{r=1}^{r'} f(S_r \cup S_0) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r]) \cup S_0) \\ \geq \sum_{r=1}^{r'} f(S^{(r)}([J]) \cup S_0) + \sum_{r=r'+1}^{J} f(S^{(r'+1)}([r]) \cup S_0) \tag{2.19}$$

if f is a submodular function, and

$$\sum_{r=1}^{r'} f(S_r \cup S_0) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r]) \cup S_0) \\ = \sum_{r=1}^{r'} f(S^{(r)}([J]) \cup S_0) + \sum_{r=r'+1}^{J} f(S^{(r'+1)}([r]) \cup S_0) \tag{2.20}$$

if f is a modular function.

We shall also need the following lemma, for which a proof is provided in Appendix A.2.

Lemma 3. *Let U and T be two nonempty subsets of $[K]$. Write, without loss of generality, that $T = \{t_1, \ldots, t_{|T|}\}$ where $1 \leq t_1 < t_2 < \cdots < t_{|T|} \leq K$. Let q and r_q be two integers such that $1 \leq q \leq |U|$, $1 \leq r_q \leq |T|$, and $S^{(q)}(U) \subseteq S^{(r_q)}(T)$. We have*

$$\sum_{r=1}^{|T|} f(S_{t_r}) + r_q f(S^{(q)}(U)) \\ \geq \sum_{r=1}^{r_q} \big(f(S^{(r)}(T)) + f(S_{t_r} \cap S^{(q)}(U)) \big) + \sum_{r=r_q+1}^{|T|} f(S_{t_r} \cap (S^{(q)}(U) \cup S^{(r_q+1)}(\{t_1, \ldots, t_r\}))) \tag{2.21}$$

if f is a submodular function, and

$$\sum_{r=1}^{|T|} f(S_{t_r}) + r_q f(S^{(q)}(U)) = \sum_{r=1}^{r_q} \big(f(S^{(r)}(T)) + f(S_{t_r} \cap S^{(q)}(U))\big) + \sum_{r=r_q+1}^{|T|} f(S_{t_r} \cap (S^{(q)}(U) \cup S^{(r_q+1)}(\{t_1, \ldots, t_r\}))) \tag{2.22}$$

if f is a modular function.

For specific functions, let $\mathsf{Z}_S := (\mathsf{Z}_i : i \in S)$ be a collection of jointly distributed random variables, and let $H(\mathsf{Z}_S)$ be the joint (Shannon) entropy of Z_S. Then, it is well known [12, Ch. 14.A] that $H_{\mathsf{Z}} : 2^S \to \mathbb{R}^+$ where

$$H_{\mathsf{Z}}(S') := H(\mathsf{Z}_{S'}), \quad \forall S' \subseteq S \tag{2.23}$$

is a *submodular* function. Furthermore, it is straightforward to verify that the rate function $R(\cdot)$ (for a given rate tuple R_I) and the capacity function $C(\cdot)$, defined in (2.2) and (2.3) respectively, are *modular* functions.

2.3 Generalized Cut-Set Bounds Relating Three Basic Cuts of the Network

2.3.1 Main Result

Theorem 1. *Consider a broadcast network with a collection of independent messages W_I collocated at the source node s and $K \geq 3$ sink nodes t_k, $k = 1, \ldots, K$. For any $k = 1, \ldots, K$, let W_{I_k} be the intended messages for the sink node t_k, and let A_k be a basic*

cut that separates the source node s from the sink node t_k. We have

$$R(I_i \cup I_j \cup I_k) + R(I_i \cap I_j) \leq C(A_i \cup A_j \cup A_k) + C(A_i \cap A_j), \tag{2.24}$$

$$R(I_i \cup I_j \cup I_k) + R((I_i \cap I_j) \cup (I_i \cap I_k) \cup (I_j \cap I_k)) \leq C(A_i \cup A_i \cup A_k) + C((A_i \cap A_j) \cup (A_i \cap A_k) \cup (A_j \cap A_k)), \tag{2.25}$$

$$R(I_i \cup I_j \cup I_k) + R(I_i \cup I_j) + R(I_i \cap I_j \cap I_k) \leq C(A_i \cup A_j \cup A_k) + C(A_i \cup A_j) + C(A_i \cap A_j \cap A_k), \tag{2.26}$$

and $$2R(I_i \cup I_j \cup I_k) + R(I_i \cap I_j \cap I_k) \leq 2C(A_i \cup A_j \cup A_k) + C(A_i \cap A_j \cap A_k) \tag{2.27}$$

for any achievable rate tuple R_I and any three distinct integers i, j, and k from $[K]$.

Note that the left-hand sides of the generalized cut-set bounds (2.24)–(2.27) are *weighted* sum rates with integer weights on the rates of the messages from $\mathsf{W}_{I_i \cup I_j \cup I_k}$. Figure 2.2 illustrates the weight distributions for the generalized cut-set bounds (2.24)–(2.27).

2.3.2 Proof of Theorem 1

Let $(n, \{\mathsf{X}_a : a \in A\})$ be an *admissible* code with block length n, where X_a is the message transmitted over the arc a. By the independence bound [1, Th 2.6.6] and the link-capacity constraints, we have

$$H_{\mathsf{X}}(A') \leq \sum_{a \in A'} H(\mathsf{X}_a) \leq n \sum_{a \in A'} C_a = nC(A'), \quad \forall A' \subseteq A. \tag{2.28}$$

For notational simplicity, in this proof we shall assume *perfect* recovery of the messages at each of the sink nodes. It should be clear from the proof that by applying the well-known

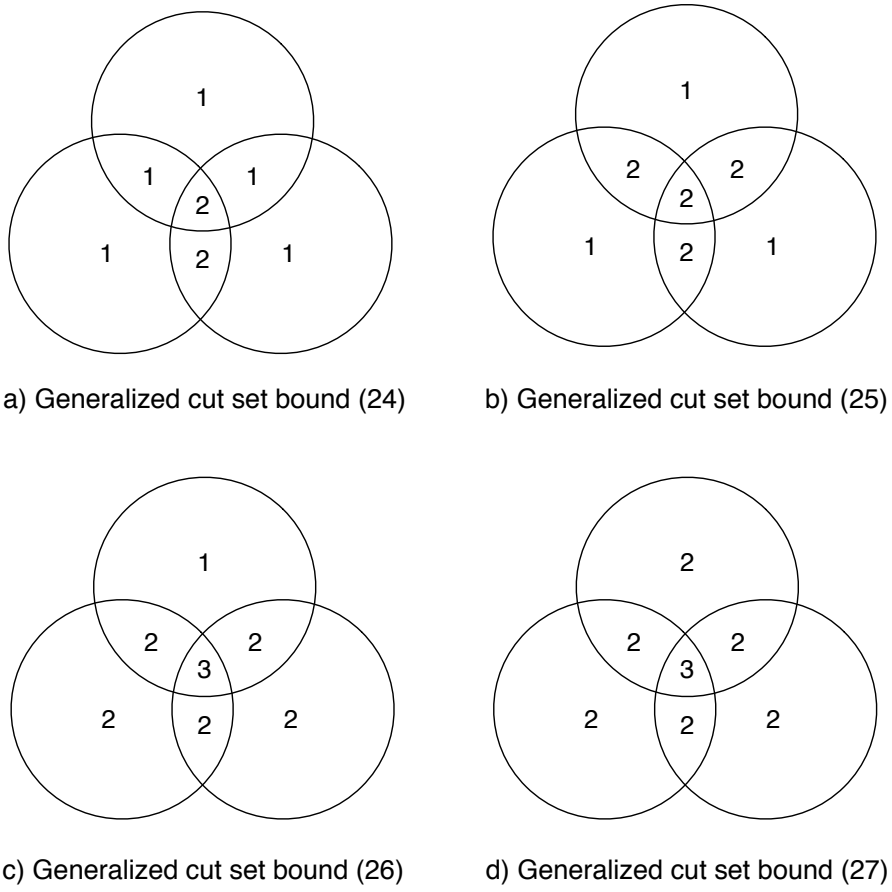

Figure 2.2: The weight distributions for the generalized cut-set bounds (2.24)–(2.27). Here, each circle represents the set of the messages intended for a particular sink node. The number within each separate area indicates the weight for the rates of the messages represented by the area.

Fano's inequality [1, Th 2.10.1], the results also hold for *asymptotically perfect* recovery. By the perfect recovery requirement, for any nonempty subset $U \subseteq [K]$ the collection of the messages $\mathsf{W}_{\cup_{k\in U} I_k}$ must be a *function* of the messages $\mathsf{X}_{\cup_{k\in U} A_k}$ transmitted over the s-t_U cut $\cup_{k\in U} A_k$. We thus have

$$H_{\mathsf{W}}(\cup_{k\in U} I_k) \leq H_{\mathsf{X}}(\cup_{k\in U} A_k), \quad \forall U \subseteq [K]. \tag{2.29}$$

Proof of (2.24). Let $U = \{i, j, k\}$ in (2.29). Denote by

$$I_{\mathsf{X}}(A_i; A_j) := I(\mathsf{X}_{A_i}; \mathsf{X}_{A_j}) \tag{2.30}$$

the mutual information between X_{A_i} and X_{A_j}. We have

$$\begin{aligned}
H_{\mathsf{W}}(I_i \cup I_j \cup I_k) &\leq H_{\mathsf{X}}(A_i \cup A_j \cup A_k) && (2.31)\\
&= H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j|A_i) + H_{\mathsf{X}}(A_k|A_i \cup A_j) && (2.32)\\
&= H_{\mathsf{X}}(A_i) + (H_{\mathsf{X}}(A_j) - I_{\mathsf{X}}(A_i; A_j)) + (H_{\mathsf{X}}(A_k) - I_{\mathsf{X}}(A_k; A_i \cup A_j)) && (2.33)\\
&= H_{\mathsf{X}}(A_i) + (H_{\mathsf{X}}(A_j) - I_{\mathsf{X},\mathsf{W}}(A_i, I_i; A_j, I_j)) + \\
&\quad (H_{\mathsf{X}}(A_k) - I_{\mathsf{X},\mathsf{W}}(A_k, I_k; A_i \cup A_j, I_i \cup I_j)) && (2.34)\\
&\leq H_{\mathsf{X}}(A_i) + (H_{\mathsf{X}}(A_j) - H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j)) + \\
&\quad (H_{\mathsf{X}}(A_k) - H_{\mathsf{X},\mathsf{W}}(A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j))) && (2.35)
\end{aligned}$$

where (2.34) follows from the fact that: 1) W_{I_i} and W_{I_j} are functions of X_{A_i} and X_{A_j} respectively so we have $I_{\mathsf{X}}(A_i; A_j) = I_{\mathsf{X},\mathsf{W}}(A_i, I_i; A_j, I_j)$; and 2) W_{I_k} and $\mathsf{W}_{I_i \cup I_j}$ are functions of X_{A_k} and $\mathsf{X}_{A_i \cup A_j}$ respectively so we have $I_{\mathsf{X}}(A_k; A_i \cup A_j) = I_{\mathsf{X},\mathsf{W}}(A_k, I_k; A_i \cup A_j, I_i \cup I_j)$, and (2.35) follows from the fact that

$$\begin{aligned}
I_{\mathsf{X},\mathsf{W}}(A_i, I_i; A_j, I_j) &\geq I_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j; A_i \cap A_j, I_i \cap I_j) && (2.36)\\
&= H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j) && (2.37)
\end{aligned}$$

and

$$
\begin{aligned}
&I_{\mathsf{X},\mathsf{W}}(A_k, I_k; A_i \cup A_j, I_i \cup I_j) \\
&\quad \geq I_{\mathsf{X},\mathsf{W}}(A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j); A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j)) && (2.38)\\
&\quad = H_{\mathsf{X},\mathsf{W}}(A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j)). && (2.39)
\end{aligned}
$$

Note that we trivially have

$$
H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j) \geq H_{\mathsf{W}}(I_i \cap I_j) \tag{2.40}
$$

$$
\text{and} \quad H_{\mathsf{X},\mathsf{W}}(A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j)) \geq H_{\mathsf{X}}(A_k \cap (A_i \cup A_j)). \tag{2.41}
$$

Substituting (2.40) and (2.41) into (2.35) gives

$$
H_{\mathsf{W}}(I_i \cup I_j \cup I_k) + H_{\mathsf{W}}(I_i \cap I_j) \leq H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k) - H_{\mathsf{X}}(A_k \cap (A_i \cup A_j)) \tag{2.42}
$$

$$
\begin{aligned}
&\leq H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k \setminus (A_i \cup A_j)) && (2.43)\\
&\leq n\left(C(A_i) + C(A_j) + C(A_k \setminus (A_i \cup A_j))\right) && (2.44)\\
&= n\left(C(A_i \cup A_j \cup A_k) + C(A_i \cap A_j)\right) && (2.45)
\end{aligned}
$$

where (2.43) follows from the independence bound

$$
H_{\mathsf{X}}(A_k) \leq H_{\mathsf{X}}(A_k \cap (A_i \cup A_j)) + H_{\mathsf{X}}(A_k \setminus (A_i \cup A_j)); \tag{2.46}
$$

(2.44) follows from (2.28) for $A' = A_i$, A_j, and $A_k \setminus (A_i \cup A_j)$; and (2.45) follows from

the fact that the capacity function $C(\cdot)$ is a modular function. Substituting

$$H_{\mathsf{W}}(I_i \cup I_j \cup I_k) = nR(I_i \cup I_j \cup I_k) \tag{2.47}$$

$$\text{and} \quad H_{\mathsf{W}}(I_i \cap I_j) = R(I_i \cap I_j) \tag{2.48}$$

into (2.45) and dividing both sides of the inequality by n complete the proof of (2.24). $\square$

We note here that if we had directly bounded from above the right-hand side of (2.31) by $nC(A_i \cup A_j \cup A_k)$ using the independence bound, it would have led to the standard cut-set bound

$$R(I_i \cup I_j \cup I_k) \leq C(A_i \cup A_j \cup A_k). \tag{2.49}$$

But the use of the independence bound would have implied that all messages transmitted over $A_i \cup A_j \cup A_k$ are *independent*, which may not be the case in the presence of multicast messages.

Proof of (2.25). Applying the two-way submodularity (2.4) of the Shannon entropy with $\mathsf{Z} = (\mathsf{X}, \mathsf{W})$, $S_1 = (A_i \cap A_j, I_i \cap I_j)$, and $S_2 = (A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j))$, we have

$$\begin{aligned}
&H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j) + H_{\mathsf{X},\mathsf{W}}(A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j)) \\
&\quad \geq H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j \cap A_k, I_i \cap I_j \cap I_k) + \\
&\qquad H_{\mathsf{X},\mathsf{W}}((A_i \cap A_j) \cup (A_i \cap A_k) \cup (A_j \cap A_k), (I_i \cap I_j) \cap (I_i \cup I_k) \cap (I_j \cup I_k)) \qquad (2.50) \\
&\quad \geq H_{\mathsf{X}}(A_i \cap A_j \cap A_k) + H_{\mathsf{W}}((I_i \cap I_j) \cup (I_i \cap I_k) \cup (I_j \cap I_k)). \qquad (2.51)
\end{aligned}$$

Substituting (2.51) into (2.35) gives

$$\begin{aligned}
&H_{\mathsf{W}}(I_i \cup I_j \cup I_k) + H_{\mathsf{W}}((I_i \cap I_j) \cup (I_j \cap I_k) \cup (I_k \cap I_i)) \\
&\quad \le H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k) - H_{\mathsf{X}}(A_i \cap A_j \cap A_k) && (2.52)\\
&\quad \le H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k \setminus (A_i \cap A_j)) && (2.53)\\
&\quad \le n\left(C(A_i) + C(A_j) + C(A_k \setminus (A_i \cap A_j))\right) && (2.54)\\
&\quad = n\left(C(A_i \cup A_j \cup A_k) + C((A_i \cap A_j) \cup (A_i \cap A_k) \cup (A_j \cap A_k))\right) && (2.55)
\end{aligned}$$

where (2.53) follows from the independence bound

$$H_{\mathsf{X}}(A_k) \le H_{\mathsf{X}}(A_k \cap (A_i \cap A_j)) + H_{\mathsf{X}}(A_k \setminus (A_i \cap A_j)); \tag{2.56}$$

(2.54) follows from (2.28) for $A' = A_i$, A_j, and $A_k \setminus (A_i \cap A_j)$; and (2.55) follows from the fact that the capacity function $C(\cdot)$ is a modular function. Substituting (2.47) and

$$H_{\mathsf{W}}((I_i \cap I_j) \cup (I_i \cap I_k) \cup (I_j \cap I_k)) = nR((I_i \cap I_j) \cup (I_i \cap I_k) \cup (I_j \cap I_k)) \tag{2.57}$$

into (2.55) and dividing both sides of the inequality by n complete the proof of (2.25). $\square$

Proof of (2.26). By the *symmetry* among i, j, and k in (2.35), we have

$$\begin{aligned}
H_{\mathsf{W}}(I_i \cup I_j \cup I_k) \le H_{\mathsf{X}}(A_i) &+ (H_{\mathsf{X}}(A_k) - H_{\mathsf{X},\mathsf{W}}(A_i \cap A_k, I_i \cap I_k)) + \\
&(H_{\mathsf{X}}(A_j) - H_{\mathsf{X},\mathsf{W}}(A_j \cap (A_i \cup A_k), I_j \cap (I_i \cup I_k)))\,. \qquad (2.58)
\end{aligned}$$

Also note that

$$H_{\mathsf{W}}(I_i \cup I_j) \leq H_{\mathsf{X}}(A_i \cup A_j) \tag{2.59}$$

$$= H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j|A_i) \tag{2.60}$$

$$= H_{\mathsf{X}}(A_i) + (H_{\mathsf{X}}(A_j) - I_{\mathsf{X}}(A_i; A_j)) \tag{2.61}$$

$$= H_{\mathsf{X}}(A_i) + (H_{\mathsf{X}}(A_j) - I_{\mathsf{X},\mathsf{W}}(A_i, I_i; A_j, I_j)) \tag{2.62}$$

$$\leq H_{\mathsf{X}}(A_i) + (H_{\mathsf{X}}(A_j) - H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j))\,. \tag{2.63}$$

Adding (2.58) and (2.63) gives

$$\begin{aligned} &H_{\mathsf{W}}(I_i \cup I_j \cup I_k) + H_{\mathsf{W}}(I_i \cup I_j) \\ &\quad \leq 2H_{\mathsf{X}}(A_i) + 2H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k) - H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j) - \\ &\qquad H_{\mathsf{X},\mathsf{W}}(A_i \cap A_k, I_i \cap I_k) - H_{\mathsf{X},\mathsf{W}}(A_j \cap (A_i \cup A_k), I_j \cap (I_i \cup I_k)). \end{aligned} \tag{2.64}$$

Applying the two-way submodularity (2.4) of the Shannon entropy with $\mathsf{Z} = (\mathsf{X}, \mathsf{W})$, $S_1 = (A_i \cap A_j, I_i \cap I_j)$, and $S_2 = (A_i \cap A_k, I_i \cap I_k)$, we have

$$\begin{aligned} &H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j) + H_{\mathsf{X},\mathsf{W}}(A_i \cap A_k, I_i \cap I_k) \\ &\quad \geq H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j \cap A_k, I_i \cap I_j \cap I_k) + H_{\mathsf{X},\mathsf{W}}(A_i \cap (A_j \cup A_k), I_i \cap (I_j \cup I_k)) \end{aligned} \tag{2.65}$$

$$\geq H_{\mathsf{W}}(I_i \cap I_j \cap I_k) + H_{\mathsf{X}}(A_i \cap (A_j \cup A_k)). \tag{2.66}$$

Note that we trivially have

$$H_{\mathsf{X},\mathsf{W}}(A_j \cap (A_i \cup A_k), I_j \cap (I_i \cup I_k)) \geq H_{\mathsf{X}}(A_j \cap (A_i \cup A_k)). \tag{2.67}$$

Substituting (2.66) and (2.67) into (2.64), we have

$$
\begin{aligned}
&H_{\mathsf{W}}(I_i \cup I_j \cup I_k) + H_{\mathsf{W}}(I_i \cup I_j) + H_{\mathsf{W}}(I_i \cap I_j \cap I_k) \\
&\quad \le 2H_{\mathsf{X}}(A_i) + 2H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k) - H_{\mathsf{X}}(A_i \cap (A_j \cup A_k)) - H_{\mathsf{X}}(A_j \cap (A_i \cup A_k)) \qquad (2.68) \\
&\quad \le H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k) + H_{\mathsf{X}}(A_i \setminus (A_j \cup A_k)) + H_{\mathsf{X}}(A_j \setminus (A_i \cup A_k)) \qquad (2.69) \\
&\quad \le n\left(C(A_i) + C(A_j) + C(A_k) + C(A_i \setminus (A_j \cup A_k)) + C(A_j \setminus (A_i \cup A_k))\right) \qquad (2.70) \\
&\quad = n\left(C(A_i \cup A_j \cup A_k) + C(A_i \cup A_j) + C(A_i \cap A_j \cap A_k)\right) \qquad (2.71)
\end{aligned}
$$

where (2.69) follows from the independence bounds

$$
\begin{aligned}
H_{\mathsf{X}}(A_i) &\le H_{\mathsf{X}}(A_i \cap (A_j \cup A_k)) + H_{\mathsf{X}}(A_i \setminus (A_j \cup A_k)) \qquad (2.72) \\
\text{and } H_{\mathsf{X}}(A_j) &\le H_{\mathsf{X}}(A_j \cap (A_i \cup A_k)) + H_{\mathsf{X}}(A_j \setminus (A_i \cup A_k)); \qquad (2.73)
\end{aligned}
$$

(2.70) follows from (2.28) for $A' = A_i,\ A_j,\ A_k,\ A_i \setminus (A_j \cup A_k)$, and $A_j \setminus (A_i \cup A_k)$; and (2.71) follows from the fact that the capacity function $C(\cdot)$ is a modular function. Substituting (2.47),

$$
\begin{aligned}
H_{\mathsf{W}}(I_i \cup I_j) &= nR(I_i \cup I_j), \qquad (2.74) \\
\text{and} \quad H_{\mathsf{W}}(I_i \cap I_j \cap I_k) &= nR(I_i \cap I_j \cap I_k) \qquad (2.75)
\end{aligned}
$$

into (2.71) and dividing both sides of the inequality by n complete the proof of (2.26). $\square$

Proof of (2.27). Adding (2.35) and (2.58), we have

$$\begin{aligned}2H_{\mathsf{W}}(I_i \cup I_j \cup I_k) \leq\ & 2H_{\mathsf{X}}(A_i) + 2H_{\mathsf{X}}(A_j) + 2H_{\mathsf{X}}(A_k) - H_{\mathsf{X},\mathsf{W}}(A_i \cap A_j, I_i \cap I_j) - \\ & H_{\mathsf{X},\mathsf{W}}(A_i \cap A_k, I_i \cap I_k) - H_{\mathsf{X},\mathsf{W}}(A_j \cap (A_i \cup A_k), I_j \cap (I_i \cup I_k)) - \\ & H_{\mathsf{X},\mathsf{W}}(A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j)). \end{aligned} \tag{2.76}$$

Note that we trivially have

$$H_{\mathsf{X},\mathsf{W}}(A_k \cap (A_i \cup A_j), I_k \cap (I_i \cup I_j)) \geq H_{\mathsf{X}}(A_k \cap (A_i \cup A_j)). \tag{2.77}$$

Substituting (2.66), (2.67), and (2.77) into (2.76), we have

$$\begin{aligned} & 2H_{\mathsf{W}}(I_i \cup I_j \cup I_k) + H_{\mathsf{W}}(I_i \cap I_j \cap I_k) \\ & \quad \leq 2H_{\mathsf{X}}(A_i) + 2H_{\mathsf{X}}(A_j) + 2H_{\mathsf{X}}(A_k) - H_{\mathsf{X}}(A_i \cap (A_j \cup A_k)) - \\ & \qquad H_{\mathsf{X}}(A_j \cap (A_i \cup A_k)) - H_{\mathsf{X}}(A_k \cap (A_i \cup A_j)) \qquad (2.78) \\ & \quad \leq H_{\mathsf{X}}(A_i) + H_{\mathsf{X}}(A_j) + H_{\mathsf{X}}(A_k) + H_{\mathsf{X}}(A_i \setminus (A_j \cup A_k)) + \\ & \qquad H_{\mathsf{X}}(A_j \setminus (A_i \cup A_k)) + H_{\mathsf{X}}(A_k \setminus (A_i \cup A_j)) \qquad (2.79) \\ & \quad \leq n\,(C(A_i) + C(A_j) + C(A_k) + C(A_i \setminus (A_j \cup A_k)) + \\ & \qquad C(A_j \setminus (A_i \cup A_k)) + C(A_k \setminus (A_i \cup A_j))) \qquad (2.80) \\ & \quad = n\,(2C(A_i \cup A_j \cup A_k) + C(A_i \cap A_j \cap A_k)) \qquad (2.81) \end{aligned}$$

where (2.79) follows from the independence bounds (2.46), (2.72), and (2.73); (2.80) follows from (2.28) for $A' = A_i$, A_j, A_k, $A_i \setminus (A_j \cup A_k)$, $A_j \setminus (A_i \cup A_k)$ and $A_k \setminus (A_i \cup A_j)$; and (2.81) follows from the fact that the capacity function $C(\cdot)$ is a modular function. Substituting (2.47) and (2.75) into (2.81) and dividing both sides of the inequality by n complete the proof of (2.27). □

We have thus completed the proof of Theorem 1.

2.4 Generalized Cut-Set Bounds Relating K Basic Cuts of the Network

2.4.1 Main Results

Theorem 2. *Consider a broadcast network with a collection of independent messages W_I collocated at the source node s and $K \geq 3$ sink nodes t_k, $k = 1, \ldots, K$. For any $k = 1, \ldots, K$, let W_{I_k} be the intended messages for the sink node t_k, and let A_k be a basic cut that separates the source node s from the sink node t_k. Let G, U and T be nonempty subsets of $[K]$ such that*

$$A^{(1)}(G) \supseteq A^{(1)}(U). \tag{2.82}$$

Let Q be a subset of $\{2, \ldots, |U|\}$, and let $(r_q : q \in Q)$ be a sequence of integers from $[|T|]$ and such that

$$A^{(q)}(U) \subseteq A^{(r_q)}(T) \quad \textit{and} \quad I^{(q)}(U) \subseteq I^{(r_q)}(T), \quad \forall q \in Q. \tag{2.83}$$

We have

$$\begin{aligned} R(I^{(1)}(G)) + \sum_{r \in \{2,\ldots,|U|\} \setminus Q} R(I^{(r)}(U)) + \sum_{q \in Q} \sum_{r=1}^{r_q} \alpha_Q(q,r) R(I^{(r)}(T)) \\ \leq C(A^{(1)}(G)) + \sum_{r \in \{2,\ldots,|U|\} \setminus Q} C(A^{(r)}(U)) + \sum_{q \in Q} \sum_{r=1}^{r_q} \alpha_Q(q,r) C(A^{(r)}(T)) \end{aligned} \tag{2.84}$$

for any achievable rate tuple R_I, where

$$\alpha_Q(q,r) = \begin{cases} 0, & \text{if } r \in Q \\ \frac{\prod_{\{p\in Q: p<r\}}(p-1)\prod_{\{p\in Q: r<p\le r_q\}} p}{r_q \prod_{\{p\in Q: p\le r_q\}}(p-1)}, & \text{if } r \notin Q \end{cases} \tag{2.85}$$

for any $q \in Q$ and $r \in [r_q]$.

Note that the generalized cut-set bound (2.84) involves a number of parameters: G, U, T, Q, and $(r_q : q \in Q)$. Specifying these parameters to certain choices will lead to potentially weaker but more applicable generalized cut-set bounds. More specifically, let $G = U = T$ and $r_q = q-1$ for any $q \in Q$. By the ordering in (2.7), the condition in (2.83) is satisfied (the condition in (2.82) holds trivially with an equality). Thus, by Theorem 2 we have

$$\begin{aligned} \sum_{r\in[|U|]\setminus Q} R(I^{(r)}(U)) + \sum_{q\in Q}\sum_{r=1}^{q-1} \alpha_Q(q,r) R(I^{(r)}(U)) \\ \le \sum_{r\in[|U|]\setminus Q} C(A^{(r)}(U)) + \sum_{q\in Q}\sum_{r=1}^{q-1} \alpha_Q(q,r) C(A^{(r)}(U)) \end{aligned} \tag{2.86}$$

for any achievable rate tuple R_I, where

$$\alpha_Q(q,r) = \begin{cases} 0, & \text{if } r \in Q \\ \frac{\prod_{\{p\in Q: p<r\}}(p-1)\prod_{\{p\in Q: r<p\le q-1\}} p}{\prod_{\{p\in Q: p\le q\}}(p-1)}, & \text{if } r \notin Q \end{cases} \tag{2.87}$$

for any $q \in Q$ and $r \in [q-1]$. A proper simplification of (2.86) leads to the following corollary. See Appendix A.3 for the details of the simplification procedure.

Corollary 4. *Consider a broadcast network with a collection of independent messages W_I collocated at the source node s and $K \ge 3$ sink nodes t_k, $k = 1, \ldots, K$. For any $k = 1, \ldots, K$, let W_{I_k} be the intended messages for the sink node t_k, and let A_k be a basic*

cut that separates the source node s from the sink node t_k. Let U be a nonempty subset of $[K]$, and let Q be a subset of $\{2,\ldots,|U|\}$. We have

$$\sum_{r=1}^{|U|}\beta_Q(r)R(I^{(r)}(U)) \le \sum_{r=1}^{|U|}\beta_Q(r)C(A^{(r)}(U)) \tag{2.88}$$

for any achievable rate tuple R_I, where $\beta_Q(r)=1$ for any $r\in[|U|]$ if $Q=\emptyset$, and

$$\beta_Q(r)=\begin{cases} 0, & \text{if } r\in Q \\ \prod_{\{q\in Q:q<r\}}(q-1)\prod_{\{q\in Q:q>r\}}q, & \text{if } r\notin Q \end{cases} \tag{2.89}$$

for any $r\in[|U|]$ if $Q\neq\emptyset$.

The generalized cut-set bound (2.88) can be further specified by letting $Q=\{2,\ldots,m\}$ for $m=1,\ldots,|U|$ (note that $Q=\emptyset$ when $m=1$). For this particular choice of Q, we have

$$\beta_Q(r)=\begin{cases} m!, & r=1 \\ 0, & r=2,\ldots,m \\ (m-1)!, & r=m+1,\ldots,|U|. \end{cases} \tag{2.90}$$

Substituting (2.90) into (2.88) immediately leads to the following corollary.

Corollary 5. *Consider a broadcast network with a collection of independent messages W_I collocated at the source node s and $K\ge 3$ sink nodes t_k, $k=1,\ldots,K$. For any $k=1,\ldots,K$, let W_{I_k} be the intended messages for the sink node t_k, and let A_k be a basic cut that separates the source node s from the sink node t_k. Let U be a nonempty subset of*

$[K]$. *We have*

$$mR(I^{(1)}(U)) + \sum_{r=m+1}^{|U|} R(I^{(r)}(U)) \leq mC(A^{(1)}(U)) + \sum_{r=m+1}^{|U|} C(A^{(r)}(U)) \tag{2.91}$$

for any achievable rate tuple R_I *and any* $m = 1, \ldots, |U|$.

Now, the generalized cut-set bound (2.27) can be recovered from Corollary 5 by setting $U = \{1, 2, 3\}$ and $m = 2$ in (2.91); the generalized cut-set bound (2.25) can be recovered from Corollary 4 by setting $U = \{1, 2, 3\}$ and $Q = \{3\}$ such that

$$\beta_Q(r) = \begin{cases} 3, & r = 1, 2 \\ 0, & r = 3; \end{cases} \tag{2.92}$$

the generalized cut-set bound (2.24) can be recovered from Theorem 2 by setting $G = \{i, j, k\}$, $U = \{i, j\}$ (so $A^{(1)}(G) \supseteq A^{(1)}(U)$) and $Q = \emptyset$; and finally, the generalized cut-set bound (2.26) can be recovered from Theorem 2 by setting $G = U = \{i, j, k\}$ (so $A^{(1)}(G) = A^{(1)}(U)$), $T = \{i, j\}$, $Q = \{2\}$, and $r_2 = 1$ such that

$$\begin{aligned} A^{(2)}(U) &= (A_i \cap A_j) \cup (A_i \cap A_k) \cup (A_j \cap A_k) \subseteq A_i \cup A_j = A^{(r_2)}(T), \\ I^{(2)}(U) &= (I_i \cap I_j) \cup (I_i \cap I_k) \cup (I_j \cap I_k) \subseteq I_i \cup I_j = I^{(r_2)}(U), \end{aligned} \tag{2.93}$$

and $\alpha_Q(2, 1) = 1$.

2.4.2 Proof of Theorem 2

Let $(n, \{\mathsf{X}_a : a \in A\})$ be an *admissible* code with block length n, where X_a is the message transmitted over the arc a. Similar to the proof of Theorem 1, we shall assume perfect recovery of the messages at each of the sink nodes. As such, for any nonempty subset $U \subseteq [K]$ the messages $\mathsf{W}_{\cup_{k\in U} I_k}$ must be *functions* of the messages $\mathsf{X}_{\cup_{k\in U} A_k}$ transmitted over the s-t_U cut $\cup_{k\in U} A_k$.

Let us first consider the case where $Q = \emptyset$. Note that

$$
\begin{aligned}
&H_{\mathsf{W}}(I^{(1)}(G)) \\
&\quad \leq H_{\mathsf{X}}(A^{(1)}(G)) && (2.94)\\
&\quad \leq H_{\mathsf{X}}(A^{(1)}(U)) + H_{\mathsf{X}}(A^{(1)}(G) \setminus A^{(1)}(U)) && (2.95)\\
&\quad = H_{\mathsf{X},\mathsf{W}}(A^{(1)}(U), I^{(1)}(U)) + H_{\mathsf{X}}(A^{(1)}(G) \setminus A^{(1)}(U)) && (2.96)\\
&\quad \leq \sum_{k\in U} H_{\mathsf{X},\mathsf{W}}(A_k, I_k) - \sum_{r=2}^{|U|} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(U), I^{(r)}(U)) + H_{\mathsf{X}}(A^{(1)}(G) \setminus A^{(1)}(U)) \\
& && (2.97)\\
&\quad = \sum_{k\in U} H_{\mathsf{X}}(A_k) - \sum_{r=2}^{|U|} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(U), I^{(r)}(U)) + H_{\mathsf{X}}(A^{(1)}(G) \setminus A^{(1)}(U)) && (2.98)\\
&\quad \leq n\left(\sum_{k\in U} C(A_k) + C(A^{(1)}(G) \setminus A^{(1)}(U))\right) - \sum_{r=2}^{|U|} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(U), I^{(r)}(U)) && (2.99)\\
&\quad = n\left(\sum_{r=1}^{|U|} C(A^{(r)}(U)) + C(A^{(1)}(G) \setminus A^{(1)}(U))\right) - \sum_{r=2}^{|U|} H_{\mathsf{W}}(I^{(r)}(U)) && (2.100)\\
&\quad = n\left(C(A^{(1)}(G)) + \sum_{r=2}^{|U|} C(A^{(r)}(U))\right) - \sum_{r=2}^{|U|} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(U), I^{(r)}(U)) && (2.101)
\end{aligned}
$$

where (2.94) and (2.96) follow from the fact that the messages $\mathsf{W}_{I^{(1)}(U)}$ are functions of $\mathsf{X}_{A^{(1)}(U)}$; (2.95) follows from the independence bound on entropy; (2.97) follows from the standard multiway submodularity (2.9); (2.98) follows from the fact that the messages W_{I_k} are functions of X_{A_k} so we have $H_{\mathsf{X},\mathsf{W}}(A_k, I_k) = H_{\mathsf{X}}(A_k)$ for any $k \in U$; (2.99) follows from the link capacity constraints; (2.100) follows from the fact that the capacity function $C(\cdot)$ is a modular function so we have $\sum_{k\in U} C(A_k) = \sum_{r=1}^{|U|} C(A^{(r)}(U))$; and (2.101) follows from the fact that the capacity function $C(\cdot)$ is a modular function and the assumption (2.82) so we have $C(A^{(1)}(G)) = C(A^{(1)}(U)) + C(A^{(1)}(G) \setminus A^{(1)}(U))$.

Rearranging the terms in (2.101) gives

$$H_{\mathsf{W}}(I^{(1)}(G)) + \sum_{r=2}^{|U|} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(U), I^{(r)}(U)) \leq n \left(C(A^{(1)}(G)) + \sum_{r=2}^{|U|} C(A^{(r)}(U)) \right). \tag{2.102}$$

Further note that

$$H_{\mathsf{W}}(I^{(1)}(G)) = nR(I^{(1)}(G)) \tag{2.103}$$

$$\text{and} \quad H_{\mathsf{X},\mathsf{W}}(A^{(r)}(U), I^{(r)}(U)) \geq H_{\mathsf{W}}(I^{(r)}(U)) = nR(I^{(r)}(U)), \quad \forall r = 2, \ldots, |U|. \tag{2.104}$$

Substituting (2.103) and (2.104) into (2.101) and dividing both sides of the inequality by n, we have

$$R(I^{(1)}(G)) + \sum_{r=2}^{|U|} R(I^{(r)}(U)) \leq C(A^{(1)}(G)) + \sum_{r=2}^{|U|} C(A^{(r)}(U)) \tag{2.105}$$

for any achievable rate tuple R_I. This completes the proof of (2.84) for $Q = \emptyset$.

Next, assume that $Q \neq \emptyset$. Write, without loss of generality, that $Q = \{q_1, \ldots, q_{|Q|}\}$ where

$$2 \leq q_1 < q_2 < \cdots < q_{|Q|} \leq |U|. \tag{2.106}$$

By Lemma 3, for any two integers q' and $r_{q'}$ such that $1 \leq q' \leq |U|$, $1 \leq r_{q'} \leq |T|$,

$A^{(q')}(U) \subseteq A^{(r_{q'})}(T)$, and $I^{(q')}(U) \subseteq I^{(r_{q'})}(T)$ we have

$$\begin{aligned}
&\sum_{r=1}^{r_{q'}} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) - r_{q'} H_{\mathsf{X},\mathsf{W}}(A^{(q')}(U), I^{(q')}(U)) \\
&\quad \leq \sum_{r=1}^{|T|} H_{\mathsf{X},\mathsf{W}}(A_{t_r}, I_{t_r}) - \sum_{r=1}^{r_{q'}} H_{\mathsf{X},\mathsf{W}}(A_{t_r} \cap A^{(q')}(U), I_{t_r} \cap I^{(q')}(U)) - \\
&\qquad \sum_{r=r_{q'}+1}^{|T|} H_{\mathsf{X},\mathsf{W}}(A_{t_r} \cap (A^{(q')}(U) \cup A^{(r_{q'}+1)}(\{t_1, \ldots, t_r\})), \\
&\qquad\qquad I_{t_r} \cap (I^{(q')}(U) \cup I^{(r_{q'}+1)}(\{t_1, \ldots, t_r\}))) \qquad (2.107) \\
&\quad \leq \sum_{r=1}^{|T|} H_{\mathsf{X}}(A_{t_r}) - \sum_{r=1}^{r_{q'}} H_{\mathsf{X}}(A_{t_r} \cap A^{(q')}(U)) - \\
&\qquad \sum_{r=r_{q'}+1}^{|T|} H_{\mathsf{X}}(A_{t_r} \cap (A^{(q')}(U) \cup A^{(r_{q'}+1)}(\{t_1, \ldots, t_r\}))) \qquad (2.108) \\
&\quad \leq \sum_{r=1}^{r_{q'}} H_{\mathsf{X}}(A_{t_r} \setminus A^{(q')}(U)) + \sum_{r=r_{q'}+1}^{|T|} H_{\mathsf{X}}(A_{t_r} \setminus (A^{(q')}(U) \cup A^{(r_{q'}+1)}(\{t_1, \ldots, t_r\})) \\
&\qquad\qquad (2.109) \\
&\quad \leq n \left(\sum_{r=1}^{r_{q'}} C(A_{t_r} \setminus A^{(q')}(U)) + \sum_{r=r_{q'}+1}^{|T|} C(A_{t_r} \setminus (A^{(q')}(U) \cup A^{(r_{q'}+1)}(\{t_1, \ldots, t_r\}))) \right) \\
&\qquad\qquad (2.110) \\
&\quad = n \left(\sum_{r=1}^{|T|} C(A_{t_r}) - \sum_{r=1}^{r_{q'}} C(A_{t_r} \cap A^{(q')}(U)) - \right. \\
&\qquad \left. \sum_{r=r_{q'}+1}^{|T|} C(A_{t_r} \cap (A^{(q')}(U) \cup A^{(r_{q'}+1)}(\{t_1, \ldots, t_r\}))) \right) \qquad (2.111) \\
&\quad = n \left(\sum_{r=1}^{r_{q'}} C(A^{(r)}(T)) - r_{q'} C(A^{(q')}(U)) \right) \qquad (2.112)
\end{aligned}$$

where (2.108) follows from the fact that the messages $\mathsf{W}_{I_{t_r}}$ are functions of $\mathsf{X}_{A_{t_r}}$ so we have

$H_{\mathsf{X},\mathsf{W}}(A_{t_r}, I_{t_r}) = H_{\mathsf{X}}(A_{t_r})$ for any $r \in [|U|]$ and the trivial inequalities

$$H_{\mathsf{X},\mathsf{W}}(A_{t_r} \cap A^{(q')}(U), I_{t_r} \cap I^{(q')}(U)) \geq H_{\mathsf{X}}(A_{t_r} \cap A^{(q')}(U)), \quad \forall r \in [r_{q'}] \tag{2.113}$$

and

$$\begin{aligned} &H_{\mathsf{X},\mathsf{W}}(A_{t_r} \cap (A^{(q')}(U) \cup A^{(r_{q'}+1)}(\{t_1, \ldots, t_r\})), I_{t_r} \cap (I^{(q')}(U) \cup I^{(r_{q'}+1)}(\{t_1, \ldots, t_r\}))) \\ &\quad \geq H_{\mathsf{X}}(A_{t_r} \cap (A^{(q')}(U) \cup A^{(r_{q'}+1)}(\{t_1, \ldots, t_r\}))); \end{aligned} \tag{2.114}$$

(2.109) follows from the independence bound on entropy; (2.110) follows from the link-capacity constraints; and (2.111) and (2.112) follow from the fact that the capacity function $C(\cdot)$ is a modular function. Letting $r_{q'} = q' = q_j$ and $U = T$ in (2.112), we have

$$\begin{aligned} &\sum_{r=1}^{q_j} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) - q_j H_{\mathsf{X},\mathsf{W}}(A^{(q_j)}(T), I^{(q_j)}(T)) \\ &\qquad \leq n \left(\sum_{r=1}^{q_j} C(A^{(r)}(T)) - q_j C(A^{(q_j)}(T)) \right). \end{aligned} \tag{2.115}$$

Let

$$n_Q(q, r) := \prod_{\{p \in Q : p < r\}} (p-1) \prod_{\{p \in Q : r < p \leq r_q\}} p \tag{2.116}$$

$$\text{and} \quad d_Q(q) := \prod_{\{p \in Q : p \leq r_q\}} (p-1) \tag{2.117}$$

for any $q \in Q$ and $r \in [r_q]$, and let $Q_i := \{q \in Q : q \leq r_{q_i}\}$. Note that $n_Q(q, r)$ and $d_Q(q)$ are always positive. Multiplying both sides of (2.115) by $n_Q(q_i, q_j)$ and then summing

over all $q_j \in Q_i$, we have

$$\sum_{j=1}^{|Q_i|} n_Q(q_i, q_j) \left(\sum_{r=1}^{q_j} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) - q_j H_{\mathsf{X},\mathsf{W}}(A^{(q_j)}(T), I^{(q_j)}(T)) \right) \leq n \left(\sum_{j=1}^{|Q_i|} n_Q(q_i, q_j) \left(\sum_{r=1}^{q_j} C(A^{(r)}(T)) - q_j C(A^{(q_j)}(T)) \right) \right). \tag{2.118}$$

Note that

$$\sum_{j=1}^{|Q_i|} n(q_i, q_j) \sum_{r=1}^{q_j} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) = \sum_{r=1}^{q_{|Q_i|}} \left(\sum_{j=j(r)}^{|Q_i|} n(q_i, q_j) \right) H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) \tag{2.119}$$

where

$$j(r) := \begin{cases} 1, & \text{for } 0 < r \leq q_1 \\ 2, & \text{for } q_1 < r \leq q_2 \\ \vdots & \\ |Q_i|, & \text{for } q_{|Q_i|-1} < r \leq q_{|Q_i|}. \end{cases} \tag{2.120}$$

We can thus rewrite (2.118) as

$$\sum_{r=1}^{q_{|Q_i|}} \left(\sum_{j=j(r)}^{|Q_i|} n(q_i, q_j) - r n_Q(q_i, r) 1_{\{r \in Q_i\}} \right) H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) \leq n \left(\sum_{r=1}^{q_{|Q_i|}} \left(\sum_{j=j(r)}^{|Q_i|} n(q_i, q_j) - r n_Q(q_i, r) 1_{\{r \in Q_i\}} \right) C(A^{(r)}(T) \right). \tag{2.121}$$

Furthermore, letting $q' = q_i$ and $r_{q'} = r_{q_i}$ in (2.112) and multiplying both sides of the

inequality by $d_Q(q_i)$, we have

$$\sum_{r=1}^{r_{q_i}} d_Q(q_i) H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) - r_{q_i} d_Q(q_i) H_{\mathsf{X},\mathsf{W}}(A^{(q_i)}(U), I^{(q_i)}(U)) \\ \leq n \left(\sum_{r=1}^{r_{q_i}} d_Q(q_i) C(A^{(r)}(T)) - r_{q_i} d_Q(q_i) C(A^{(q_i)}(U)) \right). \tag{2.122}$$

Adding (2.121) and (2.122) gives

$$\sum_{r=1}^{r_{q_i}} n'_Q(q_i, r) H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) - r_{q_i} d_Q(q_i) H_{\mathsf{X},\mathsf{W}}(A^{(q_i)}(U), I^{(q_i)}(U)) \\ \leq n \left(\sum_{r=1}^{r_{q_i}} n'_Q(q_i, r) C(A^{(r)}(T)) - r_{q_i} d_Q(q_i) C(A^{(q_i)}(U)) \right) \tag{2.123}$$

where

$$n'_Q(q_i, r) = \begin{cases} \sum_{j=j(r)}^{|Q_i|} n_Q(q_i, q_j) - r n_Q(q_i, r) 1_{\{r \in Q_i\}} + d_Q(q_i), & \text{if } 1 \leq r \leq q_{|Q_i|} \\ d_Q(q_i), & \text{if } q_{|Q_i|} < r \leq r_{q_i}. \end{cases} \tag{2.124}$$

By (2.120), when $q_{m-1} < r \leq q_m$ for some $m = 1, \ldots, |Q_i|$ ($q_0 := 0$ for convenience),

we have $j(r) = m$ and hence

$$\sum_{j=j(r)}^{|Q_i|} n_Q(q_i, q_j) = \sum_{j=m}^{|Q_i|} n_Q(q_i, q_j) \tag{2.125}$$

$$= \sum_{j=m}^{|Q_i|} \left(\prod_{l=1}^{j-1} (q_l - 1) \prod_{l=j+1}^{|Q_i|} q_l \right) \tag{2.126}$$

$$= \sum_{j=m}^{|Q_i|} \left(\prod_{l=1}^{j-1} (q_l - 1) \prod_{l=j}^{|Q_i|} q_l - \prod_{l=1}^{j} (q_l - 1) \prod_{l=j+1}^{|Q_i|} q_l \right) \tag{2.127}$$

$$= \sum_{j=m}^{|Q_i|} \left(\prod_{l=1}^{j-1} (q_l - 1) \prod_{l=j}^{|Q_i|} q_l \right) - \sum_{j=m+1}^{|Q_i|+1} \left(\prod_{l=1}^{j-1} (q_l - 1) \prod_{l=j}^{|Q_i|} q_l \right) \tag{2.128}$$

$$= \prod_{l=1}^{m-1} (q_l - 1) \prod_{l=m}^{|Q_i|} q_l - \prod_{l=1}^{|Q_i|} (q_l - 1) \tag{2.129}$$

$$= \prod_{l=1}^{m-1} (q_l - 1) \prod_{l=m}^{|Q_i|} q_l - d_Q(q_i). \tag{2.130}$$

Therefore, when $r = q_m$ for some $m \in [|Q_i|]$ we have

$$\sum_{j=m}^{|Q_i|} n_Q(q_i, q_j) - q_m n_Q(q_i, q_m) + d_Q(q_i) = \prod_{l=1}^{m-1} (q_l - 1) \prod_{l=m}^{|Q_i|} q_l - q_m \prod_{l=1}^{m-1} (q_l - 1) \prod_{l=m+1}^{|Q_i|} q_l \tag{2.131}$$

$$= 0; \tag{2.132}$$

when $q_{m-1} < r < q_m$ for some $m \in [|Q_i|]$ we have

$$\sum_{j=m}^{|Q_i|} n_Q(q_i, q_j) + d_Q(q_i) = \prod_{l=1}^{m-1} (q_l - 1) \prod_{l=m}^{|Q_i|} q_l \tag{2.133}$$

$$= r_{q_i} d_Q(q_i) \alpha_Q(q_i, r); \tag{2.134}$$

and when $q_{|Q_i|} < r \leq r_{q_i}$ we have $\alpha_Q(q_i, r) = 1/r_{q_i}$ and hence

$$d_Q(q_i) = r_{q_i} d_Q(q_i) \alpha_Q(q_i, r). \tag{2.135}$$

Combining (2.132), (2.134), and (2.135), we conclude that

$$n'_Q(q_i, r) = r_{q_i} d_Q(q_i) \alpha_Q(q_i, r), \quad \forall r \in [r_{q_i}]. \tag{2.136}$$

Dividing both sides of (2.123) by $r_{q_i} d(q_i)$ and then summing over all $q_i \in Q$, we have

$$\begin{aligned} &\sum_{q \in Q} \sum_{r=1}^{r_q} \alpha_Q(q, r) H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) - \sum_{q \in Q} H_{\mathsf{X},\mathsf{W}}(A^{(q)}(U), I^{(q)}(U)) \\ &\qquad \leq n \left(\sum_{q \in Q} \sum_{r=1}^{r_q} \alpha_Q(q, r) C(A^{(r)}(T)) - \sum_{q \in Q} C(A^{(q)}(U)) \right). \end{aligned} \tag{2.137}$$

Adding (2.102) and (2.137), we have

$$\begin{aligned} &H_{\mathsf{W}}(I^{(1)}(G)) + \sum_{r \in \{2,\ldots,|U|\} \setminus Q} H_{\mathsf{X},\mathsf{W}}(A^{(r)}(U), I^{(r)}(U)) + \sum_{q \in Q} \sum_{r=1}^{r_q} \alpha_Q(q, r) H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) \\ &\leq n \left(C(A^{(1)}(G)) + \sum_{r \in \{2,\ldots,|U|\} \setminus Q} C(A^{(r)}(U)) + \sum_{q \in Q} \sum_{r=1}^{r_q} \alpha_Q(q, r) C(A^{(r)}(T)) \right). \end{aligned} \tag{2.138}$$

Note that we trivially have

$$H_{\mathsf{X},\mathsf{W}}(A^{(r)}(T), I^{(r)}(T)) \geq H_{\mathsf{W}}(I^{(r)}(T)) = nR(I^{(r)}(T)), \quad \forall q \in Q \text{ and } r \in [r_q]. \tag{2.139}$$

Substituting (2.103), (2.104), and (2.139) into (2.138) and dividing both sides of the inequality by n complete the proof of (2.84) for $Q \neq \emptyset$.

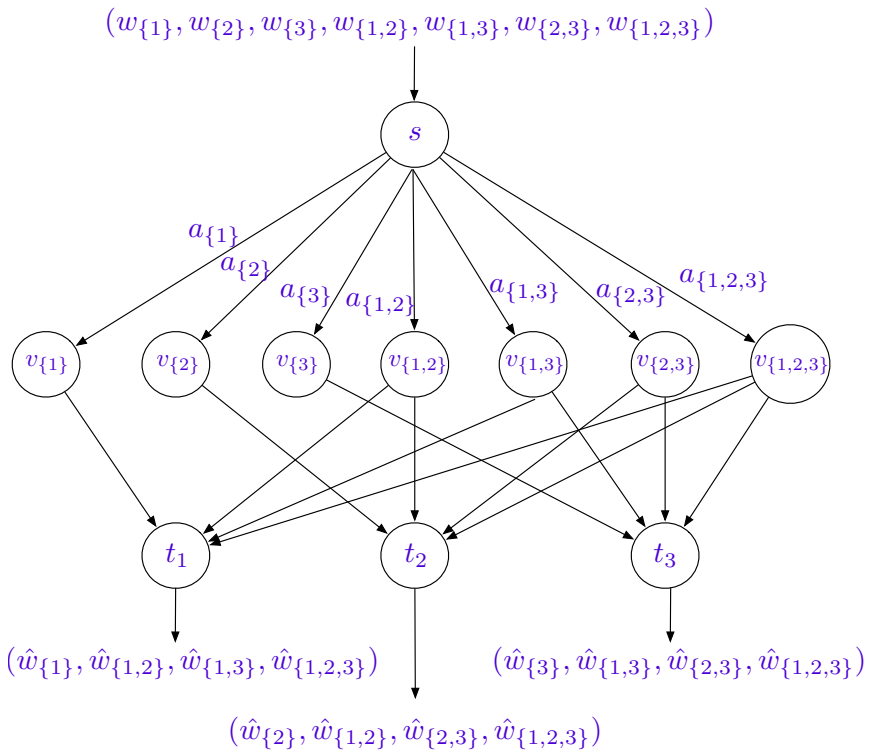

Figure 2.3: Illustration of the general combination network with $K = 3$ sink nodes and a complete message set.

We have thus completed the proof of Theorem 2.

2.5 Applications to Combination Networks

To demonstrate the tightness of the generalized cut-set bounds, let us consider a special class of broadcast networks known as *combination* networks [3]. A combination network is a broadcast network that consists of three layers of nodes (see Figure 2.3 for an illustration). The top layer consists of a single source node s, and the bottom layer consists of K sink nodes t_k, $k = 1, \ldots, K$. The middle layer consists of $2^K - 1$ intermediate nodes, each connecting to the source node s and a nonempty subset of sink nodes. While the links from the source node s to the intermediate nodes may have finite capacity, the links from the intermediate nodes to the sink nodes are all assumed to have *infinite* capacity. More

specifically, denote by v_U the intermediate node that connects to the nonempty subset U of sink nodes and a_U the link that connects the source node s to the intermediate node v_U. The link capacity for a_U is denoted by C_U. Note that when $C_U = 0$, the intermediate node v_U can be effectively removed from the network. By construction, the only interesting combinatorial structure for combination networks is cut. Therefore, combination networks provide an ideal set of problems to understand the strength and the limitations of the generalized cut-set bounds.

In Figure 2.3 we illustrate a general combination network with $K = 3$ sink nodes and a general message set that consists of a total of seven independent messages

$$(\mathsf{W}_{\{1\}}, \mathsf{W}_{\{2\}}, \mathsf{W}_{\{3\}}, \mathsf{W}_{\{1,2\}}, \mathsf{W}_{\{1,3\}}, \mathsf{W}_{\{2,3\}}, \mathsf{W}_{\{1,2,3\}}),$$

where the message W_U, $U \subseteq \{1, 2, 3\}$, is intended for all sink nodes t_k, $k \in U$. This network coding problem was first introduced and solved by Grokop and Tse [2] in the context of characterizing the *latency* capacity region [14] of the general broadcast channel with three receivers. More specifically, it was shown in [2] that the capacity region of the network is given by the set of nonnegative rate tuples

$$(R_{\{1\}}, R_{\{2\}}, R_{\{3\}}, R_{\{1,2\}}, R_{\{2,3\}}, R_{\{1,3\}}, R_{\{1,2,3\}})$$

satisfying

$$R_{\{1\}}+R_{\{1,2\}}+R_{\{1,3\}}+R_{\{1,2,3\}} \leq C_{\{1\}}+C_{\{1,2\}}+C_{\{1,3\}}+C_{\{1,2,3\}}, \tag{2.140}$$

$$R_{\{2\}}+R_{\{1,2\}}+R_{\{2,3\}}+R_{\{1,2,3\}} \leq C_{\{2\}}+C_{\{1,2\}}+C_{\{2,3\}}+C_{\{1,2,3\}}, \tag{2.141}$$

$$R_{\{3\}}+R_{\{1,3\}}+R_{\{2,3\}}+R_{\{1,2,3\}} \leq C_{\{3\}}+C_{\{1,3\}}+C_{\{2,3\}}+C_{\{1,2,3\}}, \tag{2.142}$$

$$\begin{aligned} R_{\{1\}}+R_{\{2\}}+&R_{\{1,2\}}+R_{\{2,3\}}+R_{\{1,3\}}+R_{\{1,2,3\}} \\ &\leq C_{\{1\}}+C_{\{2\}}+C_{\{1,2\}}+C_{\{2,3\}}+C_{\{1,3\}}+C_{\{1,2,3\}}, \end{aligned} \tag{2.143}$$

$$\begin{aligned} R_{\{2\}}+R_{\{3\}}+&R_{\{1,2\}}+R_{\{2,3\}}+R_{\{1,3\}}+R_{\{1,2,3\}} \\ &\leq C_{\{2\}}+C_{\{3\}}+C_{\{1,2\}}+C_{\{2,3\}}+C_{\{1,3\}}+C_{\{1,2,3\}}, \end{aligned} \tag{2.144}$$

$$\begin{aligned} R_{\{1\}}+R_{\{3\}}+&R_{\{1,2\}}+R_{\{2,3\}}+R_{\{1,3\}}+R_{\{1,2,3\}} \\ &\leq C_{\{1\}}+C_{\{3\}}+C_{\{1,2\}}+C_{\{2,3\}}+C_{\{1,3\}}+C_{\{1,2,3\}}, \end{aligned} \tag{2.145}$$

$$\begin{aligned} R_{\{1\}}+R_{\{2\}}+R_{\{3\}}+&R_{\{1,2\}}+R_{\{2,3\}}+R_{\{1,3\}}+R_{\{1,2,3\}} \\ &\leq C_{\{1\}}+C_{\{2\}}+C_{\{3\}}+C_{\{1,2\}}+C_{\{2,3\}}+C_{\{1,3\}}+C_{\{1,2,3\}}, \end{aligned} \tag{2.146}$$

$$\begin{aligned} R_{\{1\}}+R_{\{2\}}+R_{\{3\}}+&2R_{\{1,2\}}+R_{\{2,3\}}+R_{\{1,3\}}+2R_{\{1,2,3\}} \\ &\leq C_{\{1\}}+C_{\{2\}}+C_{\{3\}}+2C_{\{1,2\}}+C_{\{2,3\}}+C_{\{1,3\}}+2C_{\{1,2,3\}}, \end{aligned} \tag{2.147}$$

$$\begin{aligned} R_{\{1\}}+R_{\{2\}}+R_{\{3\}}+&R_{\{1,2\}}+2R_{\{2,3\}}+R_{\{1,3\}}+2R_{\{1,2,3\}} \\ &\leq C_{\{1\}}+C_{\{2\}}+C_{\{3\}}+C_{\{1,2\}}+2C_{\{2,3\}}+C_{\{1,3\}}+2C_{\{1,2,3\}}, \end{aligned} \tag{2.148}$$

$$R_{\{1\}} + R_{\{2\}} + R_{\{3\}} + R_{\{1,2\}} + R_{\{2,3\}} + 2R_{\{1,3\}} + 2R_{\{1,2,3\}} \leq C_{\{1\}} + C_{\{2\}} + C_{\{3\}} + C_{\{1,2\}} + C_{\{2,3\}} + 2C_{\{1,3\}} + 2C_{\{1,2,3\}}, \tag{2.149}$$

$$R_{\{1\}} + R_{\{2\}} + R_{\{3\}} + 2R_{\{1,2\}} + 2R_{\{2,3\}} + 2R_{\{1,3\}} + 2R_{\{1,2,3\}} \leq C_{\{1\}} + C_{\{2\}} + C_{\{3\}} + 2C_{\{1,2\}} + 2C_{\{2,3\}} + 2C_{\{1,3\}} + 2C_{\{1,2,3\}}, \tag{2.150}$$

$$R_{\{1\}} + 2R_{\{2\}} + 2R_{\{3\}} + 2R_{\{1,2\}} + 2R_{\{2,3\}} + 2R_{\{1,3\}} + 3R_{\{1,2,3\}} \leq C_{\{1\}} + 2C_{\{2\}} + 2C_{\{3\}} + 2C_{\{1,2\}} + 2C_{\{2,3\}} + 2C_{\{1,3\}} + 3C_{\{1,2,3\}}, \tag{2.151}$$

$$2R_{\{1\}} + R_{\{2\}} + 2R_{\{3\}} + 2R_{\{1,2\}} + 2R_{\{2,3\}} + 2R_{\{1,3\}} + 3R_{\{1,2,3\}} \leq 2C_{\{1\}} + 2C_{\{2\}} + C_{\{3\}} + 2C_{\{1,2\}} + 2C_{\{2,3\}} + 2C_{\{1,3\}} + 3C_{\{1,2,3\}}, \tag{2.152}$$

$$2R_{\{1\}} + 2R_{\{2\}} + R_{\{3\}} + 2R_{\{1,2\}} + 2R_{\{2,3\}} + 2R_{\{1,3\}} + 3R_{\{1,2,3\}} \leq 2C_{\{1\}} + 2C_{\{2\}} + C_{\{3\}} + 2C_{\{1,2\}} + 2C_{\{2,3\}} + 2C_{\{1,3\}} + 3C_{\{1,2,3\}}, \tag{2.153}$$

$$2R_{\{1\}} + 2R_{\{2\}} + 2R_{\{3\}} + 2R_{\{1,2\}} + 2R_{\{2,3\}} + 2R_{\{1,3\}} + 3R_{\{1,2,3\}} \leq 2C_{\{1\}} + 2C_{\{2\}} + 2C_{\{3\}} + 2C_{\{1,2\}} + 2C_{\{2,3\}} + 2C_{\{1,3\}} + 3C_{\{1,2,3\}}. \tag{2.154}$$

From the converse viewpoint, the inequalities (2.140)–(2.146) follow directly from the standard cut-set bounds (2.1) by considering the following three basic cuts:
$A_1 = \{a_{\{1\}}, a_{\{1,2\}}, a_{\{1,3\}}, a_{\{1,2,3\}}\}$, $A_2 = \{a_{\{2\}}, a_{\{1,2\}}, a_{\{2,3\}}, a_{\{1,2,3\}}\}$, and
$A_3 = \{a_{\{3\}}, a_{\{2,3\}}, a_{\{1,3\}}, a_{\{1,2,3\}}\}$. For the inequalities (2.147)–(2.154), the proof provided in [2] was problem-specific and appears to be rather hand-crafted. With the general-

ized cut-set bounds now in place, however, it is clear that the inequalities (2.147)–(2.149) follow directly from (2.24); the inequality (2.150) follows directly from (2.25); the inequalities (2.151)–(2.153) follow directly from (2.26); and the inequality (2.154) follows directly from (2.27). Thus, the standard and the generalized cut-set bounds together provide an *exact* characterization of the capacity region of the general combination network with three sink nodes and a complete message set.

Next, let us consider the general combination network with K sink nodes and *symmetrical* link capacity constraints [14]:

$$C_U = C_{|U|}, \quad \forall U \subseteq [K] \tag{2.155}$$

i.e., the link-capacity constraint for arc a_U depends on the subset U only via its cardinality. Assume that the source s has access to a set of $K+1$ independent messages $(\mathsf{W}_1, \ldots, \mathsf{W}_K, \mathsf{W}_0)$, where W_k, $k = 1, \ldots, K$, is a private message intended only for the sink node t_k, and W_0 is a common message intended for all K sink nodes in the network. For this communication scenario, note that $A_k = \{a_U : U \ni k\}$ is a basic cut that separates the source node s from the sink node t_k for each $k = 1, \ldots, K$. Applying Corollary 5 with $U = [K]$, we have

$$KR_0 + mR_{sp} \leq m \sum_{r=1}^{K} \binom{K}{r} C_r + \sum_{r=m+1}^{K} \sum_{j=r}^{K} \binom{K}{j} C_j \tag{2.156}$$

$$= m \sum_{r=1}^{K} \binom{K}{r} C_r + \sum_{r=m+1}^{K} (r-m) \binom{K}{r} C_r \tag{2.157}$$

for any achievable rate tuple $(R_0, R_1, \ldots, R_K)$ and any $m = 1, \ldots, K$, where $R_{sp} = \sum_{k=1}^{K} R_k$ is the sum of the private rates. It is clear that the outer bound given by the

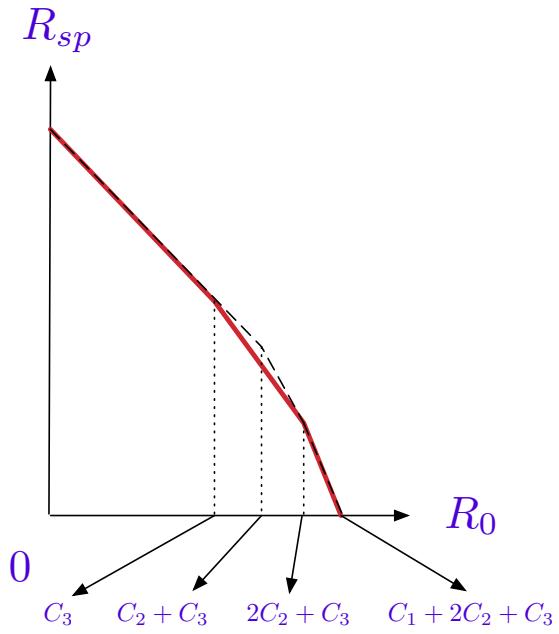

Figure 2.4: Capacity v.s. cut-set outer regions for $K = 3$ sinks. The boundary of the capacity region is illustrated by solid lines, while the boundary of the cut-set outer region is illustrated by dashed lines.

inequality (2.157) for $m = 1, \ldots, K$ has exactly $K + 1$ corner points:

$$\left(\sum_{i=r}^{K} \left(\begin{array}{c} K-1 \\ i-1 \end{array} \right) C_i, \sum_{i=1}^{r-1} \left(\begin{array}{c} K \\ i \end{array} \right) C_i \right), \quad r = 1, \ldots, K+1.$$

The achievability of these corner points was proved in [14]. Therefore, the generalized cut-set bounds also provide a *tight* characterization of the common-v.s.-sum-private capacity region of the general symmetrical combination network.

Finally, let us make an explicit comparison between the common-v.s.-sum-private capacity region of the general symmetrical combination network and the outer region given by *just* the standard cut-set bounds for the case of $K = 3$ sink nodes. For $K = 3$, the common-v.s.-sum-private capacity region of the network is given by all nonnegative

(R_0, R_{sp}) pairs satisfying

$$\begin{aligned} 3R_0 + R_{sp} &\leq 3C_1 + 6C_2 + 3C_3, \\ 3R_0 + 2R_{sp} &\leq 6C_1 + 6C_2 + 3C_3, \\ \text{and} \quad R_0 + R_{sp} &\leq 3C_1 + 3C_2 + C_3. \end{aligned} \tag{2.158}$$

The standard cut-set bounds, in this case, are given by

$$\begin{aligned} R_0 + R_1 &\leq C_1 + 2C_2 + C_3, \\ R_0 + R_2 &\leq C_1 + 2C_2 + C_3, \\ R_0 + R_3 &\leq C_1 + 2C_2 + C_3, \\ R_0 + R_1 + R_2 &\leq 2C_1 + 3C_2 + C_3, \\ R_0 + R_1 + R_3 &\leq 2C_1 + 3C_2 + C_3, \\ R_0 + R_3 + R_2 &\leq 2C_1 + 3C_2 + C_3, \\ R_0 + R_1 + R_2 + R_3 &\leq 2C_1 + 3C_2 + C_3. \end{aligned} \tag{2.159}$$

Substituting $R_1 = R_{sp} - R_2 - R_3$ into (2.159) and using Fourier-Motzkin elimination to eliminate R_2 and R_3 from the inequalities in (2.159), we may explicitly write the outer region given by just the standard cut-set bounds as the nonnegative (R_0, R_{sp}) pairs satisfying

$$\begin{aligned} 3R_0 + R_{sp} &\leq 3C_1 + 6C_2 + 3C_3, \\ 2R_0 + R_{sp} &\leq 3C_1 + 5C_2 + 2C_3, \\ \text{and} \quad R_0 + R_{sp} &\leq 3C_1 + 3C_2 + C_3. \end{aligned} \tag{2.160}$$

In Figure 2.4 we illustrate the rate regions constrained by (2.158) and (2.160), respectively. Clearly, even for the case with only $K = 3$ sink nodes, the standard cut-set bounds alone are *not* tight, while the generalized cut-set bounds provide a precise characterization of the common-v.s.-sum-private capacity region.

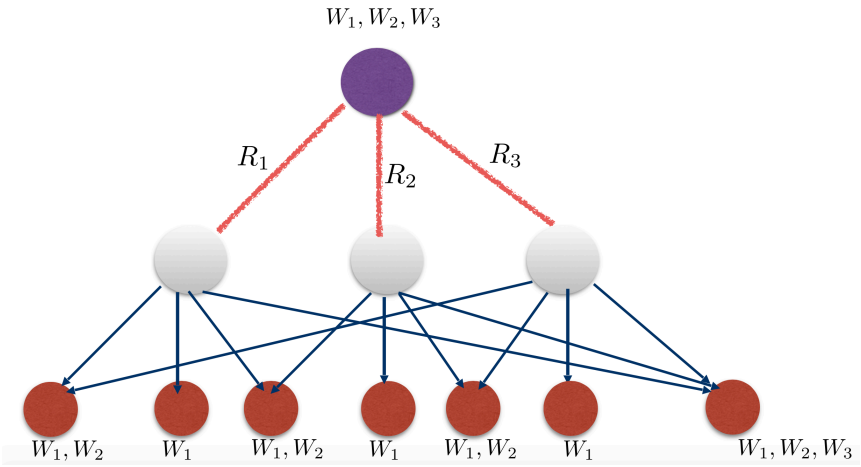

Figure 2.5: Symmetrical 3-level Diversity coding.

2.6 An Alternative Proof for Symmetrical Multilevel Diversity Coding

Here we show an alternative proof for the Lemma.1 in [15]. The underlying broadcast network shown in Figure 2.5. We assume that $H(W_i) = K_i$

Lemma 6. *For any constants* $K_1, K_2, K_3 \geq 0$ *and* (R_1, R_2, R_3) *we have*

$$R \geq K_i \tag{2.161}$$

$$R_i + R_j \geq 2K_1 + K + 2 \tag{2.162}$$

$$2R_i + Ri \oplus 1 + R_{i\oplus 2} \geq 4K_1 + 2K_2 + K_3 \tag{2.163}$$

$$R_1 + R_2 + R_3 \geq 3K_1 + \frac{3}{2} + K_3 \tag{2.164}$$

Proof. We assume the code on the link with capacity R_i is X_i. Considering 3 cuts, cut $C_1 : \{R_1\}$, $C_2 : \{R_2\}$ and $C_3 : \{R_1, R_2\}$, we can write generalized cut-set bound as

follows:

$$H(W_{C_1\cup C_2\cup C_3}) + H(W_{C_1} \cap W_{C_2}) \leq H(C_1 \cup C_2 \cup C_3) + H(C_1 \cap C_2) \tag{2.165}$$
$$H(W_1, W_2) + H(W_1 \cap W_1) \leq H(X_1, X_2) + H(\emptyset)$$
$$2K_1 + K_2 \leq R_1 + R_2$$

The second inequality can be prove by layering idea as we talked. Considering 3 cuts, cut $C_1 : \{X_1, X_2\}$, $C_2 : \{X_2, X-3\}$ and $C_3 : \{X_1, X_2, X_3\}$, and 3 cuts in the second layer $C_1' : \{X_1\}$, $C_2' : \{X_2\}$ and $C_3' : \{X_3\}$ we have:

$$H(W_{C_1\cup C_2\cup C_3}) \leq H(C_1) + H(C_2) - H(W_{C_1} \cap W_{C_2}) \tag{2.166}$$
$$H(W_{C_1\cup C_2\cup C_3}) + H(W_{C_1} \cap W_{C_2})$$
$$\leq H(C'1) + H(C'2) - H(W_{C_1'} \cap W_{C_2'}) + H(C_2') + H(C_3') - H(W_{C_2'} \cap W_{C_3'})$$
$$H(W_1, W_2, W_3) + H(W_1, W_2)$$
$$\leq H(X_1) + H(X_2) - H(W_1) + H(X_2) + H(X_3) - H(W_1)$$
$$\Rightarrow (4K_1 + 2K_2 + K_3) \leq 2R_2 + R_1 + R_3$$

In the similar fashion we can prove the last inequality. Now we consider the cuts as $C_1 : \{X_1, X_2\}$, $C_2 : \{X_2, X_3\}$ and $C_3 : \{X_1, X_3\}$, and 3 cuts in the second layer $C_1' : \{X_1\}$,

$C_2' : \{X_2\}$ and $C_3' : \{X_3\}$ we have:

$$H(W_{C_1 \cup C_2 \cup C_3}) \leq H(C_1) + H(C_2) + H(C_3) - H(W^{(2)}(C_1, C_2, C_3)) - H(W_{C_1 \cap C_2 \cap C_3}) \tag{2.167}$$

$$H(W_1, W_2, W_3) + H(W_1, W_2) + H(W_1, W_2) \leq H(C_1) + H(C_2) + H(C_3)$$

$$H(W_1, W_2, W_3) + H(W_1, W_2) + H(W_1, W_2) \leq H(C_1') + H(C_2') - H(W_1) + H(C_2') + H(C_3') - H(W_1) + H(C'1) + H(C'3) - H(W_1)$$

$$6K_1 + 3K_2 + 2K_3 \leq 2(R_1 + R_2 + R_3)$$

where $W^{(2)}(C_1, C_2, C_3) = (W_{C_1} \cap W_{C_2}) \cup (W_{C_1} \cap W_{C_3}) \cup (W_{C_3} \cap W_{C_2})$ □

2.7 Case $K = 4$

The new inequalities are as follows which are the coefficient in the form of

A:

$(K_1, K_2, K_3, K_4, R_1, R_2, R_3, R_4)$:

8.0000 4.0000 2.0000 1.0000 -4.0000 -2.0000 -1.0000 -1.0000

8.0000 4.0000 2.0000 1.0000 -4.0000 -1.0000 -2.0000 -1.0000

8.0000 4.0000 2.0000 1.0000 -4.0000 -1.0000 -1.0000 -2.0000

8.0000 4.0000 2.0000 1.0000 -2.0000 -4.0000 -1.0000 -1.0000

8.0000 4.0000 2.0000 1.0000 -2.0000 -1.0000 -4.0000 -1.0000

8.0000 4.0000 2.0000 1.0000 -2.0000 -1.0000 -1.0000 -4.0000

8.0000 4.0000 2.0000 1.0000 -1.0000 -4.0000 -2.0000 -1.0000

8.0000 4.0000 2.0000 1.0000 -1.0000 -4.0000 -1.0000 -2.0000

8.0000 4.0000 2.0000 1.0000 -1.0000 -2.0000 -4.0000 -1.0000

8.0000 4.0000 2.0000 1.0000 -1.0000 -2.0000 -1.0000 -4.0000

8.0000 4.0000 2.0000 1.0000 -1.0000 -1.0000 -4.0000 -2.0000

8.0000 4.0000 2.0000 1.0000 -1.0000 -1.0000 -2.0000 -4.0000

B:

6.0000 3.0000 2.0000 1.0000 -2.0000 -2.0000 -1.0000 -1.0000
6.0000 3.0000 2.0000 1.0000 -2.0000 -1.0000 -2.0000 -1.0000
6.0000 3.0000 2.0000 1.0000 -2.0000 -1.0000 -1.0000 -2.0000
6.0000 3.0000 2.0000 1.0000 -1.0000 -2.0000 -2.0000 -1.0000
6.0000 3.0000 2.0000 1.0000 -1.0000 -2.0000 -1.0000 -2.0000
6.0000 3.0000 2.0000 1.0000 -1.0000 -1.0000 -2.0000 -2.0000

C:

6.0000 3.0000 2.0000 1.0000 -2.0000 -2.0000 -1.0000 -1.0000
6.0000 3.0000 2.0000 1.0000 -2.0000 -1.0000 -2.0000 -1.0000
6.0000 3.0000 2.0000 1.0000 -2.0000 -1.0000 -1.0000 -2.0000
6.0000 3.0000 2.0000 1.0000 -1.0000 -2.0000 -2.0000 -1.0000
6.0000 3.0000 2.0000 1.0000 -1.0000 -2.0000 -1.0000 -2.0000
6.0000 3.0000 2.0000 1.0000 -1.0000 -1.0000 -2.0000 -2.0000

D:

4.5000 2.2500 1.5000 1.0000 -1.0000 -1.5000 -1.0000 -1.0000
4.5000 2.2500 1.5000 1.0000 -1.0000 -1.0000 -1.5000 -1.0000
4.5000 2.2500 1.5000 1.0000 -1.0000 -1.0000 -1.0000 -1.5000
4.5000 2.2500 1.5000 1.0000 -1.5000 -1.0000 -1.0000 -1.0000

E:

4.0000 2.0000 1.3333 1.0000 -1.0000 -1.0000 -1.0000 -1.0000

Plus the inequalities from case 3:

1.0000 0 0 0 -1.0000 0 0 0

1.0000 0 0 0 0 -1.0000 0 0

1.0000 0 0 0 0 0 -1.0000 0

1.0000 0 0 0 0 0 0 -1.0000

1.0000 0.5000 0 0 -0.5000 -0.5000 0 0

1.0000 0.5000 0 0 -0.5000 0 -0.5000 0

1.0000 0.5000 0 0 -0.5000 0 0 -0.5000

1.0000 0.5000 0 0 0 -0.5000 -0.5000 0

1.0000 0.5000 0 0 0 -0.5000 0 -0.5000

1.0000 0.5000 0 0 0 0 -0.5000 -0.5000

1.0000 0.5000 0.2500 0 -0.5000 -0.2500 -0.2500 0

1.0000 0.5000 0.2500 0 -0.5000 -0.2500 0 -0.2500

1.0000 0.5000 0.2500 0 -0.5000 0 -0.2500 -0.2500

1.0000 0.5000 0.2500 0 -0.2500 -0.5000 -0.2500 0

1.0000 0.5000 0.2500 0 -0.2500 -0.5000 0 -0.2500

1.0000 0.5000 0.2500 0 -0.2500 -0.2500 -0.5000 0

1.0000 0.5000 0.2500 0 -0.2500 -0.2500 0 -0.5000

1.0000 0.5000 0.2500 0 -0.2500 0 -0.5000 -0.2500

1.0000 0.5000 0.2500 0 -0.2500 0 -0.2500 -0.5000

1.0000 0.5000 0.2500 0 0 -0.5000 -0.2500 -0.2500

1.0000 0.5000 0.2500 0 0 -0.2500 -0.5000 -0.2500

1.0000 0.5000 0.2500 0 0 -0.2500 -0.2500 -0.5000

1.0000 0.5000 0.3333 0 -0.3333 -0.3333 -0.3333 0

1.0000 0.5000 0.3333 0 -0.3333 -0.3333 0 -0.3333

1.0000 0.5000 0.3333 0 -0.3333 0 -0.3333 -0.3333

1.0000 0.5000 0.3333 0 0 -0.3333 -0.3333 -0.3333

each group are essentially the same inequality up to a permutation.

Proof. Lets first see the following inequality:

$$H(X_1, X_2) \leq H(X_1) + (X_2) - H(W_1) \tag{2.168}$$

which translate to the following

$$H(X_i, X_j) \leq R_i + R_j - K_1 \tag{2.169}$$

$$\begin{aligned} H(X_i, X_j, X_j) \leq & H(X_i, X_j) \\ & + H(X_i, X_k) + H(X_j, X_k) - H(X_i, X_j, X_k) - H(W_1, W_2) \end{aligned} \tag{2.170}$$

$$2H(X_i, X_j, X_j) \leq 2R_i + 2R_j + 2R_k - 4K_1 - K_2 \tag{2.171}$$

where the last inequality is just replacement of (2.169), and

$$H(X_i, X_j, X_j) \leq H(X_i, X_j) + H(X_i, X_k) - H(W_1, W_2) \tag{2.172}$$

$$\leq 2R_i + R_j + R_k - 3K_1 - K_2 \tag{2.173}$$

again the last inequality is just replacement of (2.169), and

A:

It is with 2 cover:

$$\begin{aligned} H(W_1, W_2, W_3, W_4) &\leq H(X_1, X_2, X_3) + H(X_1, X_2, X_4) - H(W_1, W_2, W_3) \\ &\leq 2R_1 + R_2 + R_3 - 3K_1 - K_2 + 2R_1 \\ &+ R_2 + R_3 - 3K_1 - K_2 - K_1 - K_2 - K_3 \end{aligned} \tag{2.174}$$

B:

It is with 2 cover:

$$\begin{aligned} H(W_1, W_2, W_3, W_4) &\leq H(X_1, X_2, X_3) + H(X_1, X_2, X_4) - H(W_1, W_2, W_3) \\ &\leq R_1 + R_2 + R_3 - 2K_1 - 0.5K_2 + R_1 + R_2 \\ &+ R_3 - 2K_1 - 0.5K_2 - K_1 - K_2 - K_3 \end{aligned} \tag{2.175}$$

C:

This is 3 cover

$$
\begin{aligned}
H(W_1, W_2, W_3, W_4) &\leq H(X_1, X_2, X_3) + H(X_1, X_2, X_4) + H(X_1, X_3, X_4) \\
&\quad - H(W_1, W_2, W_3) - H(W_1, W_2, W_3, W_4) \\
&\leq 2R_1 + R_2 + R_3 - 3K_1 - K_2 + 2R_1 + R_2 + R_4 \\
&\quad - 3K_1 - K_2 + 2R_1 + R_3 + R_4 - 3K_1 \\
&\quad - K_2 - K_1 - K_2 - K_3 - (K_1 + K_2 + K_3 + K_4)
\end{aligned} \tag{2.176}
$$

D:

This is 3 cover

$$
\begin{aligned}
H(W_1, W_2, W_3, W_4) &\leq H(X_1, X_2, X_3) + H(X_1, X_2, X_4) + H(X_1, X_3, X_4) \\
&\quad - H(W_1, W_2, W_3) - H(W_1, W_2, W_3, W_4) \\
&\leq R_1 + R_2 + R_3 - 2K_1 - .5K_2 + R_1 + R_2 + R_3 - 2K_1 - 0.5K_2 \\
&\quad + R_1 + R_3 + R_4 - 2K_1 - 0.5K_2 \\
&\quad - K_2 - K_1 - K_2 - K_3 - (K_1 + K_2 + K_3 + K_4)
\end{aligned} \tag{2.177}
$$

E:

This is 4 cover

$$
\begin{aligned}
H(W_1, W_2, W_3, W_4) &\leq H(X_1, X_2, X_3) + H(X_1, X_2, X_4) + H(X_1, X_3, X_4) \\
&\quad + H(X_2, X_3, X_4) - H(W_1, W_2, W_3) - 2H(W_1, W_2, W_3, W_4) \\
&\leq R_1 + R_2 + R_3 - 2K_1 - .5K_2 + R_1 + R_2 + R_3 - 2K_1 - .5K_2) \\
&\quad + R_1 + R_3 + R_4 - 2K_1 - .5K_2 + R_2 + R_3 + R_4 - 2K_1 - .5K_2 \\
&\quad - K_2 - K_1 - K_2 - K_3 - 2(K_1 + K_2 + K_3 + K_4)
\end{aligned} \tag{2.178}
$$

2.8 Concluding Remarks

This chapter considered the problem of coding over broadcast networks with multiple (multicast) messages and more than two sink nodes. The standard cut-set bounds, which are known to be loose in general, are closely related to union as a specific set operation to combine different basic cuts of the network. A new set of network coding bounds (termed as *generalized* cut-set bounds), which relate the basic cuts of the network via a variety of set operations (not just the union), were established via the submodularity of the Shannon entropy. It was shown that the generalized cut-set bounds (together with the standard cut-set bounds) provide a precise characterization of the capacity region of the general combination network with three sink nodes and the common-v.s.-sum-private capacity region of the general symmetrical combination network (with arbitrary number of sink nodes).

Our ongoing work focuses primarily on further understanding the strength and the limitations of the generalized cut-set bounds established in this chapter. In particular, it would be interesting to see whether the generalized cut-set bounds are tight for the *symmetrical* capacity region of the general symmetrical combination network, which was recently characterized by Tian [14].

3. LATENCY CAPACITY REGION FOR GENERAL BROADCAST CHANNEL*

3.1 Introduction

The capacity region of broadcast channel has been studied in various different settings [1, 16]. The general broadcast channel with complete message set represents a communication scenario in which the source node transmits a different message to each distinct subset of receivers. More specifically, for the general broadcast channel with the source node s and K receivers $\{t_k : k \in [K]\}$, a *complete* message set at the source node s consists of $2^K - 1$ independent messages pertaining to each non-empty subset of receivers:

$$\mathcal{W} = \{w_{\mathcal{U}} : \emptyset \neq \mathcal{U} \subseteq [K]\}$$

where the message $w_{\mathcal{U}}$ is intended for all sink nodes from $\{t_k : k \in \mathcal{U}\}$. Thus, the set of the messages intended for the sink nodes t_k is given by:

$$\mathcal{W}_k = \{w_{\mathcal{U}} : k \in \mathcal{U} \subseteq [K]\}, \quad \forall k \in [K]. \tag{3.1}$$

With a slight abuse of notation, we shall denote the rate of the message $w_{\mathcal{U}}$ by $R_{\mathcal{U}}$ (instead of the more consistent notation $R_{w_{\mathcal{U}}}$).

Instead of asking the most fundamental question regarding the characterization of the capacity region for this channel, we are interested in a slightly different, but still fundamental concept known as *multicast capacity region* for a broadcast channel.

In this set up, suppose that the achievability of a rate tuple $\mathbf{C} := (C_{\mathcal{U}} : \mathcal{U} \subseteq [K])$ in the

*

$2^K - 1$ dimensional capacity region is given. The fundamental question of interest is that what are the set of all points that their achievability can be inferred just by knowing the achievability of the rate tuple $\mathbf{C}$)? The closure of all such achievable rate tuples is referred to as $\mathbf{C}$-*multicast region* for a broadcast channel. It is trivial to see that all the rate tuples that are element-wise marginalized by $\mathbf{C}$, belong to the $\mathbf{C}$-multicast region. For simplicity, through the rest of the chapter, we may drop the term $\mathbf{C}$ and we will refer to this region as multicast region.

It has been shown that the multicast region of the broadcast channel with complete message set, is independent of the quality of the broadcast channel [2]. Furthermore, it can be *essentially* transformed into a network coding problem over a *combination network* [3]. In this regard, the capacity region of the associated network coding problem is the multicast region of a broadcast channel. The topology of the corresponding combination network consists of three layers of nodes. The top layer consists of the source node s and the bottom layer consists of all the K users of the original broadcast channel $\{t_k : k \in [K]\}$. The middle layer consists of $2^K - 1$ intermediate nodes, each connecting to the source node s and a nonempty subset of sink nodes. While the links from the source node s to the intermediate nodes have finite capacities, the links from the intermediate nodes to the sink nodes are all assumed to have *infinite* capacities. More specifically, denote by $v_{\mathcal{U}}$ the intermediate node that connects to the nonempty subset $\mathcal{U}$ of sink nodes and $a_{\mathcal{U}}$ the arc that connects the source node s to the intermediate node $v_{\mathcal{U}}$. The link capacity for $a_{\mathcal{U}}$ is assigned as $C_{\mathcal{U}} \in \mathcal{R}_+$ which is a component in the achievable rate tuple $\mathbf{C}$, pertaining the subset of receivers $\mathcal{U}$. Note that when $C_{\mathcal{U}} = 0$, the intermediate node $v_{\mathcal{U}}$ can be effectively removed from the network. The associate combination network for a 3-user broadcast channel is illustrated in Figure 3.1.

For the case where the number of users is $K \leq 3$, the entire $\mathbf{C}$-multicast region for an achievable rate tuple $\mathbf{C}$ has been established [2]. However, for the cases where number of

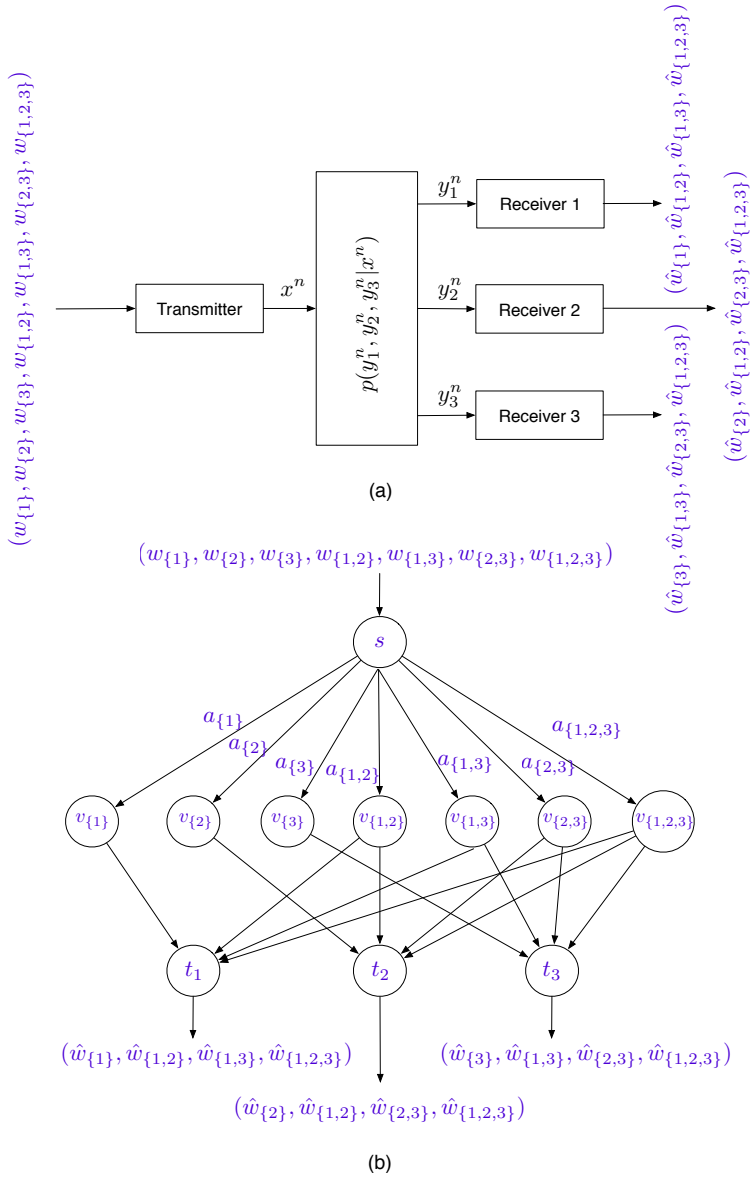

Figure 3.1: Associated combination network for a 3-user broadcast channel with complete message set

users, $K > 3$ the **C**-multicast region is still *unkown*. For the case where $K \geq 3$, Tian [14] considered a *symmetrical* setting, where

$$R_{\mathcal{U}} = R_{|\mathcal{U}|}, \quad \forall \emptyset \neq \mathcal{U} \subseteq [K] \tag{3.2}$$

for some nonnegative rate tuple $(R_k : k \in [K])$. That is, for any nonempty subset $\mathcal{U}$ of $[K]$, the rate $R_{\mathcal{U}}$ of the message $w_{\mathcal{U}}$ depend on the subset $\mathcal{U}$ only via its cardinality. We refer to this symmetrical setting as **C**-*symmetrical multicast region* for a broadcast channel, for any achievable rate tuple $\mathbf{C} := (C_1, C_2, \ldots, C_K)$. This assumption reduces the dimension of the region to K. For this setting, Tian [14] characterized the symmetrical multicast region. Tian's approach is based on rate splitting and pairwise rate transfer argument where a parametric description of symmetrical multicast region is given by using the rate splitting *parameters*.

In this chapter, we are interested in finding a polyhedral description of this region. One naive approach would be eliminating the rate splitting parameters from Tian's formulation, using Fourier-Motzkin elimination. However, the large number of the parameters that should be eliminated, makes this approach significantly inefficient and almost impossible. Our approach is to show that a set of *generalized cut-set bounds* proposed in Corollary 4, explicitly describes the capacity region. The main difficulty here is to establish an equivalency between two different description of a polytope. Unlike the traditional network information theory problems, where the dimension of the rate region is relatively small, establishing such an equivalency in higher dimensions is not an easy task and in general, there is not a unique framework for that. We use polyhedral combinatorics techniques to show that every maximal vector in the polytope given by *Generalized Cut-Set bounds* is achievable using a successive encoding scheme.

3.2 Main Results

Tian [14] showed that the symmetrical multicasting region is given by the set of nonnegative rate tuples $(R_k : k \in [K])$ satisfying:

$$R_j \leq \sum_{i=1}^{K} \phi_{i,j} r_{i,j}, \quad \forall j \in [K] \tag{3.3}$$

for some nonnegative reals $(r_{i,j} : i, j \in [K])$ satisfying

$$\sum_{j=1}^{K} r_{i,j} \leq C_i, \quad \forall i \in [K] \tag{3.4}$$

where

$$\phi_{i,j} := \begin{cases} \begin{pmatrix} K-i \\ j-i \end{pmatrix}^{-1} \begin{pmatrix} j-1 \\ j-i \end{pmatrix}, & \text{if } i < j \\ \begin{pmatrix} i \\ i-j \end{pmatrix}^{-1} \begin{pmatrix} K-j \\ i-j \end{pmatrix}, & \text{if } i > j \\ 1, & \text{if } i = j. \end{cases} \tag{3.5}$$

Our main result here is a precise *polyhedral* description of the symmetrical multicast region for a K-user broadcast channel , as summarized in the following theorem.

Theorem 3. *The $(C_k : k \in [K])$-symmetrical multicast region of a K-user broadcast channel is given by the set of rate tuples $(R_k : k \in [K])$ satisfying*

$$\left\{ \sum_{j=1}^{K} d_{\mathcal{Q}}(j) R_j \leq \sum_{j=1}^{K} d_{\mathcal{Q}}(j) C_j, \quad \forall \mathcal{Q} \subseteq [K] - \{1\} \right\} \tag{3.6}$$

where

$$d_{\mathcal{Q}}(j) := \begin{pmatrix} K \\ j \end{pmatrix} \sum_{r=1}^{j} \beta_{\mathcal{Q}}(r). \tag{3.7}$$

The converse part of the theorem follows directly from Corollary 4 where it has been shown that the generalized cut-set bounds are in fact an upper bound on the associated combination network. Consequently, they are an upper bound on the multicast region of

the broadcast network. More specifically

$$R(\mathcal{W}^{(r)}) = \sum_{\{\mathcal{U}\subseteq[K]:|\mathcal{U}|\geq r\}} R_{\mathcal{U}} = \sum_{j=r}^{K} \left(\begin{array}{c} K \\ j \end{array} \right) R_j \tag{3.8}$$

and

$$C(\mathcal{A}^{(r)}) = \sum_{\{\mathcal{U}\subseteq[K]:|\mathcal{U}|\geq r\}} R_{\mathcal{U}} = \sum_{j=r}^{K} \left(\begin{array}{c} K \\ j \end{array} \right) C_j, \quad \forall r \in [K] \tag{3.9}$$

for the above symmetrical setting and the choice of following simple cuts:

$$\mathcal{A}_k = \{a_{\mathcal{U}} \colon k \in \mathcal{U} \subseteq [K]\}, \quad \forall k \in [K]. \tag{3.10}$$

The forward part of the theorem follows from a characterization of the *maximum* vectors in the rate region (3.6) and a *successive* encoding scheme. The details of the proof are provided in Section 3.2.1.

We conclude our discussions on combination networks with the following comparison between our result and that of Tian's [14]:

- Tian's approach in [14] is *converse-centric* in that the forward part of the theorem is directly built on a *rate splitting* scheme, and the main challenge there was to prove the converse result *without* relying on a polyhedral description of the rate region.

- By comparison, our approach is *forward-centric* in that the converse part of the theorem follows directly from the generalized cut-set bounds (established systematically). The onus of the proof is on the forward part, where a *successive encoding* scheme (rather than rate splitting) is used.

- The problem of establishing the equivalence between Tian's characterization of

the symmetrical capacity region and ours is prototypical in polyhedral combinatorics [17]. However, we have not been able to establish such an equivalence using *conventional* polyhedral combinatorics techniques.

3.2.1 Achievability Proof of Theorem 3

We only need to show that every symmetrical rate tuple $(R_k : k \in [K])$ in the rate region (3.6) is achievable. The scheme that we shall consider is *successive encoding*, which we describe as follows.

For any $j \in [K]$ let $\boldsymbol{e}_j$ be a vector in $\mathcal{R}^K$ such that $e_{j,i} = 0$ for any $i \neq j$ and $e_{j,j} = 1$. For any $j \in [K-1]$ define

$$\boldsymbol{v}_j^+ := \phi_{j+1,j}\boldsymbol{e}_j - \boldsymbol{e}_{j+1} \quad \text{and} \quad \boldsymbol{v}_j^- := \phi_{j,j+1}\boldsymbol{e}_{j+1} - \boldsymbol{e}_j. \tag{3.11}$$

By using a maximum-distance separable (MDS) code, it can be shown [14] that symmetrical rate vector

$$\boldsymbol{R} = \boldsymbol{C} + \lambda_j \boldsymbol{v}_j^*$$

is achievable for any $j \in [K-1]$, $\lambda_j \in \mathcal{R}_+$, and $* \in \{+, -\}$ such that $\boldsymbol{R} \geq \boldsymbol{0}$, where $\boldsymbol{R} := (R_1, R_2, \ldots, R_K)$ and $\boldsymbol{C} := (C_1, C_2, \ldots, C_K)$. Further note that the achievability of the rate vector $\boldsymbol{R}$ induces a *virtual* symmetrical combination network[1] with symmetrical link-capacity constraints $\boldsymbol{R}$, for which the aforementioned MDS code can be applied again. By successively applying MDS codes over (virtual) symmetrical combination networks,

[1]The difference between a virtual combination network and an actual combination network is that while the links in an actual network are always reliable, there is a nonzero probability that the links in a virtual network are in outage. The outage probability, however, diminishes as the block length of the utilized MDS code increases.

any symmetrical rate vector

$$\boldsymbol{R} = \boldsymbol{C} + \sum_{j=1}^{K-1} \lambda_j \boldsymbol{v}_j^{*(j)} \tag{3.12}$$

for any $(\lambda_j : j \in [K-1]) \in \mathcal{R}_+^{K-1}$, $* : [K-1] \to \{+, -\}$, and permutation $\pi : [K-1] \to [K-1]$ such that

$$\begin{aligned} \boldsymbol{R}_i &:= \boldsymbol{R}_{i-1} + \lambda_{\pi(i)} \boldsymbol{v}_{\pi(i)}^{*(\pi(i))} \\ &= \boldsymbol{C} + \sum_{j=1}^{i} \lambda_{\pi(j)} \boldsymbol{v}_{\pi(j)}^{*(\pi(j))} \geq \boldsymbol{0}, \quad \forall i \in [K-1] \end{aligned} \tag{3.13}$$

is achievable, where $\boldsymbol{R}_0 := \boldsymbol{C}$. Our goal next is to show that any *maximum* rate vector[2] in the rate region (3.6) can be represented in the form of (3.12) and satisfying all constraints in (3.13).

Towards the above goal, let us first note that for any $\mathcal{Q} \subseteq [K] - \{1\}$, the corresponding constraint in (3.6) can be equivalently written as:

$$\sum_{j=1}^{K} d_{\mathcal{Q}}(j)(R_j - C_j) \leq 0.$$

Therefore, any rate tuple $\boldsymbol{R}$ in the rate region (3.6) can be written as $\boldsymbol{C} + \boldsymbol{x}$ for some vector $-\boldsymbol{C} \leq \boldsymbol{x} \in \mathcal{C}$, where

$$\mathcal{C} := \left\{ \boldsymbol{x} \in \mathcal{R}^K : \sum_{j=1}^{K} d_Q(j) x_j \leq 0 \right\}. \tag{3.14}$$

Furthermore, a maximum rate tuple $\boldsymbol{R}$ in the rate region (3.6) can be written as $\boldsymbol{C} + \boldsymbol{x}$ for some vector $-\boldsymbol{C} \leq \boldsymbol{x} \in \mathcal{C}$, where $\boldsymbol{x}$ is maximal in $\mathcal{C}$.

[2]For any given $\mathcal{P} \subseteq \mathcal{R}^K$, a vector $\boldsymbol{x} \in \mathcal{P}$ is said to be *maximal* in $\mathcal{P}$ if any $\boldsymbol{y} \in \mathcal{P}$ such that $\boldsymbol{y} \geq \boldsymbol{x}$ must satisfy $\boldsymbol{y} = \boldsymbol{x}$.

The following proposition provides a characterization of the maximum vectors in $\mathcal{C}$.

Proposition 1. *For any maximum vector $\boldsymbol{x} \in \mathcal{C}$, there exists a function $* : [K-1] \to \{+,-\}$ such that $\boldsymbol{x}$ can be written as a conic combination of the vectors from $\{\boldsymbol{v}_j^{*(j)} : j \in [K-1]\}$.*

By Proposition 1, any maximum rate vector in the rate region (3.6) can be represented in the form of (3.12). Our next proposition shows that a permutation $\pi : [K-1] \to [K-1]$ can be found such that all constraints in (3.13) are satisfied.

Proposition 2. *Let $\boldsymbol{x}$ be a vector in $\mathcal{R}^K$ such that*

$$-\boldsymbol{C} \leq \boldsymbol{x} = \sum_{j=1}^{K-1} \lambda_j \boldsymbol{v}_j^{*(j)} \tag{3.15}$$

for some $\boldsymbol{C} \geq 0$, $* : [K-1] \to \{+,-\}$, and $(\lambda_j : j \in [K-1]) \in \mathcal{R}_+^{K-1}$. Then, there exists a permutation $\pi : [K-1] \to [K-1]$ such that

$$\boldsymbol{x}_i := \sum_{j=1}^{i} \lambda_{\pi(j)} \boldsymbol{v}_{\pi(j)}^{*(\pi(j))} \geq -\boldsymbol{C}, \quad \forall i \in [K-1]. \tag{3.16}$$

Combining the results of Propositions 1 and 2 proves that any maximum rate vector in the rate region (3.6) can be achieved by a successive encoding scheme. By definition, the symmetrical capacity region is a *compact* set. It thus follows that any rate vector in the rate region (3.6) is achievable. For the remaining part of this section, we shall complete the proof of Theorem 3 by proving Propositions 1 and 2.

3.2.1.1 Proof of Proposition 1

Let us begin with the following two lemmas.

Lemma 7. *For any $\mathcal{Q} \subseteq [K] - \{1\}$ and any $j \in [K-1]$, we have*

$$\frac{d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} = \begin{cases} \phi_{j+1,j}^{-1}, & \textit{if } j+1 \in \mathcal{Q} \\ \phi_{j,j+1}, & \textit{if } j+1 \notin \mathcal{Q}. \end{cases} \tag{3.17}$$

Proof. Fix $j \in [K-1]$. When $j+1 \in \mathcal{Q}$, by the definition of $\beta_{\mathcal{Q}}(r)$ in (2.89) we have $\beta_{\mathcal{Q}}(j+1) = 0$. In this case, we have

$$\sum_{r=1}^{j+1} \beta_{\mathcal{Q}}(r) = \sum_{r=1}^{j} \beta_{\mathcal{Q}}(r)$$

and hence

$$\begin{aligned} \frac{d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} &= \frac{\begin{pmatrix} K \\ j \end{pmatrix} \sum_{r=1}^{j} \beta_{\mathcal{Q}}(r)}{\begin{pmatrix} K \\ j+1 \end{pmatrix} \sum_{r=1}^{j+1} \beta_{\mathcal{Q}}(r)} = \frac{\begin{pmatrix} K \\ j \end{pmatrix}}{\begin{pmatrix} K \\ j+1 \end{pmatrix}} \\ &= \frac{j+1}{K-j} = \frac{1}{\phi_{j+1,j}}. \end{aligned}$$

When $j+1 \notin \mathcal{Q}$, let us show that we always have

$$\frac{\sum_{r=1}^{j} \beta_{\mathcal{Q}}(r)}{\sum_{r=1}^{j+1} \beta_{\mathcal{Q}}(r)} = \frac{j}{j+1} \tag{3.18}$$

and hence

$$\frac{d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} = \frac{\begin{pmatrix} K \\ j \end{pmatrix} \sum_{r=1}^{j} \beta_{\mathcal{Q}}(r)}{\begin{pmatrix} K \\ j+1 \end{pmatrix} \sum_{r=1}^{j+1} \beta_{\mathcal{Q}}(r)}$$

$$= \frac{\begin{pmatrix} K \\ j \end{pmatrix} j}{\begin{pmatrix} K \\ j+1 \end{pmatrix} (j+1)} = \frac{j}{K-j} = \phi_{j,j+1}.$$

To prove (3.18), let us consider the following two cases separately.

Case 1: $\mathcal{Q} = \emptyset$. In this case, by our convention $\beta_{\mathcal{Q}}(r) = 1$ for any $r \in [K]$. We thus have

$$\sum_{r=1}^{j} \beta_{\mathcal{Q}}(r) = j, \quad \forall j \in [K-1]$$

and hence (3.18).

Case 2: $\mathcal{Q} \neq \emptyset$. In this case, let us write $\beta_{\mathcal{Q}}(r)$ more explicitly as:

$$\beta_{\mathcal{Q}}(r) = \begin{cases} \prod_{l=1}^{K} q_t, & \text{if } \ 1 \leq r < q_1 \\ \prod_{l=1}^{K} (q_t - 1), & \text{if } \ q_{|\mathcal{Q}|} < r \leq K \\ \prod_{l=1}^{t-1} (q_t - 1) \prod_{l=t}^{K} q_t, & \text{if } \ q_{t-1} < r < q_t \\ & \text{for some } t \in [|\mathcal{Q}|] - \{1\}. \end{cases}$$

When $j+1 < q_1$, we have

$$\frac{\sum_{r=1}^{j} \beta_{\mathcal{Q}}(r)}{\sum_{r=1}^{j+1} \beta_{\mathcal{Q}}(r)} = \frac{j \prod_{l=1}^{K} q_t}{(j+1) \prod_{l=1}^{K} q_t} = \frac{j}{j+1}.$$

When $j+1 > q_1$, let t^* be the largest $t \in [|\mathcal{Q}|]$ such that $q_t < j+1$. Then, for any $q_{t^*} \leq m \leq j+1$, we have

$$\begin{aligned}\sum_{r=1}^{m} \beta_{\mathcal{Q}}(r) = (q_1 - 1) \prod_{l=1}^{K} q_l \\ + \sum_{t=2}^{t^*} \left((q_t - q_{t-1} - 1) \prod_{l=1}^{t-1} (q_l - 1) \prod_{l=t}^{K} q_l \right) \\ + (m - q_{t^*}) \prod_{l=1}^{t^*} (q_l - 1) \prod_{l=t^*+1}^{K} q_l. \end{aligned} \tag{3.19}$$

The second term on the right-hand side of (3.19) can be simplified as follows:

$$\begin{aligned}&\sum_{t=2}^{t^*} \left((q_t - q_{t-1} - 1) \prod_{l=1}^{t-1} (q_l - 1) \prod_{l=t}^{K} q_l \right) \\ &= \sum_{t=2}^{t^*} \left(\prod_{l=1}^{t} (q_l - 1) \prod_{l=t}^{K} q_l \right) - \sum_{t=2}^{t^*} \left(\prod_{l=1}^{t-1} (q_l - 1) \prod_{l=t-1}^{K} q_l \right) \\ &= \sum_{t=2}^{t^*} \left(\prod_{l=1}^{t} (q_l - 1) \prod_{l=t}^{K} q_l \right) - \sum_{t=1}^{t^*-1} \left(\prod_{l=1}^{t} (q_l - 1) \prod_{l=t}^{K} q_l \right) \\ &= \prod_{l=1}^{t^*} (q_l - 1) \prod_{l=t^*}^{K} q_l - (q_1 - 1) \prod_{l=1}^{K} q_l. \end{aligned} \tag{3.20}$$

Substituting (3.20) into (3.19), we have

$$\begin{aligned}
&\sum_{r=1}^{m} \beta_{\mathcal{Q}}(r) \\
&= \prod_{l=1}^{t^*}(q_l - 1) \prod_{l=t^*}^{K} q_l + (m - q_{t^*}) \prod_{l=1}^{t^*}(q_l - 1) \prod_{l=t^*+1}^{K} q_l \\
&= m \left(\prod_{l=1}^{t^*}(q_l - 1) \prod_{l=t^*+1}^{K} q_l \right)
\end{aligned}$$

for any $q_{t^*} \leq m \leq j+1$. Note that $q_{t^*} < j+1$ implies that $j \geq q_{t^*}$. We thus have

$$\frac{\sum_{r=1}^{j} \beta_{\mathcal{Q}}(r)}{\sum_{r=1}^{j+1} \beta_{\mathcal{Q}}(r)} = \frac{j \left(\prod_{l=1}^{t^*}(q_l - 1) \prod_{l=t^*+1}^{K} q_l \right)}{(j+1) \left(\prod_{l=1}^{t^*}(q_l - 1) \prod_{l=t^*+1}^{K} q_l \right)} = \frac{j}{j+1}.$$

Combining the above two cases completes the proof of (3.18) and hence the entire lemma. □

Lemma 8. *For any $\mathcal{Q} \subseteq [K] - \{1\}$, define $\boldsymbol{d}_{\mathcal{Q}} := (d_{\mathcal{Q}}(1), d_{\mathcal{Q}}(2), \ldots, d_{\mathcal{Q}}(K))$ and $*_{\mathcal{Q}}, \overline{*}_{\mathcal{Q}} : [K-1] \to \{+, -\}$ by:*

$$*_{\mathcal{Q}}(j) := \begin{cases} +, & j+1 \in \mathcal{Q} \\ -, & j+1 \notin \mathcal{Q} \end{cases} \tag{3.21}$$

and

$$\overline{*}_{\mathcal{Q}}(j) := \begin{cases} -, & j+1 \in \mathcal{Q} \\ +, & j+1 \notin \mathcal{Q}. \end{cases} \tag{3.22}$$

Then, we have

$$\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^{*_{\mathcal{Q}}(j)} \rangle = 0 \quad \textit{and} \quad \langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^{\overline{*}_{\mathcal{Q}}(j)} \rangle < 0, \quad \forall j \in [K-1] \tag{3.23}$$

where $\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v} \rangle$ denotes the inner product between the vectors $\boldsymbol{d}_{\mathcal{Q}}$ and $\boldsymbol{v}$.

Proof. Fix $\mathcal{Q} \subseteq [K] - \{1\}$. Note that for any $j \in [K-1]$, we have

$$\begin{aligned}\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^+ \rangle &= \phi_{j+1,j} d_{\mathcal{Q}}(j) - d_{\mathcal{Q}}(j+1) \\ &= d_{\mathcal{Q}}(j+1) \left(\frac{\phi_{j+1,j} d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} - 1 \right).\end{aligned}$$

By Lemma 7, when $j+1 \in \mathcal{Q}$, we have

$$\frac{d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} = \phi_{j+1,j}^{-1}$$

and hence

$$\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^+ \rangle = 0.$$

When $j+1 \notin \mathcal{Q}$, we have

$$\frac{d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} = \phi_{j,j+1}$$

and hence

$$\begin{aligned}\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^+ \rangle &= d_{\mathcal{Q}}(j+1) \left(\phi_{j+1,j} \phi_{j,j+1} - 1 \right) \\ &= d_{\mathcal{Q}}(j+1) \left(\frac{K-j}{j+1} \frac{j}{K-j} - 1 \right) \\ &= -\frac{d_{\mathcal{Q}}(j+1)}{j+1} \\ &< 0.\end{aligned}$$

Similarly, for any $j \in [K-1]$, we have

$$\begin{aligned}\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^- \rangle &= -d_{\mathcal{Q}}(j) + \phi_{j,j+1} d_{\mathcal{Q}}(j+1) \\ &= d_{\mathcal{Q}}(j) \left(\frac{\phi_{j,j+1} d_{\mathcal{Q}}(j+1)}{d_{\mathcal{Q}}(j)} - 1 \right)\end{aligned}$$

By Lemma 7, when $j+1 \in \mathcal{Q}$, we have

$$\frac{d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} = \phi_{j+1,j}^{-1}$$

and hence

$$\begin{aligned}\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^- \rangle &= d_{\mathcal{Q}}(j) \left(\phi_{j,j+1} \phi_{j+1,j} - 1 \right) \\ &= d_{\mathcal{Q}}(j) \left(\frac{j}{K-j} \frac{K-j}{j+1} - 1 \right) \\ &= -\frac{d_{\mathcal{Q}}(j)}{j+1} \\ &< 0.\end{aligned}$$

When $j+1 \notin \mathcal{Q}$, we have

$$\frac{d_{\mathcal{Q}}(j)}{d_{\mathcal{Q}}(j+1)} = \phi_{j,j+1}$$

and hence

$$\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{v}_j^- \rangle = 0.$$

We have thus completed the proof of the lemma. □

Lemma 9. *For any $\mathcal{Q} \subseteq [K] - \{1\}$, the set of vectors*

$$\left\{ \boldsymbol{v}_1^{*_{\mathcal{Q}}(1)}, \boldsymbol{v}_2^{*_{\mathcal{Q}}(2)}, \ldots, \boldsymbol{v}_{K-1}^{*_{\mathcal{Q}}(K-1)}, \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}}(1)} \right\}$$

are linearly independent.

Proof. Let $\mathcal{Q}$ be a subset of $[K] - \{1\}$ and consider it fixed. Assume that

$$\sum_{j=1}^{K-1} \lambda_j \boldsymbol{v}_j^{*_{\mathcal{Q}}(j)} + \overline{\lambda}_1 \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}}(1)} = \mathbf{0} \tag{3.24}$$

for some collection of K reals $(\lambda_1, \lambda_2, \ldots, \lambda_{K-1}, \overline{\lambda}_1)$. Our goal is to show that (3.24) implies that

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{K-1} = \overline{\lambda}_1 = 0.$$

Let us first show that $\lambda_k = 0$ for any $k \in [K-1] - \{1\}$. By the assumption (3.24), we have

$$\begin{aligned} 0 &= \left\langle \boldsymbol{e}_K, \sum_{j=1}^{K-1} \lambda_j \boldsymbol{v}_j^{*_{\mathcal{Q}}(j)} + \overline{\lambda}_1 \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}}(1)} \right\rangle \\ &= \lambda_{K-1} \left\langle \boldsymbol{e}_K, \boldsymbol{v}_{K-1}^{*_{\mathcal{Q}}(K-1)} \right\rangle \\ &= \begin{cases} -\lambda_{K-1}, & \text{if } \; *_{\mathcal{Q}}(K-1) = + \\ \phi_{K-1,K} \lambda_{K-1}, & \text{if } \; *_{\mathcal{Q}}(K-1) = - \end{cases} \end{aligned}$$

which implies that $\lambda_{K-1} = 0$ in both cases. Next, assume that $\lambda_{j+1} = 0$ for some $j \geq 2$.

By the assumption (3.24), we have

$$
\begin{aligned}
0 &= \left\langle \boldsymbol{e}_{j+1}, \sum_{j=1}^{K-1} \lambda_j \boldsymbol{v}_j^{*_{\mathcal{Q}}(j)} + \overline{\lambda}_1 \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}}(1)} \right\rangle \\
&= \lambda_j \left\langle \boldsymbol{e}_{j+1}, \boldsymbol{v}_j^{*_{\mathcal{Q}}(j)} \right\rangle + \lambda_{j+1} \left\langle \boldsymbol{e}_{j+1}, \boldsymbol{v}_{j+1}^{*_{\mathcal{Q}}(j+1)} \right\rangle \\
&= \lambda_j \left\langle \boldsymbol{e}_{j+1}, \boldsymbol{v}_j^{*_{\mathcal{Q}}(j)} \right\rangle \\
&= \begin{cases} -\lambda_j, & \text{if } \ *_{\mathcal{Q}}(j-1) = + \\ \phi_{j,j+1}\lambda_j, & \text{if } \ *_{\mathcal{Q}}(j-1) = - \end{cases}
\end{aligned}
$$

which implies that that $\lambda_j = 0$ in both cases. Through the above induction, we conclude that $\lambda_k = 0$ for any $k \in [K-1] - \{1\}$.

Next, let us show that we have $\lambda_1 = \overline{\lambda}_1 = 0$ as well. By the assumption (3.24) and using the fact that $\lambda_k = 0$ for any $k \in [K-1] - \{1\}$, we have

$$
\begin{aligned}
\mathbf{0} &= \sum_{j=1}^{K-1} \lambda_j \boldsymbol{v}_j^{*_{\mathcal{Q}}(j)} + \overline{\lambda}_1 \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}}(1)} \\
&= \lambda_1 \boldsymbol{v}_1^{*_{\mathcal{Q}}(1)} + \overline{\lambda}_1 \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}}(1)} = \\
&\begin{cases} \left(\lambda_1 \phi_{2,1} - \overline{\lambda}_1\right) \boldsymbol{e}_1 + \left(\overline{\lambda}_1 \phi_{1,2} - \lambda_1\right) \boldsymbol{e}_2, & \text{if } \ *_{\mathcal{Q}}(1) = + \\ \left(\overline{\lambda}_1 \phi_{2,1} - \lambda_1\right) \boldsymbol{e}_1 + \left(\lambda_1 \phi_{1,2} - \overline{\lambda}_1\right) \boldsymbol{e}_2, & \text{if } \ *_{\mathcal{Q}}(1) = -. \end{cases}
\end{aligned} \tag{3.25}
$$

Note that

$$
\phi_{2,1}\phi_{1,2} = \frac{2}{(K-1)^2} \neq 1
$$

so (3.25) implies that $\lambda_1 = \overline{\lambda}_1 = 0$ in both cases. We have thus completed the proof of the lemma. □

We are now ready to prove the proposition. Let $\boldsymbol{x}$ be a maximum vector in $\mathcal{C}$. By the

definition of $\mathcal{C}$, we have

$$\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{x} \rangle \leq 0, \quad \forall \mathcal{Q} \subseteq [K] - \{1\}. \tag{3.26}$$

Since $\boldsymbol{x}$ is maximal in $\mathcal{C}$, there must exist a subset $\mathcal{Q} \subseteq [K] - \{1\}$ such that $\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{x} \rangle = 0$. Otherwise, let

$$\delta := \min_{\mathcal{Q} \subseteq [K]-\{1\}} \left[-\frac{\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{x} \rangle}{d_{\mathcal{Q}}(1)} \right] > 0$$

and we have

$$\langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{x} + \delta \boldsymbol{e}_1 \rangle = \langle \boldsymbol{d}_{\mathcal{Q}}, \boldsymbol{x} \rangle + \delta d_{\mathcal{Q}}(1) \leq 0, \quad \forall \mathcal{Q} \subseteq [K] - \{1\}$$

$$\text{and} \quad \boldsymbol{x} + \delta \boldsymbol{e}_1 > \boldsymbol{x}$$

violating the maximality of $\boldsymbol{x}$ in $\mathcal{C}$.

Assume that

$$\langle \boldsymbol{d}_{\mathcal{Q}'}, \boldsymbol{x} \rangle = 0 \tag{3.27}$$

for some $\mathcal{Q}' \subseteq [K] - \{1\}$. By Lemma 9, the set of vectors

$$\left\{ \boldsymbol{v}_1^{*_{\mathcal{Q}'}(1)}, \boldsymbol{v}_2^{*_{\mathcal{Q}'}(2)}, \ldots, \boldsymbol{v}_{K-1}^{*_{\mathcal{Q}'}(K-1)}, \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}'}(1)} \right\}$$

are linearly independent and hence span the entire $\mathcal{R}^K$. Let

$$\boldsymbol{x} = \sum_{j=1}^{K} \lambda_j \boldsymbol{v}_j^{*_{\mathcal{Q}'}(j)} + \overline{\lambda}_1 \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}'}(1)}.$$

By Lemma 8, we have

$$\begin{aligned} &\langle \boldsymbol{d}_{\mathcal{Q}'}, \boldsymbol{v}_j^{*_{\mathcal{Q}'}(j)} \rangle = 0, \quad \forall j \in [K-1] \\ \text{and} \quad &\langle \boldsymbol{d}_{\mathcal{Q}'}, \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}'}(1)} \rangle < 0. \end{aligned}$$

Combined with (3.27), we have

$$\begin{aligned} 0 = \langle \boldsymbol{d}_{\mathcal{Q}'}, \boldsymbol{x} \rangle &= \sum_{j=1}^{K-1} \lambda_j \langle \boldsymbol{d}_{\mathcal{Q}'}, \boldsymbol{v}_j^{*_{\mathcal{Q}'}(j)} \rangle + \overline{\lambda}_1 \langle \boldsymbol{d}_{\mathcal{Q}'}, \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}'}(1)} \rangle \\ &= \lambda_1 \langle \boldsymbol{d}_{\mathcal{Q}'}, v_1^{\overline{*}_{\mathcal{Q}'}(1)} \rangle \end{aligned}$$

implying that $\overline{\lambda}_1 = 0$ and hence

$$x = \sum_{j=1}^{K} \lambda_j \boldsymbol{v}_j^{*_{\mathcal{Q}'}(j)}.$$

To show that $\lambda_1 \geq 0$, consider the subset $\mathcal{Q}'' \subseteq [K] - \{1\}$ for which $j+1 \in \mathcal{Q}''$ for any $j > 1$ if and only if $j+1 \in Q'$ and $2 \in \mathcal{Q}''$ if and only if $2 \notin \mathcal{Q}'$. We thus have

$$*_{\mathcal{Q}''}(j) := \begin{cases} \overline{*}_{\mathcal{Q}'}(1), & \text{if } \ j = 1 \\ *_{\mathcal{Q}'}(j), & \text{if } \ j \neq 1 \end{cases}$$

$$\overline{*}_{\mathcal{Q}''}(j) := \begin{cases} *_{\mathcal{Q}'}(1), & \text{if } \ j = 1 \\ \overline{*}_{\mathcal{Q}'}(j), & \text{if } \ j \neq 1 \end{cases}$$

for any $j \in [K-1]$. Again by Lemma 8, we have

$$\begin{aligned} &\langle \boldsymbol{d}_{\mathcal{Q}''}, \boldsymbol{v}_j^{*_{\mathcal{Q}''}(j)} \rangle = 0, \quad \forall j \in [K-1] \\ \text{and} \quad &\langle \boldsymbol{d}_{Q''}, \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}''}(1)} \rangle < 0. \end{aligned}$$

Combined with (3.26), we have

$$\begin{aligned} 0 \geq \langle \boldsymbol{d}_{\mathcal{Q}''}, \boldsymbol{x} \rangle &= \sum_{j=1}^{K} \lambda_j \langle \boldsymbol{d}_{\mathcal{Q}''}, \boldsymbol{v}_j^{*_{\mathcal{Q}'}(j)} \rangle \\ &= \lambda_1 \langle \boldsymbol{d}_{\mathcal{Q}''}, \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}''}(1)} \rangle + \sum_{j=2}^{K} \lambda_j \langle \boldsymbol{d}_{\mathcal{Q}''}, \boldsymbol{v}_j^{*_{\mathcal{Q}''}(j)} \rangle \\ &= \lambda_1 \langle \boldsymbol{d}_{\mathcal{Q}''}, \boldsymbol{v}_1^{\overline{*}_{\mathcal{Q}''}(1)} \rangle \end{aligned}$$

and hence

$$\lambda_1 \geq 0.$$

Similarly, we can show that $\lambda_j \geq 0$ for any $j \in [K-1]$. We have thus completed the proof of the proposition.

3.2.1.2 Proof of Proposition 2

Let us first note that, to find a permutation $\pi : [K-1] \to [K-1]$ that satisfies (3.16), it suffices to find a permutation $\pi : [K-1] \to [K-1]$ such that for *any* $k \in [K]$, $v_{\pi(j),k}^{*(\pi(j))} < 0$ implies that $v_{\pi(i),k}^{*(\pi(i))} \leq 0$ for *all* $i > j$. This can be seen as follows. Fix $k \in [K]$. Let $j(k)$ be the *smallest* $j \in [K-1]$ such that $v_{\pi(j),k}^{*(\pi(j))} < 0$ (if it exists). Then, we have $v_{\pi(j),k}^{*(\pi(j))} \geq 0$

for any $j < j(k)$ and hence

$$x_{i,k} = \sum_{j=1}^{i} \lambda_{\pi(j)} v_{\pi(j),k}^{*(\pi(j))} \geq 0 \geq -C_k, \quad \forall i < j(k).$$

Furthermore, since $v_{\pi(j),k}^{*(\pi(j))} \leq 0$ for any $j > j(k)$, we have

$$x_{i,k} = x_{i+1,k} - \lambda_{\pi(i+1)} v_{\pi(i+1),k}^{*(\pi(i+1))} \geq x_{i+1,k}, \quad \forall i \geq j(k).$$

We thus have

$$x_{j(k),k} \geq x_{j(k)+1,k} \geq \cdots \geq x_{K-1,k} \geq -C_k.$$

To find a permutation $\pi : [K-1] \to [K-1]$ such that for any $k \in [K]$, $v_{\pi(j),k}^{*(\pi(j))} < 0$ implies that $v_{\pi(i),k}^{*(\pi(i))} \leq 0$ for *all* $i > j$, let us construct a directed graph with the vertex set given by $\{v_j : j \in [K-1]\}$. For each $j \in [K-1]$, we shall draw an arc from vertex v_{j+1} to vertex v_j if $*(j) = +$ or an arc from vertex v_j to vertex v_{j+1} if $*(j) = -$. We say that two vertices v_i and v_j are *adjacent* if $|i-j| \leq 1$. Note that all arcs in the graph are between adjacent vertices and between each pair of adjacent vertices there is one and only one arc. Therefore, the constructed graph is *acyclic* and there exists a *topological* order for the vertices of the graph.

For each $j \in [K-1]$, denote by a_j the arc between the vertices v_j and v_{j+1}. For each $j \in [K-1]$, denote the starting and ending vertices of a_j by $v^+(a_j)$ and $v^-(a_j)$, respectively. By the construction of the graph, for each $j \in [K-1]$, we have $v^+(a_j) = v_j$ if $*(j) = -$ and $v^+(a_j) = v_{j+1}$ if $*(j) = +$. For any given order of the vertex set $[K-1]$, an oder of the arc set $\{a_j : j \in [K-1]\}$ is said to be *compatible* with the order of the vertex set if for any two arcs a_i and a_j, we have $a_i \prec a_j$ if $v^+(a_i) \prec v^+(a_j)$.

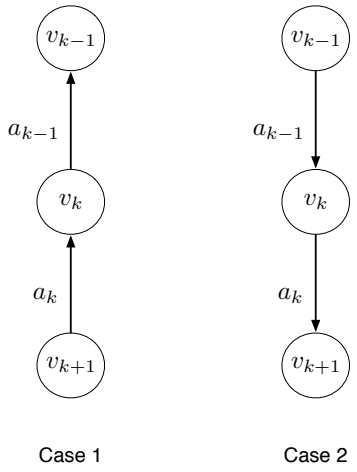

Figure 3.2: Two cases for $v_{k-1,k}^{*(k-1)} v_{k,k}^{*(k)} < 0$.

Now consider an order of the arc set $\{a_j : j \in [K-1]\}$ that is compatible with a topological order of the vertex set $\{v_j : j \in [K-1]\}$. Denote such an order by a permutation $\pi : [K-1] \to [K-1]$, i.e., $\pi(i) \leq \pi(j)$ if and only if $a_i \preceq a_j$. It remains to show that such a permutation π satisfies the desired property that for any $k \in [K]$, $v_{\pi(j),k}^{*(\pi(j))} < 0$ implies that $v_{\pi(i),k}^{*(\pi(i))} \leq 0$ for *all* $i > j$.

Fix $k \in [K]$. Note that among all $j \in [K-1]$, $v_{j,k}^{*(j)} \neq 0$ only when $j = k-1 > 0$ and $j = k < K$. Therefore, we only need to consider the cases where $v_{k-1,k}^{*(k-1)} v_{k,k}^{*(k)} < 0$ (see Figure 3.2 for an illustration) and show that $\pi(k) < \pi(k-1)$ if $v_{k-1,k}^{*(k-1)} < 0 < v_{k,k}^{*(k)}$ and $\pi(k-1) < \pi(k)$ if $v_{k,k}^{*(k)} < 0 < v_{k-1,k}^{*(k-1)}$.

Case 1: $v_{k-1,k}^{*(k-1)} < 0$ and $v_{k,k}^{*(k)} > 0$. By the definition of $\boldsymbol{v}_j^+$ and $\boldsymbol{v}_j^-$, we have $*(k-1) = *(k) = +$. We thus have $v^+(a_{k-1}) = v_k$ and $v^+(a_k) = v_{k+1}$. By the topological order of the vertex set, we have $v_{k+1} \prec v_k$, which implies that $a_k \prec a_{k-1}$ and hence $\pi(k) < \pi(k-1)$.

Case 2: $v_{k-1,k}^{*(k-1)} > 0$ and $v_{k,k}^{*(k)} < 0$. By the definition of $\boldsymbol{v}_j^+$ and $\boldsymbol{v}_j^-$, we have $*(k-1) = *(k) = -$. We thus have $v^+(a_{k-1}) = v_{k-1}$ and $v^+(a_k) = v_k$. By the topological

order of the vertex set, we have $v_{k-1} \prec v_k$, which implies that $a_{k-1} \prec a_k$ and hence $\pi(k-1) < \pi(k)$.

Combining the above two cases completes the proof of the proposition.

3.3 Conclusion

This chapter gives the polyhedral description of the multicast capacity region of a broadcast channel. It is shown that a comprehensive notion of the cut-set bound, so called generalized cut-set bound, characterizes entire capacity region. Furthermore, we show that every maximum rate vector can be achieved by a simple successive encoding scheme. Although this chapter, shed a light on the structure of the symmetric multicast capacity polytope for a broadcast channel, the multicast capacity region of the non-symmetric version remains an open problem.

4. ON THE AVERAGE ENTROPY REGIONS*

4.1 Introduction

Let $\mathcal{I}_n = \{1, \cdots, n\}$ and assume a fixed order among all $2^n - 1$ subsets of $\mathcal{I}_n$. A length-$(2^n - 1)$ vector $(h_{\mathcal{A}} : \emptyset \neq \mathcal{A} \subseteq \mathcal{I}_n)$ is called entropic if a set of n jointly distributed discrete random variables $(X_1, \cdots, X_n)$ can be found such that $h_{\mathcal{A}} = H(X_{\mathcal{A}}) \triangleq H(X_i : i \in \mathcal{A})$ for any $\emptyset \neq \mathcal{A} \subseteq \mathcal{I}_n$. The collection of all length-$(2^n - 1)$ entropic vectors is called the entropy region for n random variables and is usually denoted by Γ_n^*. It has been shown that even for $n = 3$, the entropy region Γ_n^* does not have nice topological properties. For the purposes of characterizing unconstrained information inequalities and network coding capacity regions, however, it is sufficient to study $\overline{\Gamma}_n^*$, the closure of Γ_n^*. Unlike Γ_n^*, it is known that $\overline{\Gamma}_n^*$ is a convex cone for any $n \in \mathbb{N}$ [12, Th. 15.5].

A natural outer bound on $\overline{\Gamma}_n^*$ is the set of length-$(2^n - 1)$ vectors that are constrained only by the Shannon type inequalities. We shall call this outer bound the Shannon entropy region for n random variables and denote it by Γ_n. In its most compact form, Γ_n can be described using a total of

$$n + \binom{n}{2} 2^{n-2}$$

linear inequalities [12, Eq. 14.12]. The existence of the Zhang-Yeung non-Shannon type inequality implies that $\overline{\Gamma}_n^* \subsetneqq \Gamma_n$ for $n \geq 4$ [18, Th. 4] (although they coincide for $n = 1, 2, 3$). In fact, unlike the Shannon entropy region Γ_n which is polyhedral for any

*Part of this chapter is reprinted, with permission, from [J. Chen, A. Salimi, T. Liu, and C. Tian, "Orbit-Entropy Cones and Extremal Pairwise Orbit-Entropy Inequalities," in the *Proceedings of the 2016 IEEE International Symposium on Information Theory*, pp.2614-2618.]

$n \in \mathbb{N}$, the boundary of $\overline{\Gamma}_n^*$ is known to be curved for $n \geq 4$ [19, Th. 1]. An explicit characterization of $\overline{\Gamma}_n^*$ for $n \geq 4$ is widely considered out of reach.

Fortunately, to solve any specific case of network coding, it often suffices to understand a (low-dimensional) projection of $\overline{\Gamma}_n^*$ (instead of acquiring the full knowledge of $\overline{\Gamma}_n^*$). Indeed, the Shannon type inequalities have been shown to yield exact characterizations of fundamental information-theoretic limits for many highly non-trivial problems (particularly those with certain symmetric structures [14, 20–25]), which suggests that the projections of $\overline{\Gamma}_n^*$ that are relevant to the problems under the consideration in fact coincide with those of Γ_n.

A natural questions arises: Under what kind of projections does the gap between $\overline{\Gamma}_n^*$ and Γ_n vanish? In this work we shall focus on the projections induced by permutation groups over $\mathcal{I}_n$ since such groups capture many important symmetry constraints encountered in engineering problems; the projected versions of $\overline{\Gamma}_n^*$ and Γ_n will be referred to as the orbit-entropy cones and the Shannon orbit-entropy cones, respectively. Note that both $\overline{\Gamma}_n^*$ and Γ_n, in their original forms, suffer from the curse of dimensionality. The essence of the new formulation is dimensionality reduction via symmetry considerations, which alleviates analytical difficulties and computational complexities.

4.1.1 Group Action and Orbits

Let G be a permutation group over $\mathcal{I}_n$. The group action of G on the subsets of $\mathcal{I}_n$ (induced by that on the elements of $\mathcal{I}_n$) is given by

$$g(\mathcal{A}) = \{g(a) : a \in \mathcal{A}\}, \quad g \in G, \mathcal{A} \subseteq \mathcal{I}_n.$$

In this work, we focus on the following permutation groups.

1. Partitioned symmetry group $S_{\mathcal{N}_1} \times \cdots \times S_{\mathcal{N}_q}$: This group is the product of the

symmetric groups $S_{\mathcal{N}_i}$ on $\mathcal{N}_i$, $i \in \mathcal{I}_k$, where $\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_q$ form a partition of $\mathcal{I}_n$. The action of $g \triangleq (s_{\mathcal{N}_1}, s_{\mathcal{N}_2}, \ldots, s_{\mathcal{N}_q}) \in S_{\mathcal{N}_1} \times \cdots \times S_{\mathcal{N}_q}$ on a subset $\mathcal{A}$ of $\mathcal{I}_n$ is given by

$$g(\mathcal{A}) = \bigcup_{i=1}^{q} s_{\mathcal{N}_i}(\mathcal{A} \cap \mathcal{N}_i).$$

2. Cyclic group C_n: This group consists of n elements g_i, $i \in \mathcal{I}_n$, and the action of g_i on a subset $\mathcal{A}$ of $\mathcal{I}_n$ is given by

$$g_i(\mathcal{A}) = \{(a+i)_n : a \in \mathcal{A}\},$$

where $(\cdot)_n$ means modulo n.

For each subset $\mathcal{A}$ of $\mathcal{I}_n$, we refer to the collection of distinct sets $g(\mathcal{A})$, $g \in G$, as an orbit. For an orbit $\mathcal{O}$, its cardinality is denoted by $|\mathcal{O}|$; the elements of $\mathcal{O}$ all have the same cardinality, which is denoted by $\ell_{\mathcal{O}}$. The orbits of G form a partition of $2^{\mathcal{I}_n}$. Counting the number of orbits of a permutation group G, denoted by $\omega(G)$, is a classical problem in combinatorics. It is easy to see that

$$\omega(S_{\mathcal{N}_1} \times \cdots \times S_{\mathcal{N}_q}) = \prod_{i=1}^{q} (|\mathcal{N}_i| + 1);$$

moreover, the number of orbits $\mathcal{O}$ of $S_{\mathcal{N}_1} \times \cdots \times S_{\mathcal{N}_q}$ with $\ell_{\mathcal{O}} = k$ is given by $N_q(k)$, $k \in \{0\} \cup \mathcal{I}_n$, which can be computed using the following recursive formula

$$\begin{aligned} N_0(k) &= \mathbb{I}(k = 0), \\ N_i(k) &= \sum_{j=0}^{|\mathcal{N}_i|} N_{i-1}(k-j), \quad k \in \{0\} \cup \mathcal{I}_n, i \in \mathcal{I}_q, \end{aligned}$$

where $\mathbb{I}(\cdot)$ is the indicator function. It is also well known that the number of orbits of C_n is given by

$$\omega(C_n) = \frac{1}{n}\sum_{d|n}\phi(d)2^{\frac{n}{d}},$$

where $\phi(\cdot)$ is Euler's totient function, and the sum is over divisors of n; moreover, the number of orbits $\mathcal{O}$ with $\ell_{\mathcal{O}} = k$ is given by

$$\frac{1}{n}\sum_{d|\gcd(k,n-k)}\phi(d)\binom{\frac{n}{d}}{\frac{k}{d}}, \quad k \in \{0\} \cup \mathcal{I}_n,$$

where the sum is over common divisors of k and $n-k$.

Let $\mathcal{O}_1, \cdots, \mathcal{O}_{\omega(G)-1}$ be the collection of all distinct non-empty[1] orbits of G. For n jointly distributed discrete random variables $(X_1, \cdots, X_n)$, the orbit-entropies are defined as

$$H_{\mathcal{O}_k} = \frac{1}{|\mathcal{O}_k|}\sum_{\mathcal{A}\in\mathcal{O}_k} H(X_{\mathcal{A}}), \quad k \in \mathcal{I}_{\omega(G)-1}.$$

It is instructive to consider the following examples.

1. The non-empty orbits of C_3 are given by

$$\mathcal{O}_1 = \{\{1\},\{2\},\{3\}\},$$
$$\mathcal{O}_2 = \{\{1,2\},\{2,3\},\{1,3\}\},$$
$$\mathcal{O}_3 = \{\{1,2,3\}\},$$

[1] An orbit is said to be empty if its only element is the empty set.

and we have

$$
\begin{aligned}
H_{\mathcal{O}_1} &= \frac{1}{3}[H(X_1) + H(X_2) + H(X_3)], \\
H_{\mathcal{O}_2} &= \frac{1}{3}[H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3)], \\
H_{\mathcal{O}_3} &= H(X_1, X_2, X_3).
\end{aligned}
$$

2. The non-empty orbits of $S_{\{1,2\}} \times S_{\{3\}}$ are given by

$$
\begin{aligned}
\mathcal{O}_1 &= \{\{1\}, \{2\}\}, \\
\mathcal{O}_2 &= \{\{1, 2\}\}, \\
\mathcal{O}_3 &= \{\{3\}\}, \\
\mathcal{O}_4 &= \{\{1, 3\}, \{2, 3\}\}, \\
\mathcal{O}_5 &= \{\{1, 2, 3\}\},
\end{aligned}
$$

and we have

$$
\begin{aligned}
H_{\mathcal{O}_1} &= \frac{1}{2}[H(X_1) + H(X_2)], \\
H_{\mathcal{O}_2} &= H(X_1, X_2), \\
H_{\mathcal{O}_3} &= H(X_3), \\
H_{\mathcal{O}_4} &= \frac{1}{2}[H(X_1, X_3) + H(X_2, X_3)], \\
H_{\mathcal{O}_5} &= H(X_1, X_2, X_3).
\end{aligned}
$$

For any length-(2^n-1) vector $(h_{\mathcal{A}}:\emptyset\neq\mathcal{A}\subseteq\mathcal{I}_n)$, the orbit-averages are defined as

$$h_{\mathcal{O}_k}=\frac{1}{|\mathcal{O}_k|}\sum_{\mathcal{A}\in\mathcal{O}_k}h_{\mathcal{A}},\quad k\in\mathcal{I}_{\omega(G)-1}.$$

We shall call the above mapping from $(h_{\mathcal{A}}:\emptyset\neq\mathcal{A}\subseteq\mathcal{I}_n)$ to $(h_{\mathcal{O}_1},\cdots,h_{\mathcal{O}_{\omega(G)-1}})$ the projection induced by G and denote it by P_G. It is clear that the orbit-entropy region (i.e., the set of all orbit-entropy vectors $(H_{\mathcal{O}_1},\cdots,H_{\mathcal{O}_{\omega(G)-1}})$) is given by $P_G\Gamma_n^*$. The focus of this work, however, is not $P_G\Gamma_n^*$, but rather $P_G\overline{\Gamma}_n^*$ and $P_G\Gamma_n$, which will be referred to as the orbit-entropy cone and the Shannon orbit-entropy cone, respectively.

Let S_n be the symmetric group over $\mathcal{I}_n$. A set Θ of length-(2^n-1) vectors is said to be permutation symmetric if $\mathbf{h}_g\in\Theta$ for any $\mathbf{h}\in\Theta$ and any $g\in S_n$, where $\mathbf{h}=(h_{\mathcal{A}}:\emptyset\neq\mathcal{A}\subseteq\mathcal{I}_n)$ and $\mathbf{h}_g=(h_{g(\mathcal{A})}:\emptyset\neq\mathcal{A}\subseteq\mathcal{I}_n)$ Note that both Γ_n^* (and hence $\overline{\Gamma}_n^*$) and Γ_n are permutation symmetric for any $n\in\mathbb{N}$. The following result implies that characterizing $P_G\overline{\Gamma}_n^*$ ($P_G\Gamma_n$) is equivalent to characterizing the set of vectors in $\overline{\Gamma}_n^*$ (Γ_n) satisfying the symmetry constraints induced by G.

Proposition 3. *For any convex, permutation symmetric set Θ of length-(2^n-1) vectors and any permutation group G over $\mathcal{I}_n$, we have $P_G\Theta=P_G\Theta'$, where*
$\Theta'=\{\mathbf{h}\in\Theta:h_{\mathcal{A}}=h_{\mathcal{A}'}$ *for all $\mathcal{A},\mathcal{A}'$ in the same orbit of G*$\}$.

Proof. Clearly we have $P_G\Theta\supseteq P_G\Theta'$. To show the opposite inclusion, consider $P_G\mathbf{h}$ for some arbitrary $\mathbf{h}\in\Theta$. Since Θ is permutation symmetric, it follows that $\mathbf{h}_g\in\Theta$ for any $g\in G$. By the convexity of Θ, the group average $\frac{1}{|G|}\sum_{g\in G}\mathbf{h}_g\in\Theta$. Furthermore, by the Lagrange theorem,

$$\frac{1}{|G|}\sum_{g\in G}h_{g(\mathcal{A})}=\frac{1}{|\mathcal{O}|}\sum_{\mathcal{A}\in\mathcal{O}}h_{\mathcal{A}}=h_{\mathcal{O}}$$

for any non-empty orbit $\mathcal{O}$ of G and any $\mathcal{A} \in \mathcal{O}$. Therefore, we have

$$\frac{1}{|G|}\sum_{g\in G}\mathbf{h}_g \in \Theta',$$
$$P_G\Big(\frac{1}{|G|}\sum_{g\in G}\mathbf{h}_g\Big) = P_G\mathbf{h}.$$

This completes the proof of the opposite inclusion $P_G\Theta \subseteq P_G\Theta'$. □

4.2 Partitioned Symmetry Groups

4.2.1 $q = 1$

As a warm-up exercise, we first consider the case $q = 1$, i.e., the symmetric group S_n. In a certain sense, P_{S_n} is the simplest projection among those induced by permutation groups. Indeed, every permutation group G over $\mathcal{I}_n$ is a subgroup of S_n; as a consequence, one can obtain P_{S_n} from P_G via a further projection.

It is easy to see that the non-empty orbits of S_n are given by $\mathcal{O}_i = \{\mathcal{A} \subseteq \mathcal{I}_n : |\mathcal{A}| = i\}$, $i \in \mathcal{I}_n$. The following result (which was also obtained independently in [26]) provides a complete characterization of $P_{S_n}\overline{\Gamma}_n^*$ and $P_{S_n}\Gamma_n$.

Proposition 4. $P_{S_n}\overline{\Gamma}_n^* = P_{S_n}\Gamma_n = \Pi_n$, *where* Π_n *is the set of length-*n *vectors* $(h_{\mathcal{O}_1}, \cdots, h_{\mathcal{O}_n})$ *satisfying*

$$\begin{aligned}
&2h_{\mathcal{O}_1} - h_{\mathcal{O}_2} \geq 0,\\
&2h_{\mathcal{O}_i} - h_{\mathcal{O}_{i-1}} - h_{\mathcal{O}_{i+1}} \geq 0, \quad i = 2, \cdots, n-1,\\
&h_{\mathcal{O}_n} - h_{\mathcal{O}_{n-1}} \geq 0,
\end{aligned}$$

or equivalently, Π_n *is the convex polyhedral cone generated by the vectors* $\mathbf{r}_i \triangleq (r_{i,1}, \cdots, r_{i,n})$, $i \in \mathcal{I}_n$, *with* $r_{i,k} = \min\{i, k\}$, $i \in \mathcal{I}_n$, $k \in \mathcal{I}_n$.

Proof. Let us define $X_S = \{X_i, \quad i \in S\}$. Because of symmetry, each set of elemental inequalities in the form of $I(X_i; X_j|X_S)$ where $S \subseteq [K] \setminus \{i, j\}$ and $|S| = k$.

$$I(X_i; X_j|X_S) = h_{X_S,X_i} + h_{X_S,X_j} - 2h_{X_S} \tag{4.1}$$

and therefore, the projection to symmetric-sum region will be

$$-h_{\mathcal{O}_{i-2}} + 2h_{\mathcal{O}_{i-1}} - h_{\mathcal{O}_i} \geq 0. \tag{4.2}$$

and the projection of the elemental inequality $H(X_i|X_{S\setminus\{i\}}) \geq 0$ will be $h_{\mathcal{O}_K} - h_{\mathcal{O}_{K-1}} \geq 0$.

Next it is easy to verify that all the extreme rays satisfy the supporting hyperplanes of the cone described by inequalities in the Proposition4. In order to complete the proof, we need to show that these extreme rays are The proof is by induction. It is easy to check that it holds for K=2. Consider that the result holds for any K random variables. From elemental inequalities, we can see that by adding the $K+1$th random variable, the conditional entropy constraint will be changed to

$$h_{\mathcal{O}_{K+1}} \geq h_{\mathcal{O}_K} \tag{4.3}$$

and we will have one more extra constraint

$$-h_{\mathcal{O}_{K-1}} + 2h_{\mathcal{O}_K} - h_{\mathcal{O}_{K+1}} \geq 0 \tag{4.4}$$

Therefore, the $K+1$ hyperplanes are as follows

$$h_{\mathcal{O}_{K+1}} - h_{\mathcal{O}_K} = 0 \tag{4.5}$$

$$-h_{\mathcal{O}_{K-1}} + 2h_{\mathcal{O}_K} - h_{\mathcal{O}_{K+1}} = 0 \tag{4.6}$$

$$-h_{\mathcal{O}_{i-2}} + 2h_{\mathcal{O}_{i-1}} - h_{\mathcal{O}_i} = 0, \quad \text{for} \quad i \in \{2, 3, \ldots, K\} \tag{4.7}$$

Moreover, any extremal ray of this cone should be at the intersection of at least K hyperplanes. Therefore, any K selection of these hyperplanes will identify a ray (which may not be extremal). We will have 3 different cases:

Case 1: Both (4.5),(4.6) *are in our selection*

In this case, intersection of these hyperplanes is

$$h_{\mathcal{O}_k} - h_{\mathcal{O}_{k-1}} = 0 \tag{4.8}$$

which together with any $K-2$ will identify a ray in K dimension (reduces the problem to the case with K random variables). Moreover, form induction hypothesis we know that intersection of these vectors is given by vectors in (4.5), (4.6),(4.7). Furthermore, we know from (4.5) tha the $K+1$th coordinate of the ray, will be exactly the same as the Kth coordinate. This will give us $K-1$ rays of the induction hypothesis. The last ray in the induction hypothesis is the ray with selecting all hyperplanes in (4.7) which will be taken care of in the next case.

Case 2: (4.5) *is in selection but not* (4.6)

As we mentioned before, in this case we have the last ray of induction hypothesis, i.e.

selecting (4.7). Furthermore, we know from (4.5) tha the $K+1$th coordinate of the ray, will be exactly the same as the Kth coordinate.

Case 2: (4.5) *is not in selection but* (4.6) *is*:

It is easy to verify that in this case the extreme ray will be $(1, 2, \ldots, K+1)$. This completes the proof. □

Note that the extreme rays $\{\boldsymbol{r}_i = (r_{i,1}, \ldots, r_{i,n}) : i \in \mathcal{N}_n\}$ of $P_T\Gamma_n$ can all be realized by a total-average projection of *uniform matroids* [27]. Since all matroids are known to be entropic, we conclude that

$$P_T\Gamma_n \subseteq P_T cl(\Gamma_n^*) \tag{4.9}$$

and hence

$$cl(P_T\Gamma_n^*) = P_T\Gamma_n. \tag{4.10}$$

4.3 Cyclic Groups

4.3.1 Orbit-Entropy Cones

Since $C_n = S_n$ for $n \leq 3$, it suffices to consider $n \geq 4$. One can readily verify that the non-empty orbits of C_4 are given by

$$\mathcal{O}_1 = \{\{1\}, \{2\}, \{3\}, \{4\}\},$$
$$\mathcal{O}_2 = \{\{1,2\}, \{2,3\}, \{3,4\}, \{1,4\}\},$$
$$\mathcal{O}_3 = \{\{1,3\}, \{2,4\}\},$$
$$\mathcal{O}_4 = \{\{1,2,3\}, \{2,3,4\}, \{1,3,4\}, \{1,2,3\}\},$$
$$\mathcal{O}_5 = \{\{1,2,3,4\}\},$$

and the non-empty obits of C_5 are given by

$$\begin{aligned}
\mathcal{O}_1 &= \{\{1\},\{2\},\{3\},\{4\},\{5\}\},\\
\mathcal{O}_2 &= \{\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{1,5\}\},\\
\mathcal{O}_3 &= \{\{1,3\},\{2,4\},\{3,5\},\{1,4\},\{2,5\}\},\\
\mathcal{O}_4 &= \{\{1,2,3\},\{2,3,4\},\{3,4,5\},\{1,4,5\},\{1,2,5\}\},\\
\mathcal{O}_5 &= \{\{1,2,4\},\{2,3,5\},\{1,3,4\},\{2,4,5\},\{1,3,5\}\},\\
\mathcal{O}_6 &= \{\{1,2,3,4\},\{2,3,4,5\},\{1,3,4,5\},\{1,2,4,5\},\\
&\qquad \{1,2,3,5\}\},\\
\mathcal{O}_7 &= \{\{1,2,3,4,5\}\}.
\end{aligned}$$

The following results provide a complete characterization of $P_{C_n}\overline{\Gamma}_n^*$ and $P_{C_n}\Gamma_n$ for $n = 4, 5$. The proofs follow the same general strategy as that of Proposition 4 and are omitted.

Proposition 5. $P_{C_4}\overline{\Gamma}_4^* = P_{C_4}\Gamma_4 = \Omega_4$*, where* Ω_4 *is the set of length-five vectors* $(h_{\mathcal{O}_1}, h_{\mathcal{O}_2}, h_{\mathcal{O}_3}, h_{\mathcal{O}_4}, h_{\mathcal{O}_5})$ *satisfying*

$$\begin{aligned}
&2h_{\mathcal{O}_1} - h_{\mathcal{O}_2} \geq 0,\\
&2h_{\mathcal{O}_1} - h_{\mathcal{O}_3} \geq 0,\\
&2h_{\mathcal{O}_2} - h_{\mathcal{O}_4} - h_{\mathcal{O}_1} \geq 0,\\
&h_{\mathcal{O}_2} + h_{\mathcal{O}_3} - h_{\mathcal{O}_4} - h_{\mathcal{O}_1} \geq 0,\\
&2h_{\mathcal{O}_4} - h_{\mathcal{O}_5} - h_{\mathcal{O}_2} \geq 0,\\
&2h_{\mathcal{O}_4} - h_{\mathcal{O}_5} - h_{\mathcal{O}_3} \geq 0,\\
&h_{\mathcal{O}_5} - h_{\mathcal{O}_4} \geq 0,
\end{aligned}$$

or equivalently, Ω_4 is the convex polyhedral cone generated by the following six vectors:

$$(1,1,1,1,1),\quad (1,2,2,2,2),\quad (1,2,2,3,3),$$
$$(1,2,2,3,4),\quad (1,2,1,2,2),\quad (1,\tfrac{3}{2},2,2,2).$$

Proof. As before, it is sufficient to show that for any r from the above set of six vectors, a positive scalar β can be found such that $\beta r \in P_C\Gamma_4$. Note that for the first four vectors in the set we have $r_2 = r_3$, which is the projection of the orbits $\mathcal{O}_\in$ and $\mathcal{O}_\ni$. Thus, to prove for these four vectors, it thus suffices to use the MDS codes as we did for $G = S_n$. Next, to prove for $r = (1,2,1,2,2)$, let U_1 and U_2 be two independent uniform variables over a finite field $\mathbb{F}$. Let $X_1 = X_3 = U_1$ and $X_2 = X_4 = U_2$. We have

$$H(\mathsf{X}_S) = \begin{cases} \log|\mathbb{F}|, & \text{for } S \in O_1 \\ 2\log|\mathbb{F}|, & \text{for } S \in O_2 \\ \log|\mathbb{F}|, & \text{for } S \in O_3 \\ 2\log|\mathbb{F}|, & \text{for } S \in O_4 \\ 2\log|\mathbb{F}|, & \text{for } S \in O_5 \end{cases} \tag{4.11}$$

We thus conclude that the length-15 vector
$boldsymbolh = (h_S : \emptyset \neq S \subseteq \mathcal{N}_4)$ where

$$h_S = \begin{cases} \log|\mathbb{F}|, & \text{for } S \in O_1 \\ 2\log|\mathbb{F}|, & \text{for } S \in O_2 \\ \log|\mathbb{F}|, & \text{for } S \in O_3 \\ 2\log|\mathbb{F}|, & \text{for } S \in O_4 \\ 2\log|\mathbb{F}|, & \text{for } S \in O_5 \end{cases} \tag{4.12}$$

is entropic and hence

$$(\log|\mathbb{F}|)\boldsymbol{r} = P_C\boldsymbol{h} \in P_C\Gamma_4 \tag{4.13}$$

Finally, to prove for $r = (1, 3/2, 2, 2, 2)$, let U_1, U_2, U_3 and U_4 be four independent uniform variables over a finite field F. Let $X_1 = (U_1, U_2)$, $X_2 = (U_2, U_3)$, $X_3 = (U_3, U_4)$ and $X_4 = (U_4, U_1)$. We have

$$H(\mathsf{X}_S) = \begin{cases} 2\log|\mathbb{F}|, & \text{for } S \in O_1 \\ 3\log|\mathbb{F}|, & \text{for } S \in O_2 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_3 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_4 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_5 \end{cases} \tag{4.14}$$

We thus conclude that the length-15 vector
$boldsymbolh = (h_S : \emptyset \neq S \subseteq \mathcal{N}_4)$ where

$$h_S = \begin{cases} 2\log|\mathbb{F}|, & \text{for } S \in O_1 \\ 3\log|\mathbb{F}|, & \text{for } S \in O_2 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_3 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_4 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_5 \end{cases} \tag{4.15}$$

is entropic and hence

$$(2\log|\mathbb{F}|)\boldsymbol{r} = P_C\boldsymbol{h} \in P_C\Gamma_4 \tag{4.16}$$

This completes the proof. □

Remark 1. *Note that the coding scheme that we considered for proving for* $r = (1, 3/2, 2, 2, 2)$ *is vector linear rather than scalar linear as we used before.*

Proposition 6. $P_{C_5}\overline{\Gamma}_5^* = P_{C_5}\Gamma_5 = \Omega_5$*, where* Ω_5 *is the set of length-seven vectors* $(h_{\mathcal{O}_1}, h_{\mathcal{O}_2}, h_{\mathcal{O}_3}, h_{\mathcal{O}_4}, h_{\mathcal{O}_5}, h_{\mathcal{O}_6}, h_{\mathcal{O}_7})$ *satisfying*

$$
\begin{aligned}
&2h_{\mathcal{O}_1} - h_{\mathcal{O}_2} \geq 0,\\
&2h_{\mathcal{O}_1} - h_{\mathcal{O}_3} \geq 0,\\
&2h_{\mathcal{O}_2} - h_{\mathcal{O}_1} - h_{\mathcal{O}_4} \geq 0,\\
&h_{\mathcal{O}_2} + h_{\mathcal{O}_3} - h_{\mathcal{O}_1} - h_{\mathcal{O}_4} \geq 0,\\
&2h_{\mathcal{O}_3} - h_{\mathcal{O}_1} - h_{\mathcal{O}_5} \geq 0,\\
&h_{\mathcal{O}_2} + h_{\mathcal{O}_3} - h_{\mathcal{O}_1} - h_{\mathcal{O}_5} \geq 0,\\
&2h_{\mathcal{O}_4} - h_{\mathcal{O}_2} - h_{\mathcal{O}_6} \geq 0,\\
&h_{\mathcal{O}_4} + h_{\mathcal{O}_5} - h_{\mathcal{O}_2} - h_{\mathcal{O}_6} \geq 0,\\
&2h_{\mathcal{O}_5} - h_{\mathcal{O}_3} - h_{\mathcal{O}_6} \geq 0,\\
&h_{\mathcal{O}_4} + h_{\mathcal{O}_5} - h_{\mathcal{O}_3} - h_{\mathcal{O}_6} \geq 0,\\
&2h_{\mathcal{O}_6} - h_{\mathcal{O}_4} - h_{\mathcal{O}_7} \geq 0,\\
&2h_{\mathcal{O}_6} - h_{\mathcal{O}_5} - h_{\mathcal{O}_7} \geq 0,\\
&h_{\mathcal{O}_7} - h_{\mathcal{O}_6} \geq 0,
\end{aligned}
$$

or equivalently, Ω_5 *is the convex polyhedral cone generated by the following eleven vec-*

tors:

$$r_1 = (1,1,1,1,1,1,1), \quad r_2 = (1,2,2,2,2,2,2), \quad r_3 = (1,2,2,3,3,3,3),$$
$$r_4 = (1,2,2,3,3,4,4), \quad r_5 = (1,2,2,3,3,4,5), \quad r_6 = (1,\tfrac{3}{2},2,2,2,2,2),$$
$$r_7 = (1,2,\tfrac{3}{2},2,2,2,2), \quad r_8 = (1,2,2,\tfrac{5}{2},3,3,3), \quad r_9 = (1,2,2,3,\tfrac{5}{2},3,3)$$
$$r_{10} = (1,2,\tfrac{3}{2},\tfrac{5}{2},2,\tfrac{5}{2},\tfrac{5}{2}), \quad r_{11} = (1,\tfrac{3}{2},2,2,\tfrac{5}{2},\tfrac{5}{2},\tfrac{5}{2}).$$

Proof. To show that the cyclic group C_5 is Shannon, it suffices to show that all eleven extreme rays of $P_{C_5}\Gamma_5$ are in $P_{C_5}cl(\Gamma_5^*)$. It is clear that the extreme rays $\boldsymbol{r}_i$, $i \in \mathcal{N}_5$, can be realized by a cyclic projection of *uniform matroids* [27] and are hence in $P_{C_5}cl(\Gamma_5^*)$. So we only need to show that $\boldsymbol{r}_i$, $i = 6,7,8,9$, are in $P_{C_5}cl(\Gamma_5^*)$.

To show that $\boldsymbol{r}_7 \in P_{C_5}cl(\Gamma_5^*)$, let U_i, $i \in \mathcal{N}_4$, be four independent uniform variables over a finite field $\mathbb{F}$ and

$$\mathsf{X}_1 := (\mathsf{U}_1, \mathsf{U}_2 + \mathsf{U}_3) \tag{4.17}$$
$$\mathsf{X}_2 := (\mathsf{U}_2, \mathsf{U}_3 + \mathsf{U}_4) \tag{4.18}$$
$$\mathsf{X}_3 := (\mathsf{U}_3, \mathsf{U}_1) \tag{4.19}$$
$$\mathsf{X}_4 := (\mathsf{U}_4, \mathsf{U}_2 + \mathsf{U}_3) \tag{4.20}$$
$$\mathsf{X}_5 := (\mathsf{U}_4 + \mathsf{U}_1, \mathsf{U}_3 + \mathsf{U}_4). \tag{4.21}$$

It is straightforward to verify that

$$H(\mathsf{X}_S) = \begin{cases} 2\log|\mathbb{F}|, & \text{for } S \in O_1 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_2 \\ 3\log|\mathbb{F}|, & \text{for } S \in O_3 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_4 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_5 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_6 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_7 \end{cases} \tag{4.22}$$

completing the proof that $\boldsymbol{r}_7 \in P_{C_5} cl(\Gamma_5^*)$.

To show that $\boldsymbol{r}_9 \in P_{C_5} cl(\Gamma_5^*)$, let U_i, $i \in \mathcal{N}_6$, be six independent uniform variables over a finite field $\mathbb{F}$ and

$$\mathsf{X}_1 = (\mathsf{U}_1, \mathsf{U}_6) \tag{4.23}$$

$$\mathsf{X}_2 = (\mathsf{U}_2, \mathsf{U}_4 + \mathsf{U}_5) \tag{4.24}$$

$$\mathsf{X}_3 = (\mathsf{U}_3, \mathsf{U}_5 + \mathsf{U}_6) \tag{4.25}$$

$$\mathsf{X}_4 = (\mathsf{U}_4, \mathsf{U}_1 + \mathsf{U}_5) \tag{4.26}$$

$$\mathsf{X}_5 = (\mathsf{U}_2 + \mathsf{U}_3, \mathsf{U}_3 + \mathsf{U}_5). \tag{4.27}$$

It is straightforward to verify that

$$H(\mathsf{X}_S) = \begin{cases} 2\log|\mathbb{F}|, & \text{for } S \in O_1 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_2 \\ 4\log|\mathbb{F}|, & \text{for } S \in O_3 \\ 6\log|\mathbb{F}|, & \text{for } S \in O_4 \\ 5\log|\mathbb{F}|, & \text{for } S \in O_5 \\ 6\log|\mathbb{F}|, & \text{for } S \in O_6 \\ 6\log|\mathbb{F}|, & \text{for } S \in O_7 \end{cases} \tag{4.28}$$

completing the proof that $\boldsymbol{r}_9 \in P_{C_5} cl(\Gamma_5^*)$.

By symmetry, the cases for $\boldsymbol{r}_6$ and $\boldsymbol{r}_8$ follows from that for $\boldsymbol{r}_7$ and $\boldsymbol{r}_9$, respectively. To prove the Proposition, we need to prove the achievability for this ray $(1, 2, 3/2, 5/2, 2, 5/2, 5/2)$. It seems that this is easy to see:

$$X_1 = U_1, U_3$$

$$X_2 = U_2, U_4$$

$$X_3 = U_3, U_5$$

$$X_4 = U_1, U_4$$

$$X_5 = U_5, U_2$$

We have thus completed the proof of the theorem. □

For cyclic groups C_n for $n > 5$, we can generate the hyperplanes using the code provided in the B.1.

4.4 Conclusion

We have proposed a general framework for studying $\overline{\Gamma}_n^*$ under symmetry constraints, and have obtained some initial results. In particular, it is shown that the gap between $\overline{\Gamma}_n^*$ and Γ_n may vanish under suitable projections. It is of considerable interest to develop a conceptual and systematic method for identifying such projections.

5. ON THE REPRESENTABILITY OF INTEGER POLYMATROIDS: APPLICATIONS IN LINEAR CODE CONSTRUCTION*

t has been shown that there is a duality between the linear network coding solution and the entropic vectors induced by collection of subspaces in a vector space over a finite field (dubbed linearly constructed entropic vectors). The region of all linearly constructed vectors, coincides with the set of all representable polymatroids. For any integer polymatroid, there is an associated matroid, which uniquely identifies the polymatroid. We conjecture that the representability of the underlying matroid is a sufficient condition for integer polymatroids to be linearly representable. We prove that the conjecture holds for representation over real numbers. Furthermore, we show that any real-valued submodular function (such as Shannon entropy) can be approximated (arbitrarily close) by an integer polymatroid.

5.1 Introduction

Let $f : 2^{[n]} \to \mathbb{R}$ be a real valued set function, where $[n] = \{1, 2, \ldots, n\}$. The function f is submodular if for every $S, T \subseteq [n]$

$$f(S) + f(T) \geq f(S \cup T) + f(S \cap T), \tag{5.1}$$

Submodularity has a rich combinatorial structure. Submodular functions play a key role in many combinatorial optimization problems, and have many applications in economics and engineering. In information theory, many problems are directly related to submodular function analysis, since Shannon entropy of collection of random variables is known

*

to be a submodular function. Specifically, for a collection of jointly distributed discrete random variables $\{X_1, X_2, \ldots, X_n\}$, the joint entropy of a collection of random variables $X_S := (X_i, i \in [n])$, denoted by $H(X_S, S \subseteq [n])$, is submodular. For a particular joint distribution of $(X_1, \ldots, X_n)$, the entropy of all subsets of these random variables can be expressed by a $2^n - 1$ dimensional vector $(H(X_S), S \subseteq [n])$. The region of all such vectors known as the *entropy region* and denoted by Γ_n^*. It has been shown that the closure of this region $\bar{\Gamma}^*$ is convex; however, characterization of this region for $n > 3$ is one of the well-known open problems in network information theory [12], which is closely related to the capacity region of the general network coding problem.

Many network coding capacity regions and entropy region can be upper-bounded by exploiting just the submodularity of entropy function. These upper bounds are often termed as *polymatroidal upper bounds* [12]. It has been shown that these outer bounds are not tight when $n > 3$. Many techniques exist for constructing the corresponding lower bounds. One of the most important classes, which we term as *linear construction of entropy vector* or simply *linear network coding solution*, relies on building a subspace of a vector space over a finite field, and we denote this region by Γ_n^L. However, it has been shown that linear solutions are not sufficient to characterize the entire entropy region or achieve the network coding capacity. When $n = 4$, the region Γ_n^L can be characterized by the *Ingleton inequality* and Shannon inequalities [27, 28]. The exact characterization of this region for five random variables, is given by a set of inequalities known as DFZ, together with Shannon and Ingleton inequalities [29]. In general, exact characterization of Γ_n^L is equivalent to the space of all *linearly representable integer polymatroids* [30].

From a combinatorial point of view, integer polymatroids are closely related to matroids. In this chapter, we ask the following question: Given an integer polymatroid, is it possible to infer its representability from that of a particular matroid? We conjecture that this is possible and in fact, we prove the statement for representation over $\mathbb{R}$. If the con-

jecture is true, the results concerning the representation of matroids, would be sufficient to analyze the representability of integer polymatroids. A necessary and sufficient condition for representability of integer polymatroids is given in Section 5.3. Furthermore, we show that for any $\epsilon > 0$, any submodular function f, can be approximated by a rational-valued submodular function $\hat{f}_Q$ such that for every set S, we have $|f(S) - \hat{f}_Q(S)| < \epsilon$; we refer to this as ϵ-approximation. Since any rational-valued submodular function can be considered as a properly scaled integer polymatroid, this approximation suggests that any submodular function can be approximated by an integer polymatroid. In Sections 5.4 and 5.5, we discuss the implication of these results in some network information theory problems.

5.2 Preliminaries

For a given submodular function f, we define the following distance between two arbitrary subsets of the ground set, $S, T \subseteq E$

$$D_f(S,T) = f(S) + f(T) - f(T \cup S) - f(T \cap S). \tag{5.2}$$

By definition of submodularity, $D_f(S,T) \geq 0$ for any submodular function and $D_f(S,T) \leq 0$ for any supermodular function. This distance is not an interesting object by itself, since $D_f(S,T) = 0$ for every $S \subseteq T$ (or $T \subseteq S$). However, if we take out these special subsets, the minimum value that $D_f(S,T)$ can take, becomes informative and is defined as

$$\Delta_f = \min_{S \subseteq E} \min_{i,j \in E \setminus S} |D_f(S+i, S+j)| \tag{5.3}$$

Remark 2. *It is easy to verify some of the properties of Δ_f. For example, if f and g are both submodular functions, we have*

$$\Delta_{f+g} \geq \Delta_f + \Delta_g, \tag{5.4}$$

and when f and g are both supermodular functions, we have

$$\Delta_{f+g} \leq \Delta_f + \Delta_g. \tag{5.5}$$

This particular property defines a special class of submodualr functions as follows.

Definition 1. *A submodular (supermodular) function f, defined on the ground set E, is called strictly submodular (supermodular) if $\Delta_f > 0$.*

In the following we show an example of a strictly submodular function.

Example 1. *The set function $\log(1 + |S|)$ is strictly submodular. One way to prove this is by contradiction. Assume that it is not strictly submodular and, therefore, there exist $T, S \neq \emptyset$ and $T \nsubseteq S$, $S \nsubseteq T$ such that*

$$\begin{aligned}
0 =& f(S) + f(T) - f(S \cup T) - f(S \cap T) \\
=& \log(1 + |S|) + \log(1 + |T|) - \log(1 + |S \cup T|) \\
& - \log(1 + |S \cap T|) \\
=& \log(1 + |S||T| + |S| + |T|) \\
& - \log(1 + |S \cup T||S \cap T| + |S \cup T| + |S \cap T|).
\end{aligned}$$

Since $|S| + |T| = |S \cup T| + |S \cap T|$, this implies that

$$|S \cup T||S \cap T| = |S||T| \tag{5.6}$$

Assume $|S| = |S \cup T| - x$ *and* $|T| = |S \cap T| + x$*. Therefore, we have*

$$\begin{aligned} |S||T| &= (|S \cup T| - x)(|S \cap T| + x) \\ &= |S \cap T||S \cup T| + x|S \cup T| - x|S \cap T| - x^2 \end{aligned}$$

which implies either $x = 0$ *or* $x = |S \cup T| - |S \cap T|$*. The former condition on* x *implies that* $T \subseteq S$ *and the latter implies that* $S \subseteq T$*, which contradicts our assumption.*

5.2.1 Integer Polymatroids

A *Polymatroid* P is a pair (E, f), where E is a non-empty ground set and f is a set function satisfying the following conditions:

- f is submodular: $f(S) + f(T) \geq f(S \cap T) + f(S \cup T)$, for $S, T \subseteq E$
- Nondecreasing: $f(S) \geq f(T)$, $S \supseteq T$
- Normalized: $f(\emptyset) = 0$

When f is an integer-valued set function, $P = (E, f)$ is called *integer polymatroid.*

In a way akin to representable matroids, we can define the representability of integer polymatroids as follows:

Definition 2. *An integer polymatroid* (E, f) *on the ground set* E *is representable, if there exists a collection of subspaces* V_e*,* $e \in E$*, such that for every* $S \subseteq E$*, we have* $rank(\cup_{e \in S} V_e) = f(S)$*.*

5.3 Main Results

5.3.1 Representability of Integer Polymatroids

Although integer polymatroids are interesting combinatorial objects by nature, they have a matroid structure. Moreover, it has been shown by Helgason [31], that every in-

teger polymatroid can be constructed by a matroid. Therefore, all problems in integer polymatroids are matroid problems. More specifically, let f be an integer-valued, nondecreasing submodular set function on E, with $f(\emptyset) = 0$. For each element of ground set $e \in E$, we assign a set X_e with the size of $f(\{e\})$. Now the ground set for the new matroid that we construct will be $X = \bigcup_{e \in S} X_e$, where $\bigcup$ denotes the disjoint union operation.

Theorem 4. *Helgason [31]: $\mathcal{M} = (X, r)$ is a matroid, where the rank function of a matroid is given by*

$$r(U) = \min_{T \in S}(|U \setminus \bigcup_{s \in T} X_e| + f(T)). \quad \forall U \subseteq X \tag{5.7}$$

It is easy to check that the rank function of the original integer polymatroid has not been changed. Namely,

$$f(T) = r(\bigcup_{e \in T} X_e). \tag{5.8}$$

The interesting observation here is that the integer polymatroid is defined on the ground set E, where the rank function of the matroid $r(.)$ is defined on a larger ground set than E, namely $X = \bigcup_{e \in E} X_e$ with cardinality $|X| = \sum_{e \in E} f(e)$. The construction of an integer polymatroid from matroids is not unique. In the next section, we explain the notion of extending the ground set of a polymatroid and how using this extension, it is possible to construct a matroid.

5.3.2 Extending Integer Polymatroids

Lovász [32], showed that it is possible to extend the ground set of an integer polymatroid by "adding new element e' to the element e in the ground set"[1]. Specifically, "adding

[1]Lovász [32], gave this a geometric interpretation and called "adding a point x on an element of integer polymatroid y in general position". Here we adopt his notation and for simplicity we refer to this as "adding x on an element y in the ground set of integer polymatroid".

e' to $e \in E$" means constructing a new integer polymatroid $(E \cup \{e'\}, f)$, where the value of f remains the same on the subsets of E and

$$f(T + e') = \begin{cases} f(T), & \text{if } f(T + e) = f(T) \\ f(T) + 1, & \text{if } f(T + e) > f(T). \end{cases} \tag{5.9}$$

Similarly, we can continue adding elements and eventually construct an integer polymatroid $(E \cup X, f)$, where $X = \bigcup_{e \in S} X_e$ and elements of X_e have been added to element $e \in E$. The following theorem, gives the explicit construction of a matroid.

Theorem 5. *Lovász [32]: Let $(E \cup X, f)$ be the extended polymatroid defined above. Then $\mathcal{M} = (X, r)$ is a matroid where $r(U) = f(U)$, for all $U \subseteq E$. Moreover,*

$$r(U) = \min_{T \in E}(|U \setminus \bigcup_{e \in T} X_e| + f(T)), \quad \forall U \subseteq X \tag{5.10}$$

which is identical to (5.7)*.*

Through the chapter, we will refer to this special construction of matroids as *expanding-construction*.

5.3.3 Representation of Integer Polymatroid

The following theorem gives the necessary and sufficient condition for the representability of an integer polymatroid over real numbers $\mathbb{R}$.

Theorem 6. *An integer polymatroid is representable over $\mathbb{R}$, if and only if the underlying matroid using the expanding-construction, is representable over $\mathbb{R}$.*

Proof. Assume that (E, f) is an integer polymatroid and its associated matroid using the described expanding construction is $(\cup_{e \in E} X_e, r)$. One direction trivially holds: if the underlying matroid $(\cup_{e \in E} X_e, r)$ is representable, one can take the subspace generated by

the span of the vectors associated with each X_e and therefore, by definition, the integer matroid is, indeed, representable.

To verify the other direction, we assume that there exist a collection of vector spaces V_e for every $e \in E$ and the goal is to show that the matroid derived using expanding-construction is representable. Assume that the subspaces for the integer polymatroid are given as $S = \{S_1, S_2, \ldots, S_{|E|}\}$ and define $r_i := rank(S_i)$.

The outline of the proof is as follows: Similarly to the expanding-construction, we start with the ground set S, and at each step, we "add a vector to subspace $S_i \in S$", where the definition will be made precise later. We continue adding elements and, for each $S_i \in S$, we add r_i vectors, which are denoted by $\mathcal{V}_i = \{V_1^{(i)}, \ldots, V_{r_i}^{(i)}\}$. Eventually the ground set will be $S \cup \bigcup_{i \in [|E|]} \mathcal{V}_i$ and we show that the rank function of any collection of these vectors satisfies (5.7). Therefore, we conclude that the vectors $\bigcup_{i \in [|E|]} \mathcal{V}_i$ are a linear representation of the expanding-construction matroid.

In order to complete the proof, we need to explain how we add vector $V_j^{(i)}$ to subspace S_i. Assume that we want to add a vector $V_j^{(i)}$ for $j \leq r_i$ to S_i. First, we define the following set

$$\mathcal{T}_{i,j}^* := S \cup \mathcal{V}_1 \cup \cdots \cup \mathcal{V}_{i-1} \cup \{V_1^{(i)} \ldots V_{j-1}^{(i)}\}. \tag{5.11}$$

It is easy to verify that $|\mathcal{T}_{i,j}^*| = |E| + \sum_{k=1}^{i-1} r_k + (j-1)$. We define

$$\mathcal{T}_{i,j} = \{T | T \subseteq \mathcal{T}_{i,j}^*, \quad rank(T \cup S_i) > rank(T)\}. \tag{5.12}$$

To accomplish the construction of the linear matroid, we pick a vector $V_j^{(i)}$ for $j \leq r_i$ such

that:

$$V_j^{(i)} \in S_i \setminus \bigcup_{T \in \mathcal{T}_{i,j}} span(T). \tag{5.13}$$

In order to choose the vector $V_j^{(i)}$, we need to make sure that $S_i \setminus \bigcup_{T \in \mathcal{T}_{i,j}} span(T) \neq \emptyset$.

Claim 1. *For all $i \in [|E|]$, we have $S_i \setminus \bigcup_{T \in \mathcal{T}_{i,j}} span(T) \neq \emptyset$.*

Proof. First, note that $S_i \not\subset span(T)$ for all $T \in \mathcal{T}_{i,j}$; otherwise $rank(T \cup S_i) = rank(T)$. This implies that $T \notin \mathcal{T}_{i,j}$; however, this contradicts our assumption that we started with $T \in \mathcal{T}_{i,j}$. The only possibility is $S_i \supset \bigcup_{T \in \mathcal{T}_{i,j}} span(T)$. However, since we assumed that all subspaces are in $\mathbb{R}$, we know that it is not possible to write a subspace in $\mathbb{R}$ as countable union of subspaces that do not include it. □

With this choice of vectors, once we add a new vector $V_j^{(i)}$ to the ground set of the integer polymatroid, since the chosen vector is not in the $\bigcup_{T \in \mathcal{T}_{i,j}} span(T)$, we have

$$rank(T + V_j^{(i)}) = \begin{cases} rank(T), & \text{if } T \notin \mathcal{T}_{i,j} \\ rank(T) + 1, & \text{if } T \in \mathcal{T}_{i,j}. \end{cases} \tag{5.14}$$

Without loss of generality, we continue this construction in $|E|$ steps, starting from S_1 up to $S_{|E|}$, and at each step i, we add r_i new vectors to the ground set. On the other hand, the rank function given in (5.14) is identical to (5.9) , and therefore, this proves that the collection of vectors $\mathcal{V} = \cup_{i=1}^{|E|} \mathcal{V}_i$ is a vector matroid, isomorphic to the matroid that we obtain from an expanding-construction; and this completes the proof. □

This proof cannot be directly generalized to finite fields since we used a unique property of the vector spaces in $\mathbb{R}$. Namely, a vector space over $\mathbb{R}$ cannot be decomposed

into a countable union of proper subspaces. However, we posit the following conjecture, suggested by our results,

Conjecture 1. *The integer polymatroid is representable, if and only if the underlying matroid using the expanding-construction, is representable.*

5.3.4 Approximation of Submodular Function With Rank Function of a Matroid

Recall that the rank function of an integer polymatroid is submodular; Therefore, it might seem redundant to approximate a submodular function with another submodular function. However, as we discussed in the previous sections, any scaled integer polymatroids can be constructed by matroids. Observe that when a submodular function takes rational values, it can be considered as an integer-valued submodular function with proper scaling, which is the lowest common denominator of all the function values. On the other hand, when the submodular function takes real values, this construction is impossible. However, in this case, we can approximate any submodular function (with proper scaling) by a matroid. This approximation is not only just of mathematical interest, but also useful in certain information theoretic problems.

Theorem 7. *Suppose that $f : 2^E \to \mathbb{R}$ is a real-valued submodular function over the ground set E. For every $\epsilon > 0$, there exist a polymatroid (E, f_Q) where $f_Q : 2^E \to \mathbb{Q}$, satisfying*

$$0 < f(T) - f_Q(T) < \epsilon \tag{5.15}$$

for all $T \subseteq E$.

Proof. We consider two cases; For the first case, we assume that the function is strictly submodular ($\Delta_f > 0$). For the second case, we argue that, when $\Delta_f = 0$, we can construct

a strictly submodular function $\tilde{f}$, which is properly close to the original function f, namely $f(T) - \tilde{f}(T) \leq \epsilon$, for any $T \subseteq E$.

Case 1: We consider that $f(.)$ is strictly submodular, namely $\Delta_f > 0$, and define $\epsilon^* := \min(\epsilon, \frac{\Delta_f}{2})$. Then we are guaranteed to find a rational number $f_Q(T)$ such that $f_Q(T) \in [f(T) - \epsilon^*, f(T)]$. Assume

$$f(T) - \epsilon^* \leq f_Q(T) = f(T) - \epsilon_T \leq f(T). \tag{5.16}$$

Lemma 10. *The set function f_Q defined on the ground set E is submodular.*

Proof. First, observe that, by our assumption $\Delta_f > 0$ and $\epsilon_T \leq \epsilon^*$ for every $T \subseteq E$, we have

$$\begin{aligned}
\Delta_{f_Q} =& \min_{S \subseteq E} \min_{i,j \in E \setminus S} f_Q(S+i) + f_Q(S+j) \\
& \qquad - f_Q(S) - f_Q(S+i+j) \\
\geq & \min_{S \subseteq E} \min_{i,j \in E \setminus S} f(S+i) + f(S+j) \\
& \qquad - f(S) - f(S+i+j) \\
& + \min_{S \subseteq E} \min_{i,j \in E \setminus S} \epsilon_{S+i} + \epsilon_{S+j} - \epsilon_S - \epsilon_{S+i+j} \\
\overset{a}{=} & \Delta_f + \epsilon_{S^*+i^*} + \epsilon_{S^*+j^*} - \epsilon_{S^*} - \epsilon_{S^*+i^*+j^*} \\
\geq & \Delta_f - \epsilon_{S^*} - \epsilon_{S^*+i^*+j^*} \\
\geq & \Delta_f - 2\frac{\Delta_f}{2} \\
\geq & 0
\end{aligned}$$

where S^*, i^* and j^* in (a) are the optimal solutions for the second minimization. □

Case 2: Consider that $\Delta_f = 0$; then we can construct a strictly submodular function

as follows

$$\tilde{f}(T) = f(T) - \gamma g(T), \tag{5.17}$$

where γ is small enough and $g(T)$ is a strictly supermodular function. We have the following two facts.

Fact 1. *The set function $\tilde{f}(T)$ defined over ground set E is strictly submodular.*

Fact 2. *We have $\Delta_{\tilde{f}} \geq \Delta_f + \gamma\Delta_g = \gamma\Delta_g$.*

Both are the immediate consequence of (5.4). We can choose γ to be arbitrary small, such that $\epsilon^* = \frac{\gamma\Delta_g}{2} < \epsilon$. Now with this choice of $\tilde{f}$, we have a strictly submodular function and similarly to the the previous case and (5.16), with $\tilde{f}(.)$ and ϵ^*, we are guaranteed to find rational-valued submodular function, which is an ϵ-approximation of f. This completes the proof. □

Similarly we can approximate any submodular function from above.

Corollary 11. *Suppose that $f : 2^E \to \mathbb{R}$ is a real-valued submodular function over the ground set E. For every $\epsilon > 0$, there exists a polymatroid (E, f_Q) where $f_Q : 2^E \to \mathbb{Q}$, satisfying*

$$0 < f_Q(T) - f(T) < \epsilon \tag{5.18}$$

for all $T \subseteq E$.

Proof. The proof is similar to Theorem 7, except that we need $g(T)$ to be a non-negative strictly submodular function. □

5.4 Implication in Information Theory: Fractional Network Coding Solution

We consider a network with the underlying topology as a capacitated directed acyclic graph (DAG) $((\mathcal{V}, \mathcal{A}), (C_a : a \in \mathcal{A}))$. Here, $\mathcal{V}$ and $\mathcal{A}$ are the node and the arc sets of the graph with unit edge capacities. A set of distinct nodes $\mathcal{S} \subset \mathcal{V}$ called source nodes, which have the access to a subset of message set $\mathcal{W} = \{w_1, \ldots, w_m\}$. There is also, a distinct subset of nodes $\mathcal{T} \subset \mathcal{V}$ called sink nodes. Associated with each sink node there is a subset of message set $\mathcal{W}$ as demand.

We assume that a vector of k_i symbols of message w_i for every $i \in [m]$ at a source node are encoded and a code $(n, \{\mathsf{x}_a : a \in \mathcal{A}\})$ with block length n is transmitted over the arc a.

Definition 3. *A network has $(k_1, \ldots, k_m, n)$ fractional linear solution if there exists a set of linear encoding and decoding operations at each node of the network and decoders at sink nodes, such that each sink node can perfectly decode its demanded messages.*

When $k = n = 1$, the linear network coding solution is called a scalar-linear solution. It has been shown that every scalar linear solvable network is a matroidal network. For the special case where $k = n$, the solution is called a vector linear solution. From the definition above, the following lemma is immediate [33].

Lemma 12. *If a network has a $(k_1, \ldots, k_m, n)$ fractional network coding solution over Σ, with independent messages uniformly distributed over Σ, the following hold:*

1. *For any collection of source messages $H(\cup_{i \in I} w_i) = \sum_{i \in I} k_i$, where $I \subseteq [m]$.*
2. *$H(X_a) \leq n$ for every $a \in \mathcal{A}$*
3. *$H(\cup_{a \in In(v)} X_a) = H(\cup_{a \in In(v)} X_a, \cup_{a \in Out(v)} X_a)$*

In the work by Dougherty et. al. [33], it has been shown that, if the network has a scalar linear solution over finite field Σ, with messages $w_1, \ldots, w_m$ and the links that carry the symbols $\{\mathsf{x}_a : a \in \mathcal{A}\}$), finding a scalar linear solution is equivalent to finding a mapping $T : \mathcal{W} \cup_{a\in\mathcal{A}} X_a \to E$ where E is the ground set of a *representable matroid* $\mathcal{M} = (E, r)$, such that the three conditions in Lemma 12 are satisfied with $H(\cdot) = r(\cdot)$. The converse was established by Médard and Kim [34]. Similarly, we can have the following result for fractional linear solutions.

Lemma 13. *Assume that the network has a $(k_1, \ldots, k_m, n)$ fractional linear solution over finite field Σ, with messages $w_1, \ldots, w_m$ and the links that carry the symbols $\{\mathsf{x}_a : a \in \mathcal{A}\}$). Then finding a fractional linear solution is equivalent to finding a mapping $T : \mathcal{W} \cup_{a\in\mathcal{A}} X_a \to E$, where E is the ground set of a representable integer polymatroid (E, f), such that the three conditions in Lemma 12 are satisfied with $H(\cdot) = f(\cdot)$.*

The proof is similar to the proof of Lemma 12. A similar statement has been proved in (Theorem. 3, [35]), where they introduced the notion of *discrete polymatroidal network*, and showed that the network has a fractional linear solution if and only if it is discrete polymatroidal with respect to a representable discrete polymatroid.

5.5 Implication in Information Theory: Constructing Entropic Vectors

Assume that we are given an integer vector (or rational [2]) in Γ_n. Naturally, this vector defines an integer polymatroid since entropy is a submodular function. Therefore, if it is representable over some finite field $\mathbb{F}$, we can conclude that the vector is indeed entropic and it can be constructed using linear mappings. The main problem here is that checking whether an integer polymatroid is representable, is not an easy task. However, if our conjecture is true, we can construct the associated expanded-matroid and check its representability. Therefore, one can use the extensive literature on representability of ma-

[2] A rational vector can be transformed to an integer vector with a proper scaling.

troids. Moreover, if we conclude that the underlying matroid is not representable, we can claim that nonlinear transformation (or nonlinear code in the case of network coding) is inevitable.

Furthermore, if the given vector is not integral (or rational), then using Theorem 7, we are guaranteed to find an integer polymatroid, which is arbitrarily close to the desired vector. We should then be able to study the approximated integer polymatroid to see whether it is representable or not.

5.6 Conclusion

In this chapter, we studied the representability of polymatroids. We showed that the representability of an integer polymatroid is a necessary and sufficient condition for the representability of the underlying expanding-construction matroid over reals. Moreover, we showed that it is always possible to approximate a polymatroid with an integer polymatroid.

6. CONCLUSION AND FUTURE DIRECTION

An explicit characterization of the capacity region of the general network coding problem is one of the best known open problems in information theory. A simple set of bounds that are often used in the literature to show that certain rate tuples are infeasible are based on the graph-theoretic notion of cut. The standard cut-set bounds, however, are known to be loose in general when there are multiple messages to be communicated in the network.

This dissertation focused on broadcast networks, for which the standard cut-set bounds are closely related to union as a specific set operation to combine different simple cuts of the network. A new set of explicit network coding bounds, which combine different simple cuts of the network via a variety of set operations (not just the union), were established via their connections to extremal inequalities for submodular functions. The tightness of these bounds were demonstrated via applications to combination networks.

The generalized cut-set bounds proposed in this thesis are specifically targeted at broadcast networks and are complementary to the PdE bounds in the family of cut-based network coding bounds. It is also worth mentioning that the generalized cut-set bounds proposed in this paper are a special case of the LP bounds by Yeung [12, Ch. 21] and hence are *not* tight for general broadcast network coding problems [36]. One future direction is to focus on further understanding the strength and limitations of the generalized cut-set bounds via concrete broadcast network coding problems.

This dissertation gives the polyhedral description of the latency capacity of a broadcast channel. It is shown that a comprehensive notion of the cut-set bound, so called generalized cut-set bound, characterizes entire capacity region. Furthermore, it is shown that every maximum rate vector can be achieved by a simple successive encoding scheme. Although this dissertation, shed a light on the structure of the symmetric latency capacity

polytope for a broadcast channel, the latency capacity region of the non-symmetric version remains an open problem. Tian's approach in [14] is *converse-centric* in that the forward part of the theorem is directly built on a rate splitting scheme, and the main challenge there was to prove the converse result *without* relying on a polyhedral description of the rate region. By comparison, our approach is *forward-centric* in that the converse part of the theorem follows directly from the generalized cut-set bounds (established systematically). The onus of the proof is on the forward part, where a successive encoding scheme (rather than rate splitting) is used.

It is known that there is a direct relation between network coding solutions and characterization of entropy region. Specifically, entropy inequalities play a central role in proving converse coding theorems for network information theoretic problems. This thesis also studied new aspects of entropy inequalities. First, inequalities relating average joint entropies rather than entropies over individual subsets were studied. It was shown that the closures of the average entropy regions where the averages are over all subsets of the same size and all sliding windows of the same size respectively are identical. This implies that that averaging over sliding windows always suffices as far as unconstrained entropy inequalities are concerned. Therefore, the aforementioned fact on the monotonicity of average joint entropy per element is a universal truth rather than an isolated curious observation.

Second, the existence of non-Shannon type inequalities [18] was one of the most significant discoveries in information theory during the last twenty years. Under total symmetry, however, it was known that all non-Shannon type inequalities are implied by Shannon type inequalities [12]. Mathematically, the total symmetry can be represented using the symmetry groups S_n. In the second part of this thesis, the existence of non-Shannon type inequalities under partial symmetry was studied, where the partial symmetry was represented using the subgroups of S_n. This naturally led to the notion of Shannon and

non-Shannon groups, based on which a complete classification of all permutation groups over four elements was established. With five random variables, it was shown that there are no non-Shannon type inequalities under cyclic symmetry.

There are several directions that one may consider exploring in the future. Perhaps the most straightforward extension is to consider the cyclic groups C_n for $n \geq 6$. Note that even though the cases where $n = 4$ and 5 have been resolved in this thesis, the techniques that we used rely on a "brute-force" calculation of the extreme rays of $P_{C_n}\Gamma_n$ and have a complexity that grows exponentially with n. A new representation which can further expose the structure of $P_{C_n}\Gamma_n$ may be needed in order to make progress.

Another direction of interest is to understand which partial symmetry is particularly relevant to engineering and whether non-Shannon type inequalities exist under those partial symmetry. The modern development of distributed storage systems provides several examples [22, 37] where there is symmetry built into the design principles and requirements.

Note that with symmetry not only non-Shannon type inequalities may completely disappear (dominated by the Shannon type inequalities), the number of independent Shannon type inequalities may also be substantially reduced. For example, without any symmetry the total number of independent Shannon type inequalities over n variables is

$$n + \binom{n}{2} 2^{n-2}.$$

By comparison, under total symmetry the total number of independent Shannon type inequalities over n variables is only n. Therefore, partial symmetry can potentially provide *huge* advantages when a computational approach is utilized for characterizing the fundamental limits of complex information systems [23].

Finally, it is shown shown that the representability of an integer polymatroid is a necessary and sufficient condition for the representability of the underlying expanding-construction matroid over reals. Moreover, it is always possible to approximate a polymatroid with an integer polymatroid. One interesting future direction would be investigating the conjecture we proposed, that this necessary and sufficient condition can be valid over any finite field.

REFERENCES

[1] T. M. Cover and J. A. Thomas, "Elements of information theory 2nd edition," 2006.

[2] L. Grokop and D. N. Tse, "Fundamental constraints on multicast capacity regions," *arXiv preprint arXiv:0809.2835*, 2008.

[3] C. K. Ngai and R. W. Yeung, "Network coding gain of combination networks," in *Information Theory Workshop, 2004. IEEE*. IEEE, 2004, pp. 283–287.

[4] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, no. 3, pp. 399–404, 1956.

[5] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[6] S.-Y. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[7] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking (TON)*, vol. 11, no. 5, pp. 782–795, 2003.

[8] E. Erez and M. Feder, "Capacity region and network codes for two receivers multicast with private and common data," in *Workshop on Coding, Cryptography and Combinatorics*, 2003.

[9] C. Ngai and R. Yeung, "Multisource network coding with two sinks," in *Communications, Circuits and Systems, 2004. ICCCAS 2004. 2004 International Conference on*, vol. 1. IEEE, 2004, pp. 34–37.

[10] A. Ramamoorthy and R. D. Wesel, "The single source two terminal network with network coding," *arXiv preprint arXiv:0908.2847*, 2009.

[11] G. Kramer and S. A. Savari, "Edge-cut bounds on network coding rates," *Journal of Network and Systems Management*, vol. 14, no. 1, pp. 49–67, 2006.

[12] R. W. Yeung, *Information theory and network coding*. Springer Science & Business Media, 2008.

[13] N. J. Harvey, R. Kleinberg, and A. R. Lehman, "On the capacity of information networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, no. SI, pp. 2345–2364, 2006.

[14] C. Tian, "Latent capacity region: a case study on symmetric broadcast with common messages," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3273–3285, 2011.

[15] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1059–1064, 1997.

[16] T. M. Cover, "Comments on broadcast channels," *IEEE Transactions on information theory*, vol. 44, no. 6, pp. 2524–2530, 1998.

[17] M. Goemans, "Lecture notes on linear programming and polyhedral combinatorics," Massachusetts Institute of Technology, 18.433: Combinatorial Optimazation, 2009.

[18] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1440–1452, 1998.

[19] F. Matus, "Infinitely many information inequalities," in *2007 IEEE International Symposium on Information Theory*. IEEE, 2007, pp. 41–44.

[20] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1059–1064, 1997.

[21] R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 609–621, 1999.

[22] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, 2010.

[23] C. Tian, "Rate region of the (4, 3, 3) exact-repair regenerating codes," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 1426–1430.

[24] J. Jiang, N. Marukala, and T. Liu, "Symmetrical multilevel diversity coding and subset entropy inequalities," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 84–103, 2014.

[25] Z. Xiao, J. Chen, Y. Li, and J. Wang, "Distributed multilevel diversity coding," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6368–6384, 2015.

[26] Q. Chen and R. W. Yeung, "Partition-symmetrical entropy functions," *arXiv preprint arXiv:1407.7405*, 2014.

[27] A. W. Ingleton, "Representation of matroids," *Combinatorial Mathematics and Its Applications*, vol. 23, 1971.

[28] D. Hammer, A. E. Romashchenko, A. Shen, and N. K. Vereshchagin, "Inequalities for shannon entropies and kolmogorov complexities," in *Computational Complexity, 1997. Proceedings., Twelfth Annual IEEE Conference on (Formerly: Structure in Complexity Theory Conference)*. IEEE, 1997, pp. 13–23.

[29] R. Dougherty, C. Freiling, and K. Zeger, "Linear rank inequalities on five or more variables," *arXiv preprint arXiv:0910.0284*, 2009.

[30] T. Chan, A. Grant, and D. Pfluger, "Truncation technique for characterizing linear polymatroids," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6364–6378, 2011.

[31] T. Helgason, "Aspects of the theory of hypermatroids," in *Hypergraph Seminar*. Springer, 1974, pp. 191–213.

[32] L. Lovász, "Flats in matroids and geometric graphs," *Combinatorial surveys*, pp. 45–86, 1977.

[33] R. Dougherty, C. Freiling, and K. Zeger, "Networks, matroids, and non-shannon information inequalities," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 1949–1969, 2007.

[34] A. Kim and M. Médard, "Scalar-linear solvability of matroidal networks associated with representable matroids," in *2010 6th International Symposium on Turbo Codes & Iterative Information Processing*. IEEE, 2010, pp. 452–456.

[35] V. T. Muralidharan and B. S. Rajan, "Linear network coding, linear index coding and representable discrete polymatroids," *arXiv preprint arXiv:1306.1157*, 2013.

[36] T. Chan and A. Grant, "Mission impossible: Computing the network coding capacity region," in *2008 IEEE International Symposium on Information Theory*. IEEE, 2008, pp. 320–324.

[37] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A survey on network codes for distributed storage," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 476–489, 2011.

APPENDIX A

SOME PROOFS FOR SECTION 2

A.1 Proof of Lemma 1

Fix two integers r' and J such that $0 < r' < J \leq K$. Let

$$T_r := \begin{cases} \emptyset, & \text{for } r = 1, \ldots, r' \\ S^{(r'+1)}([r]), & \text{for } r = r'+1, \ldots, J, \end{cases} \tag{A.1}$$

and let $G_r := S_r \cup T_r$ for $r = 1, \ldots, J$. By the standard multiway submodularity (2.9) and modularity (2.10) we have

$$\sum_{r=1}^{r'} f(S_r) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r])) = \sum_{r=1}^{J} f(G_r) \geq \sum_{r=1}^{J} f(G^{(r)}([J])) \tag{A.2}$$

if f is a submodular function, and

$$\sum_{r=1}^{r'} f(S_r) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r])) = \sum_{r=1}^{J} f(G_r) = \sum_{r=1}^{J} f(G^{(r)}([J])) \tag{A.3}$$

if f is a modular function. Next, we shall show that

$$G^{(r)}([J]) = \begin{cases} S^{(r)}([J]), & \text{for } r = 1, \ldots, r' \\ S^{(r'+1)}([J - r + r' + 1]), & \text{for } r = r'+1, \ldots, J. \end{cases} \tag{A.4}$$

We shall consider the following two cases separately.

Case 1: $r \in [r']$. Note that $S_r \subseteq G_r$ for any $r \in [J]$, so we have $S^{(r)}([J]) \subseteq G^{(r)}([J])$ for any $r \in [J]$. On the other hand, since $T_r \subseteq S^{(r'+1)}([J])$ for all $r \in [J]$, we have

$G_r \subseteq S_r \cup S^{(r'+1)}([J])$ and hence $G^{(r)}([J]) \subseteq S^{(r)}([J]) \cup S^{(r'+1)}([J])$ for all $r \in [J]$. Since $S^{(r)}([J]) \supseteq S^{(r'+1)}([J])$ for all $r \in [r']$, we have $G^{(r)}([J]) \subseteq S^{(r)}([J])$ for all $r \in [r']$. We thus conclude that $G^{(r)}([J]) = S^{(r)}([J])$ for all $r \in [r']$.

Case 2: $r \in \{r'+1, \ldots, J\}$. For this case, we have the following fact.

Fact 3. *For any $r \in \{r'+1, \ldots, J\}$, we have*

$$G^{(r)}([J]) = \cup_{m=1}^{\min\{r, r'+2\}} \left(S^{(m-1)}([J-r+m-1]) \cap T_{J-r+m} \right). \tag{A.5}$$

Proof. Fix $r \in \{r'+1, \ldots, J\}$. By definition,

$$G^{(r)}([J]) = \cup_{\{U \subseteq [J]: |U|=r\}} \cap_{k \in U} G_k. \tag{A.6}$$

Fix $U \subseteq [J]$ such that $|U| = r$. We have

$$\begin{aligned}
\cap_{k \in U} G_k &= \cap_{k \in U} (S_k \cup T_k) && \text{(A.7)} \\
&= \cup_{U' \subseteq U} \left((\cap_{k \in U'} S_k) \cap (\cap_{k \in U \setminus U'} T_k) \right) && \text{(A.8)} \\
&= \left(\cup_{U' \subset U} \left((\cap_{k \in U'} S_k) \cap T_{\bar{k}(U')} \right) \right) \cup (\cap_{k \in U} S_k) && \text{(A.9)}
\end{aligned}$$

where $\bar{k}(U')$ is the *smallest* integer in $U \setminus U'$, and (A.9) follows from the fact that

$$T_1 \subseteq T_2 \subseteq \cdots \subseteq T_J. \tag{A.10}$$

Write, without loss of generality, that $U = \{u_1, \ldots, u_r\}$ where $1 \leq u_1 < u_2 < \cdots < u_r \leq J$. Fix $\bar{k}(U') = u_m$ for some $m \in [r]$. Then we must have $U' \supseteq \{u_1, \ldots, u_{m-1}\}$ for

any such U'. We thus have from (A.9) that

$$\cap_{k\in U} G_k = \left(\cup_{m=1}^{r} \left((\cap_{l=1}^{m-1} S_{u_l}) \cap T_{u_m}\right)\right) \cup (\cap_{l=1}^{r} S_{u_l}) \,. \tag{A.11}$$

The right-hand side of (A.11) can be further simplified based on the following two observations. First, for any $r \in \{r'+1, \ldots, J\}$ we have $u_r \geq r \geq r'+1$ and hence

$$T_{u_r} = S^{(r'+1)}([u_r]) \supseteq \cap_{l=1}^{r'+1} S_{u_l} \supseteq \cap_{l=1}^{r} S_{u_l}. \tag{A.12}$$

We thus have

$$\cap_{l=1}^{r} S_{u_l} \subseteq (\cap_{l=1}^{r-1} S_{u_l}) \cap T_{u_r} \tag{A.13}$$

and hence

$$\cap_{k\in U} G_k = \cup_{m=1}^{r} \left((\cap_{l=1}^{m-1} S_{u_l}) \cap T_{u_m}\right) . \tag{A.14}$$

Second, since $u_{r'+2} \geq r'+2$, we have

$$\cap_{l=1}^{r'+1} S_{u_l} \subseteq S^{(r'+1)}([u_{r'+2}]) = T_{u_{r'+2}} \tag{A.15}$$

and hence

$$(\cap_{l=1}^{r'+1} S_{u_l}) \cap T_{u_{r'+2}} = \cap_{l=1}^{r'+1} S_{u_l}. \tag{A.16}$$

It follows that for any $m \geq r'+2$, we have

$$(\cap_{l=1}^{m-1} S_{u_l}) \cap T_{u_m} \subseteq \cap_{l=1}^{r'+1} S_{u_l} = (\cap_{l=1}^{r'+1} S_{u_l}) \cap T_{u_{r'+2}}. \tag{A.17}$$

Substituting (A.17) into (A.14), we have

$$\cap_{k\in U} G_k = \cup_{m=1}^{\min\{r,r'+2\}} \left((\cap_{l=1}^{m-1} S_{u_l}) \cap T_{u_m} \right). \tag{A.18}$$

Finally, substituting (A.18) into (A.6), we have

$$G^{(r)}([J]) = \cup_{\{U\subseteq[J]:|U|=r\}} \left(\cup_{m=1}^{\min\{r,r'+2\}} \left((\cap_{l=1}^{m-1} S_{u_l}) \cap T_{u_m} \right) \right) \tag{A.19}$$

$$= \cup_{m=1}^{\min\{r,r'+2\}} \left(\cup_{\{U\subseteq[J]:|U|=r\}} \left((\cap_{l=1}^{m-1} S_{u_l}) \cap T_{u_m} \right) \right) \tag{A.20}$$

for any $r \in \{r'+1, \ldots, J\}$. Note that for any $U \subseteq [J]$ such that $|U| = r$, the largest numerical value that u_m can assume is $J - r + m$ for any $m \in [r]$. By the ordering in (A.10), for any $m = 1, \ldots, r$ we have

$$\cup_{\{U\subseteq[J]:|U|=r\}} \left((\cap_{l=1}^{m-1} S_{u_l}) \cap T_{J-r+m} \right) = \left(\cup_{\{1\le u_1<u_2<\cdots<u_{m-1}\le J-r+m-1\}} \cap_{l=1}^{m-1} S_{u_l} \right) \cap T_{J-r+m} \tag{A.21}$$

$$= S^{(m-1)}([J-r+m-1]) \cap T_{J-r+m}. \tag{A.22}$$

Substituting (A.22) into (A.20) completes the proof of the fact. □

Further note that for any $r \in \{r'+1, \ldots, J\}$ we have

$$T_{J-r+m} \subseteq S^{(r'+1)}([J-r+m]) \subseteq S^{(r')}([J-r+m-1]) \subseteq S^{(m-1)}([J-r+m-1]) \tag{A.23}$$

for any $2 \le m \le r'+1$. When $r = r'+1$, substituting (A.10) and (A.23) into Fact 3 we

have

$$G^{(r)}([J]) = \cup_{m=1}^{r} T_{J-r+m} = T_J = S^{(r'+1)}([J]). \tag{A.24}$$

When $r \in \{r'+2, \ldots, J\}$, by Fact 3 we have

$$G^{(r)}([J]) = \cup_{m=1}^{r'+2} \left(S^{(m-1)}([J-r+m-1]) \cap T_{J-r+m} \right) \tag{A.25}$$

$$= \left(\cup_{m=1}^{r'+1} \left(S^{(m-1)}([J-r+m-1]) \cap T_{J-r+m} \right) \right) \cup \left(S^{(r'+1)}([J-r+r'+1]) \cap T_{J-r+r'+2} \right) \tag{A.26}$$

$$= \left(\cup_{m=1}^{r'+1} T_{J-r+m} \right) \cup \left(S^{(r'+1)}([J-r+r'+1]) \cap T_{J-r+r'+2} \right) \tag{A.27}$$

$$= T_{J-r+r'+1} \cup \left(S^{(r'+1)}([J-r+r'+1]) \cap T_{J-r+r'+2} \right) \tag{A.28}$$

$$= S^{(r'+1)}([J-r+r'+1]) \cup \left(S^{(r'+1)}([J-r+r'+1]) \cap S^{(r'+1)}([J-r+r'+2]) \right) \tag{A.29}$$

$$= S^{(r'+1)}([J-r+r'+1]) \tag{A.30}$$

where (A.27) follows from (A.23), and (A.28) follows from the ordering in (A.10). Combining (A.24) and (A.30) completes the proof of (A.4) for $r \in \{r'+1, \ldots, J\}$.

Finally, substituting (A.4) into (A.2) and (A.3) we have

$$\sum_{r=1}^{r'} f(S_r) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r])) \geq \sum_{r=1}^{r'} f(S^{(r)}([J])) + \sum_{r=r'+1}^{J} f(S^{(r'+1)}([J-r+r'+1])) \tag{A.31}$$

$$= \sum_{r=1}^{r'} f(S^{(r)}([J])) + \sum_{r=r'+1}^{J} f(S^{(r'+1)}([r])) \tag{A.32}$$

if f is a submodular function, and

$$\sum_{r=1}^{r'} f(S_r) + \sum_{r=r'+1}^{J} f(S_r \cup S^{(r'+1)}([r])) = \sum_{r=1}^{r'} f(S^{(r)}([J])) + \sum_{r=r'+1}^{J} f(S^{(r'+1)}([r])) \tag{A.33}$$

if f is a modular function. This completes the proof of Lemma 1.

A.2 Proof of Lemma 3

Without loss of generality, we may assume that $T = [|T|]$ such that $t_r = r$ for all $r = 1, \ldots, |T|$. Under this assumption, the inequality (2.21) can be written as

$$\begin{aligned}
&\sum_{r=1}^{|T|} f(S_r) + r_q f(S^{(q)}(U)) \\
&\qquad \geq \sum_{r=1}^{r_q} \big(f(S^{(r)}(T)) + f(S_r \cap S^{(q)}(U)) \big) + \sum_{r=r_q+1}^{|T|} f(S_r \cap (S^{(q)}(U) \cup S^{(r_q+1)}([r]))).
\end{aligned} \tag{A.34}$$

Assume that f is a modular function. By the two-way submodularity (2.4) we have

$$\begin{aligned}
&\sum_{r=1}^{|T|} f(S_r) + r_q f(S^{(q)}(U)) \\
&\quad = \sum_{r=1}^{r_q} \big(f(S_r) + f(S^{(q)}(U))\big) + \sum_{r=r_q+1}^{|T|} \big(f(S_r) + f(S^{(q)}(U) \cup S^{(r_q+1)}([r]))\big) - \\
&\qquad \sum_{r=r_q+1}^{|T|} f(S^{(q)}(U) \cup S^{(r_q+1)}([r])) \qquad \text{(A.35)} \\
&\quad \geq \sum_{r=1}^{r_q} \big(f(S_r \cap S^{(q)}(U)) + f(S_r \cup S^{(q)}(U))\big) + \\
&\qquad \sum_{r=r_q+1}^{|T|} \big(f(S_r \cap (S^{(q)}(U) \cup S^{(r_q+1)}([r]))) + f(S_r \cup (S^{(q)}(U) \cup S^{(r_q+1)}([r])))\big) - \\
&\qquad \sum_{r=r_q+1}^{|T|} f(S^{(q)}(U) \cup S^{(r_q+1)}([r])). \qquad \text{(A.36)}
\end{aligned}$$

Applying Corollary 2 with $r' = r_q$, $J = |T|$, and $S_0 = S^{(q)}(U)$, we have

$$\begin{aligned}
&\sum_{r=1}^{r_q} f(S_r \cup S^{(q)}(U)) + \sum_{r=r_q+1}^{|T|} f(S_r \cup S^{(r_q+1)}([r]) \cup S^{(q)}(U)) \\
&\qquad \geq \sum_{r=1}^{r_q} f(S^{(r)}(T) \cup S^{(q)}(U)) + \sum_{r=r_q+1}^{|T|} f(S^{(r_q+1)}([r]) \cup S^{(q)}(U)) \qquad \text{(A.37)} \\
&\qquad = \sum_{r=1}^{r_q} f(S^{(r)}(T)) + \sum_{r=r_q+1}^{|T|} f(S^{(r_q+1)}([r]) \cup S^{(q)}(U)) \qquad \text{(A.38)}
\end{aligned}$$

where (A.38) follows from the assumption $S^{(r_q)}(T) \supseteq S^{(q)}(U)$ such that $S^{(r)}(T) \supseteq S^{(q)}(U)$ for any $r = 1, \ldots, r_q$. Substituting (A.38) into (A.36) completes the proof of (A.34) and hence that of (2.21).

When f is a modular function, both inequalities (A.36) and (A.37) hold with an equality. This completes the proof of (2.22) and hence that of the entire corollary.

A.3 Proof of Corollary 4

Note that when $Q = \emptyset$, $\beta_Q(r) = 1$ for all $r \in [|U|]$. In this case, the corollary follows directly from (2.86). Now, assume that Q is nonempty. Write, without loss of generality, that $Q = \{q_1, \ldots, q_{|Q|}\}$ where

$$1 =: q_0 < q_1 < q_2 < \cdots < q_{|Q|} \leq |U|. \tag{A.39}$$

Note that

$$\sum_{q \in Q} \sum_{r=1}^{q-1} \alpha_Q(q, r) R(I^{(r)}(U)) = \sum_{r=1}^{q_{|Q|}-1} \beta'_Q(r) R(I^{(r)}(U)) \tag{A.40}$$

where

$$\beta'_Q(r) = \sum_{l=m}^{|Q|} \alpha_Q(q_l, r) \tag{A.41}$$

for any $q_{m-1} \leq r < q_m$ for some $m \in [|Q|]$. When $r = q_m$ for some $m \in [|Q| - 1]$, by (2.87) and (A.41) we have $\alpha_Q(q_l, r) = 0$ for any $l = m, \ldots, |Q|$ and hence

$$\beta'_Q(r) = 0. \tag{A.42}$$

When $q_{m-1} < r < q_m$ for some $m \in [|Q|]$, by (2.87) and (A.41) we have

$$\alpha_Q(q_l, r) = \frac{\prod_{t=1}^{m-1}(q_t - 1) \prod_{t=m}^{l-1} q_t}{\prod_{t=1}^{l}(q_t - 1)} \tag{A.43}$$

for any $l = m, \ldots, |Q|$ and hence

$$\beta'_Q(r) = \sum_{l=m}^{|Q|} \frac{\prod_{t=1}^{m-1}(q_t - 1)\prod_{t=m}^{l-1} q_t}{\prod_{t=1}^{l}(q_t - 1)} \tag{A.44}$$

$$= \frac{\prod_{t=1}^{m-1}(q_t - 1)}{\prod_{t=1}^{|Q|}(q_t - 1)} \sum_{l=m}^{|Q|} \left(\prod_{t=m}^{l-1} q_t \prod_{t=l+1}^{|Q|} (q_t - 1) \right) \tag{A.45}$$

$$= \frac{\prod_{t=1}^{m-1}(q_t - 1)}{\prod_{t=1}^{|Q|}(q_t - 1)} \sum_{l=m}^{|Q|} \left((q_l - (q_l - 1)) \prod_{t=m}^{l-1} q_t \prod_{t=l+1}^{|Q|} (q_t - 1) \right) \tag{A.46}$$

$$= \frac{\prod_{t=1}^{m-1}(q_t - 1)}{\prod_{t=1}^{|Q|}(q_t - 1)} \sum_{l=m}^{|Q|} \left(\prod_{t=m}^{l} q_t \prod_{t=l+1}^{|Q|} (q_t - 1) - \prod_{t=m}^{l-1} q_t \prod_{t=l}^{|Q|} (q_t - 1) \right) \tag{A.47}$$

$$= \frac{\prod_{t=1}^{m-1}(q_t - 1)}{\prod_{t=1}^{|Q|}(q_t - 1)} \left(\prod_{t=m}^{|Q|} q_t - \prod_{t=m}^{|Q|} (q_t - 1) \right) \tag{A.48}$$

$$= \frac{\prod_{t=1}^{m-1}(q_t - 1)\prod_{t=m}^{|Q|} q_t}{\prod_{t=1}^{|Q|}(q_t - 1)} - 1 \tag{A.49}$$

$$= \frac{\beta_Q(r)}{\prod_{t=1}^{|Q|}(q_t - 1)} - 1, \tag{A.50}$$

where (A.50) follows from the fact that

$$\beta_Q(r) = \prod_{t=1}^{m-1}(q_t - 1) \prod_{t=m}^{|Q|} q_t, \quad \forall q_{m-1} < r < q_m \tag{A.51}$$

by the definition (2.89) of $\beta_Q(r)$.

By (A.40), (A.42), and (A.50), the left-hand side of (2.86) can be simplified as

$$\sum_{r\in[|U|]\setminus Q} R(I^{(r)}(U)) + \sum_{q\in Q}\sum_{r=1}^{q-1} \alpha_Q(q,r) R(I^{(r)}(U))$$

$$= \sum_{r\in[|U|]\setminus Q} R(I^{(r)}(U)) + \sum_{r=1}^{q_{|Q|}-1} \beta'_Q(r) R(I^{(r)}(U)) \tag{A.52}$$

$$= \sum_{r\in[|U|]\setminus Q} R(I^{(r)}(U)) + \sum_{r\in[q_{|Q|}]\setminus Q} \left(\frac{\beta_Q(r)}{\prod_{t=1}^{|Q|}(q_t-1)} - 1 \right) R(I^{(r)}(U)) \tag{A.53}$$

$$= \frac{1}{\prod_{t=1}^{|Q|}(q_t-1)} \left(\sum_{r\in[q_{|Q|}]\setminus Q} \beta_Q(r) R(I^{(r)}(U)) + \left(\prod_{t=1}^{|Q|}(q_t-1) \right) \sum_{r=q_{|Q|}+1}^{|U|} R(I^{(r)}(U)) \right) \tag{A.54}$$

$$= \frac{1}{\prod_{t=1}^{|Q|}(q_t-1)} \left(\sum_{r=1}^{q_{|Q|}} \beta_Q(r) R(I^{(r)}(U)) + \sum_{r=q_{|Q|}+1}^{|U|} \beta_Q(r) R(I^{(r)}(U)) \right) \tag{A.55}$$

$$= \frac{1}{\prod_{t=1}^{|Q|}(q_t-1)} \sum_{r=1}^{|U|} \beta_Q(r) R(I^{(r)}(U)), \tag{A.56}$$

where (A.55) follows from the facts that $\beta_Q(r) = 0$ for all $r \in Q$ and that

$$\beta_Q(r) = \prod_{t=1}^{|Q|}(q_t - 1), \quad \forall r \geq q_{|Q|} + 1 \tag{A.57}$$

by the definition (2.89) of $\beta_Q(r)$.

Similarly, the right-hand side of (2.86) can be simplified as

$$\sum_{r\in[|U|]\setminus Q} C(A^{(r)}(U)) + \sum_{q\in Q}\sum_{r=1}^{q-1} \alpha_Q(q,r) C(A^{(r)}(U)) = \frac{1}{\prod_{t=1}^{|Q|}(q_t-1)} \sum_{r=1}^{|U|} \beta_Q(r) C(A^{(r)}(U)). \tag{A.58}$$

Substituting (A.56) and (A.58) into (2.86) and multiplying both sides of the inequality by

$\prod_{t=1}^{|Q|}(q_t - 1)$ complete the proof of Corollary 4.

APPENDIX B

PYTHON CODE FOR CYCLIC GROUP

B.1 Python Code For Generating Inequalities For Projection Under Cyclic Group

```
import numpy as np
import csv
import itertools

## SET VARIABLES
#n is number of random variables
n=5
fieldsize=n
##for i in range(1,n+1):
##    for j in range(i+1,n+1):
##        print str(i)+str(j)

def cycshift(strelement,fieldsize):
    l=str()
    element=list(strelement)
    for i in range(len(element)):
        if int(element[i])==fieldsize:
            element[i]=1
        else:
            element[i]=int(element[i])+1
        element.sort()
    for i in element:
```

```
        l+=str(i)
    return l


""" The next function, strcheck(s,liststr) is to check
      if the elements of s is in the list of liststr; if so i
      t will put 1 in that element of the list other wise 0"""

def strcheck(s,liststr):
    lists=list(s)
    output=[]
    for i in range(len(liststr)):

        """preset that elements of 's' are not in the list element i
        temp=1

        for j in range(len(s)):
            if lists[j] in liststr[i]:
                temp*=1
            else:
                temp*=0
        output.append(temp)
    return output



print strcheck('er',['qer','ety','eq','re'])
```

```
def union(a, b):
    """ return the union of two lists """
    lsa=list(a)
    lsb=list(b)
    ls=list(set(lsa) | set(lsb))
    intls=[int(l) for l in ls]
    intls.sort()
    g=str()
    for l in intls:
        g+=str(l)
    return g



"""the output shows that if string element
        is in the lists of the listlist(which is a list of list)"""
'''findmx('1',[['31'],['1','13']])
 returns [1, 1]'''
def findmx(element,listlist):
    outls=[]
    for i in range(len(listlist)):
        if sum(strcheck(element,listlist[i]))>=1:
            outls.append(1)
        else:
            outls.append(0)
```

```
    return outls


def findx(element,listlist):
    outls=[]
    for i in range(len(listlist)):
        if sum(strcheck(element,listlist[i]))==1:
            out= i
            break;
        else:
            out=-1
    return out

##
##        else:
##            return -1
##

    ## Code to print combinations-----------------------------------
#print cycshift('125',5)
l=str()
initlist=[]
for i in range(1,n+1):
    initlist.append(str(i))
    #l+=str(i)
    print l
Hlist=[]
```

```
wind=1
temp=[]
Comblist=[[],initlist]


"""""first make the first element of the combination as a string"""""
for e in range(1,wind+1):
    l+=str(e)

itr=1

while True:
        temp=[]
        for i in Comblist[itr]:
          #print i
          for j in range(int(list(i).pop())+1,n+1):

            temp.append(i+str(j))

        if temp==[]:
            break;
```

```
            Comblist.append(temp)
            #print Comblist
            itr+=1
print Comblist
##-------------------------------------------------------------------
# next is to alternate in elements to get all combinations


##-----Improving the combination to Print Orbits-------
HOrbits=[[],[[],initlist]]
ct=0
for i in range(2,len(Comblist)):
    Comblist_i=Comblist[i][:]
    Comblist_i.reverse()
    HOrbits.append([[]])

    while Comblist_i!=[]:
        temp=[]
        newelement=Comblist_i.pop()



        temp.append(newelement)
        #newelement=cycshift(newelement,n)
        flg=True
        while True:

            shifel=cycshift(newelement,n)
```

```
            if ( shifel in Comblist_i):
                newelement=shifel
                temp.append(shifel)
                Comblist_i.remove(shifel)
            else:
                break;
        HOrbits[i].append(temp)

print HOrbits

print findmx('12',HOrbits[3])

### ------------ Print Inequalities ---------------
l=0
orbitleaders=[]
for i in HOrbits:

    """Here I just get one element from each orbit as"""
    if i!=[]:
        orbitleaders.append([item[0] for item in i if item!=[]])
                                    #Orbit leaser or coset.
                                    # Remember that i[0]=[]
        l+=len(i)-1


print orbitleaders
temp_ineq=[0]*l
```

```
newinequality=temp_ineq[:]
inequlaities=[]
                    ##_____H(X_i|X_{rest})>=0---------
newinequality[len(temp_ineq)-1]=1
newinequality[len(temp_ineq)-2]=-1

inequlaities.append(newinequality)
print inequlaities
newinequality=temp_ineq[:]

                    ##---I(X_i,X_j)=>0--------------
for i in range(1,len(HOrbits[2])):
    newinequality[0]=1
    newinequality[i]=-1
    inequlaities.append(newinequality)
    newinequality=temp_ineq[:]
print inequlaities

                    ##---I(X_i,X_j|X_k,...)

p=orbitleaders[:]
Comblist.remove([])
HOrbits.remove([])
ind=0
for i in range(len(p)-2):
    '''here we choose each element of orbitleaders'''
```

```
    ind+=len(p[i])
    for j in range(len(p[i])):
        UP=[]
        for k in range(len(strcheck(p[i][j],Comblist[i+1]))):
            if strcheck(p[i][j],Comblist[i+1])[k]!=0:
                UP.append(Comblist[i+1][k])
            #UP contains elements with one more letter
            for item1 in UP:
                for item2 in UP:

                    if (item1!=item2):
                        newinequality=temp_ineq[:]
                        newinequality[ind+findx(item1,HOrbits[i+1])-1]+=1
                        newinequality[ind+findx(item2,HOrbits[i+1])-1]+=1
                        newinequality[ind+len(p[i+1])
                        +findx(union(item1,item2),HOrbits[i+2])-1]=-1
                        newinequality[ind-len(p[i])+j]=-1
                        if newinequality not in inequlaities:

                            inequlaities.append(newinequality)

print inequlaities
```