2013

Network coding for multiple unicast over directed acyclic networks

Shurui Huang
Iowa State University

Network coding for multiple unicast over directed acyclic networks

by

Shurui Huang

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Program of Study Committee:
Aditya Ramamoorthy, Major Professor
Zhengdao Wang
Lei Ying
Nicola Elia
Dan Nordman

Iowa State University

Ames, Iowa

2013

DEDICATION

I would like to dedicate this dissertation to my husband Hao Chen, my daughter Adalyn Chen, and my parents Zhenye Huang and Lihua Ma. Without my husband and my parents' support I would not have been able to complete this work.

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vi

ACKNOWLEDGEMENTS viii

ABSTRACT x

CHAPTER 1. INTRODDUCTION 1

1.1 Network coding for multiple unicast 2

1.2 Dissertation outline 4

CHAPTER 2. BACKGROUND AND RELATED WORK 6

CHAPTER 3. NETWORK CODING FOR THREE UNICAST SESSIONS 10

3.1 Preliminaries 10

3.2 Network coding for three unicast sessions - Infeasible instances 13

3.3 Network coding for three unicast sessions - Feasible instances 16

3.3.1 Code assignment procedure for instances with connectivity level [1 3 3] 19

3.3.2 Code assignment procedure for instances with connectivity level [2 2 4] 21

3.3.3 Code assignment procedure for instances with connectivity level [1 2 5] 27

3.4 Simulation results 40

3.5 Conclusions 42

CHAPTER 4. NETWORK CODING FOR TWO UNICAST SESSIONS . 44

4.1 System model 44

4.2 Achievable rate region for given $k_{12-1}, k_{12-2}, k_{1-1}, k_{2-2}, k_{1-2}$, and k_{2-1} 45

4.2.1 Low interference case - $k_{1-2} + k_{2-1} \leq \min(k_{12-1}, k_{12-2})$ 49

4.2.2 High interference case - $k_{1-2} + k_{2-1} > \min(k_{12-1}, k_{12-2})$ 54

4.2.3 Increasing the achievable rate region by modifying the graph 59

4.3 Achievable rate region for given $k_{1-12}, k_{2-12}, k_{1-1}, k_{2-2}, k_{1-2}$, and k_{2-1} 60

4.4 Comparison with existing results 61

4.5 Conclusions 64

CHAPTER 5. CONCLUSIONS AND FUTURE WORKS 65

5.1 Contributions 65

5.2 Future work 66

APPENDIX A. PROOF OF LEMMA 3.2.4 67

APPENDIX B. CLAIM B.0.1 71

APPENDIX C. LEMMA C.0.2 72

APPENDIX D. LEMMA D.0.3 73

APPENDIX E. LEMMA E.0.4 74

APPENDIX F. LEMMA F.0.5 75

BIBLIOGRAPHY 76

LIST OF TABLES

Table 3.1 Proportions of networks with differences and performance improvement 42

Table 4.1 dimension and rank of matrices 47

LIST OF FIGURES

Figure 1.1 The butterfly network, where there is no routing solution but there exists a network coding solution. 2

Figure 3.1 (a) An example of [2 2 2] connectivity network without a network coding solution. (b) An example of [1 1 3] connectivity network without a network coding solution. 13

Figure 3.2 An example of [2 3] connectivity network, rate $\{2, 1\}$ cannot be supported. 14

Figure 3.3 (a) An instance of network where there are several pairs of neighboring overlap segments. E_1 and E_3 are neighboring overlap segments along P_{21}, E_1 and E_2 are neighboring overlap segments along P_{12}. E_1 and E_4 are not overlap segments along any paths. (b) A network with connectivity level [2 4] and rate $\{2, 1\}$. The coloring of the different paths helps us to show that a linear network coding solution exists. . . 23

Figure 3.4 (a) Subgraph G' when P'_{11} overlap with P'_{21}. (b) Subgraph G' when P'_{11} overlap with both P'_{21} and P'_{22}. 29

Figure 3.5 Possible subgraphs G' when P'_{11} does not overlap with either P'_{21} or P'_{22}. 34

Figure 3.6 Figures (a) and (b) denote possible subgraphs $\tilde{G}$ obtained after the graph modification procedure for G. Figure (c) shows an example of the overlap between the red $s_3 - t_3$ paths and P'_{22}. 38

Figure 3.7 a) Level-1 network. b) Level-2 network. c) Level-3 network. d) Level-4 network. 40

Figure 4.1 (a) An example of C_{t_1} and C_{t_2} when the multicast region shaded is pentagonal. (b) Another example where the multicast region is rectangular. 46

Figure 4.2 The achievable rate region for the low interference case. For each point in the shaded grey area, both terminals can recover both the sources. In the hatched grey area and the hatched white area, for a given rate point, its x-coordinate is the rate for $s_1 - t_1$ and its y-coordinate is the rate for $s_2 - t_2$; the terminals are not guaranteed to decode both sources in this region. The union of the hatched white region, the hatched gray region and the gray region is the final extended rate region for the low interference case. 54

Figure 4.3 (a) High interference case where $k_{1-1} \leq k_{1-2}$ and $k_{2-2} \leq k_{2-1}$. (b) High interference case where $k_{1-1} \geq k_{1-2}$ and $k_{2-2} \leq k_{2-1}$. 55

Figure 4.4 (a) The extended rate region for the high interference case from point Q_1. (b) The final extended rate region for the case of high interference. 58

Figure 4.5 An example of a network where a larger achievable rate region can be achieved by removing edges e_1 and e_2. 60

Figure 4.6 An example of a network where the achievable rate regions are different using the original result and the reversibility result. All edges are unit capacity. 62

Figure 4.7 An example of a high interference network when our scheme can achieve a higher rate pair compared to many other schemes. 64

Figure A.1 An example where t_1 at decode at rate n_1, but t_2 cannot decode at rate $n_2 - 3n_1/2 + 1$. 68

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this dissertation. First and foremost, Dr. Aditya Ramamoorthy for his guidance, patience and support throughout this research and the writing of this dissertation. He gave me the opportunity to join the group in the most difficult time of my life. His broad knowledge of various communication systems, deep understanding of network coding, strong mathematical skills guided me into the wonderful research world. Without his encouragement and help, I could never complete my graduate education and start a professional career.

I would like to thank my committee members Dr. Zhengdao Wang, Dr. Nicola Elia, Dr. Lei Ying, and Dr. Dan Nordman for evaluating my dissertation and helping me improve my research work. All the knowledge that I have learned from you broaden my horizons and will guide me throughout my career pathway.

Special thanks to Dr. Namrata Vaswani and Dr. Dionysios Aliprantis. The rewarding experiences of being the teaching assistant in their classes have made me a more responsible individual.

I thank all the past and present members of Communication Systems and Networks Laboratory: Dr. Naveen Kumar, Dr. Cuizhu Shi, Dr. Shizheng Li, Wenyu Wang, Ziqi Zhang, Zhao Song, Ardhendu S Tripathy. I have learned a lot from these members through group meetings and insightful discussions. The friendly atmosphere in the group has always made my life more enjoyable.

I was able to enjoy my life at Iowa State University because of my wonderful friends, including Lei Ke, Kun Qiu, Wei Lu, Chenlu Qiu, Shan Zhou, Renliang Gu, Chong Li, Jinchun

Zhan, Meng Zhang, Yue Xu, Kang Gui, and graduate students in the power group. I hope our friendship will last a lifetime.

Finally, I could have never made it this far without my parents Zhenye Huang, Lihua Ma, and my husband Hao Chen. I thank them for their unconditional love and support. Their emotional and financial support has helped me go through the hardest time of my life and also given me the strength to complete this work.

ABSTRACT

In a network that supports multiple unicast, there are several source terminal pairs; each source wishes to communicate with its corresponding terminal. Multiple unicast connections form bulk of the traffic over both wired and wireless networks. Thus, network coding schemes that can help improve network throughput for multiple unicasts are of considerable interest. In this dissertation, we consider the multiple unicast problem over directed acyclic networks with unit-capacity edges when there are three source terminal pairs and two source terminal pairs. For three unicast problem, we assume that the three $s_i - t_i$ pairs wish to communicate at unit-rate via network coding. We define the connectivity level vector $[k_1\ k_2\ k_3]$ such that there exist k_i edge-disjoint paths between s_i and t_i. We attempt to classify networks based on the connectivity level. We identify certain feasible and infeasible connectivity levels $[k_1\ k_2\ k_3]$ for unit rate transmission. For the feasible cases, we construct schemes based on linear network coding. For the infeasible cases, we provide counter-examples, i.e., instances of graphs where the multiple unicast cannot be supported under any (potentially nonlinear) network coding scheme.

For two unicast problem, we assume that we only know certain minimum cut values for the network, e.g., mincut(S_i, T_j), where $S_i \subseteq \{s_1, s_2\}$ and $T_j \subseteq \{t_1, t_2\}$ for different subsets S_i and T_j. Based on these values, we propose an achievable rate region for this problem using linear network codes. Towards this end, we begin by defining a multicast region where both sources are multicast to both the terminals. Following this we enlarge the region by appropriately encoding the information at the source nodes, such that terminal t_i is only guaranteed to decode information from the intended source s_i, while decoding a linear function of the other source. The rate region depends upon the relationship of the different cut values in the network.

CHAPTER 1. INTRODDUCTION

In the past decade, network coding has emerged as an alternative to routing in data transmission in both wired and wireless networks. In a traditional router network, each intermediate node duplicates, stores, and forwards the receiving packets. Although this simple scheme is easy to be implemented and widely used in the communication network, its inherent weakness as viewing the packets as commodity flow but not as information packets has greatly limited the capability of the network. Instead of simply forwarding the received packets at the intermediate node, a node in a network coded system processes the incoming flows in multiple different operations: combining, extracting, copying, and forwarding. Because network coding can use the network resources more efficiently, it has advantages over routing in various aspects, such as increasing the throughput, reducing the resource usage, and improving network robustness. In this chapter, we will briefly introduce several basic ideas of network coding.

It is well known that the maximum rate that one source terminal pair can achieve is equal to the minimum cut value of their connection, and this rate can be achieved by routing [1]. However, for more general network connections in which the terminals require certain subsets of messages available at the sources, routing cannot achieve the optimum solution in general. A well known example is the case of multicast for butterfly network shown in Fig. 1.1. In this network, s needs to transmit X_1 and X_2 to both t_1 and t_2 where X_1 and X_2 are independent with $H(X_1) = H(X_2) = 1$. The capacity of each link is 1. The link $v_3 - v_4$ acts as a bottleneck under routing. However, if we transmit $X_1 + X_2$ on $v_3 - v_4$, both terminals can be satisfied. The above example shows that the throughput of the network is increased by utilizing encoding and decoding in the network.

The properties of network coding have been extensively studied for the multicast network

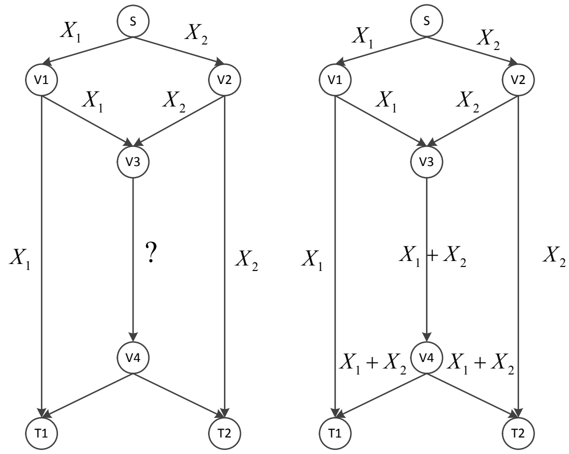


Figure 1.1 The butterfly network, where there is no routing solution but there exists a network coding solution.

in which a source S needs to transmit the same set of information to multiple terminals $t_1, \ldots, t_n$. It has been shown that rate h can be simultaneously supported for each $S - t_i$ pair by linear network codes if the min-cut value between S and each receiver is greater than or equal to h [2]. An algebraic approach [3] for network coding based multicast has been proposed demonstrating that the messages received at each terminal is the source messages multiplied by a transfer matrix with rank h. By inverting the transfer matrix, each terminal can recover the source messages at rate h. A polynomial time deterministic code assignment procedure for multicast network has been studied in [4]. Furthermore, a distributed code assignment scheme is suggested in [5]. It is proved that the multicast capacity can be achieved with high probability if the linear code coefficients are chosen randomly from a large enough field. As for the cost consideration, it is mentioned in [6] that the minimum cost multicast connections can be identified by solving a polynomial-time solvable optimization problem with a decentralized algorithm.

1.1 Network coding for multiple unicast

A multiple unicast network is defined as a network in which there are several source terminal pairs, and each source wishes to communicate with its corresponding terminal. Since multiple unicast networks compose a large amount of real-world network, network coding schemes that can help improve network throughput for multiple unicasts have received intensive research efforts. However, it is well recognized that the design of constructive network coding schemes for multiple unicasts is a hard problem, since at each terminal there exists undesired interference from other sessions. Furthermore, it is proved in [7] that there are instances of network where linear network coding is insufficient.

For undirected multiple unicast network, it has been conjectured by Li and Li [8] that network coding does not provide any advantages over routing. For directed acyclic network, because network coding can achieve higher throughput than routing in a butterfly network, the work of [9] forms a linear program to find the achievable rate region by packing multiple butterfly structures in the original graph. The works of [10] and [11] propose a sufficient and necessary condition on the network structure for two unicast session unit rate transmission. It is pointed out that besides the two edge disjoint paths structure and the butterfly structure, there exists another basic structure that can support unit rate transmission, namely, the grail structure. For non-unit rate two session unicast problem, an achievable rate region is constructed given the min-cut value between each source and terminal pair [12]. The second part of this thesis extends their achievable region given more cut values of the network. A recent work of [13] by Das et al. considers the multiple unicast problem with an interference alignment approach (proposed in [14]). For three unicast problem, under certain algebraic conditions, if the min-cut value for each source terminal pair is 1, then rate 1/2 can be achieved simultaneously. Some further study of interference alignment approach is presented in [15] and [16]. For the outer bound of the capacity region, the authors in [17] propose an outer bound for general networks. This bound is hard to evaluate even for small networks due to the large number of inequalities involved. An improved GNS bound over network sharing bound has been suggested in [18]. It is proved that the GNS bound is the tightest bound that can be realized using only edge-cut bounds. For two unicast session, the work of [19] also proposes an outer bound that can be achieved by certain network structures using the cut-set bound.

In this dissertation, we consider linear network coding schemes for multiple unicast over directed acyclic networks with unit capacity edges. Specifically, we focus on the cases when there are three unicast sessions and when there are two unicast sessions. For the three unicast problem, there are source-terminal pairs denoted $s_i - t_i, i = 1, \ldots, 3$, such that the maximum flow from s_i to t_i is k_i. Each source contains a unit-entropy message that needs to be communicated to the corresponding terminal. We characterize several feasible and infeasible values of the connectivity level vector $[k_1\ k_2\ k_3]$ for unit rate transmission. For the feasible connectivity level vectors, we construct schemes based on linear network coding. For the infeasible connectivity level vectors, we provide instances of graphs where the multiple unicast cannot be supported under any (potentially nonlinear) network coding scheme. For two unicast problems, our aim is to find the achievable region assuming that we only know certain minimum cut values for the network, e.g., $\text{mincut}(S_i, T_j)$, where $S_i \subseteq \{s_1, s_2\}$ and $T_j \subseteq \{t_1, t_2\}$ for different subsets S_i and T_j. We classify networks according to the relationship of the different cut values of the network. To find the achievable region, we first find a multicast region where both sources can be multicast to the terminals. Subsequently, this region is extended according to the specific class that the network belongs to. In both two unicast network and three unicast networks, our achievability scheme uses random linear network coding (or modified random linear network coding) and appropriate precoding at the sources.

1.2 Dissertation outline

The remainder of this dissertation is organized as follows:

Chapter 2 presents some background knowledge of multiple unicast network and discusses several related works.

Chapter 3 considers the multiple unicast problem with three source-terminal pairs over directed acyclic networks with unit capacity edges. The network coding model and the three unicast problem formulation are first introduced. Next, several infeasible connectivity level vectors for unit rate transmission are discussed with instances of graphs. Then the achievable schemes for several feasible connectivity level vectors are presented. Finally, some simulation results are shown to demonstrate that by packing our unit rate schemes, the throughput of some multiple unicast network with higher capacity edges can be improved. Part of this work has appeared in [20] [21] and a revised version has been accepted for journal publication [22].

Chapter 4 investigates the multiple unicast problem with two source-terminal pairs over directed acyclic networks with unit capacity edges. The network coding system model is first presented, followed by the precise problem formulation for the two unicast problem. Then our proposed achievable rate region is derived according to the different cut values. The comparison between our achievable region and existing literature is also provided. The content of this chapter has appeared in [23] and a revised version has been accepted for journal publication [24].

Finally, Chapter 5 summarizes our contributions and presents the ongoing and future work.

CHAPTER 2. BACKGROUND AND RELATED WORK

In a multiple unicast connection, there are several source terminal pairs; each source wishes to communicate with its corresponding terminal at certain rate. The achievable region for the multiple unicast problem has been investigated for both directed acyclic networks [9] [10] [17] [25] and undirected networks [8] in previous work. For directed acyclic network, several works study the achievable region by identifying some special structures of the network. For example, because the butterfly network shows an increment of throughput by network coding over routing, the work of [9] attempts to increase the throughput by packing multiple butterfly structures within the original graph using a linear optimization approach. A similar but distributed scheme is suggested by Ho et al. in [26] which proposes back pressure algorithms for finding achievable rates using XOR operation between pairs of flows. For two unicast sessions, besides the butterfly structure and the two edge disjoint paths structure, there exists another basic structure (grail structure) that supports unit rate transmission [10]. By analyzing the three basic structures, the work of [10](also see [11]) proposes a necessary and sufficient condition on the network structure such that unit rate transmission is guaranteed for two unicast sessions. Instead of analyzing the network with combinatorial approaches, the work of [27], provides an information theoretic characterization for directed acyclic networks. The rate on each edge should satisfy certain inequalities which are derived from link connection patterns. Hence, several bounds for the transmission rate can be generated. However, in practice, evaluating these bounds becomes computationally infeasible even for small networks because of the large number of inequalities that are involved. As for the undirected networks, there is open conjecture as to whether there is any advantage to using network coding as compared to routing [8].

Multiple unicast in the presence of link faults and errors, under certain restricted (though realistic) network topologies has been studied in [28] [29]. The underlying idea is to transmit redundant network coded information over protection paths such that multiple unicast can be simultaneously protected.

For the outer bound of the capacity region for the unicast network, an explicit outer bound (Network Sharing bound) for multiple unicast problem is found in [30]. By analyzing the constraints on the side information at the terminal, the Network Sharing bound provides significant improvement over min-cut bound. A more improved outer bound (GNS bound) is proposed in [18], and proved to be tight in some special structured network. It is also suggested that the GNS bound is the tightest outer bound that can be realized using only edge-cut bounds. Price and Javidi [19] also characterize an outer bound of the rate region in two unicast session network using cut-set bound, and provided a class of network structure in which the outer bound is the exact capacity region. By combining graph theoretic and information theoretic techniques, the work of [17] proposes another outer bound that consists of a series of information inequalities derived from the network structure. However, this bound is hard to evaluate even for small sized networks due to the large number of inequalities involved in the characterization.

Some recent work deals with the case of three unicast sessions, which is also the focus of Chapter 3 of the dissertation. The work of [13] and [15] use the technique of interference alignment (proposed in [14]) for multiple unicast. Roughly speaking they use random linear network coding and design appropriate precoding matrices at the source nodes that allow undesired interference at a terminal to be aligned. However, their approach requires several algebraic conditions to be satisfied in the network. It does not appear that these conditions can be checked efficiently. There has been a deeper investigation of these conditions in [16]. This dissertation is closest in spirit to these papers. Specifically, we also examine network coding for the three-unicast problem. However, the problem setting is somewhat different. Considering networks with unit capacity edges and given the maximum-flow k_i between each source (s_i) - terminal (t_i) pair we attempt to either design a network code that allows unit-rate communication between each source-terminal pair, or demonstrate an instance of a network where unit-rate communication is impossible. Our achievability schemes for unit rate are useful since they can be packed into networks with higher capacity edges. Furthermore, these schemes require vector network coding over at most two time units, unlike the work of [13] and [15], that require a significantly higher level of time-expansion.

At the same time, several works have focused on the case of two unicast networks. For instance, by examining every edge on a path that connects a source and terminal, the work of [10] (see also [11]) presented a necessary and sufficient condition on the network structure for the existence of a network coding solution that supports unit rate transmission for each $s_i - t_i$ connection. These works further pointed out that if a two unicast network can support unit rate transmission, an XOR coding scheme suffices. Reference [12] considered directed acyclic networks and proposed an achievable rate region for non-unit rate two unicast problem based on the number of edge disjoint paths for each $s_i - t_i$ connection. Their result suggested that if the rate at one session needs to be increased by h, the rate at the other session needs to be decreased by $2h$. In this dissertation we also propose an achievable region for the two-unicast problem using linear network codes based on some of the cut values. We consider directed acyclic networks with unit capacity edges and assume that we know certain minimum cut values for the network, e.g., mincut(S_i, T_j), where $S_i \subseteq \{s_1, s_2\}$ and $T_j \subseteq \{t_1, t_2\}$ for different subsets S_i and T_j. To find the achievable region, we first find a multicast region where both sources can be multicast to the terminals. Subsequently, this region is extended according to the relationship of the different cut values of the network. Our achievability scheme uses random linear network coding and appropriate precoding at the sources. The achievable region in [12] is contained in our achievable region given that we have more cut values. The following results have appeared since the publication of our preliminary conference paper [23]. The work of [31] treats the two unicast problem as an instance of a linear deterministic interference channel and finds a network code that uses random linear network coding. By applying the Han-Kobayashi scheme as splitting the information flow as the common part and the private part, they derive the achievable region in terms of the rank of transmission matrices. Their region contains our proposed achievable region. The authors in [32] also derive an achievable region by exploiting the equivalence with deterministic interference channels; their region is completely specified by the cut values in the network (in contrast, in certain cases our region is specified in terms of the rank of matrices that depend on the network code). However, for some networks our scheme achieves a larger region.

CHAPTER 3. NETWORK CODING FOR THREE UNICAST SESSIONS

3.1 Preliminaries

We represent the network as a directed acyclic graph $G = (V, E)$. Each edge $e \in E$ has unit capacity and can transmit one symbol from a finite field of size q per unit time (we are free to choose q large enough). If a given edge has higher capacity, it can be treated as multiple unit capacity edges. A directed edge e between nodes i and j is represented as (i, j), so that $head(e) = j$ and $tail(e) = i$. A path between two nodes i and j is a sequence of edges $\{e_1, e_2, \ldots, e_k\}$ such that $tail(e_1) = i, head(e_k) = j$ and $head(e_i) = tail(e_{i+1}), i = 1, \ldots, k-1$. The network contains a set of n source nodes s_i and n terminal nodes $t_i, i = 1, \ldots n$. Each source node s_i observes a discrete integer-entropy source, that needs to be communicated to terminal t_i. Without loss of generality, we assume that the source (terminal) nodes do not have incoming (outgoing) edges. If this is not the case one can always introduce an artificial source (terminal) node connected to the original source (terminal) node by an edge of sufficiently large capacity that has no incoming (outgoing) edges.

We now discuss the network coding model under consideration in this paper. For the sake of understanding the model, suppose for now that each source has unit-entropy, denoted by X_i (as will be evident, in the sequel we work with integer entropy sources). In scalar linear network coding, the signal on an edge (i, j) is a linear combination of the signals on the incoming edges of i or the source signals at i (if i is a source). We shall only be concerned with networks that are directed acyclic and can therefore be treated as delay-free networks [3]. Let Y_{e_i} (such that $tail(e_i) = k$ and $head(e_i) = l$) denote the signal on edge $e_i \in E$. Then, we have

$$Y_{e_i} = \sum_{\{e_j | head(e_j)=k\}} f_{j,i} Y_{e_j} \text{ if } k \in V \backslash \{s_1, \ldots, s_n\}, \text{ and}$$

$$Y_{e_i} = \sum_{j=1}^{n} a_{j,i} X_j \quad \text{where } a_{j,i} = 0 \text{ if } X_j \text{ is not observed at } k.$$

The coefficients $a_{j,i}$ and $f_{j,i}$ are from the operational field. Note that since the graph is directed acyclic, it is equivalently possible to express Y_{e_i} for an edge e_i in terms of the sources X_j's. If $Y_{e_i} = \sum_{k=1}^{n} \beta_{e_i,k} X_k$ then we say that the global coding vector of edge e_i is $\boldsymbol{\beta}_{e_i} = [\beta_{e_i,1} \ \cdots \ \beta_{e_i,n}]$. We shall also occasionally use the term coding vector instead of global coding vector in this paper. We say that a node i (or edge e_i) is downstream of another node j (or edge e_j) if there exists a path from j (or e_j) to i (or e_i).

Vector linear network coding is a generalization of the scalar case, where we code across the source symbols in time, and the intermediate nodes can implement more powerful operations. Formally, suppose that the network is used over T time units. We treat this case as follows. Source node s_i now observes a vector source $[X_i^{(1)} \ \ldots \ X_i^{(T)}]$. Each edge in the original graph is replaced by T parallel edges. In this graph, suppose that a node j has a set of β_{inc} incoming edges over which it receives a certain number of symbols, and β_{out} outgoing edges. Under vector network coding, node j chooses a matrix of dimension $\beta_{out} \times \beta_{inc}$. Each row of this matrix corresponds to the local coding vector of an outgoing edge from j.

Note that the general multiple unicast problem, where edges have different capacities and the sources have different entropies can be cast in the above framework by splitting higher capacity edges into parallel unit capacity edges and a higher entropy source into multiple, collocated unit-entropy sources. This is the approach taken below.

An instance of the multiple unicast problem is specified by the graph G and the source terminal pairs $s_i - t_i, i = 1, \ldots, n$, and is denoted $< G, \{s_i - t_i\}_1^n, \{R_i\}_1^n >$, where the integer rates R_i denote the entropy of the i^{th} source. The $s_i - t_i$ connections will be referred to as sessions that we need to support.

Let the sources at s_i be denoted as $X_{i1}, \ldots, X_{iR_i}$. The instance is said to have a scalar linear network coding solution if there exist a set of linear encoding coefficients for each node in V such that each terminal t_i can recover $X_{i1}, \ldots, X_{iR_i}$ using the received symbols at its input edges. Likewise, it is said to have a vector linear network coding solution with vector length T if the network employs vector linear network codes and each terminal t_i can recover

$[X_{i1}^{(1)} \ \ldots \ X_{i1}^{(T)}], \ldots, [X_{iR_i}^{(1)} \ \ldots \ X_{iR_i}^{(T)}]$. If the instance has either a scalar or a vector network coding solution, we say that it is feasible.

We will also be interested in examining the existence of a routing solution, wherever possible. In a routing solution, each edge carries a copy of one of the sources, i.e., each coding vector is such that at most one entry takes the value 1, all others are 0. Scalar (vector) routing solutions can be defined in a manner similar to scalar (vector) network codes. We now define some quantities that shall be used throughout the paper.

Definition 3.1.1 *Connectivity level. The connectivity level for source-terminal pair $s_i - t_i$ is said to be β if the maximum flow between s_i and t_i in G is β. The connectivity level of the set of connections $s_1 - t_1, \ldots, s_n - t_n$ is the vector [max-flow($s_1 - t_1$) max-flow($s_2 - t_2$) $\ldots$ max-flow($s_n - t_n$)].*

In this work our aim is to characterize the feasibility of the multiple unicast problem based on the connectivity level of the $s_i - t_i$ pairs. The questions that we seek to answer are of the following form - suppose that the connectivity level is $[k_1 \ k_2 \ \ldots \ k_n]$. Does any instance always have a linear (scalar or vector) network coding solution? If not, is it possible to demonstrate a counter-example, i.e, an instance of a graph G and $s_i - t_i$'s such that recovering the i-th source at t_i for all i is impossible under linear (or nonlinear) strategies?

We conclude this section by observing that a multiple unicast instance $< G, \{s_i - t_i\}_1^n, \{1, 1, \ldots, 1\} >$ with connectivity level $[n \ n \ \ldots \ n]$ is always feasible. Let $X_i, i = 1, \ldots, n$ denote the i-th unit entropy source. We employ vector routing over n time units. Source s_i observes $[X_i^{(1)} \ \ldots \ X_i^{(n)}]$ symbols. Each edge e in the original graph G is replaced by n parallel edges, $e^1, e^2, \ldots, e^n$. Let G_α represent the subgraph of this graph consisting of edges with superscript α. It is evident that max-flow$(s_\alpha - t_\alpha) = n$ over G_α. Thus, we transmit $X_\alpha^{(1)}, \ldots, X_\alpha^{(n)}$ over G_α using routing, for all $\alpha = 1, \ldots, n$. It is clear that this strategy satisfies the demands of all the terminals. In general, though a network with the above connectivity level may not be able to support a scalar routing solution.

3.2 Network coding for three unicast sessions - Infeasible instances

It is clear based on the discussion above that for three unicast sessions if the connectivity level is [3 3 3], then a vector routing solution always exists. We investigate counter-examples for certain connectivity levels in this section.

Lemma 3.2.1 *There exist multiple unicast instances with three unicast sessions,* $< G, \{s_i - t_i\}_{i=1}^3, \{1, 1, 1\} >$ *such that the connectivity levels* [2 2 2] *and* [1 1 3] *are infeasible.*

proof: The examples are shown in Figs. 3.1(a) and 3.1(b). In Fig. 3.1(a), the cut specified by the set of nodes $\{s_1, s_2, s_3, v_1, v_2\}$ has a value of two, while it needs to support a sum rate of three. Similarly in Fig. 3.1(b), the cut $\{s_1, s_2, v_1\}$ has a value of one, but needs to support a rate of two. ■

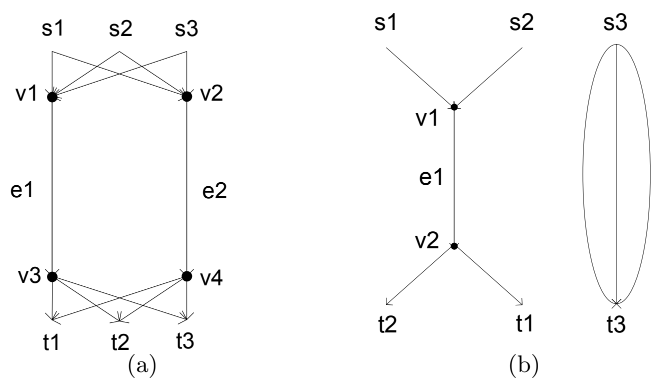


Figure 3.1 (a) An example of [2 2 2] connectivity network without a network coding solution. (b) An example of [1 1 3] connectivity network without a network coding solution.

While the cutset bound is useful in the above cases, there exist certain connectivity levels for which a cut set bound is not tight enough. We now present such an instance in Fig. 3.2. This instance was also presented in [12], though the authors did not provide a formal proof of this fact.

Lemma 3.2.2 *There exists a multiple unicast instance, with two sessions* $< G, \{s_1 - t_1, s_2 - t_2\}, \{2, 1\} >$ *with connectivity level* [2 3] *that is infeasible.*

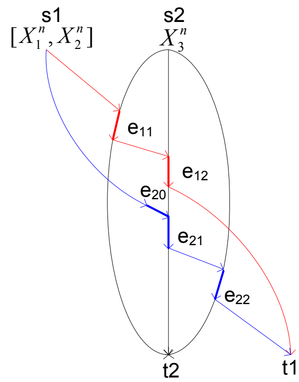


Figure 3.2 An example of [2 3] connectivity network, rate $\{2, 1\}$ cannot be supported.

proof: The graph instance is shown in Fig. 3.2. Assume that in n time units, s_1 observes two vector sources $[X_1^{(1)} \ \dots \ X_1^{(n)}]$ and $[X_2^{(1)} \ \dots \ X_2^{(n)}]$, s_2 observes one vector source $[X_3^{(1)} \ \dots \ X_3^{(n)}]$. The sources are denoted as X_1^n, X_2^n and X_3^n and are independent. The n symbols that are transmitted over edge (i, j) are denoted by Y_{ij}^n. Suppose that the alphabet of X_i is $\mathcal{X}$. Since the entropy rates for the three sources are the same, we assume $H(X_i) = \log|\mathcal{X}| = a$. Also, since we are interested in the feasibility of the solution, we assume that the alphabet size of Y_{ij} is also the same as $\mathcal{X}$, and $H(Y_{ij}) \leq \log|\mathcal{X}| = a$ by the capacity constraint of the edge. At terminal t_1 and t_2, from Y_{11}^n, Y_{12}^n, Y_{21}^n and Y_{22}^n, we estimate X_1^n, X_2^n and X_3^n. Let the estimate be denoted as $\widehat{X}_1^n$, $\widehat{X}_2^n$ and $\widehat{X}_3^n$. Suppose that there exist network codes and decoding functions such that $P((\widehat{X}_1^n, \widehat{X}_2^n) \neq (X_1^n, X_2^n)) \to 0$ as $n \to \infty$. For successful decoding at t_1, using Fano's inequality, we have

$$H(X_1^n, X_2^n | \widehat{X}_1^n, \widehat{X}_2^n) \leq n\epsilon_n. \tag{3.1}$$

where $n\epsilon_n = 1 + 2nP_e \log(|\mathcal{X}|)$, $P_e = P((\widehat{X}_1^n, \widehat{X}_2^n) \neq (X_1^n, X_2^n))$ and $\epsilon_n \to 0$ as $n \to \infty$. The topological structure of the network implies that $\widehat{X}_1^n, \widehat{X}_2^n$ are functions of Y_{12}^n and Y_{22}^n. Hence, we have

$$\begin{aligned} H(X_1^n, X_2^n | Y_{12}^n, Y_{22}^n) &= H(X_1^n, X_2^n | \widehat{X}_1^n, \widehat{X}_2^n, Y_{12}^n, Y_{22}^n) \\ &\leq H(X_1^n, X_2^n | \widehat{X}_1^n, \widehat{X}_2^n) \leq n\epsilon_n. \end{aligned} \tag{3.2}$$

Since $H(Y_{12}^n, Y_{22}^n) \leq 2an$, using eq. (3.2) and the independence of X_1^n, X_2^n and X_3^n, by Claim B.0.1 (see Appendix), we have

$$an - n\epsilon_n \leq H(X_3^n|Y_{12}^n, Y_{22}^n) \leq an, \text{ and} \tag{3.3}$$

$$H(Y_{12}^n, Y_{22}^n|X_3^n) \geq 2an - 2n\epsilon_n. \tag{3.4}$$

Next, we have

$$\begin{aligned}
H(Y_{21}^n, Y_{22}^n) &\overset{(a)}{=} H(X_3^n, Y_{21}^n, Y_{22}^n) - H(X_3^n|Y_{21}^n, Y_{22}^n) \\
&\overset{(b)}{=} H(X_3^n, Y_{21}^n) - H(X_3^n|Y_{21}^n, Y_{22}^n) \\
&\overset{(c)}{\leq} 2an - H(X_3^n|Y_{21}^n, Y_{22}^n, Y_{20}^n, Y_{12}^n, X_1^n, X_2^n) \\
&\overset{(d)}{=} 2an - H(X_3^n|Y_{22}^n, Y_{20}^n, Y_{12}^n, X_1^n, X_2^n) \\
&\overset{(e)}{=} 2an - H(X_3^n|Y_{22}^n, X_1^n, X_2^n, Y_{12}^n) \\
&\overset{(f)}{=} 2an - H(X_3^n|Y_{22}^n, Y_{12}^n) + I(X_3^n; X_1^n, X_2^n|Y_{22}^n, Y_{12}^n) \\
&\leq 2an - H(X_3^n|Y_{22}^n, Y_{12}^n) + H(X_1^n, X_2^n|Y_{22}^n, Y_{12}^n) \\
&\overset{(g)}{\leq} 2an - an + n\epsilon_n + n\epsilon_n = an + 2n\epsilon_n,
\end{aligned} \tag{3.5}$$

where (a) follows from the chain rule, (b) holds because Y_{22}^n is a function of X_3^n and Y_{21}^n, (c) follows from the capacity constraints and the fact that conditioning reduces entropy, (d) follows as Y_{21}^n is a function of Y_{12}^n and Y_{20}^n, (e) is due to the fact that Y_{20}^n is a function of X_1^n and X_2^n, (f) follows from the definition of mutual information, and (g) is a consequence of eq. (3.2) and eq. (3.3). The above inequalities indicate that e_{21} and e_{22} need to carry the same information asymptotically for successful decoding at t_1.

From the network, we know that Y_{12}^n is a function of Y_{11}^n and X_3^n. This implies that

$$\begin{aligned}
H(Y_{11}^n, Y_{21}^n, Y_{22}^n|X_3^n) &= H(Y_{11}^n, Y_{21}^n, Y_{22}^n, X_3^n|X_3^n) \\
&\geq H(Y_{12}^n, Y_{21}^n, Y_{22}^n|X_3^n) \\
&\geq H(Y_{22}^n, Y_{12}^n|X_3^n) \overset{(a)}{\geq} 2an - 2n\epsilon_n,
\end{aligned} \tag{3.6}$$

where (a) is due to eq. (3.4). Finally, we have

$$\begin{aligned}
&H(X_3^n|Y_{11}^n, Y_{21}^n, Y_{22}^n) \\
&= H(Y_{11}^n, Y_{21}^n, Y_{22}^n|X_3^n) + H(X_3^n) - H(Y_{22}^n, Y_{21}^n, Y_{11}^n) \\
&\overset{(a)}{\geq} 2an - 2n\epsilon_n + an - H(Y_{22}^n, Y_{21}^n) - H(Y_{11}^n|Y_{22}^n, Y_{21}^n) \\
&\overset{(b)}{\geq} 3an - 2n\epsilon_n - an - 2n\epsilon_n - H(Y_{11}^n|Y_{22}^n, Y_{21}^n) \\
&\overset{(c)}{\geq} 2an - 4n\epsilon_n - an = an - 4n\epsilon_n,
\end{aligned} \tag{3.7}$$

where (a) is due to eq. (3.6), (b) is because of eq. (3.5) and (c) holds because of the capacity constraint on Y_{11}^n. This implies that t_2 cannot decode X_3^n with an asymptotically vanishing probability of error. ■

Corollary 3.2.3 *There exists a multiple unicast instance with three sessions, and connectivity level* [2 3 2] *that is infeasible.*

proof: Consider the instance $< G, \{s'_i - t'_i\}_1^3, \{1, 1, 1\} >$, where G is the graph in Fig. 3.2. The sources s'_1 and s'_3 are collocated at s_1 (in G), and the terminals t'_1 and t'_3 are collocated at t_1 (in G). Likewise, the source s'_2 and terminal t'_2 are located at s_2 and t_2 in G. The three sessions have connectivity level [2 3 2]. Based on the arguments in Lemma 3.2.2, there is no feasible solution for this instance. ■

The previous example can be generalized to an instance with two unicast sessions with connectivity level $[n_1\ n_2]$ that cannot support rates $R_1 = n_1, R_2 = n_2 - 3n_1/2 + 1$ when $n_2 \geq 3n_1/2$ and $n_1 > 1$.

Theorem 3.2.4 *For a directed acyclic graph G with two $s - t$ pairs, if the connectivity level for (s_1, t_1) is n_1, for (s_2, t_2) is n_2, where $n_2 \geq 3n_1/2$ and $n_1 > 1$, there exist instances that cannot support $R_1 = n_1$ and $R_2 = n_2 - 3n_1/2 + 1$.*

proof: Provided in the Appendix A. ■

3.3 Network coding for three unicast sessions - Feasible instances

It is evident that there exist instances with connectivity level [2 2 3] (and component-wise lower) that are infeasible. Therefore, the possible instances that are potentially feasible are

[1 3 3] and [1 2 4], or their permutations and connectivity levels that are greater than them. In the discussion below, we show that all the instances with the connectivity levels [1 3 3], [2 2 4] and [1 2 5] are feasible using linear network codes. Our work leaves out one specific connectivity level vector, namely [1 2 4] for which we have been unable to provide either a feasible network code or a network topology where communicating at unit rate is impossible.

As pointed out by the work of [3], under linear network coding, the case of multiple unicast requires (a) the transfer matrix for each source-terminal pair to have a rank that is high enough, and (b) the interference at each terminal to be zero. Under random linear network coding, it is possible to assert that the rank of any given transfer matrix from a source s_i to a terminal t_j has w.h.p. a rank equal to the minimum cut between s_i and t_j; however, in general this is problematic for satisfying the zero-interference condition.

Our strategies rely on a combination of graph-theoretic and algebraic methods. Specifically, starting with the connectivity level of the graph, we use graph theoretic ideas to argue that the transfer matrices of the different terminals have certain relationships. The identified relationships then allow us to assert that suitable precoding matrices that allow each terminal to be satisfied can be found. A combination of graph-theoretic and algebraic ideas were also used in the work of [33], where the problem of multicasting finite field sums over wired networks was considered. However, there are some crucial differences. Reference [33] considered a multicast situation; thus, the issue of dealing with interference did not exist. As will be evident, a large part of the effort in the current work is to demonstrate that the terminals can decode their intended message in the presence of the interfering messages.

We begin with the following definitions.

Definition 3.3.1 *Minimality. Consider a multiple unicast instance* $< G = (V, E), \{s_i - t_i\}_1^n, \{1, \ \ldots, \ 1\} >$*, with connectivity level* $[k_1 \ k_2 \ \ldots \ k_n]$*. The graph* G *is said to be minimal if the removal of any edge from* E *reduces the connectivity level. If* G *is minimal, we will also refer to the multiple unicast instance as minimal.*

Clearly, given a non-minimal instance $G = (V, E)$, we can always remove the non-essential edges from it, to obtain the minimal graph $G_{\min}$. This does not affect connectivity. A network code for $G_{\min} = (V, E_{\min})$ can be converted into a network code for G by simply assigning the zero coding vector to the edges in $E \backslash E_{\min}$.

Definition 3.3.2 *Overlap edge. An edge e is said to be an overlap edge for paths P_i and P_j in G, if $e \in P_i \cap P_j$.*

Definition 3.3.3 *Overlap segment. Consider a set of edges $E_{os} = \{e_1, \ldots, e_l\} \subset E$ that forms a path. This path is called an overlap segment for paths P_i and P_j if*

(i) $\forall k \in \{1, \ldots, l\}$, *$e_k$ is an overlap edge for P_i and P_j,*

(ii) none of the incoming edges into tail(e_1) are overlap edges for P_i and P_j, and

(iii) none of the outgoing edges leaving head(e_l) are overlap edges for P_i and P_j.

Our solution strategy is as follows. We first convert the original instance into another *structured* instance where each internal node has at most degree three (in-degree + out-degree). We then convert this new instance into a minimal one, and develop the network code assignment algorithm. This network code, can be converted into a network code for the original instance.

Following [34] we can efficiently construct a *structured* graph $\hat{G} = (\hat{V}, \hat{E})$ in which each internal node $v \in \hat{V}$ is of total degree at most three with the following properties.

(a) $\hat{G}$ is acyclic.

(b) For every source (terminal) in G there is a corresponding source (terminal) in $\hat{G}$.

(c) For any two edge disjoint paths P_i and P_j for one unicast session in G, there exist two *vertex* disjoint paths in $\hat{G}$ for the corresponding session in $\hat{G}$.

(d) Any feasible network coding solution in $\hat{G}$ can be efficiently turned into a feasible network coding solution in G.

In all the discussions below, we will assume that the graph G is structured. It is clear that this is w.l.o.g. based on the previous arguments.

3.3.1 Code assignment procedure for instances with connectivity level [1 3 3]

We begin by showing some basic results for two-unicast. The three unicast result follows by applying vector network coding over two time units and using the two-unicast results.

Lemma 3.3.4 *A minimal multiple unicast instance* $< G, \{s_1 - t_1, s_2 - t_2\}, \{1, m\} >$ *with connectivity level* $[1\ m+1]$ *is always feasible.*

proof: Denote the path from s_1 to t_1 as $\mathcal{P}_1 = \{P_{11}\}$, and the $m+1$ paths from s_2 to t_2 as $\mathcal{P}_2 = \{P_{21}, \ldots, P_{2m+1}\}$. The information that needs to be transmitted from s_1 is X_1, and the information that needs to be transmitted from s_2 is $X_{21}, \ldots, X_{2m}$. We assume that P_{11} overlaps with all paths in $\mathcal{P}_2$. Otherwise, if P_{11} overlaps with n paths in $\mathcal{P}_2$ where $0 \leq n < m+1$, w.l.o.g, assume they are $P_{21}, \ldots, P_{2n}$. Then $X_{2n}, \ldots, X_{2m}$ can be simply transmitted over the overlap free paths $P_{2n+1}, \ldots, P_{2m+1}$, and the problem reduces to communicating X_1 and $X_{21}, \ldots, X_{2n-1}$ over $P_{11} \cup P_{21} \cup \cdots \cup P_{2n}$, which corresponds to the statement of the theorem with m replaced by $n-1$. Hence, we focus on the case that P_{11} overlaps with all paths in $\mathcal{P}_2$.

We assume that the local coding vectors for each edge are indeterminates for now. Source s_2 uses a precoding matrix Θ; the rows of Θ specify the coding vectors on the outgoing edges of s_2. The choice of the local coding vectors and Θ is discussed below. The transmitted symbol on the outgoing edge from s_2 belonging to P_{2i} is $[\theta_{i1}\ \cdots\ \theta_{im}][X_{21}\ \cdots\ X_{2m}]^T$ where $i = 1, \ \ldots\ , m+1$. Let $\underline{\theta}_j = [\theta_{1j}\ \cdots\ \theta_{(m+1)j}]^T$ where $j = 1, \ldots, m$.

As P_{11} overlaps with all paths on $\mathcal{P}_2$, there will be many overlap segments on P_{11}. Let E_{os1} denote the overlap segment that is closest to t_1 (under the topological order imposed by the directed acyclic nature of the graph) along P_{11} and suppose that it is on P_{21}. A key observation is that E_{os1} is also the overlap segment on P_{21} that is closest to t_2. Indeed if there is another overlap segment E'_{os1} that is closer to t_2 along P_{21}, then it implies the existence of a cycle in the graph. Let the coding vectors at each intermediate node be specified by indeterminates for now.

The overall transfer matrix from the pair of sources $\{s_1, s_2\}$ to t_1 can be expressed as

$$[M_{11} \mid M_{12}] = [\alpha_1 \mid \gamma_{11}\ \cdots\ \gamma_{1(m+1)}].$$

Similarly, the transfer matrix from the pair of sources $\{s_1, s_2\}$ to t_2 can be expressed as

$$[M_{21} \mid M_{22}] = \left[\begin{array}{c|cccc} \alpha_1 & \gamma_{11} & \cdots & \gamma_{1(m+1)} \\ \alpha_2 & \gamma_{21} & \cdots & \gamma_{2(m+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m+1} & \gamma_{(m+1)1} & \cdots & \gamma_{(m+1)(m+1)} \end{array}\right].$$

The received vector at terminal t_i is therefore $[M_{i1} \mid M_{i2}] \begin{bmatrix} X_1 \\ \Theta[X_{21} \cdots X_{2m}]^T \end{bmatrix}$. The variables $\alpha_i' s$ and $\gamma_{ij}' s$ in the above matrices depend on the indeterminate local coding vectors and are therefore undetermined at this point.

We emphasize that the first row of $[M_{21} \mid M_{22}]$ is the same as $[M_{11} \mid M_{12}]$. As there exists a single path between s_1 and t_1, it is clear that α_1 is not identically zero. Similarly, as there are $m+1$ edge-disjoint paths between s_2 to t_2, we have that $\det(M_{22})$ is not identically zero. Now suppose that we employ random linear network coding at all nodes. Using the Schwartz-Zippel lemma [35], this implies that $\alpha_1 \neq 0$ and $\det(M_{22}) \neq 0$ w.h.p. We assume that $\alpha_1 \neq 0$ and $\det(M_{22}) \neq 0$ in the discussion below. Next we select θ_{ij}, $i = 1, \ldots, m+1$, $j = 1, \ldots, m$ such that they satisfy the following equation.

$$M_{22}[\underline{\theta}_1 \ \cdots \ \underline{\theta}_m] = \begin{bmatrix} 0 & \cdots & 0 \\ a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_m \end{bmatrix} \tag{3.8}$$

where $a_1, \ldots, a_m$ are non-zero values. Note that such $[\underline{\theta}_1 \ \cdots \ \underline{\theta}_m]$ can be chosen since M_{22} is full-rank.

Terminal t_1 can decode, since $M_{12}[\underline{\theta}_1 \ \cdots \ \underline{\theta}_m] = [0 \cdots 0]$ and $\alpha_1 \neq 0$, and t_2 can decode, since X_1 is available at t_2, and $rank(M_{22}[\underline{\theta}_1 \ \cdots \ \underline{\theta}_m]) = m$ (from eq. (3.8)). Finally, we note that there are $q-1$ choices for each $\underline{\theta}_j$. ■

We remark that the main issue in the above argument is to demonstrate that the choice of Θ works simultaneously for both t_1 and t_2. The observation that E_{os1} is overlap segment closest to t_1 and t_2 along P_{11} and P_{21} respectively allows us to make this argument.

The result for three unicast sessions with connectivity level [1 3 3] now follows by using vector linear network coding over two time units, as discussed below.

Theorem 3.3.5 *A multiple unicast instance with three sessions,* $< G, \{s_i - t_i\}_1^3, \{1, 1, 1\} >$ *with connectivity level at least* [1 3 3] *is feasible.*

proof: W.l.o.g. we assume that the connectivity level is exactly [1 3 3]. We use vector linear network coding over two time units. For facilitating the presentation we form a new graph G^* where each edge $e \in E$ is replaced by two parallel unit capacity edges e^1 and e^2 in G^*. The messages at source node s_i are denoted $[X_{i1}\ X_{i2}], i = 1, \ldots, 3$. Let the subgraph of G^* induced by all edges with superscript i be denoted G_i^*. In G_1^*, there exists a single $s_1 - t_1$ path and three edge disjoint $s_2 - t_2$ paths. Therefore, we can transmit X_{11} from s_1 to t_1 and $[X_{21}\ X_{22}]$ from s_2 to t_2 using the result of Lemma 3.3.4. Similarly, we use G_2^* to communicate X_{12} from s_1 to t_1 and $[X_{31}\ X_{32}]$ from s_3 to t_3. Thus, over two time units a rate of [1 1 1] can be supported. ∎

3.3.2 Code assignment procedure for instances with connectivity level [2 2 4]

Our solution approach is similar in spirit to the discussion above. In particular, we first investigate a two-unicast scenario with connectivity level [2 4] and rate requirement $\{2, 1\}$ and use that in conjunction with vector network coding to address the three-unicast with connectivity level [2 2 4].

Lemma 3.3.6 *A minimal multiple unicast instance* $< G, \{s_1 - t_1, s_2 - t_2\}, \{2, 1\} >$ *with connectivity level* [2 4] *is feasible.*

proof: Let $\mathcal{P}_1 = \{P_{11}, P_{12}\}$ denote two edge disjoint paths (also vertex disjoint due to the structured nature of G) from s_1 to t_1 and $\mathcal{P}_2 = \{P_{21}, P_{22}, P_{23}, P_{24}\}$ denote the four vertex disjoint paths from s_2 to t_2. Let the source messages at s_1 be denoted by X_1 and X_2, and the source message at s_2 by X_3. We color the edges of the graph such that each edge on P_{11} is colored red, each edge on P_{12} is colored blue and each edge on a path in $\mathcal{P}_2$ is colored black.

As the paths in $\mathcal{P}_1$ and $\mathcal{P}_2$ are vertex-disjoint, it is clear that a node with an in-degree of two is such that its outgoing edge has two colors (either *(blue, black)* or *(red, black)*). The path further downstream continues to have two colors until it reaches a node of out-degree two.

Such an overlap segment with two colors will be referred to as a *mixed color overlap segment*. We shall also use the terms *red* or *blue overlap segment* to refer to segments with colors *(red, black)* and *(blue, black)* respectively. Note that by our naming convention path P_{ij} is a path that enters terminal t_i. Under the topological order in G we can identify the overlap segment on P_{ij} that is closest to t_i. In the discussion below this will be referred to as the last overlap segment with respect to path P_{ij}. Two overlap segments E_{os1} and E_{os2} are said to be neighboring with respect to P_{ij} if there are no overlap segments between them along P_{ij}. An example of neighboring overlap segments is shown in Fig. 3.3(a).

Claim 3.3.7 *Consider two neighboring mixed color overlap segments E_{os1} and E_{os2} with respect to path $P_{1i} \in \mathcal{P}_1$. Then E_{os1} and E_{os2} cannot lie on the same path $P_{2j} \in \mathcal{P}_2$.*

proof: W.l.o.g., assume that $E_{os1} = \{e_1, \ldots, e_{k_1}\}$ and $E_{os2} = \{e'_1, \ldots, e'_{k_2}\}$ are such that e_{k_1} is upstream of e'_1. Now assume that both E_{os1} and E_{os2} are on P_{2j}. Note that $head(e_{k_1})$ has two outgoing edges, one of which belongs to P_{1i} and the other belongs to P_{2j} (denoted by e^*). We claim that e^* can be removed while the connectivity level remains the same. This is because e^* does not belong to P_{1i} and P_{2k}, $\forall k \neq j$. Moreover, after the removal, P_{2j} can be modified to the path specified as $path(s_2, head(e_{k_1})) - path(e_{k_1}, e'_1) - path(head(e'_1), t_2)$ where $path(e_{k_1}, e'_{k_2})$ is along P_{1i}. The new P_{2j} is vertex disjoint of P_{2k}, $\forall k \neq j$, since E_{os1} and E_{os2} are neighboring mixed color overlap segments along P_{1i} which means that $path(e_{k_1} - e'_1)$ is either purely blue or purely red. This contradicts the minimality of the graph. ■

Likewise, two neighboring mixed color overlap segments with respect to P_{2i}, cannot lie on the same path P_{1j}.

To explain our coding scheme, we first denote the last red (blue) overlap segment with respect to P_{11} (P_{12}) by E_r (E_b). If there is no E_r, then X_1 can be transmitted along P_{11}. According to Lemma 3.3.4, X_2 and X_3 can be transmitted to t_1 and t_2 respectively. A similar

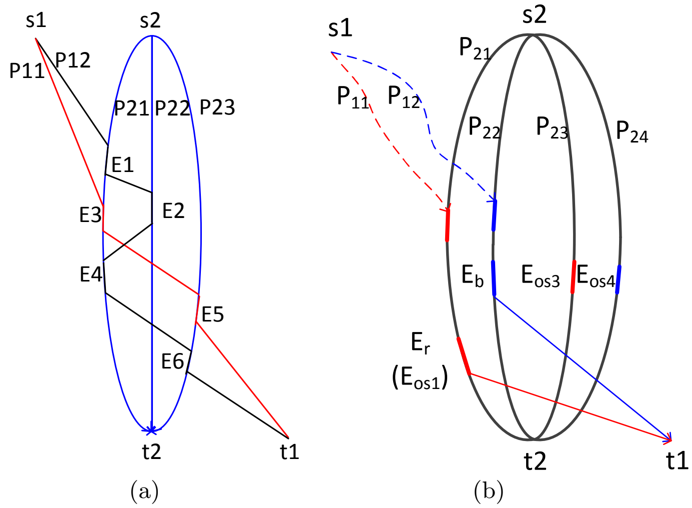


Figure 3.3 (a) An instance of network where there are several pairs of neighboring overlap segments. E_1 and E_3 are neighboring overlap segments along P_{21}, E_1 and E_2 are neighboring overlap segments along P_{12}. E_1 and E_4 are not overlap segments along any paths. (b) A network with connectivity level [2 4] and rate $\{2, 1\}$. The coloring of the different paths helps us to show that a linear network coding solution exists.

argument can be applied to the case when there is no E_b. Hence, we assume that both E_r and E_b exist. Based on their locations in G, we distinguish the following two cases.

- *Case 1: E_r and E_b are on different paths $\in \mathcal{P}_2$.*

W.l.o.g. we assume that E_r and E_b are on paths P_{21} and P_{22}. If there are no mixed color overlap segments on either P_{23} or P_{24}, X_3 can be transmitted to t_2 through the overlap free path, and X_1, X_2 can be routed to t_1. Therefore, we focus on the case that there are mixed color overlap segments on both P_{23} and P_{24}. Let E_{osi} denote the last mixed color overlap segments with respect to P_{2i}, $i = 1, \ldots, 4$ (see Fig. 3.3(b)).

Our coding scheme is as follows. Symbol X_i is transmitted over the outgoing edge from s_1 over $P_{1i}, i = 1, 2$; symbols $\theta_j X_3$ are transmitted over the outgoing edges of s_2 over P_{2j}, $j = 1, \ldots, 4$ respectively. The values of $\theta_j \in GF(q)$ will be chosen as part of the code assignment below. Let the coding vectors at each intermediate node be specified by indeterminates for now. The overall transfer matrix from the pair of sources $\{s_1, s_2\}$ to t_1 can be expressed as

$$[M_{11} \mid M_{12}] = \left[\begin{array}{cc|cccc} \alpha_1 & \beta_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \alpha_2 & \beta_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \end{array}\right],$$

such that the received vector at t_1 is $[M_{11} \mid M_{12}][X_1\ X_2 \mid \theta_1 X_3\ \ldots\ \theta_4 X_3]^T$. Recall that E_r and E_b are the last mixed color segments with respect to P_{11} and P_{12}. Thus, they carry the same information as the incoming edges of t_1 which implies that the row vectors of $[M_{11} \mid M_{12}]$ are the coding vectors on E_r and E_b respectively. Similarly, the transfer matrix from $\{s_1, s_2\}$ to the edge set $\{E_r, E_b, E_{os3}, E_{os4}\}$ can be expressed as

$$[M^e_{21} \mid M^e_{22}] = \left[\begin{array}{cc|cccc} \alpha_1 & \beta_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \alpha_2 & \beta_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \alpha_3 & \beta_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \\ \alpha_4 & \beta_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{array}\right]$$

where we use the superscript e to emphasize that these transfer matrices are to the edge set $\{E_r, E_b, E_{os3}, E_{os4}\}$ and not to the terminal t_2.

Note that the entries of the transfer matrices above are functions of the choice of the local coding vectors in the network which are indeterminate. Thus, at this point, the M_{ij} and M^e_{ij} matrices are also composed of indeterminates.

As there exist two edge disjoint paths from s_1 to $\{E_r, E_b\}$, the determinant of M_{11} is not identically zero. Similarly, since the edges E_r, E_b, E_{os3} and E_{os4} lie on different paths in $\mathcal{P}_2$, there are four edge disjoint paths from s_2 to the edge subset $\{E_r, E_b, E_{os3}, E_{os4}\}$, and the determinant of M^e_{22} is not identically zero. This implies that their product is not identically zero. Hence, by the Schwartz-Zippel lemma [35], under random linear network coding there exists an assignment of local coding vectors so that $rank(M_{11}) = 2$ and $rank(M^e_{22}) = 4$. We assume that the local coding vectors are chosen from a large enough field $GF(q)$ so that this is the case. For this choice of local coding vectors we propose a choice of $\underline{\theta} = [\theta_1\ \theta_2\ \theta_3\ \theta_4]^T$ such that the decoding is simultaneously successful at both t_1 and t_2.

Decoding at t_1: As M_{11} is a square full-rank matrix, we only need to null the interference from s_2. Accordingly, we choose $\underline{\theta}$ from the null space of M_{12}, i.e.,

$$M_{12}\underline{\theta} = 0. \tag{3.9}$$

There are at least $q^2 - 1$ such non-zero choices for $\underline{\theta}$ as M_{12} is a 2×4 matrix.

Decoding at t_2: The primary issue is that one needs to demonstrate that the choice of $\underline{\theta}$ allows both terminals to simultaneously decode. Indeed, it may be possible that our choice of $\underline{\theta}$ along with a specific network topology may make it impossible to decode at t_2. The key argument that this does not happen requires us to leverage certain topological properties of the overlap segments, that we present below.

Claim 3.3.8 *In G either one or both of the following statements hold. (i) E_r is the last overlap segment w.r.t. P_{21}. (ii) E_b is the last overlap segment w.r.t. P_{22}.*

proof: Assume that neither statement is true. This means that there is a blue overlap segment E'_b below E_r along P_{21}, and there is a red overlap segment E'_r below E_b along P_{22}. Thus, E'_r is upstream of E_r and E'_b is upstream of E_b. However, this means that edges E'_r, E_r, E'_b and E_b form a cycle, which is a contradiction. ■

In the discussion below, w.l.o.g., we assume that E_r is the last overlap segment on P_{21}. The argument above allows us to identify edges E_r, E_{os3} and E_{os4} that carry the *same symbols* as those entering t_2. We show below that the X_1 and X_2 components can be canceled by using the information on E_{os3} and E_{os4} while retaining the X_3 component.

Let $\underline{\gamma}_i$ represent the vector $[\gamma_{i1}\ \gamma_{i2}\ \gamma_{i3}\ \gamma_{i4}]^T, i = 1, \ldots, 4$ in the discussion below. Note that if $[\alpha_3\ \beta_3]$ and $[\alpha_4\ \beta_4]$ are linearly independent, there exist δ_3 and δ_4 such that

$$[\alpha_1\ \beta_1] = \delta_3[\alpha_3\ \beta_3] + \delta_4[\alpha_4\ \beta_4],$$

where δ_3 and δ_4 are not both zero. Thus, t_2 can recover $[-\underline{\gamma}_1 + \delta_3\underline{\gamma}_3 + \delta_4\underline{\gamma}_4]^T\underline{\theta}X_3$. Note that $\underline{\gamma}_1^T\underline{\theta} = 0$, by the constraint on $\underline{\theta}$ above, thus we only need to pick $\underline{\theta}$ such that $[\delta_3\underline{\gamma}_3 + \delta_4\underline{\gamma}_4]^T\underline{\theta} \neq 0$. To see that this can be done, we note that M_{22} is full rank which implies that the matrix $[\underline{\gamma}_1\ \ \underline{\gamma}_2\ \ (\delta_3\underline{\gamma}_3 + \delta_4\underline{\gamma}_4)]^T$ is full rank. Therefore, there exist at most q choices for $\underline{\theta}$ such that $[\underline{\gamma}_1\ \ \underline{\gamma}_2\ \ (\delta_3\underline{\gamma}_3 + \delta_4\underline{\gamma}_4)]^T\underline{\theta} = 0$. Hence, there are at least $q^2 - q - 1 > 0$ non-zero choices for $\underline{\theta}$ that allow decoding at t_1 and t_2 simultaneously.

If $[\alpha_3\ \beta_3]$ and $[\alpha_4\ \beta_4]$ are dependent, decoding can be performed simply by working only with the received values over E_{os3} and E_{os4} using a similar argument as above.

- *Case 2: E_r and E_b are on the same path P_{2i}.*

W.l.o.g., assume that E_b is downstream of E_r along P_{21}. Then E_b will be the last overlap segment w.r.t. P_{21}. Let E'_b denote the blue overlap segment that is a neighbor of E_b w.r.t. P_{12}. Note that E'_b cannot be on P_{21} according to Claim 3.3.7. If E'_b does not exist, it implies that there is only one blue overlap segment (namely, E_b) in the network. Therefore, there only exist red overlap segments on P_{23} and P_{24}; using Lemma 3.3.4, X_1 and X_3 can be transmitted to t_1 and t_2 respectively over $P_{11} \cup P_{23} \cup P_{24}$, and X_2 can be routed along P_{12} to t_1.

We now focus on the case when an E'_b exists and assume (w.l.o.g.) that it is on P_{22}. The main difference is that instead of using random coding over the entire graph, we modify our coding scheme such that random coding is performed over the graph except at E_b and all the edges downstream of E_b. At E_b, deterministic coding is performed such that E_b carries the same information as the incoming edge of it along P_{12}. The information on E_b is further routed to all the downstream edges of E_b. Note that by the deterministic coding, E_b carries the same information as E'_b.

Decoding at t_1: Using the arguments developed in Case 1, it is clear that X_1 and X_2 can be decoded from the information on E'_b and E_r. The code assignment ensures that E_b and E'_b carry the same information, thus t_1 is satisfied.

Decoding at t_2: In Case 1, we showed that X_3 can be decoded from the information on E_r, E_{os3} and E_{os4}. A similar argument can be made that X_3 can be decoded from the information on E'_b, E_{os3} and E_{os4}. Since E_b carries the same information as E'_b and E_b is the last overlap segment on P_{21}, terminal t_2 can decode X_3 by the information on E_b, E_{os3} and E_{os4}.

■

By using the result of Lemma 3.3.6 and the idea of vector network coding, we have the following theorem when the connectivity level is [2 2 4].

Theorem 3.3.9 *A multiple unicast instance with three sessions,* $< G, \{s_i - t_i\}_1^3, \{1,1,1\} >$ *with connectivity level at least* $[2\ 2\ 4]$ *is feasible.*

proof: It can be seen that the line of argument used in the proof of Theorem 3.3.5, namely using vector network coding over two time units and use the result of Lemma 3.3.6 gives us the desired result. ∎

3.3.3 Code assignment procedure for instances with connectivity level $[1\ 2\ 5]$

We now consider network code assignment for networks where the connectivity level is $[1\ 2\ 5]$. The code assignment in this case requires somewhat different techniques. In particular, the idea of using a two-session unicast result along with vector network coding does not work unlike the cases considered previously. At the top level, we still use random network coding followed by appropriate precoding to align the interference seen by the terminals. However, as we shall see below, we will need to depart from a purely random linear code in the network in certain situations.

As before, we consider a minimal structured graph G and let X_i be the source symbol at source node s_i for $i = 1, \dots, 3$ and $\mathcal{P}_1 = \{P_{11}\}$ denote the path from s_1 to t_1, $\mathcal{P}_2 = \{P_{21}, P_{22}\}$ denote the edge disjoint paths from s_2 to t_2, $\mathcal{P}_3 = \{P_{31}, P_{32}, P_{33}, P_{34}, P_{35}\}$ denote the edge disjoint paths from s_3 to t_3.

Our scheme operates as follows: X_1 is transmitted over the outgoing edge from s_1 along P_{11} , $\xi_i X_2$ are transmitted over the outgoing edges of s_2 along P_{2i}, $i = 1, 2$, and $\theta_j X_3$ are transmitted over the outgoing edges of s_3 along P_{3j}, $j = 1, \dots, 5$ where $\underline{\xi} = [\xi_1\ \xi_2]^T$ and $\underline{\theta} = [\theta_1\ \dots\ \theta_5]^T$ are precoding vectors chosen from a finite field with size q.

Let $M_i = [M_{i1} \mid M_{i2} \mid M_{i3}]$ denote the transfer matrix from $\{s_1, s_2, s_3\}$ to terminal t_i. Each M_{ij} corresponds to the transformation from source s_j to terminal t_i, i.e., the number of columns in M_{ij} is $1, 2$ and 5 for $j = 1, 2$ and 3 respectively. Similarly, the number of rows in M_{ij} is $1, 2$ and 5 for $i = 1, 2$ and 3 respectively.

In the discussion below we will need to refer to the individual entries of M_1 and M_2.

Accordingly, we express these matrices explicitly as follows.

$$M_1 = [M_{11} \mid M_{12} \mid M_{13}] = [\alpha_1 \mid \underline{\beta}^T \mid \underline{\gamma}^T\ ]$$
$$= [\alpha_1 \mid \beta_1\ \beta_2 \mid \gamma_1\ \gamma_2\ \gamma_3\ \gamma_4\ \gamma_5]\,,$$
$$M_2 = [M_{21} \mid M_{22} \mid M_{23}] = \left[\begin{array}{c|c|c} \alpha'_1 & \underline{\beta'}^T_1 & \underline{\gamma'}^T_1 \\ \alpha'_2 & \underline{\beta'}^T_2 & \underline{\gamma'}^T_2 \end{array}\right]$$
$$= \left[\begin{array}{c|cc|ccccc} \alpha'_1 & \beta'_{11} & \beta'_{12} & \gamma'_{11} & \gamma'_{12} & \gamma'_{13} & \gamma'_{14} & \gamma'_{15} \\ \alpha'_2 & \beta'_{21} & \beta'_{22} & \gamma'_{21} & \gamma'_{22} & \gamma'_{23} & \gamma'_{24} & \gamma'_{25} \end{array}\right],$$

where the entries of the matrices above are functions of indeterminate local coding vectors. The cut conditions imply that $\det(M_{ii})$ is not identically zero for $i = 1, \ldots, 3$, and furthermore that their product $\det(M_{11})\det(M_{22})\det(M_{33})$ is not identically zero.

Our solution proceeds as follows. We first identify a minimal structured subgraph G' of G with the following properties.

(i) There exists a path P'_{11}, from s_1 to t_1,

(ii) vertex disjoint paths P'_{21} and P'_{22} from s_2 to t_2,

(iii) path $P'_{1\to 2}$ from s_1 to t_2 and

(iv) path $P'_{2\to 1}$ from s_2 to t_1.

Again, G' is said to be minimal if the removal of any edge from it causes one of the above properties to fail. We note that it is possible that there do not exist any paths from s_1 to t_2 or from s_2 to t_1 in G. These situations are considered below.

Our analysis depends on the following topological properties of G'.

Case 1: The graph G' is such that

- there is no path from s_1 to t_2 in G', i.e., $P'_{1\to 2} = \emptyset$ (this happens only if there is no path from s_1 to t_2 in G), or
- there is no path from s_2 to t_1 in G', i.e., $P'_{2\to 1} = \emptyset$ (this happens only if there is no path from s_2 to t_1 in G), or

- there are paths $P'_{1\to 2}$ and $P'_{2\to 1}$ in G', and there are overlap segments between P'_{11} and $P'_{21} \cup P'_{22}$.

<u>*Case 2:*</u> The graph G' is such that

- there are paths $P'_{1\to 2}$ and $P'_{2\to 1}$ in G', and P'_{11} does not overlap with either P'_{21} or P'_{22}.

We emphasize that together Case 1 and Case 2 cover all the possible types of subgraphs for G'. Specifically, either $P'_{1\to 2} = \emptyset$ or $P'_{2\to 1} = \emptyset$. If both $P'_{1\to 2}$ and $P'_{2\to 1}$ exist in G', then either there are overlaps between P'_{11} and $P'_{21} \cup P'_{22}$ or there are not.

Theorem 3.3.10 *A multiple unicast instance with three sessions,* $< G, \{s_i - t_i\}_1^3, \{1,1,1\} >$*, with connectivity level* $[1\ 2\ 5]$ *is feasible.*

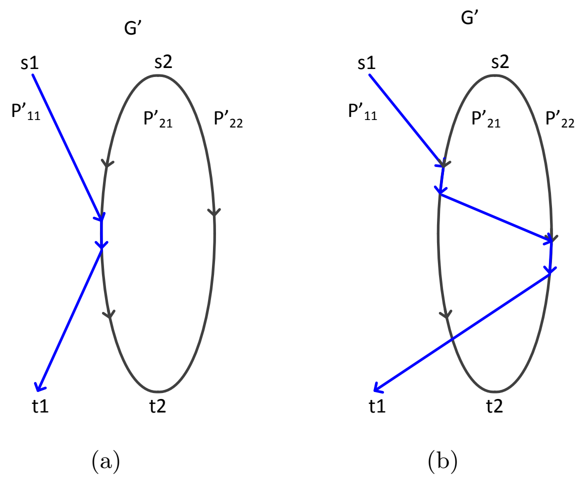


Figure 3.4 (a) Subgraph G' when P'_{11} overlap with P'_{21}. (b) Subgraph G' when P'_{11} overlap with both P'_{21} and P'_{22}.

proof: We break up the proof into two parts based on type of the subgraph G' that we can find in G.

<u>Proof when there exists a subgraph G' that satisfies the conditions of Case 1</u>

We perform random linear coding over the graph G over a large enough field. In the discussion below, we will leverage the fact that multivariate polynomials that are not identically zero, evaluate to a non-zero value w.h.p. under a uniformly random choice of the variables. This is needed at several places. By using standard union bound techniques, we can claim that our strategy works w.h.p.

In particular, in the discussion below, we assume that the matrices $M_{ii}, i = 1, \ldots, 3$ are full rank and design appropriate precoding vectors $\underline{\xi}$ and $\underline{\theta}$.

Decoding at t_1: For t_1 to decode X_1, we need to have $\alpha_1 \neq 0$ and the precoding constraints

$$[\beta_1 \ \beta_2]\underline{\xi} = 0, \text{ and} \tag{3.10}$$

$$[\gamma_1 \ \gamma_2 \ \gamma_3 \ \gamma_4 \ \gamma_5]\underline{\theta} = 0. \tag{3.11}$$

There are at least $q-1$ non-zero vectors $\underline{\xi}$ and q^4-1 non-zero vectors $\underline{\theta}$ that can be selected from the field of size q such that eq. (3.10) and eq. (3.11) are satisfied.

Decoding at t_2:

We begin by noting that since $rank(M_{22}) = 2$, $M_{22}\underline{\xi} \neq 0$, as long as $\xi \neq 0$. Next, we argue according to the topological structure of G'. The following possibilities can occur.

(i) There is no path from s_1 to t_2 in G', i.e., $P'_{1\to 2} = \emptyset$. This implies that $\alpha'_1 = \alpha'_2 = 0$ and in G, interference at t_2 only exists from s_3. Next, at least one component of $M_{22}\underline{\xi}$ will be non-zero, based on the argument above; w.l.o.g. assume that it is the first component. We choose $\underline{\theta}$ to satisfy

$$\underline{\gamma}'^{T}_{1}\underline{\theta} = 0. \tag{3.12}$$

It is evident that there are at least $q^3 - 1$ non-zero choices of $\underline{\theta}$ that satisfy the required constraints on $\underline{\theta}$ (eqs. (3.11) and (3.12)). Hence t_2 can decode.

(ii) There exists a path $P'_{1\to 2}$ from s_1 to t_2, i.e., $P'_{1\to 2} \neq \emptyset$.. This means that M_{21} is not identically zero. Here, we first align the interference from s_3 within the span of interference from s_1 by selecting an appropriate $\underline{\theta}$. We have the following lemma.

Lemma 3.3.11 *If $M_{21} \neq 0$, there exist at least $q^4 - 1$ choices for $\underline{\theta}$ such that*

$$M_{23}\underline{\theta} = cM_{21} \tag{3.13}$$

where c is some constant.

proof: First, w.l.o.g., we assume $\alpha'_2 \neq 0$. Hence, there exists a full rank 2×2 upper triangular matrix U such that $UM_{21} = [0 \ \alpha'_2]^T$. Next, define

$$[1 \ 0]UM_{23} = \underline{\widetilde{\gamma}}_1^{\prime T} \tag{3.14}$$

and choose $\underline{\theta}$ to satisfy $\underline{\widetilde{\gamma}}_1^{\prime T}\underline{\theta} = 0$ and set $c = \underline{\gamma}_2^{\prime T}\underline{\theta}/\alpha'_2$. Upon inspection, it can be verified that this implies that $UM_{23}\underline{\theta} = cUM_{21}$. As U is invertible, and there is only one linear constraint on $\underline{\theta}$, we have the required conclusion. ■

Thus, under this choice of $\underline{\theta}$, the interference from s_3 is aligned within the span of the interference from s_1 at t_2. Let $\underline{X} = [X_1 \ X_2 \ X_3]^T$. The received signal at t_2 is

$$[M_{21} \ M_{22}\underline{\xi} \ M_{23}\underline{\theta}]\underline{X} = [M_{21} \ M_{22}\underline{\xi}] \begin{bmatrix} X_1 + cX_3 \\ X_2 \end{bmatrix}. \tag{3.15}$$

The following claim concludes the decoding argument for t_2.

Claim 3.3.12 *If M_{21} is not identically zero, under random linear coding w.h.p., there exists a $\underline{\xi}$ such that rank$[M_{21} \ \ M_{22}\underline{\xi}] = 2$ and $[\beta_1 \ \beta_2]\underline{\xi} = 0$.*

proof: We will show that there exists an assignment of local coding vectors such that $\det[M_{21} \ \ M_{22}\underline{\xi}] \neq 0$. This will imply that w.h.p. under random linear coding, this property continues to hold.

Suppose that there is no path from s_2 to t_1 in G, i.e., $P'_{2\to 1} = \emptyset$ and $[\beta_1 \ \beta_2]$ is identically zero. This does not impose any constraint on $\underline{\xi}$. Next, M_{22} is full rank w.h.p. Hence, we can choose a $\underline{\xi}$ such that required condition is satisfied.

If there exists a path $P'_{2\to 1}$ from s_2 to t_1 in G', $[\beta_1 \ \ \beta_2]$ is not identically zero. W.l.o.g., we assume that β_1 is not identically zero. By Lemma C.0.2 (see Appendix), proving that $\det[M_{21} \ \ M_{22}\underline{\xi}] \neq 0$, is equivalent to checking that the determinant in (C.1) is not identically zero. Now we demonstrate that there exists a set of local coding vectors such that the determinant in (C.1) is non-zero. We consider the subgraph $G' = P'_{11} \cup P'_{21} \cup P'_{22} \cup P'_{1\to 2} \cup P'_{2\to 1}$ (identified above) - our choice of the coding vectors on all the other edges will be assigned to the zero vector. As both $P'_{1\to 2} \neq \emptyset$ and $P'_{2\to 1} \neq \emptyset$, we only consider the case where P'_{11} overlaps with $P'_{21} \cup P'_{22}$. We distinguish the following cases.

1. *P'_{11} overlaps with either P'_{21} or P'_{22}.* W.l.o.g., assume it is P'_{21}. First note that when P'_{11} overlap with one of P'_{21} and P'_{22} in G', there is a path from s_1 to t_2 and a path from s_2 to t_1 in $P'_{11} \cup P'_{21} \cup P'_{22}$. Hence, G' can be completely represented by $P'_{11} \cup P'_{21} \cup P'_{22}$. This is shown in Fig. 3.4(a). It is evident that we can choose coding coefficients such that

$$[\beta_1 \quad \beta_2] = [1 \ 0], \text{ and}$$
$$[M_{21} \quad M_{22}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.16}$$

   By substituting them into eq. (C.1), the determinant of $[M_{21} \quad M_{22}\underline{\xi}]$ is not zero.

2. *P'_{11} overlaps with both P'_{21} and P'_{22}.* Using a similar argument as above, G' can be completely represented by $P'_{11} \cup P'_{21} \cup P'_{22}$ if P'_{11} overlaps with both P'_{21} and P'_{22}. Note that there will be one overlap between P'_{11} and each of P'_{21} and P'_{22}. Otherwise, assume there are two overlaps between P'_{11} and P'_{21}, then some edges can be removed without contradicting the minimality of the graph G'. This is shown in Fig. 3.4(b). Assume P'_{11} overlap with P'_{21} first. We can find a set of coding coefficients such that

$$[\beta_1 \quad \beta_2] = [1 \ 1] \text{ and}$$
$$[M_{21} \quad M_{22}] = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \tag{3.17}$$

   By substituting them into eq. (C.1), the determinant of $[M_{21} \quad M_{22}\underline{\xi}]$ is not zero.

In both cases, therefore the required condition holds w.h.p. under random linear coding. ■

Terminal t_2 can decode since it can solve the system of equations specified by in eq. (3.15).

<u>Decoding at t_3</u>: At t_3, we need to decode X_3 in the presence of the interference from s_1 and s_2. The prior constraints on $\underline{\theta}$, namely (3.11) and (3.12) for case (i), or (3.11) and (3.13) for case (ii) allow at least $q^3 - 1$ choices for it. As M_{33} is full-rank, this implies that there are at least

$q^3 - 1$ corresponding distinct $M_{33}\underline{\theta}$ vectors. Next, for t_3 to decode X_3, from Lemma D.0.3, we need to have

$$M_{33}\underline{\theta} \notin span([M_{31} \quad M_{32}\underline{\xi}]). \tag{3.18}$$

Since there are at most q^2 vectors in $span([M_{31} \quad M_{32}\underline{\xi}])$, there are at least $q^3 - q^2 - 1 > 0$ choices for $\underline{\theta}$ such that all the required constraints on $\underline{\theta}$ are satisfied.

Proof when there exists a subgraph G' that satisfies the conditions of Case 2

As before, our overall strategy will be to use random linear network coding, however in certain cases we will need to make modifications to the code assignment. We argue based on the properties of the minimal structured subgraph G'. Recall that under Case 2, paths $P'_{1\to 2}$ and $P'_{2\to 1}$ exist and P'_{11} does not overlap with $P'_{21} \cup P'_{22}$. As the graph is structured, this implies that P'_{11}, P'_{21} and P'_{22} are all vertex disjoint.

Our first goal is to show that G' is topologically equivalent to one of the graphs shown in Figs. 3.5(a), 3.5(b) and 3.5(c). Towards this end, we color $P'_{11} \cup P'_{21} \cup P'_{22}$ black, the path $P'_{1\to 2}$ red, and the path $P'_{2\to 1}$ blue. In this process, certain edges will get a set of colors (which are a subset of $\{red, blue, black\}$). Note that there cannot be any edge that has the color $\{blue, red\}$. To see this, assume otherwise: then one could find a new path from s_1 to t_1 that overlaps $P'_{1\to 2}$ and $P'_{2\to 1}$ and delete at least one edge from P'_{11}, contradicting the minimality of G'. By similar arguments, $P'_{1\to 2}$ and $P'_{2\to 1}$ cannot overlap on $P'_{21} \cup P'_{22}$. Hence, paths $P'_{1\to 2}$ and $P'_{2\to 1}$ can only overlap if they also overlap with P'_{11}.

Next, we identify certain special edges in G'. As there is only one path going out of s_1, P'_{11} and $P'_{1\to 2}$ will overlap. A similar argument shows that P'_{11} and $P'_{2\to 1}$ will overlap. Likewise, $P'_{1\to 2}$ and $P'_{2\to 1}$ will overlap with P'_{21} or P'_{22}. Consider, the overlap between P'_{11} and $P'_{1\to 2}$. Using the minimality of G' it can be seen that there can be exactly one overlap segment between them; we identify the edge $\in P'_{11} \cap P'_{1\to 2}$ at the farthest distance from s_1, such that it has two outgoing edges belonging to exclusively P'_{11} and $P'_{1\to 2}$, and call it e_1. Similarly, we identify the edge $\in P'_{11} \cap P'_{2\to 1}$ that is closest to s_1, and call it e_3.

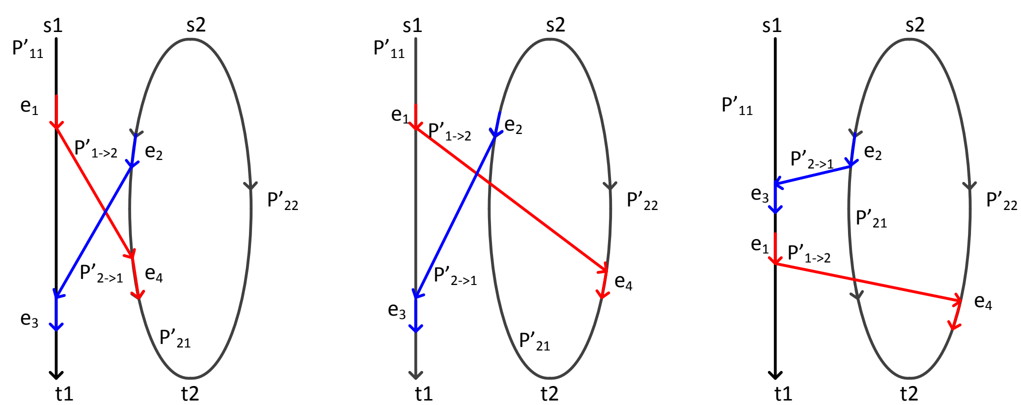
Next, consider the overlap between $P'_{1\to 2}$ and $P'_{21} \cup P'_{22}$. Once again, by minimality it holds that there is exactly one contiguous overlap segment between $P'_{1\to 2}$ and $P'_{21} \cup P'_{22}$, that can either be on P'_{21} or P'_{22}. We identify e_4 as the edge in $P'_{1\to 2} \cap (P'_{21} \cup P'_{22})$ that is closest to s_1. In a similar manner, e_2 is identified as the edge $P'_{2\to 1} \cap (P'_{21} \cup P'_{22})$ that is farthest away from s_2.

We now consider the possible orders of the edges $e_1, \ldots, e_4$. As e_1 and e_3 belong to P'_{11}, one of them has to be downstream of the other along P'_{11}. Consider the following cases.

- *e_3 is downstream of e_1 along P'_{11}.* If edges e_2 and e_4 lie on the same path $\in \{P'_{21}, P'_{22}\}$, we first note that e_4 has to be downstream of e_2 (by minimality, otherwise the segment between e_1 and e_3 along P'_{11} can be removed); the graph G' is topographically equivalent to Fig. 3.5(a). If e_2 and e_4 lie on different paths $\in \{P'_{21}, P'_{22}\}$, the graph G' is topographically equivalent to Fig. 3.5(b).

- *e_1 is downstream of e_3 along P'_{11}, or $e_1 = e_3$.* In this case e_2 and e_4 have to lie on different paths $\in \{P'_{21}, P'_{22}\}$. To see this, assume they both lie on P'_{21}: if e_4 is downstream of e_2, the minimality of G' does not hold (segment between e_2 and e_4 along P'_{21} can be removed), whereas if e_2 is downstream of e_4, the acyclicity of G' is contradicted. Therefore, the only possibility is that e_2 and e_4 lie on different paths $\in \{P'_{21}, P'_{22}\}$ and in this case G' is topographically equivalent to Fig. 3.5(c).

With the above arguments in place, it is clear that G' is topographically equivalent to one of the graphs in Fig. 3.5(a), 3.5(b) or 3.5(c).



(a) (b) (c)

Figure 3.5 Possible subgraphs G' when P'_{11} does not overlap with either P'_{21} or P'_{22}.

We now present our schemes for the different possibilities for G'. For the class of G' that fall in Fig. 3.5(a), it suffices to use the approach in the proof of Theorem 3.3.10. Namely, we use random linear network coding in the network and precoding at sources s_2 and s_3. As in this case $M_{21} \neq 0$, one needs to argue that $rank[M_{21}\ M_{22}\xi] = 2$. Following the line of argument used previously, we can do this by demonstrating a choice of local coding coefficients such that $[\beta_1\ \beta_2] = [1\ 0]$ and $[M_{21}\ \ M_{22}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. However, such an approach does not work when the subgraph G' belong to the class of graphs shown in Figs. 3.5(b) and 3.5(c). For instance, it is easy to observe that if we use random coding on Fig. 3.5(b), and precoding to cancel the X_2 component at t_1, then t_2 will receive a linear combination of X_1 and X_2 w.h.p., i.e., decoding X_2 at t_2 will fail. Accordingly, when G' looks like Fig. 3.5(b) or 3.5(c), we require a different scheme that we now present.

Modified random coding for cases in Fig 3.5(b) and Fig 3.5(c).

It is clear that the strategy of random linear network coding and precoding at the sources fails since the determinant of the matrix $[M_{21}\ M_{22}\underline{\xi}]$ is identically zero for the cases in Fig. 3.5(b) and 3.5(c). Thus, at the top level our approach is to modify the original graph G by removing certain edges and identifying a special node in G that is upstream of t_2. The transfer matrix on the two incoming edges of this special node can be expressed as $[\tilde{M}_{21}\ \tilde{M}_{22}\ \tilde{M}_{23}]$ such that the determinant of $[\tilde{M}_{21}\ \tilde{M}_{22}\underline{\xi}]$ is not identically zero. Thus, at this node it becomes possible to remove the effect of X_1 via deterministic coding. Accordingly, our strategy is to first perform random linear coding at all nodes except the special node and those that are downstream of the special node. Following this, we perform deterministic coding at the special node to cancel the effect of X_1, and random linear coding downstream of it. Finally, we argue based on the precoding constraints that each terminal can decode its desired message. In the discussion below we outline each of the steps and the corresponding analysis in a systematic manner.

Recall that based on G' (which is a subgraph of G) we have identified paths P'_{11}, P'_{21}, P'_{22} that are all vertex disjoint, paths $P'_{1\to 2}$ and $P'_{2\to 1}$ and edges $e_1, \dots, e_4$. At the outset we demonstrate that certain structures in G, need not be considered. In particular,

- if in G, there exists a path from s_1 to t_1 that has an overlap with $P'_{21} \cup P'_{22}$, it is clear that an alternate minimal subgraph G'' can be found that satisfies the conditions of Case 1.

- In G, a path from s_1 cannot have an overlap with $path(e_2 - e_3)$. To see this note that G' is a subgraph of G; therefore if path$(e_2 - e_3)$ exists in it, then it necessarily has to belong to a path P_{3i} from s_3 to t_3. We emphasize that the entire path including e_2 and e_3 have to belong to P_{3i} because by assumption all nodes in the graph have in-degree + out-degree at most 3. In a similar manner, the path from s_1 that overlaps with path$(e_2 - e_3)$ also needs to belong to path P_{3j}.If $i = j$, then it implies the existence of a path from s_1 to t_1 that has an overlap with $P'_{21} \cup P'_{22}$; however, this is explicitly ruled out by the discussion in the previous bullet. Thus, $i \neq j$; however, this is impossible since the paths P_{3i} and P_{3j} are edge disjoint.

Accordingly, in the discussion below, we will assume that the above scenarios do not occur.

Graph modification procedure for original graph G:

(i) Remove all edges downstream of e_2 on P'_{21} that have no overlap with a path from $\cup_{i=1}^{5} P_{3i}$.

(ii) Identify an edge, denoted e_{first} on P'_{22}, with the property that e_{first} is the edge closest to s_2 such that there exists a $path(s_1 - e_{first})$. Note that e_{first} exists due to the existence of path $P'_{1\to 2}$ in G.

(iii) Remove edges downstream of e_{first} while maintaining the following properties - (a) there exists a path from $e_{first} - t_2$, and (b) $max - flow(s_3 - t_3) = 5$. Rename P'_{22} to be $path(s_2 - e_{first} - t_2)$. It is important to note that after this procedure, removal of any edge downstream of e_{first} would cause either property (a) or (b) to fail.

(iv) Identify edge $e_{last} \in P'_{22}$ such that it is the edge closest to t_2 with the property that it has two incoming edges - $e'_1 \notin P'_{22}$ such that there exists $path(s_1 - e'_1)$ and $e'_2 \in P'_{22}$. Again e'_1 is guaranteed to exist as $P'_{1\to 2}$ exists in G.

As a consequence of the modification procedure, there is no overlap between $path(s_1 - e'_1)$ and P'_{22}. To see this, assume otherwise, i.e., an overlap segment, denoted E_{os} exists between $path(s_1 - e'_1)$ and P'_{22}. As e_{first} is the edge closest to s_2 such that there is a path between s_1 and e_{first}, it follows that E_{os} is downstream of e_{first} along P'_{22}. However, this contradicts the property of the modified graph after Step (iii) in the modification procedure above.

Next, note that $path(e_2 - e_3)$ has to overlap with a path from $\cup_{i=1}^{5} P_{3i}$ (as G is minimal) which means that the downstream neighboring edge of e_2 along P'_{21} cannot belong to any path in $\cup_{i=1}^{5} P_{3i}$ and will be removed in Step (i). Likewise the incoming edge of t_2 along P'_{21} will also be removed. At the end of the graph modification procedure, and using the observations made above, it is clear that we can identify a subgraph $\tilde{G}$ of G that is topologically equivalent to either Fig. 3.6(a) or 3.6(b).

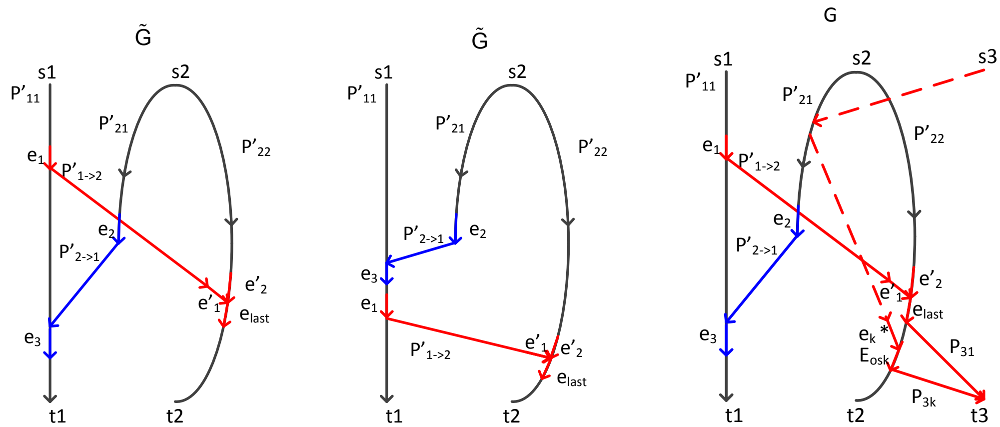
Next, we perform random linear coding over the modified graph except at edge e_{last} and all the edges downstream of e_{last}, and impose the precoding constraints $[\beta_1\ \beta_2]\underline{\xi} = 0$ and $\underline{\gamma}^T \underline{\theta} = 0$. This ensures that t_1 is satisfied. Furthermore, note that there is no path from e_{last} to t_1; therefore any code assignment on e_{last} and its downstream edges will not affect decoding at t_1.

For t_2 to decode X_2, we first demonstrate that by using deterministic coding for edge e_{last}, the X_1 component can be canceled while the X_2 component can be maintained on e_{last}. Note that e'_1 and e'_2 denote the incoming edges of e_{last}; we denote the transfer matrix to these two edges by $[\tilde{M}_{21}\ \tilde{M}_{22}\ \tilde{M}_{23}]$.

Claim 3.3.13 *For the network structures in Fig. 3.6(a) and Fig. 3.6(b), the determinant of* $[\tilde{M}_{21}\ \tilde{M}_{22}\underline{\xi}]$ *is not identically zero where* $\underline{\xi}$ *satisfies* $[\beta_1\ \beta_2]\underline{\xi} = 0$.

proof: Based on previous arguments, we have identified the subgraph $\tilde{G}$ of G that is topologically equivalent to either Fig. 3.6(a) or 3.6(b). By Lemma C.0.2, proving the claim is equivalent to showing that the determinant of eq. (C.1) is not identically zero. Based on $\tilde{G}$ it is evident that local coding vectors for the case of Fig. 3.6(a) can be chosen such that

$$[\beta_1\ \ \beta_2] = [1\ 0],\ \text{and}$$



(a) (b) (c)

Figure 3.6 Figures (a) and (b) denote possible subgraphs $\tilde{G}$ obtained after the graph modification procedure for G. Figure (c) shows an example of the overlap between the red $s_3 - t_3$ paths and P'_{22}.

$$[\tilde{M}_{21} \quad \tilde{M}_{22}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.19}$$

Similarly, for the case of Fig. 3.6(b) they can be chosen as

$$[\beta_1 \quad \beta_2] = [1 \ 0], \text{ and}$$

$$[\tilde{M}_{21} \quad \tilde{M}_{22}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.20}$$

Substituting the local coefficients into eq. (C.1) we have the required conclusion. ■

We now want to argue that t_2 can be satisfied. Note that edge e'_1 must belong to a path from $\mathcal{P}_3$, as the graph is minimal. Assume that there are k paths from $\mathcal{P}_3$ that overlap with $path(e_{last} - t_2)$; w.l.o.g. we assume that these are the paths $P_{31}, \ldots, P_{3k}$.

Next, we note that there can be at most one overlap between a path P_{3j} and $path(e_{last} - t_2)$. This is due to Step (iii) of the graph modification procedure, where we removed edges downstream of e_{first}, (and hence e_{last}) such that the $max - flow(s_3 - t_3) = 5$ and there is path between $e_{first} - t_2$. If there are multiple overlaps between P_{3j} and $path(e_{last} - t_2)$, this would mean that there exists at least one edge that was not removed by Step (iii). As depicted in Fig. 3.6(c), we denote the overlap segments as $E_{os1}, \ldots, E_{osk}$, where E_{osj} is upstream of $E_{os(j+1)}$ for $j = 1, \ldots, k-1$ along P'_{22}. Also note that the first edge of E_{os1} is e_{last}.

The next step in the code assignment is to use deterministic local coding coefficients so that the transmitted symbol on e_{last} does not have an X_1 component. Note that it is guaranteed to have an X_2 component by the Claim 3.3.13 above. Following this, we again use random linear coding on edges downstream of e_{last}. By the definition of e_{last} there is no edge $\in P'_{22}$ downstream of e_{last} that is reachable from s_1. Thus all coding vectors along P'_{22} downstream of e_{last} do not have an X_1 component. Let the coding vector on the edge $\in E_{osk}$ closest to t_2 be denoted by $[0 \mid \underline{\hat{\beta}}^T \mid \underline{\hat{\gamma}}^T]$, where it is evident that $\hat{\beta} \neq 0$ w.h.p. We enforce the precoding constraint $\underline{\hat{\gamma}}^T \underline{\theta} = 0$. This satisfies t_2.

Finally, we discuss the decoding at t_3. Consider the overlap segments $E_{os1}, \dots, E_{osk}$ discussed above. Each of these overlap segments has an incoming edge that does not lie on P'_{22} (the other has to be on P'_{22}). We denote these edges by $e^*_i, i = 1, \dots, k$, where we emphasize that $e^*_1 = e'_1$. Let the edges entering t_3 on paths $P_{3(k+1)}, \dots, P_{35}$ be denoted $e^*_{k+1}, \dots, e^*_5$. Denote the transfer matrix on the edges $e^*_1, \dots, e^*_5$ by $[\hat{M}_{31} \mid \hat{M}_{32} \mid \hat{M}_{33}]$. Note that with high probability it holds that $rank(\hat{M}_{33}) = 5$, since the max-flow from s_3 to these set of edges is 5.

Next consider the rank of the coding vectors on edges $\{e_{last}, e^*_2, e^*_3, e^*_4, e^*_5\}$. For the sake of argument suppose that we remove the row of $\hat{M}_{33}$ corresponding to e^*_1 and replace it with the corresponding row of e_{last}. As we used a deterministic code assignment for edge e_{last} the rank of the updated $\hat{M}_{33}$ may drop to four, however it will be no less than four since it has four linearly independent row vectors.

It can be seen that further random linear coding downstream of e_{last} will therefore be such that $rank(M_{33})$ (recall that $[M_{31}|M_{32}|M_{33}]$ is the transfer matrix to t_3) is at least four w.h.p. Moreover, it can be seen that the information on E_{osk} also reaches t_3, thus t_3 can decode X_2. Therefore at t_3 over the other four incoming edges we have a system of equations specified by the matrix $[\breve{M}_{31}|\breve{M}_{33}]$ (of dimension 4×6) with unknowns X_1 and X_3. Furthermore $rank(\breve{M}_{33}) \geq 3$. The constraints on $\underline{\theta}$ thus far dictate that there are $q^3 - 1$ non-zero choices for it. As shown in the appendix (cf. Lemma E.0.4) this implies that there are at least $q^2 - 1$ distinct values for $\breve{M}_{33}\underline{\theta}$. For decoding X_3 at t_3, from Lemma D.0.3, we need to have

$$\breve{M}_{33}\underline{\theta} \notin span(\breve{M}_{31}). \tag{3.21}$$

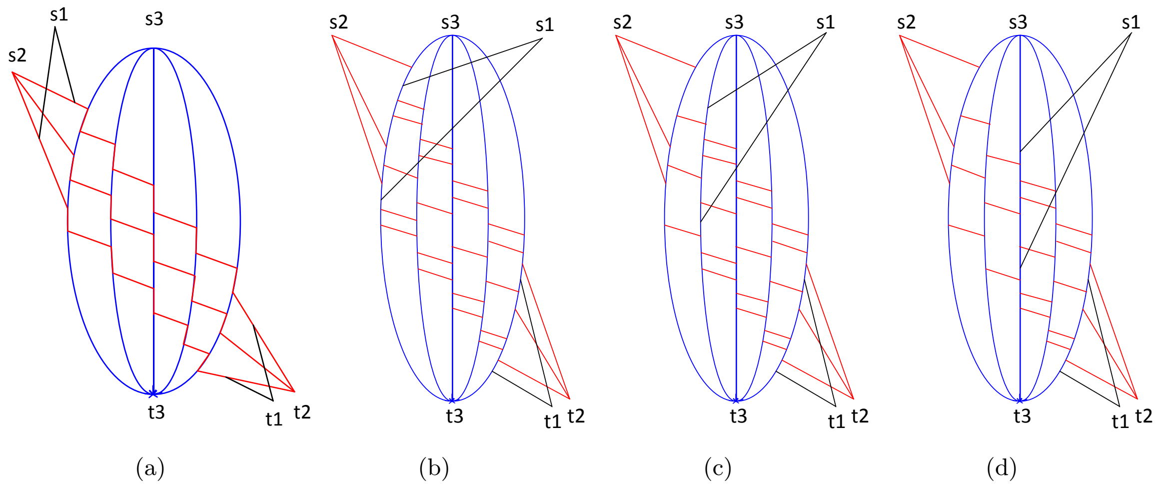


Figure 3.7 a) Level-1 network. b) Level-2 network. c) Level-3 network. d) Level-4 network.

As there are at most q vectors in the span of M_{31}, it follows that there are at least $q^2 - q - 1 > 0$ non-zero values of $\underline{\theta}$ such that t_3 can be satisfied. ■

3.4 Simulation results

Our feasibility results thus far have been for the case of unit-rate transmission over networks with unit-capacity edges. In this section, we present simulation results that demonstrate that these can also be used for networks with higher edge capacities, that can potentially support higher rates for the connections. The main idea is to pack multiple basic feasible solutions along with fractional routing solutions to achieve a higher throughput. The packing can be achieved by formulating appropriate integer linear programs. We compared these results to the case of solutions that can be achieved via pure fractional routing.

We applied our technique to several classes of networks. We did not see a benefit in the case of networks generated using random geometric graphs (this is consistent with previous results [9]). We have found that our techniques are most powerful for networks where the paths between the various $s_i - t_i$ pairs have significant overlap. Accordingly, we experimented with four classes of networks (shown in Fig. 3.7) with varying levels of overlap between the different source-terminal pairs. The level-1 network (Fig. 3.7(a)) has the maximum overlap between the $s_1 - t_1$ paths and the other paths; the overlap decreases with an increase in the level number of the network. The edge capacities in the networks were chosen randomly and independently with distributions as explained below. We conducted two sets of simulations.

- *Simulation 1.* Let C denote the edge capacity. For the level-1 network for the black edges we chose $P(C = 1) = 0.25, P(C = 2) = 0.4, P(C = 3) = 0.35$; for the other edges, $P(C = 1) = 0.15, P(C = 2) = 0.6, P(C = 3) = 0.25$. In the other networks we chose $P(C = 1) = 0.15, P(C = 2) = 0.6, P(C = 3) = 0.25$ for all the edges. Thus in this set of simulations, the maximum edge capacity is three. We generated 300 networks from these distributions and compared the performance of our schemes with pure fractional routing. The results shown in the first row of Table 3.1 indicate that the level-1 network has the maximum number of instances where a difference in the throughput was observed; both [1 2 5] and [2 2 4] structures appear here. For the other networks, the [2 2 4] structure appeared most often. The second row of Table 3.1 records the average performance improvement when there was a difference between our scheme and routing; it varies between 4.9% to 5.59%.

- *Simulation 2.* In this set of simulations we increased the average edge capacity. For the level-1 network for the black edges we chose $P(C = 5) = 0.25, P(C = 6) = 0.4, P(C = 7) = 0.35$; for the other edges, $P(C = 5) = 0.15, P(C = 6) = 0.6, P(C = 7) = 0.25$. In the other networks we chose $P(C = 5) = 0.15, P(C = 6) = 0.6, P(C = 7) = 0.25$ for all the edges. Again, we generated 300 networks from these distributions and compared the performance of our schemes with pure fractional routing. The results shown in the third row of Table 3.1 indicate that in this higher capacity simulation, the number of networks where our schemes outperform pure routing is significantly higher. For instance for the level-2 and level-3 networks more than 50% of the networks showed an increase in the throughput using our methods. Another interesting point, is that one observes an increased gap for level-3 networks compared to the other cases. The fourth row of Table 3.1 records the average performance improvement when there was a difference between our scheme and routing; it varies between 0.45% to 1.16%.

We found that though there were instances of all the structures being packed by the ILP,

Table 3.1 Proportions of networks with differences and performance improvement

Network	Level-1	Level-2	Level-3	Level-4
Simulation 1 proportions	5.33%	2.33%	1%	0
Performance improvement	5.59%	5.06%	4.90%	-
Simulation 2 proportions	47%	53%	80.67%	2.33%
Performance improvement	1.16%	1.31%	1.36%	0.45%

the majority were [2 2 4] structures. For the level-4 network, since [2 2 4] structure cannot be packed effectively, there is a significant drop in the proportions of networks that exhibit a difference with respect to routing as compared to the level-3 and level-4 networks. There were significant advantages in our approach for the case of networks with higher edge capacities as in these networks the chance of packing our basic feasible structures is higher. The average performance improvement obtained when there was a difference between our schemes and routing is not very high. We remark that the complexity of running the ILP increases with higher edge capacities and that was a limiting factor in our experiments; the performance improvement may be higher for large scale examples. Overall, our results indicate that there is a benefit to using our techniques even for networks with higher capacities, where the different source-terminal paths have a large overlap.

3.5 Conclusions

In this work we considered the three-source, three-terminal multiple unicast problem for directed acyclic networks with unit capacity edges. Our focus was on characterizing the feasibility of achieving unit-rate transmission for each session based on the knowledge of the connectivity level vector. For the infeasible instances we have demonstrated specific network topologies where communicating at unit-rate is impossible, while for the feasible instances we have designed constructive linear network coding schemes that satisfy the demands of each terminal. Our schemes are non-asymptotic and require vector network coding over at most two time units. Our work leaves out one specific connectivity level vector, namely [1 2 4] for which we have been unable to provide either a feasible network code or a network topology where communicating at unit rate is impossible. Our experimental results indicate that there are benefits to using our techniques even for networks where the edges have higher and potentially different capacities. Specifically, our basic feasible solutions can be packed along with routing to obtain a higher throughput.

CHAPTER 4. NETWORK CODING FOR TWO UNICAST SESSIONS

4.1 System model

We consider a network represented by a directed acyclic graph $G = (V, E)$. There is a source set $S = \{s_1, s_2\} \in V$ in which each source observes a random process (the processes are independent) with a discrete integer entropy, and there is a terminal set $T = \{t_1, t_2\} \in V$ in which t_i needs to uniquely recover the information transmitted from s_i at rate R_i. Each edge $e \in E$ has unit capacity and can transmit one symbol from a finite field of size q. If a given edge has a higher capacity, it can be divided into multiple parallel edges with unit capacity. Without loss of generality (W.l.o.g.), we assume that there is no incoming edge into source s_i, and no outgoing edge from terminal t_i. By Menger's theorem, the minimum cut between sets $S_{N_1} \subseteq S$ and $T_{N_2} \subseteq T$ is the number of edge disjoint paths from S_{N_1} to T_{N_2}, and will be denoted by $k_{N_1-N_2}$ where $N_1, N_2 \subseteq \{1, 2\}$. For two unicast sessions, we define the *cut vector* as the vector of the cut values k_{1-1}, k_{2-2}, k_{1-2}, k_{2-1}, k_{12-1}, k_{12-2}, k_{1-12}, k_{2-12} and k_{12-12}.

The network coding model in this work is based on [3]. Assume that source s_i needs to transmit at rate R_i. Then the random variable observed at s_i is denoted as $X_i = (X_{i1}, X_{i2}, \cdots, X_{iR_i})$, where each X_{ij} is an element of the finite field of size q denoted by $GF(q)$. For linear network codes, the signal on an edge (i, j) is a linear combination of the signals on the incoming edges on i or a linear combination of the source signals at i. Let Y_{e_n} ($tail(e_n) = k$ and $head(e_n) = l$) denote the signal on edge $e_n \in E$. Then, we have

$$Y_{e_n} = \sum_{\{e_m | head(e_m)=k\}} f_{m,n} Y_{e_m} \text{ if } k \in V \setminus \{s_1, s_2\}, \text{ and}$$

$$Y_{e_n} = \sum_{j=1}^{R_i} a_{ij,n} X_{ij} \text{ if } X_i \text{ is observed at } k.$$

The local coding vectors $a_{ij,n}$ and $f_{m,n}$ are also chosen from $GF(q)$. We can also express Y_{e_n} as $Y_{e_n} = \sum_{j=1}^{R_1} \alpha_{j,n} X_{1j} + \sum_{j=1}^{R_2} \beta_{j,n} X_{2j}$. The global coding vector of Y_{e_n} is $[\alpha_n, \beta_n] = [\alpha_{1,n}, \cdots, \alpha_{R_1,n}, \beta_{1,n}, \cdots, \beta_{R_2,n}]$. We are free to choose an appropriate value of the field size q.

In this work, we present an achievable rate region given the cut vector; namely, k_{1-1}, k_{2-2}, k_{1-2}, k_{2-1}, k_{12-1}, k_{12-2}, k_{1-12}, k_{2-12} and k_{12-12}. W.l.o.g, we assume that there are k_{i-ij} outgoing edges from s_i and k_{ij-i} incoming edges into t_i. If this is not the case one can always introduce an artificial source (terminal) node connected to the original source (terminal) node by k_{i-ij} (k_{ij-i}) edges. It can be seen that the new network has the same cut vector as the original network.

4.2 Achievable rate region for given $k_{12-1}, k_{12-2}, k_{1-1}, k_{2-2}, k_{1-2}$, and k_{2-1}

We first consider the case that a subset of the cut values in the cut vector are available, namely, $k_{12-1}, k_{12-2}, k_{1-1}, k_{2-2}, k_{1-2}$, and k_{2-1}. Suppose for now that only t_1 is interested in recovering both the random variables X_1 and X_2 which are observed at s_1 and s_2 respectively. Denote the rate from s_1 to t_1 and s_2 to t_1 as R_{11} and R_{12}. The rate pairs (R_{11}, R_{12}) are achieved via routing [36] and the corresponding capacity region C_{t_1} is given by

$$C_{t_1} = \{R_{11} \leq k_{1-1}, \;\; R_{12} \leq k_{2-1}, \;\; R_{11} + R_{12} \leq k_{12-1}\}.$$

The capacity region C_{t_2} for t_2 can be drawn in a similar manner (an example is shown in Fig. 4.1(a)). We also find the boundary points $W_{1u}, W_{1l}, W_{2u}, W_{2l}$[1] such that their coordinates are $W_{1u} = (k_{12-1} - k_{2-1}, k_{2-1}), W_{1l} = (k_{1-1}, k_{12-1} - k_{1-1}), W_{2u} = (k_{12-2} - k_{2-2}, k_{2-2}), W_{2l} = (k_{1-2}, k_{12-2} - k_{1-2})$. A simple achievable rate region for our problem can be arrived at by multicasting both sources X_1 and X_2 to both the terminals t_1 and t_2.

Lemma 4.2.1 *Rate pairs (R_1, R_2) belonging to the following set $\mathcal{B}$ can be achieved for two unicast sessions.*

$$\mathcal{B} = \{R_1 \leq \min(k_{1-2}, k_{1-1}), R_2 \leq \min(k_{2-1}, k_{2-2}), R_1 + R_2 \leq \min(k_{12-1}, k_{12-2})\}.$$

[1] subscripts l and u are meant to denote lower and upper.

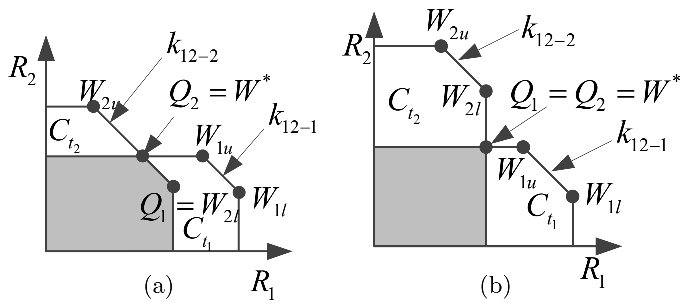


Figure 4.1 (a) An example of C_{t_1} and C_{t_2} when the multicast region shaded is pentagonal. (b) Another example where the multicast region is rectangular.

proof: We multicast both the sources to each terminal. This can be done using the multi-source multi-sink multicast result (Thm. 8 in [3]). ■

Subsequently we will refer to region $\mathcal{B}$ achieved by multicast as the *multicast region* (the grey region in Fig. 4.1(a)). It can be observed that if the cut values are such that

$$\min(k_{1-2}, k_{1-1}) + \min(k_{2-1}, k_{2-2}) \leq \min(k_{12-1}, k_{12-2}), \tag{4.1}$$

then the region is rectangular (Fig. 4.1(b)), otherwise, it is pentagonal (Fig. 4.1(a)).

We now move on to precisely formulating the problem. Let Z_i denote the received vector at t_i, X_i denote the transmitted vector at s_i, and H_{ij} denote the transfer function from s_j to t_i. Let M_i denote the encoding matrix at s_i, i.e., M_i is the transformation from X_i to the transmitted symbols on the outgoing edges from s_i. In our formulation, we will let the length of X_i to be k_{i-i}, i.e., the maximum possible. For transmission at rates R_1 and R_2, we introduce precoding matrices $V_i, i = 1, 2$ of dimension $R_i \times k_{i-i}$, so that the overall system of equations is as follows.

$$\begin{aligned} Z_1 &= H_{11}M_1V_1X_1 + H_{12}M_2V_2X_2, \\ Z_2 &= H_{21}M_1V_1X_1 + H_{22}M_2V_2X_2. \end{aligned} \tag{4.2}$$

We say that t_i can receive information at rate R_i from s_i if it can decode V_iX_i perfectly; each entry in V_i is either 0 or 1. The row dimension of the V_i's can be adjusted to obtain different rate vectors. Under random linear network coding, it can be shown that there exist local coding vectors over a large enough field such that the ranks of the different matrices

Table 4.1 dimension and rank of matrices

matrix	H_{11}	H_{12}	$[H_{11} \quad H_{12}]$	H_{21}	H_{22}	$[H_{21} \quad H_{22}]$
dimension	$k_{12-1}\times$ k_{1-12}	$k_{12-1}\times$ k_{2-12}	$k_{12-1}\times$ $(k_{1-12}+k_{2-12})$	$k_{12-2}\times$ k_{1-12}	$k_{12-2}\times$ k_{2-12}	$k_{12-2}\times$ $(k_{1-12}+k_{2-12})$
rank	k_{1-1}	k_{2-1}	k_{12-1}	k_{1-2}	k_{2-2}	k_{12-2}

in the first row of Table 4.1 are given by the corresponding entries in the third row, which correspond to the maximum possible. Furthermore, by the multi-source multi-sink multicast result [3], when $(R_1, R_2) \in \mathcal{B}$ these matrices are such that $[H_{11}M_1 \quad H_{12}M_2]$ is a full column rank matrix of dimension $k_{12-1} \times (R_1 + R_2)$, and $[H_{21}M_1 \quad H_{22}M_2]$ is a full column rank matrix of dimension $k_{12-2} \times (R_1 + R_2)$. In Table 4.1, for instance since the minimum cut between s_1 and t_1 is k_{1-1}, we know that the maximum rank of H_{11} is k_{1-1}. Using the formalism of [3], we can conclude that there is a square submatrix of H_{11} of dimension $k_{1-1} \times k_{1-1}$ whose determinant is not identically zero. Such appropriate submatrices can be found for each of the matrices in the first row of Table 4.1. This in turn implies that their product is not identically zero and therefore using the Schwartz-Zippel lemma [35], we can conclude that there exists an assignment of local coding vectors over a sufficiently large finite field so that the rank of all the matrices is simultaneously the maximum possible. While, the Schwartz-Zippel lemma requires random choice of the local coding vectors, the probability of success in the algorithm can be made arbitrarily close to one if the field size is chosen large enough, or through repeated trials, hence it runs in random polynomial time. For the rest of the paper, we assume that such a choice of local coding vectors has been made. Our arguments will revolve around appropriately modifying source encoding matrices M_1 and M_2.

Note that in general the multicast region has a pentagonal shape (see Fig. 4.1(a)). Two points on this pentagon (denoted as Q_1 and Q_2) are of specific interest. At point Q_1, we denote the achievable rate pair by (R_1^*, R_2^*) where

$$R_1^* = \min(k_{1-2}, k_{1-1}), \text{ and}$$
$$R_2^* = \min(\min(k_{2-1}, k_{2-2}), \min(k_{12-1}, k_{12-2}) - R_1^*).$$

If the region is pentagonal, then $R_1^* = \min(k_{1-2}, k_{1-1})$ and $R_2^* = \min(k_{12-1}, k_{12-2}) - R_1^*$.

Likewise at point Q_2, we denote the achievable rate pair by (R_1^{**}, R_2^{**}) where

$$R_1^{**} = \min(\min(k_{1-2}, k_{1-1}), \min(k_{12-1}, k_{12-2}) - R_2^{**}), \text{ and}$$
$$R_2^{**} = \min(k_{2-1}, k_{2-2}).$$

If the region is pentagonal, then $R_1^{**} = \min(k_{12-1}, k_{12-2}) - R_2^{**}$ and $R_2^{**} = \min(k_{2-1}, k_{2-2})$. If the region is rectangular, then $Q_1 = Q_2$, and $R_1^{*} = R_1^{**} = \min(k_{1-2}, k_{1-1})$ and $R_2^{*} = R_2^{**} = \min(k_{2-1}, k_{2-2})$. In Fig. 4.1(a), these boundary points are $Q_1 = W_{2l}$ and $Q_2 = W^*$, and the multicast region is pentagonal. Another example is shown in Fig. 4.1(b) where $Q_1 = Q_2$ and the multicast region is rectangular.

In what follows, we will present our arguments towards increasing the value of R_1 and R_2 to achieve points that are near Q_1 but do not belong to $\mathcal{B}$. In this paper we refer to $k_{1-2} + k_{2-1}$ as a measure of the interference in the network and in the subsequent discussion present achievable regions based on its value. We emphasize though that this is nomenclature used for ease of presentation. Indeed a high value of k_{1-2} does not necessarily imply that there is a lot of interference at t_2, since the network code itself dictates the amount of interference seen by t_2. The following lemma will be used extensively.

Lemma 4.2.2 *Consider a system of equations $Z = H_1X_1 + H_2X_2$, where X_1 is a vector of length l_1 and X_2 is a vector of length l_2 and $Z \in span([H_1 \quad H_2])$*[2]*. The matrix H_1 has dimension $z_t \times l_1$, and rank $l_1 - \sigma$, where $0 \leq \sigma \leq l_1$. The matrix H_2 is full rank and has dimension $z_t \times l_2$ where $z_t \geq (l_1 + l_2 - \sigma)$. Furthermore, the column spans of H_1 and H_2 intersect only in the all-zeros vectors, i.e. $span(H_1) \cap span(H_2) = \{0\}$. Then there exists a unique solution for X_2.*

proof: Because $Z \in span([H_1 \quad H_2])$, there exists X_1 and X_2 such that $Z = H_1X_1 + H_2X_2$. Now assume there is another set of X_1' and X_2' such that $Z = H_1X_1' + H_2X_2'$. This implies

$$H_1(X_1 - X_1') = H_2(X_2 - X_2'). \tag{4.3}$$

Because $span(H_1) \cap span(H_2) = \{0\}$, both sides of eq. (D.1) are zero. Furthermore, since H_2 is a full rank matrix, $X_2 = X_2'$, i.e., the solution for X_2 is unique. ■

[2] *Throughout the paper, $span(A)$ refers to the column span of A.*

We next define the achievable rate region which will be used in the rest of the paper.

Definition 4.2.3 *A rate point* (R_1, R_2) *is said to lie in the achievable rate region* $\mathcal{R}_A$ *if there exist full column rank source encoding matrices* M_1 *and* M_2 *where* $rank(M_1) = R_1$ *and* $rank(M_2) = R_2$ *such that*

$$\begin{aligned} &rank(H_{11}M_1) = rank(M_1),\ rank(H_{22}M_2) = rank(M_2),\ \textit{and} \\ &span(H_{i1}M_1) \cap span(H_{i2}M_2) = \{0\}\ \textit{for } i = 1, 2. \end{aligned} \tag{4.4}$$

The condition above will be referred in the remainder of the paper as the achievable condition.

It can be observed that the multicast region $\mathcal{B}$ is a subset of $\mathcal{R}_A$.

4.2.1 Low interference case - $k_{1-2} + k_{2-1} \leq \min(k_{12-1}, k_{12-2})$

Note that it always holds that $k_{2-1} + k_{1-1} \geq k_{12-1}$ and $k_{1-2} + k_{2-2} \geq k_{12-2}$. Together with the low interference condition, this implies that $k_{1-1} \geq k_{1-2}$ and $k_{2-2} \geq k_{2-1}$. It follows that the multicast region is a rectangle since eq. (4.1) is satisfied and $R_1^* = k_{1-2}, R_2^* = k_{2-1}$. Furthermore, $Q_1 = Q_2 = W^*$ as shown in the example in Fig. 4.1(b).

Our solution strategy is to first consider the encoding matrices M_1 and M_2 at the point Q_1, and to introduce a new encoding matrix at s_1, denoted M_1' (with $R_1^* + \delta$ columns) such that $span(H_{11}M_1') \cap span(H_{12}) = \{0\}$. As shown below, this will allow t_1 to decode from s_1 at rate $R_1^* + \delta$ and t_2 to decode from s_2 at rate R_2^*. After the modification, each t_i is guaranteed to decode at the appropriate rate from s_i. A similar argument applies for R_2^* to arrive at the achievable rate region. At the point Q_1, as both terminals can decode both sources, it holds that

$$\begin{aligned} &rank(H_{i1}M_1) = k_{1-2}, rank(H_{i2}M_2) = k_{2-1},\ \text{and} \\ &span(H_{i1}M_1) \cap span(H_{i2}M_2) = \{0\} \text{ for } i = 1, 2. \end{aligned}$$

Before stating the main result, we present the following lemma.

Lemma 4.2.4 *Rate Increase Lemma. Consider a rate point* $(R_1, R_2) \in \mathcal{R}_A$ *with corresponding matrices* M_1 *and* M_2 *such that (1)* $rank([H_{11} \quad H_{12}M_2]) = r > rank([H_{11}M_1 \quad H_{12}M_2]) =$

$R_1 + \Delta$, where $rank(H_{12}M_2) = \Delta \le R_2$ and (2) $rank([H_{21}M_1]) = rank(H_{21})$. There exist matrices M_1' and M_2' such that t_1 can decode at rate $r - \Delta$ and t_2 can decode at rate R_2.

proof: We first prove that if M_1 and M_2 satisfy Condition (1), then there exist a series of full rank matrices $\bar{M}_1^{(n)} = [\tilde{M}_1^{(n)} \quad M_1]$ of dimension $k_{1-12} \times (n + R_1)$ such that $rank([H_{11}\bar{M}_1^{(n)} \quad H_{12}M_2]) = R_1 + \Delta + n$, $0 \le n \le (r - R_1 - \Delta)$. We prove this part by induction. When $n = 0$, $\bar{M}_1^{(0)} = M_1$, $rank([H_{11}\bar{M}_1^{(0)} \quad H_{12}M_2]) = R_1 + \Delta$.

Assume that when $n = l \le r-1-R_1-\Delta$, $\bar{M}_1^{(n)}$ can be found such that $rank([H_{11}\bar{M}_1^{(l)} \quad H_{12}M_2]) = R_1 + \Delta + l$. When $n = l + 1 \le r - R_1 - \Delta$, if there does not exist an $\bar{M}_1^{(l+1)}$, all the columns in $[H_{11} \quad H_{12}M_2]$ are linear combinations of $[H_{11}\bar{M}_1^{(l)} \quad H_{12}M_2]$, which contradicts the fact that $rank([H_{11} \quad H_{12}M_2]) = r > r - 1 \ge l + R_1 + \Delta$. Hence, there must exist a series of full rank matrices $\bar{M}_1^{(n)}$ such that $rank([H_{11}\bar{M}_1^{(n)} \quad H_{12}M_2]) = R_1 + \Delta + n$ is satisfied when $0 \le n \le r - R_1 - \Delta$.

Next, we prove that t_1 can decode at rate $r - \Delta$ and t_2 can decode at rate R_2 using $M_1' = \bar{M}_1^{(r-R_1-\Delta)}$ and $M_2' = M_2$.

Decoding at t_1: Since M_1' is a full rank matrix of dimension $k_{1-12} \times (r - \Delta)$, it also satisfies (i) $rank(H_{11}M_1') = r - \Delta$ and (ii) $span(H_{11}M_1') \cap span(H_{12}M_2) = \{0\}$ because of the following argument. We have

$$\begin{aligned} r &= rank([H_{11}M_1' \quad H_{12}M_2]) \le rank([H_{11}M_1']) + rank([H_{12}M_2]) \\ &\le rank(M_1') + rank(H_{12}M_2) = r - \Delta + \Delta = r. \end{aligned}$$

Then all the inequalities become equalities and (i) and (ii) are satisfied. Then by Lemma D.0.3 and the above conditions, t_1 can decode at rate $r - \Delta$.

Decoding at t_2: From Condition (2), we have $span(H_{21}M_1) = span(H_{21})$ (see Lemma F.0.5 in the Appendix). Furthermore, since $span(M_1) \subseteq span(M_1')$, we have $span(H_{21}M_1) \subseteq span(H_{21}M_1') \subseteq span(H_{21})$. This implies that $span(H_{21}M_1) = span(H_{21}M_1') = span(H_{21})$. Furthermore, since $span(H_{21}M_1) \cap span(H_{22}M_2) = \{0\}$, we also have $span(H_{21}M_1') \cap span(H_{22}M_2) = \{0\}$. Then by Lemma D.0.3 and the fact that $H_{22}M_2$ is full rank, t_2 can decode at rate R_2. ■

Lemma 4.2.5 *If $k_{1-2}+k_{2-1} \leq \min(k_{12-1}, k_{12-2})$, the rate pair in the following region can be achieved.*

$$R_1 \leq k_{12-1} - k_{2-1}, \qquad R_2 \leq k_{12-2} - k_{1-2}.$$

proof: In this case, $(R_1^*, R_2^*) = (k_{1-2}, k_{2-1})$ is the boundary point $Q_1 = Q_2$. Let M_1 and M_2 denote the source encoding matrices at Q_1.

First, note that $rank(H_{12}M_2) = rank(H_{12}) = k_{2-1}$, which implies that $span(H_{12}) = span(H_{12}M_2)$. Therefore $rank([H_{11} \ \ H_{12}]) = rank([H_{11} \ \ H_{12} \ \ H_{12}M_2] = rank([H_{11} \ \ H_{12}M_2])$. This implies that $rank([H_{11} \ \ H_{12}M_2]) = k_{12-1} \geq k_{1-2} + k_{2-1} = rank([H_{11}M_1 \ \ H_{12}M_2])$ since by assumption $k_{1-2}+k_{2-1} \leq \min(k_{12-1}, k_{12-2})$. Moreover, $rank(H_{21}M_1) = rank(H_{21}) = k_{1-2}$. Therefore by the Rate Increase Lemma, we can achieve rate point $(R_1 = k_{12-1} - k_{2-1}, R_2 = k_{2-1})$. Using a similar argument, we can further increase R_2 such that rate pair $(k_{12-1} - k_{2-1}, k_{12-2} - k_{1-2})$ can be achieved. This region is the hatched gray region in Fig. 4.2. ■

This implies that the point $W' = (k_{12-1} - k_{2-1}, k_{12-2} - k_{1-2})$ is achievable. Also note that since we applied the Rate Increase Lemma, we have $rank([H_{11}M_1' \ \ H_{12}M_2]) = rank([H_{11} \ \ H_{12}M_2])$. Next, we consider the scenario in which rates can be traded off between the two unicast sessions.

Lemma 4.2.6 *Rate Exchange Lemma – 1-1 tradeoff. Consider a rate point $(R_1, R_2) \in \mathcal{R}_A$ with corresponding matrices M_1 and M_2.*

(a) If M_1 and M_2 satisfy (1) $rank([H_{11}M_1 \ \ H_{12}M_2]) = rank([H_{11} \ \ H_{12}M_2]) = r$, where $R_1 + R_2 \geq r$, and (2) $rank(H_{21}M_1) = rank(H_{21})$, there exist M_1' and M_2' such that t_1 can decode at rate $\min(R_1 + 1, k_{1-1})$ and t_2 can decode at rate $\max(R_2 - 1, 0)$.

(b) If M_1 and M_2 satisfy (1) $rank([H_{11} \ \ H_{12}M_2]) = r > rank([H_{11}M_1 \ \ H_{12}M_2]) = R_1 + \Delta$, where $rank(H_{12}M_2) = \Delta \leq R_2$, and (2) $rank(H_{21}M_1) < rank(H_{21})$, there exist M_1' and M_2' such that t_1 can decode at rate $\min(R_1 + 1, k_{1-1})$ and t_2 can decode at rate $\max(R_2 - 1, 0)$.

Lemma 4.2.7 *Rate Exchange Lemma – 1-2 tradeoff. Consider a rate point $(R_1, R_2) \in \mathcal{R}_A$ with corresponding matrices M_1 and M_2. If M_1 and M_2 satisfy (1) $rank([H_{11}M_1 \ \ H_{12}M_2]) =$*

$rank([H_{11}\ \ H_{12}M_2]) = r$, where $R_1 + R_2 \geq r$, and (2) $rank(H_{21}M_1) < rank(H_{21})$, there exist M_1'' and M_2'' such that t_1 can decode at rate $\min(R_1+1, k_{1-1})$ *and t_2 can decode at rate* $\max(R_2-2, 0)$.

proof: 1-1 tradeoff. We assume that $R_1 + 1 \leq k_{1-1}$ and $R_2 - 1 \geq 0$. A vector $\vec{\alpha}$ is added to M_1 to form M_1' such that $M_1' = [\vec{\alpha}\ \ M_1]$ and $rank(H_{11}M_1') = R_1 + 1$ where $H_{11}M_1'$ is of dimension $k_{12-1} \times (R_1 + 1)$.

For part (a), because of Condition (1), $H_{11}\vec{\alpha}$ will be a nonzero linear combination of the vectors in $H_{11}M_1$ and $H_{12}M_2$, i.e., $H_{11}\vec{\alpha} = H_{11}M_1\vec{\gamma}_1 + H_{12}M_2\vec{\gamma}_2$. Note that $\vec{\gamma}_1$ is unique; otherwise, assume that there exist $\vec{\gamma}_1'$ and $\vec{\gamma}_2'$ such that $H_{11}\vec{\alpha} = H_{11}M_1\vec{\gamma}_1' + H_{12}M_2\vec{\gamma}_2'$ where $\vec{\gamma}_1' \neq \vec{\gamma}_1$. If $H_{12}M_2\vec{\gamma}_2 = H_{12}M_2\vec{\gamma}_2'$ then $H_{11}M_1\vec{\gamma}_1 = H_{11}M_1\vec{\gamma}_1'$ which indicates that $H_{11}M_1$ is not full column rank. On the other hand if $H_{12}M_2\vec{\gamma}_2 \neq H_{12}M_2\vec{\gamma}_2'$, then it means that $span(H_{11}M_1) \cap span(H_{12}M_2) \neq \{0\}$. Hence, by contradiction, we have $\vec{\gamma}_1' = \vec{\gamma}_1$, which indicates that $\vec{\gamma}_1$ is unique. Then, $\vec{\beta} = H_{11}\vec{\alpha} - H_{11}M_1\vec{\gamma}_1$ is a vector which contains at least one nonzero element. Otherwise, if $\vec{\beta}$ is a zero vector, $rank(H_{11}M_1')$ will be rank R_1 which is a contradiction. Assume w.l.o.g. that the nonzero element is on the first row of $\vec{\beta}$.

Next, we select a full rank matrix U of dimension $R_2 \times (R_2 - 1)$ from the null space of the first row of $H_{12}M_2$ such that the first row of $H_{12}M_2U$ is a zero row vector. It follows that $H_{11}\vec{\alpha}$ can not be represented by a linear combination of the vectors in $H_{11}M_1$ and $H_{12}M_2U$, which indicates that $H_{11}\vec{\alpha} \notin span([H_{11}M_1\ \ H_{12}M_2U])$. Next, because $span(H_{11}M_1) \cap span(H_{12}M_2) = \{0\}$, we have $span(H_{11}M_1) \cap span(H_{12}M_2U) = \{0\}$. Finally, we conclude that $span(H_{11}M_1') \cap span(H_{12}M_2') = \{0\}$ where $M_2' = M_2U$. Hence, t_1 can decode at rate $\min(R_1 + 1, k_{1-1})$.

For part (a) if Condition (2) is satisfied, $span(H_{21}M_1) = span(H_{21})$. Using an argument similar to the one used in the proof of Lemma 4.2.4, it can be shown that $span(H_{21}M_1') = span(H_{21}) = span(H_{21}M_1)$. This implies that $span(H_{21}M_1') \cap span(H_{22}M_2') = \{0\}$ since $span(H_{22}M_2') \subseteq span(H_{22}M_2)$. Then t_2 can decode at rate $R_2 - 1$ since $rank(H_{22}M_2') = R_2 - 1$.

For part (b) if Condition (1) is satisfied, we can find an M_1' such that $rank(H_{11}M_1') = R_1 + 1$ and $span(H_{11}M_1') \cap span(H_{12}M_2) = \emptyset$. At the same time, if Condition (2) of part (b) is satisfied, $rank(H_{21}M_1') - rank(H_{21}M_1) \leq 1$. Then $rank(span(H_{21}M_1') \cap span(H_{22}M_2))$ can be as large as 1. As $H_{22}M_2$ is a full column rank matrix, we can find an M_2' by deleting one column from M_2 such that $span(H_{21}M_1') \cap span(H_{22}M_2') = \{0\}$ where M_2' is a full rank matrix of dimension $k_{2-12} \times (R_2 - 1)$. Furthermore, since $span(H_{12}M_2') \subseteq span(H_{12}M_2)$, we will have that $span(H_{11}M_1') \cap span(H_{12}M_2') = \{0\}$. With this M_1' and M_2', the rate point (R_1+1, R_2-1) can be achieved. ■

proof: 1-2 tradeoff. We assume that $R_1 + 1 \leq k_{1-1}$ and $R_2 - 2 \geq 0$.

Note that Condition (1) here is the same as in the Rate Exchange Lemma – 1-1 tradeoff – part(a). Therefore, we can find two matrices M_1' and M_2' with rank $R_1 + 1$ and $R_2 - 1$ by appending one vector to M_1 and selecting $M_2' = M_2U$ such that $rank(H_{11}M_1') = R_1 + 1$, and $span(H_{11}M_1') \cap span(H_{12}M_2') = \{0\}$ where U is a full rank matrix of dimension $R_2 \times (R_2 - 1)$ such that $rank(H_{12}M_2) - rank(H_{12}M_2U) = 1$.

If Condition (2) is satisfied, $rank(H_{21}M_1') - rank(H_{21}M_1)$ can be as large as 1. Then $rank(span(H_{21}M_1') \cap span(H_{22}M_2'))$ can be as large as 1. Because $H_{22}M_2'$ is a full column rank matrix, we can find an M_2'' by deleting one column from M_2' such that $span(H_{21}M_1') \cap span(H_{22}M_2'') = \{0\}$ where M_2'' is a full rank matrix of dimension $k_{2-12} \times (R_2 - 2)$. Furthermore, since $span(H_{12}M_2'') \subseteq span(H_{12}M_2')$, we will have that $span(H_{11}M_1') \cap span(H_{12}M_2'') = \{0\}$. Finally let $M_1'' = M_1'$. With encoding matrices M_1'' and M_2'', it can be seen that (R_1+1, R_2-2) can be achieved. ■

By applying the Rate Exchange Lemma – 1-1 tradeoff – part (a), at point $W' = (k_{12-1} - k_{2-1}, k_{12-2} - k_{1-2})$, we have the following theorem.

Theorem 4.2.8 *If $k_{1-2} + k_{2-1} \leq \min(k_{12-1}, k_{12-2})$, the following rate region (see Fig. 4.2) can be achieved.*

Region 1:

$$R_1 \leq k_{1-1}, \qquad R_2 \leq k_{2-2},$$

$$R_1 + R_2 \leq k_{12-1} - k_{2-1} + k_{12-2} - k_{1-2}.$$

proof: Note that point $W' = (R_1, R_2) = (k_{12-1} - k_{2-1}, k_{12-2} - k_{1-2})$ is achieved by using the Rate Increase Lemma. Let M_1 and M_2 be the encoding matrices at W'. Then, we have $rank([H_{11}M_1 \quad H_{12}M_2]) = rank([H_{11} \quad H_{12}M_2])$, and we further have that $rank(H_{21}M_1) =$

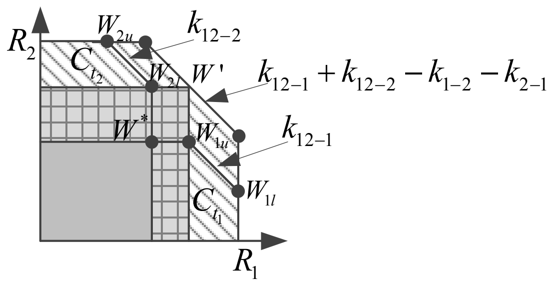


Figure 4.2 The achievable rate region for the low interference case. For each point in the shaded grey area, both terminals can recover both the sources. In the hatched grey area and the hatched white area, for a given rate point, its x-coordinate is the rate for $s_1 - t_1$ and its y-coordinate is the rate for $s_2 - t_2$; the terminals are not guaranteed to decode both sources in this region. The union of the hatched white region, the hatched gray region and the gray region is the final extended rate region for the low interference case.

$rank(H_{21}) = k_{1-2}$. Applying the Rate Exchange Lemma – 1-1 tradeoff – part (a) we have the required conclusion. ■

remark: Note that it always holds that $k_{12-1} \geq k_{1-1}$, $k_{12-2} \geq k_{2-2}$. Along with the low interference condition, we can conclude that $k_{12-1} - k_{2-1} + k_{12-2} - k_{1-2} \geq \max(k_{1-1}, k_{2-2}) \geq (k_{1-1} + k_{2-2})/2$. As $k_{1-1} + k_{2-2}$ is always an upper bound (albeit loose) on $R_1 + R_2$, this implies that our rate region is within a multiplicative gap of two of the outer bound.

4.2.2 High interference case - $k_{1-2} + k_{2-1} > \min(k_{12-1}, k_{12-2})$

Note that for the low interference case, the low interference condition implies that $k_{1-1} \geq k_{1-2}$ and $k_{2-2} \geq k_{2-1}$. However, in high interference case, there are several possibilities. We show a case where $k_{1-1} \leq k_{1-2}$ and $k_{2-2} \leq k_{2-1}$ in Fig. 4.3(a). When $k_{1-1} \geq k_{1-2}$, Fig. 4.3(b) illustrates an example where $k_{2-2} \leq k_{2-1}$, and Fig. 4.1(a) (in Section 4.2.1) illustrates an example where $k_{2-2} \geq k_{2-1}$. It can be observed here that unlike the low interference case, Q_1 may not be the same point as Q_2. In the discussion below we present rate regions by extending them from the rate points Q_1 and Q_2.

Claim 4.2.9 *When $Q_1 \neq Q_2$, the Rate Increase Lemma cannot be applied to increase the rate*

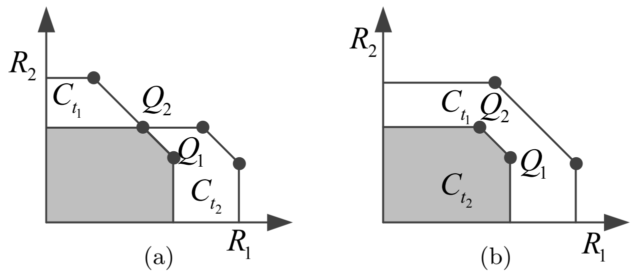


Figure 4.3 (a) High interference case where $k_{1-1} \leq k_{1-2}$ and $k_{2-2} \leq k_{2-1}$.
(b) High interference case where $k_{1-1} \geq k_{1-2}$ and $k_{2-2} \leq k_{2-1}$.

to t_2 above R_2^* at Q_1 or to increase the rate to t_1 above R_1^{**} at Q_2.

Proof: As $Q_1 \neq Q_2$, using eq. (4.1), we conclude that $\min(k_{1-2}, k_{1-1}) + \min(k_{2-1}, k_{2-2}) > \min(k_{12-1}, k_{12-2})$. Then at Q_1, $R_2^* = \min(\min(k_{2-1}, k_{2-2}), \min(k_{12-1}, k_{12-2}) - \min(k_{1-2}, k_{1-1})) < \min(k_{2-1}, k_{2-2}) \leq k_{2-1}$. Next, since $rank(H_{12}M_2) \leq rank(M_2) = R_2^* < rank(H_{12}) = k_{2-1}$, Condition (2) of the Rate Increase Lemma is not satisfied. A similar argument applies for Q_2. ∎

In view of the above claim, using our achievable strategies one can at best use the Rate Exchange Lemma to increase the rate to t_2 at Q_1 while reducing the rate to t_1. As $Q_1 \neq Q_2$, the multicast region is a pentagon and applying the 1-1 tradeoff will at most allow us to achieve the boundary between Q_1 and Q_2, while the 1-2 tradeoff achieves interior points in the multicast region. As points on the $Q_1 - Q_2$ boundary are already achieved by multicasting both sources, the region is not enlarged.

Hence, we will consider rate points (R_1, R_2) such that $R_1 > R_1^*$ and $R_2 = R_2^*$ at Q_1 (and similarly $R_1 = R_1^{**}$ and $R_2 > R_2^{**}$ at Q_2). At Q_1, if $k_{1-2} \geq k_{1-1}$, $R_1^* = k_{1-1}$, i.e. increasing R_1 is impossible since it attains its maximum. Therefore, we assume that $k_{1-2} < k_{1-1}$. By the high interference condition and the fact that $k_{1-2} + k_{2-2} \geq k_{12-2}$, we have $(R_1^*, R_2^*) = (k_{1-2}, \min(k_{12-1}, k_{12-2}) - k_{1-2})$. We begin by modifying the source encoding matrices at point Q_1, with the goal of increasing R_1 the rate to t_1 above R_1^*. Our strategy at Q_1 is similar to the one for the low interference case, namely, we attempt to trace a region of achievable rates by using the Rate Increase and Rate Exchange lemmas. The main difference is that here we also use the 1-2 tradeoff result (cf. Lemma 4.2.7). Note that in the discussion below, we present the arguments for increasing rates at Q_1 and Q_2 separately. However, if $Q_1 = Q_2$, then the arguments are still applicable.

Theorem 4.2.10 *If $k_{1-2} + k_{2-1} > \min(k_{12-1}, k_{12-2})$ and $k_{1-2} < k_{1-1}$, then the rate pair in the following region can be achieved.*

Region 2:

$$D_1 \cap (D_2 \cup D_3 \cup D_4) \quad \text{if } k_{2-1} < k_{2-2}, \text{ or}$$

$$D_1 \cap (D_2 \cup D_3) \quad \text{if } k_{2-1} \geq k_{2-2}, \text{ where}$$

$D_1 : R_1 \leq k_{1-1}$,

$D_2 : R_1 + R_2 \leq rank([H_{11} \;\; H_{12}M_2]) \quad$ *when* $R_2 \leq \min(k_{12-1}, k_{12-2}) - k_{1-2}$,

$D_3 : R_1 + 2R_2 \leq R_2^* + rank([H_{11} \;\; H_{12}M_2]) \quad$ *when* $\min(k_{12-1}, k_{12-2}) - k_{1-2} \leq R_2 \leq \min(k_{2-1}, k_{2-2})$,

$D_4 : R_1 + R_2 \leq R_2^* + rank([H_{11} \;\; H_{12}M_2]) - k_{2-1} \quad$ *when* $k_{2-1} < R_2 \leq k_{2-2}$,

where $R_2^* = \min(k_{12-1}, k_{12-2}) - k_{1-2}$, M_1 and M_2 are the encoding matrices at Q_1.

Note that in the above characterization, the rate constraints depend on $rank([H_{11} \;\; H_{12}M_2])$; we show a lower bound on $rank([H_{11} \;\; H_{12}M_2])$ in Section 4.2.2.1.

Proof: Given that $k_{1-2} + k_{2-1} > \min(k_{12-1}, k_{12-2})$ and $k_{1-2} < k_{1-1}$, we will extend the rate region from Q_1 where $R_1^* = k_{1-2}$, $R_2^* = \min(k_{12-1}, k_{12-2}) - k_{1-2}$. Let M_1 and M_2 denote the encoding matrices at Q_1. At Q_1, we first need to increase R_1 while keeping R_2 as large as possible. Suppose that we can use the Rate Increase Lemma to increase R_1. This implies that $\min(k_{12-1}, k_{12-2}) = rank([H_{11}M_1 \;\; H_{12}M_2]) < rank([H_{11} \;\; H_{12}M_2]) \leq rank([H_{11} \;\; H_{12}]) = k_{12-1}$ which implies that $\min(k_{12-2}, k_{12-1}) = k_{12-2}$. In the following discussion, we assume this is the case. By Rate Increase Lemma, we can achieve the rate point $W' = (R_1', R_2') = (rank([H_{11} \;\; H_{12}M_2]) - R_2^*, R_2^*)$. The corresponding encoding matrices are M_1' and $M_2' = M_2$.

When we want to further increase R_1 above R_1', we could use Rate Exchange Lemma – 1-1 tradeoff – part (a) repeatedly, since $rank(H_{21}M_1) = k_{1-2} = R_1^*$ and $span(M_1) \subseteq span(M_1')$, implying that $rank(H_{21}M_1') = rank(H_{21}) = k_{1-2}$. When R_1' is increased by δ, R_2' is decreased by δ where $0 \leq \delta \leq \min(R_2^*, k_{1-1} - R_1')$ ($\delta \leq k_{1-1} - R_1'$ comes from the fact that R_1' can be increased to at most k_{1-1}). Terminal t_1 can decode messages from s_1 at rate $R_1'' = R_1' + \delta$ and t_2 can decode messages from s_2 at rate $R_2'' = R_2' - \delta$. Denote the new set of encoding matrices as M_1'' and M_2''. This is shown by the line $(W', \bar{W}')$ in Fig. 4.4(a) which corresponds to D_2.

On the other hand, at W', we can increase R_2 such that $R_2 = R_2' + \delta_1$ where $0 \leq \delta_1 \leq \min(k_{2-1} - R_2^*, k_{2-2} - R_2^*)$. First note that $k_{12-2} = rank([H_{21}M_1 \ \ H_{22}M_2]) \leq rank([H_{21}M_1' \ \ H_{22}M_2']) \leq rank([H_{21}M_1' \ \ H_{22}]) \leq rank([H_{21} \ \ H_{22}]) = k_{12-2}$ which implies $rank([H_{21}M_1' \ \ H_{22}M_2']) = rank([H_{21}M_1' \ \ H_{22}])$. Then by using Rate Exchange Lemma – 1-2 tradeoff, since $rank(H_{12}) - rank(H_{12}M_2') = k_{2-1} - (\min(k_{12-1}, k_{12-2}) - k_{1-2}) > 0$ we can increase R_2' by δ_1 and decrease R_1' by $2\delta_1$, and the boundary point $(R_1' - 2\delta_1, R_2' + \delta_1)$ can be achieved where $0 \leq \delta_1 \leq \min(k_{2-1} - R_2^*, k_{2-2} - R_2^*, R_1'/2)$ which corresponds to D_3 ($\delta_1 \leq R_1'/2$ comes from the fact that R_1 should be not smaller than 0). If we have that $k_{2-1} \leq \min(k_{2-2}, R_1'/2 + R_2^*)$, we will arrive at the boundary point $W'' = (R_1'', R_2'') = (R_2^* + rank([H_{11} \ \ H_{12}M_2]) - 2k_{2-1}, k_{2-1})$. The corresponding matrices are M_1'' and M_2''. This is demonstrated by the line (W', W'') in Fig. 4.4(a).

If we have that $R_1'' \geq 0$ and $k_{2-1} < k_{2-2}$, at point W'', we can further increase R_2 such that $R_2 = R_2'' + \delta_2$ and $R_1 = R_1'' - \delta_2$ where $0 \leq \delta_2 \leq \min(k_{2-2} - k_{2-1}, R_1'')$. The corresponding encoding matrix at s_2 is M_2'''. By Rate Exchange Lemma – 1-1 tradeoff – part (a), since $rank(H_{12}) = rank(H_{12}M_2'')$, t_1 can decode at rate $R_1'' - \delta_2$, and t_2 can decode at rate $R_2'' + \delta_2$. Then W''' is achieved and the procedure is demonstrated by the line (W'', W''') in Fig. 4.4(a) which corresponds to D_4. The entire extended rate region for this case is shown in Fig. 4.4(a).

■

We next consider increasing R_2 above R_2^{**} at Q_2. If $k_{2-1} \geq k_{2-2}$, R_2 cannot be increased as $R_2^{**} = k_{2-2}$. Hence, we assume that $k_{2-1} < k_{2-2}$. A similar analysis for Q_2 results in the following region.

Corollary 4.2.11 *If $k_{1-2} + k_{2-1} > \min(k_{12-1}, k_{12-2})$ and $k_{2-1} < k_{2-2}$, then the rate pair in the following region can be achieved.*

Region 3:

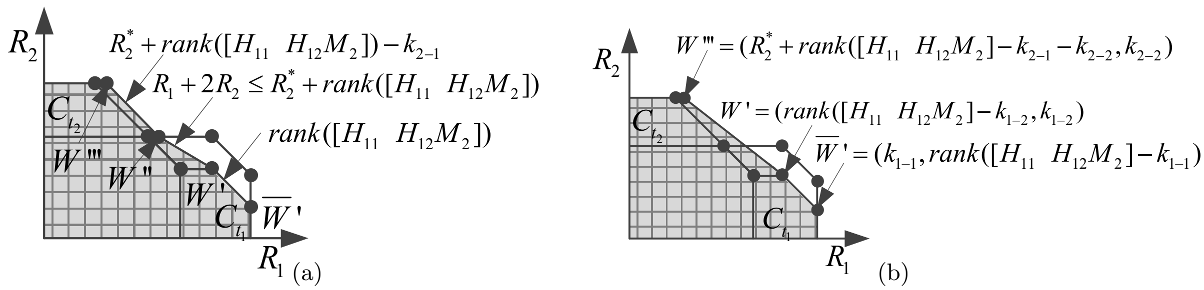


Figure 4.4 (a) The extended rate region for the high interference case from point Q_1. (b) The final extended rate region for the case of high interference.

$$D_1' \cap (D_2' \cup D_3' \cup D_4') \quad \textit{if } k_{1-2} < k_{1-1}, \textit{ or}$$

$$D_1' \cap (D_2' \cup D_3') \quad \textit{if } k_{1-2} \geq k_{1-1} \textit{ where,}$$

$D_1' : R_2 \leq k_{2-2}$,

$D_2' : R_1 + R_2 \leq rank([H_{21}M_1 \;\; H_{22}])$ *when* $R_1 \leq \min(k_{12-1}, k_{12-2}) - k_{2-1}$,

$D_3' : 2R_1 + R_2 \leq R_1^{**} + rank([H_{21}M_1 \;\; H_{22}])$ *when* $\min(k_{12-1}, k_{12-2}) - k_{2-1} \leq R_1 \leq \min(k_{1-2}, k_{1-1})$,

$D_4' : R_1 + R_2 \leq R_1^{**} + rank([H_{21}M_1 \;\; H_{22}]) - k_{1-2}$ *when* $k_{1-2} < R_1 \leq k_{1-1}$,

where $R_1^{**} = \min(k_{12-1}, k_{12-2}) - k_{2-1}$, M_1 *and* M_2 *are the encoding matrices at* Q_2.

From the above argument, the overall rate region is the convex hull of multicast region, and either Region 2 or Region 3 or both depending upon the cut conditions. For instance when $k_{1-2} < k_{1-1}$ and $k_{2-1} < k_{2-2}$ the final region is shown in Fig. 4.4(b), where boundary segment $W''' - W'$ is achieved via timesharing.

Finally, note that when $k_{1-2} \geq k_{1-1}$ and $k_{2-1} \geq k_{2-2}$, we cannot enlarge the region using our achievability schemes, i.e., the achievable region is the multicast region.

4.2.2.1 Lower bound of $rank([H_{11} \;\; H_{12}M_2])$

As before, let (R_1^*, R_2^*) denote the rate point at Q_1 and let M_1 and M_2 denote the corresponding encoding matrices. First note that $rank([H_{11} \;\; H_{12}M_2]) \geq rank(H_{11}) = k_{1-1}$ and

$rank([H_{11}\ \ H_{12}M_2]) \geq rank([H_{11}M_1\ \ H_{12}M_2]) = R_1^* + R_2^*$. Next we will also find another nontrivial lower bound of $rank([H_{11}\ \ H_{12}M_2])$ by the following lemma.

Lemma 4.2.12 *Given* $rank([H_{11}\ \ H_{12}]) = k_{12-1}$, $rank(H_{12}) = k_{2-1}$ *and* $rank([H_{12}M_2]) = l$, *we have* $rank([H_{11}\ \ H_{12}M_2]) \geq k_{12-1} - k_{2-1} + l$.

proof: By the assumed conditions, there are k_{2-1} columns in H_{12} that are linearly independent, and in H_{11}, we can find a subset of $k_{12-1} - k_{2-1}$ columns denoted H'_{11} such that $span(H'_{11}) \cap span(H_{12}) = \{0\}$ and $rank(H'_{11}) = k_{12-1} - k_{2-1}$, which further imply that $rank([H'_{11}\ H_{12}]) = k_{12-1}$.

Since $span(H_{12}M_2) \subseteq span(H_{12})$ this means that $span(H'_{11}) \cap span(H_{12}M_2) = \{0\}$. Then $rank([H'_{11}\ H_{12}M_2]) = rank(H'_{11}) + rank(H_{12}M_2) = k_{12-1} - k_{2-1} + l$. Hence, $rank([H_{11}\ H_{12}M_2]) \geq rank([H'_{11}\ H_{12}M_2]) = k_{12-1} - k_{2-1} + l$. ■

Together with the two lower bounds above, we have $rank([H_{11}\ H_{12}M_2]) \geq \max(k_{1-1}, k_{12-1} - k_{2-1} + R_2^*, R_1^* + R_2^*)$. A case where $\max(k_{1-1}, k_{12-1} - k_{2-1} + R_2^*, R_1^* + R_2^*) = k_{12-1} - k_{2-1} + R_2^*$ is shown in Fig. 4.4(b) where $R_2^* = k_{12-2} - k_{1-2}$.

4.2.3 Increasing the achievable rate region by modifying the graph

Thus far, we have presented achievable rate regions for both the low and high interference scenarios. An interesting observation about these regions is that it is possible to enlarge the regions by considering the removal of judiciously chosen edges from the network. We have noted that by removing certain edges from the network, the achievable rate region can be extended. For example, Fig. 4.5 corresponds to a scenario where $k_{1-1} = 3$, $k_{1-2} = 1$, $k_{2-1} = 3$, $k_{2-2} = 3$, $k_{12-1} = 3$ and $k_{12-2} = 3$. Hence, the sum rate $R_1 + R_2 \leq 3$ using Theorem 4.2.10. However, one can achieve the rate points $(R_1, R_2) = (1, 3)$ and $(3, 1)$ by removing edges e_1 and e_2 since k_{2-1} drops to 1 and the low interference result (cf. Theorem 4.2.8) applies. Furthermore note that the rate points $(1, 3)$ and $(3, 1)$ are not achievable by routing need network coding.

In principle, one could consider the union of the achievable rate regions obtained by removing certain subset of the edges from the network to perhaps obtain a larger region. Finding

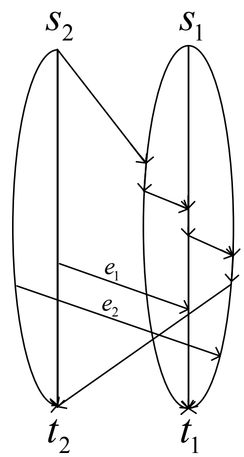


Figure 4.5 An example of a network where a larger achievable rate region can be achieved by removing edges e_1 and e_2.

such edges in a systematic manner is an interesting problem. However, we are unaware of any known algorithm for it.

4.3 Achievable rate region for given $k_{1-12}, k_{2-12}, k_{1-1}, k_{2-2}, k_{1-2}$, and k_{2-1}

We have discussed the achievable rate region given $k_{12-1}, k_{12-2}, k_{1-1}, k_{2-2}, k_{1-2}$, and k_{2-1} in the previous section. However, there are other cuts that are potentially useful in finding the achievable rate region. In this section, we will discuss the achievable rate region for given $k_{1-12}, k_{2-12}, k_{1-1}, k_{2-2}, k_{1-2}$, and k_{2-1} using the reversibility result introduced in [37]. Towards this end define the reverse of a network G as the network $G' = (V', E')$ where (1) The nodes V' and edges E' in G' are the same as in G, except the direction of edges are reversed. (2) The sources in G are the terminals in G' and vice versa.

For the double unicast problem, we will have that $s'_i = t_i$ and $t'_i = s_i$, $i = 1, 2$. Let $k_{1-12}, k_{2-12}, k_{1-1}, k_{2-2}, k_{1-2}$ and k_{2-1} denote the cut in G and let $k'_{12-1}, k'_{12-2}, k'_{1-1}, k'_{2-2}, k'_{1-2}$ and k'_{2-1} denote the cut in G'. It is evident that $k'_{12-1} = k_{1-12}$, $k'_{12-2} = k_{2-12}$, $k'_{1-1} = k_{1-1}$, $k'_{2-2} = k_{2-2}$, $k'_{1-2} = k_{2-1}$ and $k'_{2-1} = k_{1-2}$. By Theorem 4 in [37] a linear network coding solution for rate pair (R_1, R_2) in the original network G is in one-to-one correspondence with the rate pair $(R_1', R_2') = (R_1, R_2)$ in the reversed network G'. Thus, our idea is to determine an achievable rate pair in G' and then claim the existence of a corresponding rate pair in G. The process consists of substituting the corresponding cuts of the reverse network into the multicast region $\mathcal{B}$, Region 1, Region 2 and Region 3 of the original network, to obtain a new set of regions $\mathcal{B}'$, Region 1', Region 2' and Region 3'.

In the interest of avoiding repetitive arguments, we discuss the process of determining Region 2' by means of an example. For the original graph, in Region 2, $D_2 : R_1 + R_2 \leq rank([H_{11} \quad H_{12}M_2])$ when $R_2 \leq \min(k_{12-1}, k_{12-2}) - k_{1-2}$. Thus, for Region 2', the corresponding $D_2 : R_1 + R_2 \leq rank([H_{11}' \quad H_{12}'M_2'])$ when $R_2 \leq \min(k_{1-12}, k_{2-12}) - k_{2-1}$ where H_{ij}' is the transfer matrix from s_j' to t_i', and M_i' is the source encoding matrix at s_i'. The other inequalities can be determined in a similar manner.

Hence, given all possible cuts in a double unicast network, the achievable rate region is convex hull of multicast region $\mathcal{B}$, $\mathcal{B}'$ and the corresponding extended region in different cases.

In order to demonstrate the utility of considering the reversed network, consider the network shown in Fig. 4.6. It can be verified that the rate regions are different using the original result and reversibility result. with our schemes. In particular, using the reversibility result can achieve rate point (1,1) whereas the original result cannot.

4.4 Comparison with existing results

The work that is most closely related to the present paper is by [12] that also considers the double unicast problem with arbitrary rates. Assuming that $k_{2-2} \leq k_{1-1}$, the region in [12] is given by EF09 = EF09(a) $\cup$ EF09(b), where

$$\text{EF09(a)} = \{(R_1, R_2) : R_1 + 2R_2 \leq k_{1-1}, R_2 \leq k_{2-2}\}, \text{and}$$

$$\text{EF09(b)} = \{(R_1, R_2) : 2R_1 + R_2 \leq k_{2-2}, R_1 \leq k_{1-1}\}.$$

A comparison between our region and theirs indicates that our region is larger than theirs. To see this, consider the low interference case and a rate point (R_1, R_2) that lies in EF09(a). We have that $R_1 + R_2 \leq R_1 + 2R_2 \leq k_{1-1} \leq k_{12-1} - k_{2-1} + k_{12-2} - k_{1-2}$ (since $k_{1-2} + k_{2-1} \leq$

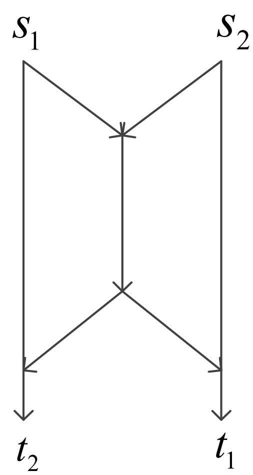


Figure 4.6 An example of a network where the achievable rate regions are different using the original result and the reversibility result. All edges are unit capacity.

$\min(k_{12-1}, k_{12-2}))$ and $R_2 \leq k_{2-2}$, i.e. (R_1, R_2) also belongs to our region.

For the high interference case, we argue as follows. Let (R_1, R_2) belong to EF09(a).

- If $k_{1-2} \leq k_{1-1}$, we show that (R_1, R_2) belongs to Region 2. Note that $R_1 + 2R_2 \leq k_{1-1} \leq rank([H_{11} \quad H_{12}M_2])$. However, the RHS of D_2 and D_3 is at least as large as $rank([H_{11} \quad H_{12}M_2])$, and for D_4 we have $R_1 + 2R_2 \leq rank([H_{11} \quad H_{12}M_2]) \leq R_2^* + rank([H_{11} \quad H_{12}M_2]) - k_{2-1} + R_2$ (since in D_4, $k_{2-1} \leq R_2 \leq k_{2-2}$) indicating that (R_1, R_2) is within Region 2.

- If $k_{1-2} > k_{1-1}$ and $k_{2-1} \geq k_{2-2}$, we have $R_1 + R_2 \leq R_1 + 2R_2 \leq k_{1-1} \leq \min(k_{1-2}, k_{12-1}) \leq \min(k_{12-2}, k_{12-1})$ which shows that (R_1, R_2) is within our multicast region.

- If $k_{1-2} > k_{1-1}$ and $k_{2-1} < k_{2-2}$, we consider different ranges for R_2. For $0 \leq R_2 \leq k_{2-1}$, $R_1 + R_2 \leq R_1 + 2R_2 \leq k_{1-1} \leq \min(k_{1-2}, k_{12-1}) \leq \min(k_{12-2}, k_{12-1})$ which implies that (R_1, R_2) is within our multicast region. On the other hand when $k_{2-1} \leq R_2 \leq k_{2-2}$, we have $k_{1-1} - 2k_{2-2} \leq R_1 \leq k_{1-1} - 2k_{2-1}$ (from the definition of EF09(a)). This implies that (R_1, R_2) belongs to Region 3. To see this we note that the relevant range of Region

3 is D_2' since $k_{1-1} - 2k_{2-1} \le \min(k_{12-1}, k_{12-2}) - k_{2-1}$. We have $R_1 + R_2 \le R_1 + 2R_2 \le k_{1-1} \le \min(k_{1-1} + k_{2-1}, \min(k_{12-1}, k_{12-2})) = R_1^{**} + R_2^{**} = rank([H_{21}M_1 \quad H_{22}M_2]) \le rank([H_{21}M_1 \quad H_{22}])$ indicating that such a point is within Region 3.

In a similar manner it can be shown that all rate points in EF09(b) are within our rate region.

The authors in [10] and [11] explore the unit-rate case $R_1 = R_2 = 1$ in detail. Such schemes can potentially be packed into networks with higher capacities. References [10,11] rely heavily on an analysis of the graph theoretic structures that are possible in double unicast networks. Thus, our scheme will in general be weaker than their approach on certain networks. Likewise the work of [9] [26] also considers the achievable rate region using network coding between pair of sources. However, there are networks where our approach is strictly better than all the above approaches. We show such an example in Fig. 4.7. In Fig. 4.7, we can achieve rates (4,2) by the argument using in Region 2, whereas it can be verified that the above schemes do not support this rate point. For instance, if $R_2 = 2$, $R_1 \le 3$ in EF09, whereas the scheme in [10] can at most achieve a rate of $(1, 2)$. Furthermore, we note that the enlargement of the achievable region by considering the removal of certain edges discussed in Section 4.2.3 also improves our region in many cases.

The following results have appeared since the submission of the present paper and the publication of our preliminary conference paper [23]. The work of [31] treats the two unicast problem as an instance of a linear deterministic interference channel and finds a network code that uses random linear network coding. Their region contains our proposed achievable region. The authors in [32] also derive an achievable region by exploiting the equivalence with deterministic interference channels; their region is completely specified by the cut values in the network (in contrast, in certain cases our region and the region in [31] is specified in terms of the rank of matrices that depend on the network code). However, for some networks our scheme achieves a larger region. As an example, if one considers the two-unicast butterfly network with $k_{1-1} = k_{2-2} = 1, k_{1-2} = k_{2-1} = 2$ and $k_{12-1} = k_{12-2} = 2$, our scheme achieves the multicast point $(1, 1)$ whereas the region in [32] is empty.

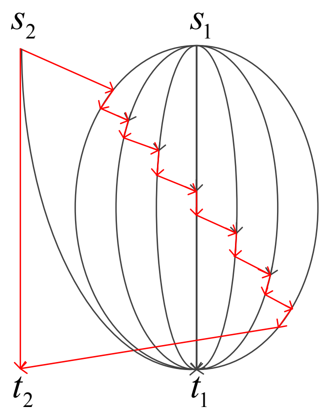


Figure 4.7 An example of a high interference network when our scheme can achieve a higher rate pair compared to many other schemes.

4.5 Conclusions

In this work, we presented an achievable rate region for the double unicast problem for directed acyclic networks with unit capacity edges. The proposed strategy combines random linear network coding along with appropriate precoding at the source nodes. Networks are classified according the relationship of the values of the cuts between various subsets of the sources and the terminals. We begin with the multicast region where both sources are multicast to both terminals and then enlarge the region by either unilaterally increasing one of the rates or trading off rates between the connections. The proposed region can potentially be enlarged by considering regions that are obtained by the judicious removal of certain edges from the network.

CHAPTER 5. CONCLUSIONS AND FUTURE WORKS

5.1 Contributions

This dissertation has focused on the multiple unicast problem over directed acyclic networks when there are three sessions and two sessions. The most significant contribution and conclusions of this work can be summarized as follows.

1. For three unicast problem, given the connectivity level vector $[k_1\ k_2\ k_3]$ where there exist k_i edge disjoint paths between s_i to t_i, we decide if unit rate transmission is feasible. For connectivity level vector [1 3 3], [2 2 4] and [1 2 5], we present constructive linear network coding schemes. For connectivity level vector [1 1 3], [2 2 2], [2 2 3], we provide instances of network that cannot support unit rate transmission. For connectivity level vector [1 2 4], we are not able to provide either a network coding solution or a network topology to demonstrate the infeasibility of unit rate transmission. The experimental results indicate that for networks where the different source terminal paths have a significant overlap, our constructive unit rate schemes can be packed along with routing to provide higher throughput as compared to a pure routing approach.

2. For two unicast problem, we assume we know certain minimum cut values for the network, e.g., mincut(S_i, T_j), where $S_i \subseteq \{s_1, s_2\}$ and $T_j \subseteq \{t_1, t_2\}$ for different subsets S_i and T_j. Based on these values, we propose an achievable rate region using linear network codes. We first define the multicast region where both sources are multicast to both terminals. Following this we enlarge the region by appropriately encoding the information at the source nodes, such that terminal t_i is only guaranteed to decode information from the intended source s_i, while decoding a linear function of the other source. We also

incorporate the techniques of removing certain edges and network inversion to further enlarge the achievable region.

5.2 Future work

Based on what has been accomplished so far in this dissertation, several suggestions for further research work are provided below:

1. For three unicast problem, we have identify certain feasible/infeasible instances with two unicast sessions, where the message entropies are different, e.g., Lemma 3.2.2 and Lemma 3.3.4. These are used to arrive at conclusion for the problem in the case of high sessions (more than three sessions). Hence, it is beneficial to analyze the achievable rate region for double unicast network, and then analyze the more general case, e.g., we are interested in given the cut value mincut(s_1, t_1), mincut(s_2, t_2), if there exists a general method to decide the achievable region.

2. For the two unicast problem, we have demonstrated that the proposed region can potentially be enlarged by considering regions that are obtained by removing certain edges from the network. However, it is not an easy problem. An intuition is to convert an original network of high interference to a corresponding low interference one since a 1-1 tradeoff can always be done in Region 1. While this is an intuition, this is not always true in every high interference network. Future work would include the investigation of systematic techniques for finding the appropriate edges to be removed.

3. For general multiple unicast problem, we have packed our three unicast unit rate schemes in a general unicast problem to increase the capacity. A nature question to ask is if we can pack our non-unit rate two unicast schemes in the graphs to increase the capacity over routing. This question is more involved since we have to divide the original graphs into subgraphs that have certain cut vectors. A future research interest could be optimizing the dividing procedure to achieve the maximum rate for each session.

APPENDIX A. PROOF OF LEMMA 3.2.4

proof: When n_1 is even, the network structure is shown in Fig. A.1.

Assume in n time units, s_1 observes n_1 independent source vectors $X_{11}^n, \ldots, X_{1n_1}^n$, s_2 observes $n_2 - 3n_1/2 + a$ independent source vectors $X_{21}^n, \ldots, X_{2m}^n$ where $m = n_2 - 3n_1/2 + a$ and a is a positive constant. For the simplicity of the proof, we assume that the alphabet of X_{1i} and X_{2j} is $\mathcal{X}$, and $H(X_{1i}) = H(X_{2j}) = 1, \forall i, j$. The n random variables that e_i carries are denoted as $Y_{e_i}^n$, or simply Y_i^n. From $Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n$, we estimate $X_{11}^n, \ldots, X_{1n_1}^n$. Let the estimate be $\widehat{X}_{11}^n, \ldots, \widehat{X}_{1n_1}^n$.

From the Fano's inequality, we shall have

$$H(X_{11}^n, \ldots, X_{1n_1}^n | \widehat{X}_{11}^n, \ldots, \widehat{X}_{1n_1}^n) \leq n\epsilon_n. \tag{A.1}$$

where $n\epsilon_n = 1 + nP_e \log(|\mathcal{X}|)$. For t_1 to decode $X_{11}^n, \ldots, X_{1n_1}^n$ asymptotically, $\epsilon_n \to 0$ as $P_e \to 0$, when $n \to \infty$, where $P_e = P((\widehat{X}_{11}^n, \ldots, \widehat{X}_{1n_1}^n) \neq (X_{11}^n, \ldots, X_{1n_1}^n))$.

Because $\widehat{X}_{11}^n, \ldots, \widehat{X}_{1n_1}^n$ are function of $Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n$, we will have

$$\begin{aligned} & H(X_{11}^n, \ldots, X_{1n_1}^n | Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \\ & = H(X_{11}^n, \ldots, X_{1n_1}^n | \widehat{X}_{11}^n, \ldots, \widehat{X}_{1n_1}^n, Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \\ & \leq H(X_{11}^n, \ldots, X_{1n_1}^n | \widehat{X}_{11}^n, \ldots, \widehat{X}_{1n_1}^n) \leq n\epsilon_n. \end{aligned} \tag{A.2}$$

Because $H(Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \leq n_1 n$, eq. (A.2) and the independence among $X_{11}^n, \ldots, X_{1n_1}^n, X_{21}^n, \ldots, X_{2m}^n$, by Claim B.0.1, we will have

$$mn - n\epsilon_n \leq H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \leq mn; \tag{A.3}$$

$$H(Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n | X_{21}^n, \ldots, X_{2m}^n) \geq n_1 n - 2n\epsilon_n. \tag{A.4}$$

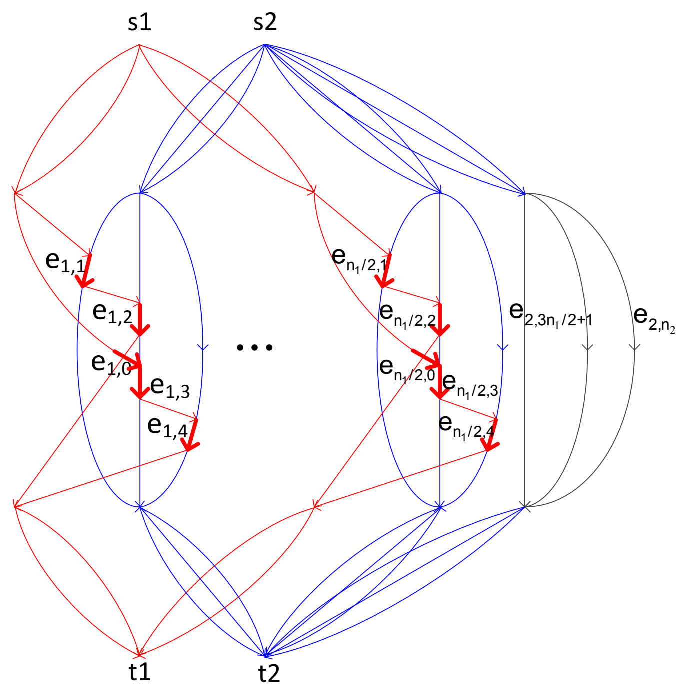


Figure A.1 An example where t_1 at decode at rate n_1, but t_2 cannot decode at rate $n_2 - 3n_1/2 + 1$.

Next, we shall have

$$
\begin{aligned}
& H(Y_{1,3}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,3}^n, Y_{n_1/2,4}^n) \\
& \overset{(a)}{=} H(X_{21}^n, \ldots, X_{2m}^n, Y_{1,3}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,3}^n, Y_{n_1/2,4}^n) - H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,3}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,3}^n, Y_{n_1/2,4}^n) \\
& \overset{(b)}{=} H(X_{21}^n, \ldots, X_{2m}^n, Y_{1,3}^n, \ldots, Y_{n_1/2,3}^n) - H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,3}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,3}^n, Y_{n_1/2,4}^n) \\
& \overset{(c)}{\leq} mn + (n_1/2)n - H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,3}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,3}^n, Y_{n_1/2,4}^n) \\
& \overset{(d)}{\leq} mn + (n_1/2)n - H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,0}^n, Y_{1,2}^n, Y_{1,3}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,0}^n, Y_{n_1/2,2}^n, Y_{n_1/2,3}^n, Y_{n_1/2,4}^n, X_{11}^n, \ldots, X_{1n_1}^n) \\
& \overset{(e)}{=} mn + (n_1/2)n - H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,0}^n, Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,0}^n, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n, X_{11}^n, \ldots, X_{1n_1}^n) \\
& \overset{(f)}{=} mn + (n_1/2)n - H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n, X_{11}^n, \ldots, X_{1n_1}^n) \\
& \overset{(g)}{=} mn + (n_1/2)n - H(X_{21}^n, \ldots, X_{2m}^n | Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \\
& \quad + I(X_{21}^n, \ldots, X_{2m}^n; X_{11}^n, \ldots, X_{1n_1}^n | Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \\
& \overset{(h)}{\leq} mn + (n_1/2)n - mn + n\epsilon_n + I(X_{21}^n, \ldots, X_{2m}^n; X_{11}^n, \ldots, X_{1n_1}^n | Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \\
& \leq mn + (n_1/2)n - mn + n\epsilon_n + H(X_{11}^n, \ldots, X_{1n_1}^n | Y_{1,2}^n, Y_{1,4}^n, \ldots, Y_{n_1/2,2}^n, Y_{n_1/2,4}^n) \\
& \overset{(i)}{\leq} mn + (n_1/2)n - mn + n\epsilon_n + n\epsilon_n = (n_1/2)n + 2n\epsilon_n
\end{aligned} \tag{A.5}
$$

(a) follows from the chain rule, (b) is because $Y^n_{1,4},\ \ldots, Y^n_{n_1/2,4}$ are functions of $X^n_{21}, \ldots, X^n_{2m}$ and $Y^n_{1,3},\ \ldots, Y^n_{n_1/2,3}$. (c) is because of the capacity constraints. (d) is because conditioning reduces entropy. (e) is because $Y^n_{1,3},\ \ldots, Y^n_{n_1/2,3}$ are functions of $Y^n_{1,2},\ \ldots, Y^n_{n_1/2,2}$ and $Y^n_{1,0},\ \ldots, Y^n_{n_1/2,0}$. (f) is because $Y^n_{1,0},\ \ldots, Y^n_{n_1/2,0}$ are functions of $X^n_{11}, \ldots, X^n_{1n_1}$. (g) follows from the mutual information definition. (h) is from eq. (A.3). (i) is from eq. (A.2).

From the network, we know that $Y^n_{1,2},\ \ldots, Y^n_{n_1/2,2}$ are functions of $Y^n_{1,1},\ \ldots, Y^n_{n_1/2,1}$ and $X^n_{21}, \ldots, X^n_{2m}$. Then

$$\begin{aligned}
&H(Y^n_{1,1}, Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,1}, Y^n_{n_1/2,3}, Y^n_{n_1/2,4}, Y^n_{2,3n_1/2+1}, \ldots, Y^n_{2,n_2}|X^n_{21}, \ldots, X^n_{2m})\\
&= H(Y^n_{1,1}, Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,1}, Y^n_{n_1/2,3}, Y^n_{n_1/2,4}, Y^n_{2,3n_1/2+1}, \ldots, Y^n_{2,n_2}, X^n_{21}, \ldots, X^n_{2m}|X^n_{21}, \ldots, X^n_{2m})\\
&\geq H(Y^n_{1,2}, Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,2}, Y^n_{n_1/2,3}, Y^n_{n_1/2,4}, Y^n_{2,3n_1/2+1}, \ldots, Y^n_{2,n_2}|X^n_{21}, \ldots, X^n_{2m})\\
&\geq H(Y^n_{1,2}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,2}, Y^n_{n_1/2,4}|X^n_{21}, \ldots, X^n_{2m}) \overset{(a)}{\geq} n_1 n - 2n\epsilon_n
\end{aligned} \tag{A.6}$$

(a) is due to eq. (A.4).

Finally, we shall have

$$\begin{aligned}
&H(X^n_{21}, \ldots, X^n_{2m}|Y^n_{1,1}, Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,1}, Y^n_{n_1/2,3}, Y^n_{n_1/2,4}, Y^n_{2,3n_1/2+1}, \ldots, Y^n_{2,n_2})\\
&= H(Y^n_{1,1}, Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,1}, Y^n_{n_1/2,3}, Y^n_{n_1/2,4}, Y^n_{2,3n_1/2+1}, \ldots, Y^n_{2,n_2}|X^n_{21}, \ldots, X^n_{2m})\\
&\quad + H(X^n_{21}, \ldots, X^n_{2m}) - H(Y^n_{1,1}, Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,1}, Y^n_{n_1/2,3}, Y^n_{n_1/2,4}, Y^n_{2,3n_1/2+1}, \ldots, Y^n_{2,n_2})\\
&\overset{(a)}{\geq} n_1 n - 2n\epsilon_n + mn - H(Y^n_{1,1}, Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,1}, Y^n_{n_1/2,3}, Y^n_{n_1/2,4}, Y^n_{2,3n_1/2+1}, \ldots, Y^n_{2,n_2})\\
&= n_1 n - 2n\epsilon_n + mn - H(Y^n_{1,1}, \ldots, Y^n_{n_1/2,1}, Y^n_{2,3n_1/2+1}, \ldots, Y_{2n_2}|Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,3}, Y^n_{n_1/2,4})\\
&\quad - H(Y^n_{1,3}, Y^n_{1,4}, \ldots, Y^n_{n_1/2,3}, Y^n_{n_1/2,4})\\
&\overset{(b)}{\geq} n_1 n - 2n\epsilon_n + mn - (n_2 - 3n_1/2 + n_1/2)n - (n_1/2)n - 2n\epsilon_n\\
&= n_1 n - 2n\epsilon_n + n_2 n - 3/2 n_1 n + an - (n_2 - 3n_1/2 + n_1/2)n - (n_1/2)n - 2n\epsilon_n = an - 4n\epsilon_n
\end{aligned} \tag{A.7}$$

(a) is because of eq. (A.6). (b) is because of eq. (A.5) and the capacity constraints.

When $n \to \infty$, for t_1 to asymptotically decode $X^n_{11}, \ldots, X^n_{1n_1}$, we shall have $\epsilon_n \to 0$. Then t_2 cannot decode $X^n_{21}, \ldots, X^n_{2m}$ asymptotically where $m = n_2 - 3n_1/2 + a$ and $a = 1$. This indicates that when t_1 decodes s_1 at rate n_1 where $n_1 \geq 2$ and n_1 is even, t_2 cannot decode the information at s_2 at rate $n_2 - 3n_1/2 + 1$.

When n_1 is odd and $n_1 > 1$, we could find a network where P_{1,n_1} is overlapped with P_{2,n_2}. The remaining network is the same as in Fig. A.1. With a similar argument, we can prove that when t_1 can decode $X_{11}^n, \ldots, X_{2n_1}^n$, X_2 cannot decode $X_{21}^n, \ldots, X_{2m}^n$ where $m = [n_2 - 1 - 3(n_1 - 1)/2] + a = n_2 - 3n_1/2 + 1/2 + a$ where $a = 1/2$, which indicates when t_1 decodes s_1 at rate n_1 where $n_1 \geq 3$ and n_1 is odd, t_2 cannot decode the information at s_2 at rate $n_2 - 3n_1/2 + 1$. ■

APPENDIX B. CLAIM B.0.1

Claim B.0.1 *For two independent random variables X_1 and X_2 with $H(X_1) = a$ and $H(X_2) = b$, if $H(X_1|Y) \leq \epsilon_n$ where Y is another random variable with $H(Y) \leq a$, then $b - \epsilon_n \leq H(X_2|Y) \leq b$, $H(Y|X_2) \geq a - 2\epsilon_n$.*

proof: Since $H(X_1) = a$ and $H(X_1|Y) \leq \epsilon_n$, we have

$$H(Y) = H(X_1, Y) - H(X_1|Y) \geq H(X_1) - H(X_1|Y) \geq a - \epsilon_n.$$

Next $H(Y) \leq a$ implies that

$$H(Y|X_1) = H(X_1|Y) + H(Y) - H(X_1) \leq \epsilon_n + a - a = \epsilon_n.$$

As X_1 and X_2 are independent and $H(X_2) = b$, we have

$$\begin{aligned} b = H(X_2) = H(X_2|X_1) &\leq H(X_2|X_1, Y) + H(Y|X_1) \\ &\leq H(X_2|X_1, Y) + \epsilon_n \leq H(X_2|Y) + \epsilon_n \leq b + \epsilon_n. \end{aligned}$$

Thus,

$$b - \epsilon_n \leq H(X_2|Y) \leq b.$$

Finally, we obtain

$$\begin{aligned} H(Y|X_2) = H(Y) - I(Y; X_2) &= H(Y) + H(X_2|Y) - H(X_2) \\ &\geq a - \epsilon_n + b - \epsilon_n - b = a - 2\epsilon_n \end{aligned}$$

■

APPENDIX C. LEMMA C.0.2

Lemma C.0.2 *If $\beta_1 \neq 0$, $det([M_{21} \;\; M_{22}\underline{\xi}])$ can be represented by*

$$\frac{\xi_2}{\beta_1} det \begin{bmatrix} \alpha_1' & -\beta_2\beta_{11}' + \beta_1\beta_{12}' \\ \alpha_2' & -\beta_2\beta_{21}' + \beta_1\beta_{22}' \end{bmatrix}. \tag{C.1}$$

where $\underline{\xi}$ satisfies $[\beta_1 \;\beta_2]\underline{\xi} = 0$.

proof: Because $\underline{\xi}$ satisfies $[\beta_1 \;\beta_2]\underline{\xi} = 0$, we can have $\xi_1 = -\beta_2\xi_2/\beta_1$. Note ξ_2 can be selected to be nonzero. To see this, if $\beta_2 = 0$, ξ_2 can be arbitrary and $\xi_1 = 0$. If $\beta_2 \neq 0$, $\xi_2 = \beta_1\xi_1/\beta_2$ can also be nonzero. By substituting ξ_1 into $[M_{21} \;\; M_{22}\underline{\xi}]$, the determinant of $[M_{21} \;\; M_{22}\underline{\xi}]$ becomes

$$\det\left[M_{21} \;\; M_{22}\begin{bmatrix} -\frac{\beta_2\xi_2}{\beta_1} \\ \xi_2 \end{bmatrix}\right] = \det \begin{bmatrix} \alpha_1' & -\frac{\beta_2\xi_2\beta_{11}'}{\beta_1} + \xi_2\beta_{12}' \\ \alpha_2' & -\frac{\beta_2\xi_2\beta_{21}'}{\beta_1} + \xi_2\beta_{22}' \end{bmatrix} = \frac{\xi_2}{\beta_1}\det \begin{bmatrix} \alpha_1' & -\beta_2\beta_{11}' + \beta_1\beta_{12}' \\ \alpha_2' & -\beta_2\beta_{21}' + \beta_1\beta_{22}' \end{bmatrix}. \tag{C.2}$$

where ξ_2/β_1 is nonzero. ∎

APPENDIX D. LEMMA D.0.3

Lemma D.0.3 *Consider a system of equations $Z = H_1X_1 + H_2X_2$, where X_1 is a vector of length l_1 and X_2 is a vector of length l_2 and $Z \in span([H_1 \quad H_2])$*[1]*. The matrix H_1 has dimension $z_t \times l_1$, and rank $l_1 - \sigma$, where $0 \le \sigma \le l_1$. The matrix H_2 is full rank and has dimension $z_t \times l_2$ where $z_t \ge (l_1 + l_2 - \sigma)$. Furthermore, the column spans of H_1 and H_2 intersect only in the all-zeros vectors, i.e. $span(H_1) \cap span(H_2) = \{0\}$. Then there exists a unique solution for X_2.*

proof: Because $Z \in span([H_1 \quad H_2])$, there exists X_1 and X_2 such that $Z = H_1X_1 + H_2X_2$. Now assume there is another set of X_1' and X_2' such that $Z = H_1X_1' + H_2X_2'$. Then we will have

$$H_1(X_1 - X_1') = H_2(X_2 - X_2'). \tag{D.1}$$

Because $span(H_1) \cap span(H_2) = \{0\}$, both sides of eq. D.1 are zero. Furthermore, since H_2 is a full rank matrix, $X_2 = X_2'$. The solution of X_2 is unique. ■

[1] *Throughout the paper, $span(A)$ refers to the column span of A.*

APPENDIX E. LEMMA E.0.4

Lemma E.0.4 *There are at least* q^2-1 *distinct values for* $\breve{M}_{33}\underline{\theta}$ *when there are* q^3-1 *distinct values for* $\underline{\theta}$*.*

proof: Since $\breve{M}_{33}$ is a 4×5 matrix with rank at least 3, we could find two vectors $\underline{\breve{\gamma}}_1$ and $\underline{\breve{\gamma}}_2$ such that the matrix $\breve{M}'_{33} = \begin{bmatrix} \breve{M}_{33} \\ \underline{\breve{\gamma}}_1 \\ \underline{\breve{\gamma}}_2 \end{bmatrix}$ and $rank(\breve{M}'_{33}) = 5$. We will have that there are q^3-1 distinct values for $\breve{M}'_{33}\underline{\theta}$. Next note that since $rank(M_{33}) \geq 4$, $\underline{\breve{\gamma}}_1$ can be selected as the coding coefficient for X_3 on E_{osk} such that $rank \begin{bmatrix} \breve{M}_{33} \\ \underline{\breve{\gamma}}_1 \end{bmatrix} \geq 4$. Since by precoding at s_3, $\underline{\breve{\gamma}}_1\underline{\theta} = 0$. Hence, by removing $\underline{\breve{\gamma}}_1\underline{\theta}$ from $\breve{M}'_{33}\underline{\theta}$, there will be q^3-1 distinct vectors, if we further remove $\underline{\breve{\gamma}}_2\underline{\theta}$ from $\breve{M}'_{33}\underline{\theta}$, there will be at least q^2-1 distinct values. Hence, there will be at least q^2-1 distinct values for $\breve{M}_{33}\underline{\theta}$. ∎

APPENDIX F. LEMMA F.0.5

Lemma F.0.5 *If* $rank(HM) = rank(H) = r$*, then* $span(HM) = span(H)$*.*

proof: First note that $span(HM) \subseteq span(H)$. Assume $span(HM) \neq span(H)$, then there is a vector $\vec{v} \in span(H)$ but not in $span(HM)$. Then,

$$rank([HM \quad \vec{v}]) = rank(HM) + 1 = r + 1 > r = rank(H)$$

However, it contradicts the fact that $rank(H) \geq rank([HM \quad \vec{v}])$, since $[HM \quad \vec{v}] \subseteq span(H)$. Hence $span(HM) = span(H)$. ■

BIBLIOGRAPHY

[1] C. Fragouli and E. Soljanin, *Network Coding Fundamentals.* Now Publishers Inc., 2007.

[2] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, "Network Information Flow," *IEEE Trans. on Info. Th.*, vol. 46, no. 4, pp. 1204–1216, 2000.

[3] R. Koetter and M. Médard, "An Algebraic approach to network coding," *IEEE/ACM Trans. on Networking*, vol. 11, no. 5, pp. 782–795, 2003.

[4] S. Jaggi, P. Sanders, P. Chou, M. Effros, S. Egner, K. Jain, and L.M.G.M.Tolhuizen, "Polynomial Time Algorithms for Multicast Network Code Construction," *IEEE Trans. on Info. Th.*, vol. 51, no. 6, pp. 1973–1982, 2005.

[5] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger, "A Random Linear Network Coding Approach to Multicast," *IEEE Trans. on Info. Th.*, vol. 28, no. 4, pp. 585–592, 1982.

[6] D. S. Lun, N. Ratnakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, "Minimum-Cost Multicast over Coded Packet Networks," *IEEE Trans. on Info. Th.*, vol. 52, pp. 2608–2623, June 2006.

[7] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. on Info. Th.*, vol. 51, no. 8, pp. 2745 – 2759, 2005.

[8] Z. Li and B. Li, "Network coding: the Case of Multiple Unicast Sessions," in *42st Allerton Conf. on Comm., Contr. and Comp.*, 2004.

[9] D. Traskov, N. Ratnakar, D. Lun, R. Koetter, and M. Medard, “Network Coding for Multiple Unicasts: An Approach based on Linear Optimization,” in *IEEE Intl. Symp. on Info. Th.*, 2006, pp. 1758–1762.

[10] C.-C. Wang and N. B. Shroff, “Pairwise Intersession Network Coding on Directed Networks,” *IEEE Trans. on Info. Th.*, vol. 56, no. 8, pp. 3879–3900, 2010.

[11] S. Shenvi and B. K. Dey, “A Simple Necessary and Sufficient Condition for the Double Unicast Problem,” in *IEEE Intl. Conf. Comm.*, 2010, pp. 1–5.

[12] E. Erez and M. Feder, “Improving the Multicommodity Flow Rate with Network Codes for Two Sources,” *IEEE J. Select. Areas Comm.*, vol. 27, no. 5, pp. 814–824, 2009.

[13] A. Das, S. Vishwanath, S. A. Jafar, and A. Markopoulou, “Network Coding for Multiple Unicasts: An Interference Alignment Approach,” in *IEEE Intl. Symp. on Info. Th.*, 2010, pp. 1878 – 1882.

[14] V. Cadambe and S. Jafar, “Interference alignment and degrees of freedom of the K user interference channel,” *IEEE Trans. on Info. Th.*, vol. 54, no. 8, pp. 3425–3441, 2008.

[15] A. Ramakrishnan, A. Das, H. Maleki, A. Markopoulou, S. Jafar, and S. Vishwanath, “Network coding for three unicast sessions: Interference alignment approaches,” in *48th Allerton Conf. on Comm., Contr. and Comp.*, 2010, pp. 1054 – 1061.

[16] J. Han, C.-C. Wang, and N. Shroff, “Analysis of Precoding-based Intersession Network Coding and The Corresponding 3-Unicast Interference Alignment Scheme,” in *49th Allerton Conf. on Comm., Contr. and Comp.*, 2011, pp. 1033–1040.

[17] N. Harvey, R. Kleinberg, and A. Lehman, “On the capacity of information networks,” *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2345–2364, 2006.

[18] S. U. Kamath, D. N. C. Tse, and V. Anantharam, “Generalized Network Sharing Outer Bound and the Two-Unicast Problem,” in *Proceedings of Netcod*, 2011, pp. 1–6.

[19] J. Price and T. Javidi, "Network Coding Games with Unicast Flows," *IEEE J. Select. Areas Comm.*, vol. 26, no. 7, pp. 1302–1316, 2008.

[20] S. Huang and A. Ramamoorthy, "A Note on the Multiple Unicast Capacity of Directed Acyclic Networks," in *IEEE Intl. Conf. Comm.*, 2011, pp. 1–6.

[21] ——, "On the multiple unicast capacity of 3-source, 3-terminal directed acyclic networks," in *Information Theory and Applications Workshop (ITA)*, 2012, pp. 152 –159.

[22] ——, "On the multiple unicast capacity of 3-source, 3-terminal directed acyclic networks," *IEEE/ACM Trans. on Networking, 2013 (accepted)*, 2013 [Online] *Available: http://arxiv.org/abs/1302.4474*.

[23] ——, "An achievable region for the double unicast problem based on a minimum cut analysis," in *IEEE Information Theory Workshop (ITW)*, 2011, pp. 120 – 124.

[24] ——, "An achievable region for the double unicast problem based on a minimum cut analysis," *IEEE Trans. on Communications, 2013 (accepted)*, 2012 [Online] *Available: http://www.ece.iastate.edu/~adityar/publications.htm*.

[25] M. Médard, M. Effros, T. Ho, and D. Karger, "On coding for non-multicast networks," 41^{st} *Allerton Conference on Communication, Control and Computing*, 2003.

[26] T. Ho, Y. Chang, and K. J. Han, "On constructive network coding for multiple unicasts," in *44th Allerton Conf. on Comm., Contr. and Comp.*, 2006.

[27] X. Yan, R. W. Yeung, and Z. Zhang, "The Capacity Region for Multi-source Multi-sink Network Coding," in *IEEE Intl. Symp. on Info. Th.*, 2007, pp. 116–120.

[28] A. E. Kamal, A. Ramamoorthy, L. Long, and S. Li, "Overlay protection against link failures using network coding," *IEEE/ACM Trans. on Networking*, vol. 19, no. 4, pp. 1071–1084, 2011.

[29] S. Li and A. Ramamoorthy, "Protection against link errors and failures using network coding in overlay networks," *IEEE Trans. on Comm.*, vol. 59, no. 2, pp. 518–528, 2011.

[30] X. Yan, J. Yang, and Z. Zhang, "An outer bound for multisource multisink network coding with minimum cost considerateion," *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2373–2385, 2006.

[31] X. Xu, Y. Zeng, Y. Guan, and T. Ho, "On the Capacity Region of Two-User Linear Deterministic Interference Channel and Its Application to Multi-Session Network Coding," *Submitted. (available at http://www.arxiv.org/abs/1207.1986).*

[32] W. Zeng, V. R. Cadambe, and M. Medard, "An Edge Reduction Lemma for Linear Network Coding and An Application to Two-Unicast Networks," in *50th Allerton Conf. on Comm., Contr. and Comp.*, 2012.

[33] A. Ramamoorthy and M. Langberg, "Communicating the sum of sources over a network," *IEEE Journal on Selected Areas in Communication: Special Issue on In-network Computation: Exploring the Fundamental Limits*, vol. 3(4), pp. 655–665, 2013.

[34] M. Langberg, A. Sprintson, and J. Bruck, "The encoding complexity of network coding," *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2368–2397, 2006.

[35] R. Motwani and P. Raghavan, *Randomized Algorithms.* Cambridge University Press, 1995.

[36] R. K. Ahuja, T. L. Maganti, and J. B. Orlin, *Network Flows:Theory, Algorithms and Applications.* Prentice-Hall, 1993.

[37] R. Koetter, M. Effros, T. Ho, and M. Médard, "Network codes as codes on graphs," in *Conf. on Information Sciences and Systems*, 2004.