FUNDAMENTALS OF MULTI-HOP MULTI-FLOW WIRELESS NETWORKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ilan Shomorony

August 2014

FUNDAMENTALS OF MULTI-HOP
MULTI-FLOW WIRELESS NETWORKS

Ilan Shomorony, Ph.D.
Cornell University 2014

The conventional design of wireless networks is based on a centralized architecture where a base station, or an access point, directly exchanges data with the end users, and communication is restricted to the one-to-many (broadcast) and many-to-one (multiple-access) *single-hop* paradigms. However, as the number of users and the data demand increase dramatically, and we move towards the future of wireless networks, *multi-hop* and *multi-flow* paradigms are expected to play a central role by enabling a *denser spatial reuse of the spectrum* and the adaptation to heterogeneous network scenarios characterized by the presence of low-power nodes, relays, and user-operated infrastructure.

A major challenge in multi-hop multi-flow wireless networks is that "interference management" and "relaying" are coupled with each other. In other words, wireless relay nodes must play a dual role: they serve as intermediate steps for multi-hop communication and as part of the mechanism that allows interference management schemes. Nonetheless, in the information theory literature, these two tasks have traditionally been addressed separately, and the fundamental principles of the "networks of the future" are currently not well understood. In this dissertation, we take a unified approach to relaying and interference management, and seek to develop tools to study the fundamentals of multi-hop multi-flow wireless networks.

In the first part of the dissertation, we study multi-hop multi-flow wireless

networks from a high-SNR, or degrees-of-freedom (DoF) perspective. We first consider multi-hop *two-unicast* networks, and characterize the DoF as a function of the network graph. Then, we consider $K \times K \times K$ wireless networks and introduce a coding scheme called *Aligned Network Diagonalization* (AND) that allows the relays to neutralize *all* the interference experienced by the destinations. This proves that $K \times K \times K$ wireless networks have K DoF and demonstrates the potential of a coupled approach to relaying and interference management.

In the second part of the dissertation, we present a characterization of the Gaussian noise as the worst-case additive noise in multi-hop multi-flow wireless networks. Besides generalizing a classical point-to-point information theory result, this provides theoretical support for the widespread adoption of Gaussian noise models and yields a tool for obtaining capacity outer bounds for Gaussian networks by considering networks with different noise statistics.

In the final part of the dissertation, we introduce new techniques to obtain capacity outer bounds for multi-hop multi-flow networks. First, we use the worst-case noise result to show that the capacity region of $K \times K \times K$ wireless networks with general connectivity can be outer-bounded by the capacity region of the same network under the truncated deterministic model. We then present a generalization of the classical cut-set bound for multi-hop multi-flow *deterministic* networks, which, besides recovering and unifying other previously known bounds, yields new applications, in both deterministic and non-deterministic settings. In particular, we obtain a rank-based bound for the capacity of linear deterministic multi-flow networks and for the DoF of AWGN multi-flow networks, which yields graph-theoretic conditions for K DoF to be achievable in $K \times K \times K$ networks with general connectivity.

BIOGRAPHICAL SKETCH

Ilan Shomorony received a B.S. degree in Mathematics and Electrical and Computer Engineering from the Worcester Polytechnic Institute, in Worcester, MA, in 2009. He completed his graduate work in Electrical and Computer Engineering at Cornell University, in Ithaca, NY, where he earned the M.S. degree in August 2012 and the Ph.D. degree in August 2014. He spent the summer of 2011 as an intern at HP Labs, in Palo Alto, CA.

In 2009, Ilan received the Olin Fellowship from the School of Electrical and Computer Engineering at Cornell University, and, in 2011, his paper "Sum Degrees-of-Freedom of Two-Unicast Wireless Networks" was a finalist in the ISIT Best Student Paper Award. In 2013, he received the Qualcomm Innovation Fellowship with Alireza Vahid for their work on "Collaborative Interference Management". Upon his graduation, he was awarded a postdoctoral fellowship from the Simons Institute for the Theory of Computing at the University of California, Berkeley.

To my father,

who would have been extremely proud of my achievement,

and to my mother,

whose strength and devotion never cease to amaze me.

ACKNOWLEDGEMENTS

The five years that led to the writing of this dissertation have been some of the most exciting yet turbulent times in my life. The thrills and the hard work associated with the start of a career in scientific research were punctuated by personal and family hardships, making the whole experience even more challenging and deepening the sense of accomplishment I get from completing this dissertation. I am extremely grateful to all the friends, colleagues, and family members who, through their care and support, made it all possible.

First and foremost, I thank my advisor, Prof. Salman Avestimehr, for his guidance and mentoring during my years in grad school. Having started at Cornell at the same time as him, I had the opportunity of experiencing first-hand his enthusiasm and passion for research, intensified by the pressure of the tenure-track process. His technical brightness and knowledge of the field were inspiring and invaluable as I took my first steps into the research world. Moreover, his teaching and research style and his approach to “picking the right problem”, “tackling it”, “cracking it”, and later “collecting the low-hanging fruits” will certainly have an everlasting impact in the researcher I have become, the professor I hope to be, and, perhaps most notably, my PowerPoint presentations.

Together with me from day one in grad school was my good friend and office-mate Alireza Vahid. The (somewhat excessive) interaction we had over five years turned us into partners both in crime and in award-winning research proposals. I thank him for cheering up my days in the office with his jokes and good spirit and blame him for wasting countless hours of my precious time.

Just like the lab becomes a grad-student’s second (or sometimes first) home, the lab-mates often become his/her second family. I wholeheartedly thank my colleagues, the entire Cornell ECE family and all my good friends who managed to turn the gloomy prison-like Rhodes Hall into a wonderful place. I extend special thanks to my research group members Sina Lashgari, Navid NaderiAlizadeh, and David Kao, the founders of the ECE party group Enrique Mallada and Nithin Michael, my personal counselor, trainer, and bodyguard Guilherme Pinto, and Rhodes Penitentiary fellow inmates Ibrahim Issa, Raphael Loca, Nina Voulis, Yiting Xie, Yücel Altuğ, Prof. Eilyan Bitar, Daniel Richter, and Tirza Routtenberg, former inmates Oliver Kosut and Frank Ciaramello, and prison escapees Rad Niazadeh, Hazal Yüksel, Silas Fong, Ronald Ilma, Amandy Nwana, and Amir Sadovnik.

In addition, I was fortunate enough to have friends outside the Cornell ECE walls. They were responsible for bringing balance and diversity to my life, making sure I never spent more than twenty hours a day in the lab, and contributing to the development of my drinking habit. I would like to thank my WPI friends for life Alvaro Soares, Shant Orchanian and Jeremy Chapman; the artist, singer, and polyglot Tatiana Abaurre Alencar, with whom I shared many precious moments; my skiing instructor Lindsay March, whose passion for all things in life was captivating; the multi-talented biologist Eliana Monteverde *y todos los latinos*; the Brazilian *pegadores* Daniel Hauagge, Bruno Abrahao and Pedro Erber; and especially my cousin, roommate, math consultant, and *cachaça*-buddy Daniel Fleischman. I would also like to offer my gratitude to Rabbi Dovid Birk and all the great friends I made at the Cornell Chabad, for inspiring me, "spiritualizing the material", and brightening up many of my days.

Last, but most importantly, I would like to express my deepest gratitude towards my beautiful family, who stood by me at every step of my life. My father Gabriel would have been the happiest of all people to see how far I have come, and I hope that my brothers André, Jonas and I will keep honoring his memory

in all of our future endeavors. I also thank my family members in Brazil, who always gave me their full support and often flew all the way to the United States, either to celebrate the happy moments or to provide a helping hand in the difficult times. Finally, I want to thank my wonderful mother Denise, a woman whose devotion to her family cannot be described in words and whose strength in the face of life's adversities is a true source of inspiration for all those who know her and especially for me.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Shannon's information theory is regarded as one of the most influential works from the twentieth century. While its applications and consequences range across a number of fields, its impact on the world of communications, setting the beginning of a *digital age*, is especially significant. In particular, the considerable progress achieved in extending Shannon's original insights to the context of the one-to-many (broadcast) and many-to-one (multiple-access) *single-hop* communication paradigms enabled a revolution in wireless communications that brought us technologies such as cellular telephony and wi-fi.

Most notably in the last few years, these techonologies have experienced an unparalleled adoption and we have witnessed a dramatic increase in wireless data traffic, caused by the success of online media streaming services and the proliferation of smart phones and tablets. In the next decade, the continuation of this trend, illustrated in Fig. 1.1, will pose significant technical challenges to the wireless industry. Given the scarcity of unused wireless spectrum, the only way to meet this ever increasing demand is to exploit a much denser *spatial reuse of the spectrum*, or a *densification* of wireless networks. This means that, instead of relying on traditional network architectures, centralized around the idea of a base station or an access point, the wireless network infrastructure should be supplemented and enhanced with the deployment of lower-power base stations and relays antennas, the employment of currently existing nodes for relaying purposes, and the utilization of user-operated infrastructure, such as residential femtocells.

As a result, new concepts in wireless communications such as multi-

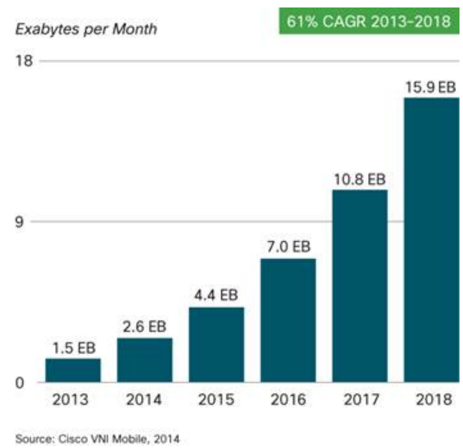


Figure 1.1: Cisco forecast for mobile data traffic until 2018.

hopping, device-to-device communications and heterogeneous networks start emerging as important components of the networks of the future. In enabling all of these advances, *multi-hop* and *multi-flow* communication paradigms are expected to play a central role. However, Shannon's original theory developed for point-to-point channels does not account for these new scenarios, and most of the research aimed at developing a *network information theory* has traditionally focused on each of these two aspects – multi-hop and multi-flow – separately.

A major challenge in studying multi-hop multi-flow wireless networks is that the problems of "interference management" and "relaying" are coupled with each other. In other words, wireless relay nodes must play a dual role: they serve as intermediate steps for multi-hop communication and as part of the mechanism that allows interference management schemes. Nonetheless, in the information theory literature, these two tasks are commonly addressed individually. The relaying problem is usually studied in the context of multi-hop single-flow wireless networks (or relay networks). For such networks, the capacity was shown in [6] to be within a constant gap of the cut-set bound, and several relaying strategies are known to achieve the capacity to within a constant gap (e.g., quantize-map-forward [6], lattice quantization followed by map-and-forward [47] and compress-and-forward [42]). On the other hand, the problem of interference management is mostly studied in the context of multi-flow single-hop wireless networks (or interference channels). While the capacity of the interference channel remains unknown (except for special cases, such as [4, 16, 22, 44, 50, 52, 60]), there has been a variety of capacity approximations derived, including constant-gap capacity approximations [11, 12, 20] and degrees-of-freedom characterizations [13, 19, 34, 35, 43, 45].

Once we consider multi-hop multi-flow wireless networks, a natural question is whether simply combining insights from these two research directions is optimal or if there are signicant performance gains to be obtained from a *coupled* approach to relaying and interference management. However, the results along this line of study are scarcer and, in spite of the current industry interest in their potential applications, our knowledge about the fundamental performance limits of multi-hop multi-flow is still very limited. The goal of this dissertation is to develop novel approaches to study the fundamental principles and unlock the potentials of multi-hop multi-flow wireless communications.

1.1 A Motivating Example

As we move to the multi-hop multi-flow paradigm, it is fundamental to understand whether a decoupled approach for relaying and interference management is optimal. To make this question clear, consider the $K \times K \times K$ wireless network, a two-hop wireless network with K source nodes, K relay nodes and K destination nodes, shown in Fig. 1.2. An approach that decouples relaying and interference management could consist of viewing the $K \times K \times K$ wireless network as the concatenation of two K-user interference channels. Then, interference management techniques designed for the K-user interference channel can be individually applied to each hop, and the relaying is simply the decoding and re-encoding operations performed by the relays. Notice that, under such a decoupled approach, the performance of the overall communication scheme is essentially limited by the performance of the scheme applied to each hop, and two-hop communication systems are conceptually similar to single-hop communication systems. But could a "coupled" scheme take advantage of additional opportunities provided by the relays and attain significantly better performance? In other words, does multi-hopping provide us with additional flexibilities that make the design of communication schemes conceptually different?

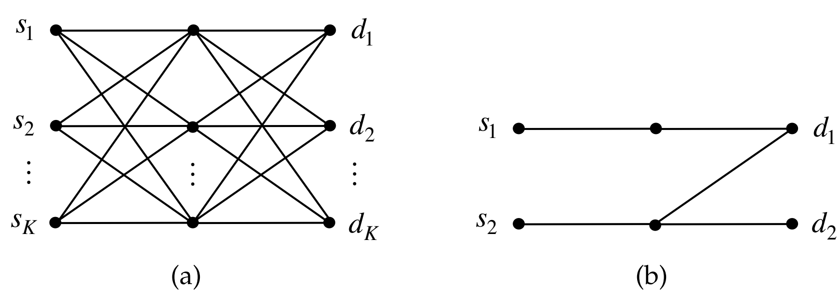


Figure 1.2: (a) A $K \times K \times K$ wireless network with fully connected hops. (b) A $2 \times 2 \times 2$ wireless network with non-fully connected hops.

Throughout this dissertation, in order to tackle this question, we will often focus on a high-SNR analysis, where our metric are the rates achieved asymptotically in the SNR, or the degrees of freedom. The reason for this choice is two-fold. First, we notice that, already from the point of view of this coarse metric, there is a large gap between the rates achieved by state-of-the-art techniques and the known outer bounds. Thus, we cannot hope for tighter capacity characterizations without first obtaining a first-order approximation provided by a degrees-of-freedom analysis. Second, we point out that, unlike an exact capacity characterization, the degrees-of-freedom characterization is usually oblivious to the specifics of noise distributions and channel gain values, and tends to be intrinsically related to structural properties of the network, such as the topology, the interference patterns, and the traffic demands. Hence, such a characterization often reveals conceptual insights about the fundamentals of communication in a given setting.

In order to compare existing coupled and decoupled approaches, Fig. 1.3 depicts the degrees-of-freedom performance of several schemes for the $K \times K \times K$ wireless network with fully connected hops from Fig. 1.2(a). The decoupled

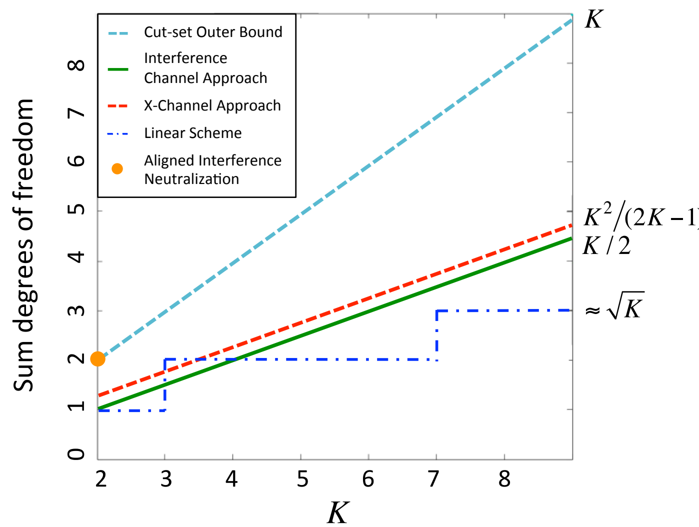


Figure 1.3: Degrees of freedom achieved by different schemes on the $K \times K \times K$ wireless network.

approach that views the $K \times K \times K$ wireless network as the concatenation of two K-user interference channels achieves $K/2$ degrees of freedom since $K/2$ degrees of freedom are achievable on a K-user interference channel, both when the channel gains are fixed and when they are time-varying [13, 45]. Another similar decoupled approach consists of viewing each hop of the $K \times K \times K$ wireless network as a K-user X-Channel (where each transmitter has a message for each receiver). This approach in fact achieves $K^2/(2K-1)$ degrees of freedom [14], which is slightly better than $K/2$. A strategy that couples relaying and interference management can be devised using the result from [48] that shows that, in an $N \times K \times N$ wireless network, scalar linear operations at the relays can neutralize the interference at all destinations as long as $K \geq N(N-1)+1$. Thus, it is possible to achieve $\max\{N : K \geq N(N-1)+1\}$ (roughly $\sqrt{K}$) degrees of freedom on the $K \times K \times K$ wireless network, by using only a subset of N source-destination pairs. As depicted in Fig. 1.3, this coupled scheme only outperforms the Interference Channel and X-Channel approaches for $K = 3$. Another coupled strategy was recently proposed for the case $K = 2$ in [24]. The proposed scheme, named Aligned Interference Neutralization, manages to achieve the cut-set bound of two degrees of freedom, and outperforms all decoupled approaches. However, in general, for $K > 2$, all known schemes fall short of the cut-set outer bound of K degrees of freedom.

Can coupled approaches bring us all the way to the cut-set bound? Or are there tighter outer bounds than the cut-set bound? The latter question appears to be particularly relevant in the case of non-fully-connected hops. For instance, it is easy to see that the $2 \times 2 \times 2$ network illustrated in Fig. 1.2(b) is essentially a Z-channel, known to admit only one degree of freedom, while the cut-set bound only implies that the degrees of freedom cannot exceed two. Therefore, $K \times K \times K$ wireless networks are a canonical example of multi-hop multi-flow networks where the gap between the state-of-the-art inner bounds and the outer bounds is very significant, and an important step in understanding how suboptimal decoupled approaches can be in general.

1.2 Overview of Contributions

In this dissertation, we address the questions raised in the last section for different classes of networks, including $K \times K \times K$ wireless networks. We make original contributions both in the form of inner bounds based on coupling relaying and interference management and in the form of new outer bounds that go beyond the classical cut-set bound.

We present our results divided into three separate parts. In the first one, we provide complete characterizations of the degrees of freedom of two kinds of multi-hop multi-flow wireless networks: *two-unicast layered networks* and $K \times K \times K$ *fully-conected wireless networks*. In the second part, we take a slight detour and present a characterization of the Gaussian noise as the worst-case additive noise in multi-hop multi-flow wireless networks. Besides generalizing a classical point-to-point information theory result for general networks, this result will be useful in the following part of the dissertation. The third and final part of the dissertation deals with new outer-bounding techniques for multi-hop multi-flow networks. Using the worst-case noise result, we will first show that the capacity region of a $K \times K \times K$ wireless network with general connectivity can be outer-bounded by the capacity region of the same network under the truncated deterministic model [6]. We then present a new general outer-bounding technique for multi-hop multi-flow *deterministic* networks. This technique is shown to be a generalization of the classical cut-set bound (for deterministic networks) and, when used in combination with the previous result, provides a new outer bound for the degrees of freedom of non-fully-connected $K \times K \times K$ (AWGN) wireless networks. This bound is tight in the case of the "adjacent-cell interference" topology, and yields graph-theoretic necessary and sufficient conditions for K degrees of freedom to be achievable for non-fully-connected hops. Next we describe each of these three parts in more detail.

I. Degrees-of-Freedom Characterizations: In Part I of this dissertation, we characterize the degrees of freedom of two important classes of multi-hop multi-flow wireless networks. First, in Chapter 2, we consider networks with two source-destination pairs, or two-unicast networks, with a layered topology, an arbitrary number of layers, and arbitrary connectivity between adjacent layers. An example is shown in Fig. 1.4. For such networks, we completely characterize

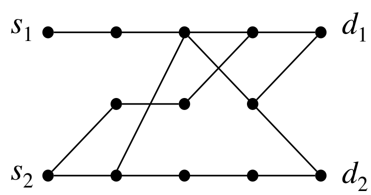


Figure 1.4: Example of a two-unicast layered wireless network.

the sum degrees of freedom as a function of the network topology and show that they can only take values 1, 3/2 and 2 (Theorem 2.1). We then extend this result and characterize the full degrees-of-freedom region, establishing that it can only take one of the five shapes shown in Fig. 1.5 (Theorem 2.2). In these degrees-of-freedom characterizations, two important new notions are introduced. The first one is the idea of *network condensation*, by which a network with an arbitrary number of layers is reduced to a network with at most four layers with the same degrees of freedom. The second one is the graph-theoretic concept of *paths with manageable interference*, which represents a first attempt at finding flow-like structures in multi-user wireless networks. In addition, we develop novel outer bounds that capture the interference structure of a given topology, in order to obtain an outer bound that is tighter than the classical cut-set bound.

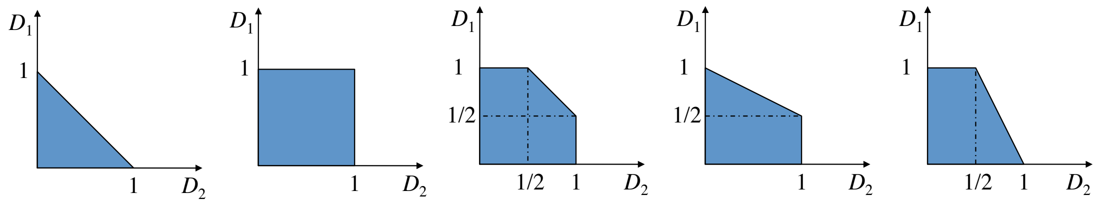


Figure 1.5: Five possible degrees-of-freedom regions for two-unicast layered wireless networks.

Extending the results from two-unicast to general K-unicast wireless networks is a difficult task. To make progress on this front, in Chapter 3, we focus on the $K \times K \times K$ wireless network, shown in Fig. 1.2. As depicted in Fig. 1.3, the best previously known communication schemes for this network achieved essentially $K/2$ degrees of freedom using approaches that decouple the tasks of relaying and interference management. However, the best known upper bound was one degree of freedom per user, or K sum degrees of freedom, obtained from the cut-set outer bound. As a result, this network represented a canonical scenario in which there was a large gap (both quantitative and conceptual) in our understanding of multi-hop multi-flow wireless networks. In Chapter 3, we are able to close this gap by showing that the cut-set bound of K degrees of freedom is in fact achievable, both when the channels are time-varying (Theorem 3.1) and when they are constant (Theorem 3.2). We introduce a new scheme called Aligned Network Diagonalization (AND), which exploits the potential of the relays for interference management in order to effectively create K parallel interference-free channels between each source and its corresponding destination, allowing each user to achieve arbitrarily close to one degree of freedom. This result demonstrates that decoupling relaying and interference management in multi-hop multi-flow wireless networks is suboptimal and a joint design can yield coding schemes that perform significantly better.

II. Robustness of Gaussian Models: In order to study the fundamental limits of communication in networks, it is important to question the meaningfulness of the network models being considered. With this objective, we challenge one of the most widespread assumptions in the stochastic modeling of wireless networks: the Gaussian models. Often motivated by the fact that, from the Central Limit Theorem, the composite effect of many (almost) independent random processes should be approximately Gaussian, such models are ubiquitous in data compression and data communication problems. The additive noise experienced by wireless receivers, for instance, is often modeled as a white Gaussian random process. Similarly, but perhaps less intuitively, data sources are also commonly modeled as Gaussian processes. While these models are formally justified in point-to-point setups as the worst-case assumptions, the same was not known to be the case in network setups and the main reason for these assumptions was analytical tractability. From a theoretical standpoint, a relevant question is: In what scenarios are these Gaussian models worst-case assumptions? And from a practical perspective, we would like to know how compression and communication schemes designed under Gaussian assumptions can be useful in non-Gaussian, practical scenarios.

In Part II of this dissertation, we answer these questions in the context of multi-hop multi-flow wireless networks (Chapter 4) and joint source-channel coding in arbitrary networks (Chapter 5). More precisely, in Theorem 4.1, we prove that for an arbitrary network with i.i.d. additive noise at all nodes, if we fix the noise variances,

$$C_{\text{Gaussian}} \subseteq C_{\text{non-Gaussian}},$$

where C_{Gaussian} and $C_{\text{non-Gaussian}}$ are respectively the capacity regions when the noise distributions are all Gaussian and when they have any other distribution.

Similar ideas allow us to show in Theorem 5.1 that, for arbitrary memoryless networks, if we fix the covariance matrix of the data sources,

$$\mathcal{DR}_{\text{Gaussian}} \subseteq \mathcal{DR}_{\text{non-Gaussian}},$$

where $\mathcal{DR}_{\text{Gaussian}}$ and $\mathcal{DR}_{\text{non-Gaussian}}$ are respectively the distortion regions when the sources are jointly Gaussian and when they have any other distribution.

We prove that the Gaussian distribution for noise and sources is indeed worst-case in these settings by providing a framework that allows coding schemes designed under Gaussian assumptions to be converted to coding schemes that achieve the same performance under arbitrary statistical assumptions. Therefore, not only do we generalize the classical information theory worst-case noise and source results to network settings, but we also establish a robustness result: there exist optimal coding schemes that are robust to non-Gaussianities in the noise and source distributions.

The main idea behind the conversion from the coding scheme designed for a Gaussian network to this new, robust coding scheme is a linear transformation that is applied to the network's inputs and outputs with the purpose of making the resulting sources or noises more Gaussian-like. Fig. 1.6 illustrates this process for the case of the data sources in network compression problems. First, each source node applies a transformation to its non-Gaussian data source with the purpose of "Gaussifying" it. More precisely, we find a sequence $Q_i^{(b)}$ of such transformations for each source such that the resulting effective sources converge in distribution to Gaussian, i.e.,

$$(X_1^{(b)}, ..., X_k^{(b)}) \xrightarrow{d} (X_1^G, ..., X_k^G) \text{ as } b \to \infty.$$

We then prove the existence of optimal coding schemes for which the above

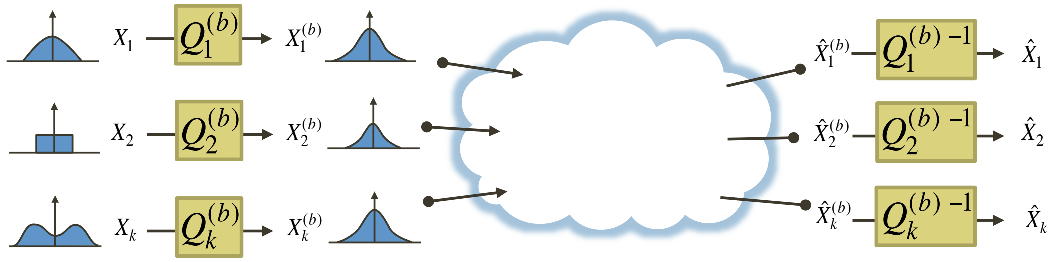


Figure 1.6: "Gaussifying" non-Gaussian data sources.

convergence in distribution implies convergence in distortion, i.e.,

$$E\left\|\mathbf{X}_i^{(b)} - \hat{\mathbf{X}}_i^{(b)}\right\|^2 \to E\left\|\mathbf{X}_i^{G} - \hat{\mathbf{X}}_i^{G}\right\|^2 \text{ as } b \to \infty.$$

This implies that we can build a new coding scheme whose achieved distortion is arbitrarily close to the distortion achieved by the original coding scheme when the sources are actually Gaussian.

III. Outer-Bounding Techniques: A classic tool in the study of network capacity is the cut-set bound [23]. This capacity outer bound is attractive due to its generality – it applies to arbitrary memoryless networks – and the fact that it is a single-letter expression. Furthermore, it is known to be tight in multicast wireline and linear deterministic networks and within a constant gap of capacity in AWGN relay networks [6]. For multi-flow networks, however, the cut-set bound is easily seen to be arbitrarily loose, even for a degrees-of-freedom analysis. Aside from the wireline case, where improvements over the cut-set bound (or min-cut) are known [28, 39, 62], most "non-cut-set" bounds are tied to specific settings (e.g., [20, 55]), and few general techniques are known.

In the third part of this dissertation, we study new ways to obtain outer bounds in multi-hop multi-flow networks. First we notice that an important consequence of the worst-case noise characterization in Chapter 4 is that it allows us to establish connections between the capacity region of networks under different models. In Chapter 6, we pursue this direction and demonstrate in Theorem 6.1 that, in $K \times K \times K$ wireless networks with general connectivity,

$$C_{\text{Gaussian}}(P) \subseteq C_{\text{Uniform}}(P) \subseteq C_{\text{Truncated}}(2P + \alpha), \tag{1.1}$$

where $C_{\text{Gaussian}}(P)$, $C_{\text{Uniform}}(P)$ and $C_{\text{Truncated}}(P)$ are respectively the capacity regions with additive Gaussian noises, additive uniform noises and under the truncated deterministic model [6] for a transmit power constraint of P, and α is a constant. Therefore, any outer bound found for a truncated deterministic $K \times K \times K$ network can be directly translated into an outer bound for the AWGN $K \times K \times K$ network. This idea will be used in Chapter 7 to derive a new outer bound for the degrees of freedom of non-fully-connected layered wireless networks.

In Chapter 7, we propose a new generalization to the cut-set bound for *deterministic* K-unicast networks. More precisely, Theorem 7.1 shows that if a rate tuple $(R_1, ..., R_K)$ is achievable, then there exists a joint distribution $p(x_V)$ on the transmit signals of the nodes in V, such that

$$\sum_{i=1}^{K} R_i \leq \sum_{j=1}^{\ell} I(X_{\Omega_j}; Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_{\Omega_{j-1}^c}), \tag{1.2}$$

for all choices of ℓ node subsets $\Omega_1, ..., \Omega_\ell$ such that $V = \Omega_0 \supseteq \Omega_1 \supseteq \Omega_2 \supseteq ... \supseteq \Omega_\ell \supseteq \Omega_{\ell+1} = \emptyset$, and $d_i \in \Omega_j \Leftrightarrow s_i \in \Omega_{j+1}$ for $j = 0, 1, ..., \ell$, $i = 1, ..., K$ and any $\ell \geq 1$. The usual cut-set bound corresponds to the case $\ell = 1$.

The intuition behind our bound comes from noticing that a coding scheme for a K-unicast network $\mathcal{N}$, when applied to a *concatenation* of multiple copies of $\mathcal{N}$, can be used to achieve the original rates while inducing essentially the same distribution on the transmit signals of each copy of $\mathcal{N}$. Hence, one should be able to apply the cut-set bound to the concatenated network with a restriction on the possible transmit signal distributions. As we show, one can in fact require the transmit signals distribution on each copy to be *the same*, which can significantly reduce the values that the mutual information terms attain.

In terms of applications of this cut-set bound generalization, we first consider linear finite-field networks. These networks have recently received attention as they allow the deterministic modeling of wireless networks and can provide insights about their AWGN counterparts. Similar to the cut-set bound used in [6], we obtain a general outer-bound expression in terms of ranks of transfer matrices. We then return our focus to $K \times K \times K$ topologies. For binary $K \times K \times K$ networks, our rank-based bound yields necessary and sufficient conditions for rate K to be achieved. Furthermore, using the relationship in (1.1), since the truncated channel model is a deterministic model, we can obtain a bound on the degrees of freedom of $K \times K \times K$ AWGN networks with general connectivity. This bound is tight in the case of the $K \times K \times K$ topology with "adjacent-cell interference" shown in Fig. 1.7 (Theorem 7.3), and allows us to establish graph-theoretic necessary and sufficient conditions for K degrees of freedom to be achievable in general topologies (Theorem 7.2). Thus we es-

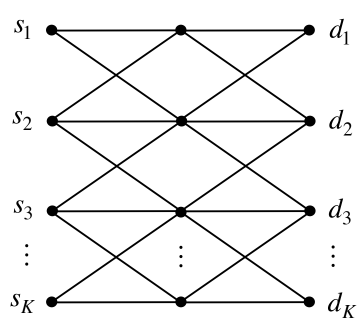


Figure 1.7: $K \times K \times K$ wireless network with adjacent-cell interference.

sentially find a graph-theoretic characterization of manageable interference for $K \times K \times K$ wireless networks, similar to what was done in Chapter 2 for two-unicast networks.

1.3 Definitions and Problem Formulation

A K-unicast wireless network $\mathcal{N} = (G, L)$ consists of a directed graph $G = (V, E)$, where V is the node set and $E \subset V \times V$ is the edge set, and a set of K source-destination pairs $L = \{(s_1, d_1), ..., (s_K, d_K)\} \subset V \times V$. For instance, in a two-unicast Gaussian networks, $L = \{(s_1, d_1), (s_2, d_2)\}$, for some vertices $s_1, s_2, d_1, d_2 \in V$. We will say that a given network is *layered* if the vertex set V can be partitioned into r subsets $V_1, V_2, ..., V_r$ (called layers) in such a way that $E \subset \bigcup_{i=1}^{r-1} V_i \times V_{i+1}$. We will also in general let $\mathcal{S} = \{s_1, ..., s_K\}$ be the set of sources and $\mathcal{D} = \{d_1, ..., d_K\}$ be the set of destinations. For a node $v \in V$, we will let $\mathcal{I}(v) \triangleq \{u \in V : (u, v) \in E\}$ (the input nodes) and $\mathcal{O}(v) \triangleq \{u \in V : (v, u) \in E\}$ (the output nodes).

A real-valued channel gain $h_{i,j}[t]$ is associated with each edge $(i, j) \in E$ at time t. Under *fast fading*, the channel gains $\{h_{i,j}[t]\}_{t=0}^{\infty}$ for $i, j \in \mathcal{V}$ are assumed to be mutually independent i.i.d random processes each obeying an absolutely continuous distribution with finite variance. Under *slow fading*, channel gains are assumed to be constant throughout the communication block.

At time t, each node $i \in V$ transmits a real-valued signal $X_i[t]$. The signal received by node j at time t is given by

$$Y_j[t] = \sum_{i \in \mathcal{I}(j)} h_{i,j}[t] X_i[t] + Z_j[t], \text{ for } t = 1, 2, \dots, \tag{1.3}$$

where $Z_j[t]$ is the i.i.d. additive noise at node j, which is independent of all transmit signals up to time t and of the additive noise processes at all other nodes. We will be mostly interested in the case where $Z_j[t]$ is a zero-mean unit-variance Gaussian discrete-time white noise process associated with node j. If the block of communication has length n, we will use X_i^n to represent the vector $(X_i[0], ..., X_i[n-1])$ and if A is a subset of the nodes, $X_A[t] = (X_i[t] : i \in A)$. Also, if we have a set of nodes named $v_1, v_2, ..., v_m$, when clear from context we will write X_i instead of X_{v_i} to simplify the notation.

Definition 1.1 *A coding scheme C with block length $n \in \mathbb{N}$ and rate tuple $\mathbf{R} = (R_1, ..., R_K) \in \mathbb{R}^K$ for a K-unicast additive noise wireless network consists of:*

1. *An encoding function $f_i : \{1, ..., 2^{nR_i}\} \to \mathbb{R}^n$ for each source s_i, $i = 1, ..., K$, where each codeword $f_i(w_i)$, $w_i \in \{1, ..., 2^{nR_i}\}$, satisfies an average power constraint of P.*
2. *Relaying functions $r_v^{(t)} : \mathbb{R}^{t-1} \to \mathbb{R}$, for $t = 0, ..., n-1$, for each node $v \in V$ that is not a source, satisfying the average power constraint*

$$\frac{1}{n} \sum_{t=0}^{n-1} \left[r_v^{(t)}(y_0, ..., y_{t-1}) \right]^2 \leq P,$$

for all $(y_0, ..., y_{n-1}) \in \mathbb{R}^n$.

3. *A decoding function $g_i : \mathbb{R}^n \to \{1, ..., 2^{nR_i}\}$ for each destination d_i, $i = 1, ..., K$.*

We will assume that instantaneous channel state information is available at all nodes. This means that the encoding, relaying and decoding functions, as defined in Definition 1.1, may depend not only on the node's received signals but also on all channel state information available up to time t, $\mathcal{H}^{(t)} = \left\{ h_{S_i,V_j}[\tau], h_{V_i,D_j}[\tau] : i, j \in \{1, ..., K\}, 0 \leq \tau \leq t \right\}$. This dependence is omitted in Definition 1.1 for simplicity.

Definition 1.2 *The error probability of a coding scheme C (as defined in Definition 1.1), is given by*

$$P_{\text{error}}(C) = \Pr\left[\bigcup_{i=1}^{K} \{W_i \neq g_i(Y_{d_i}[0], ..., Y_{d_i}[n-1])\} \right],$$

where we assume that each W_i is chosen independently and uniformly at random from $\{1, ..., 2^{nR_i}\}$, that source s_i transmits codeword $f_i(W_i)$ over the n time steps, and relay v transmits $r_v^{(t)}(Y_v[0], ..., Y_v[t-1])$ at time $t = 1, ..., n-1$.

Definition 1.3 *A rate tuple $\mathbf{R} = (R_1, ..., R_K)$ is said to be achievable for a given K-unicast wireless network if there exists a sequence of coding schemes C_n with rate tuple $\mathbf{R}$ and blocklength n, for which $P_{\text{error}}(C_n) \to 0$, as $n \to \infty$. The sequence of coding schemes C_n, $n = 1, 2, ...$, is then said to achieve rate tuple $\mathbf{R}$.*

Definition 1.4 *The capacity region $C(P)$ of a K-unicast wireless network is the closure of the set of achievable rate tuples, and the sum-capacity is defined as*

$$C_\Sigma(P) = \max_{(R_1,...,R_K)\in C(P)} \sum_{i=1}^{K} R_i.$$

Definition 1.5 *The sum degrees of freedom of a K-unicast wireless network are given by*

$$D_\Sigma = \lim_{P\to\infty} \frac{C_\Sigma(P)}{\frac{1}{2}\log P}.$$

Definition 1.6 *The degrees-of-freedom region of a K-unicast wireless network is*

$$\mathbf{D} = \Bigg\{ (D_1, ..., D_K) \in \mathbb{R}_+^K : \forall\, w_1, ..., w_K \in \mathbb{R}_+, \\ \sum_{i=1}^{K} w_i D_i \le \lim_{P\to\infty} \left(\sup_{(R_1,...,R_K)\in C(P)} \frac{w_1 R_1 + ... + w_K R_K}{\frac{1}{2}\log P} \right) \Bigg\}. \tag{1.4}$$

Part I

Degrees-of-Freedom Characterizations

As an exact characterization of the Shannon capacity is still an unrealistic goal for the vast majority of multi-hop multi-flow wireless networks, most of the research on the fundamental limits of communication in these settings has sought alternative ways to obtain insights about the design of optimal coding schemes. Some of the recent noteworthy approaches include the study of deterministic models of wireless networks that mimic the behavior of their stochastic counterparts [6, 12], attempts at finding constant-gap capacity approximations [6, 20, 61], investigations of the Diversity-Multiplexing tradeoff [29, 71, 72], and characterizations of degrees of freedom and generalized degrees of freedom [13, 19, 34, 35, 43, 45].

In this part of the dissertation, we will study multi-hop multi-flow wireless networks from a degrees-of-freedom perspective. This metric can be understood as an asymptotic high-SNR capacity approximation and is particularly interesting in multi-flow (i.e., multi-user) networks, where the performance at the high-SNR regime is essentially limited by the amount of interference between the users. Hence, a degrees-of-freedom characterization aims to capture the ability of the users to perform interference management schemes and can be thought of as the gain that carefully designed coding schemes can obtain over a simple time-sharing of the network resources among the users.

We approach the characterization of the degrees of freedom of multi-hop multi-flow wireless networks from two complementary angles. First we consider networks that can have an arbitrary number of hops but only two flows, i.e., two-unicast multi-hop networks. Then we consider networks that have only two hops, but an arbitrary number of flows, i.e., two-hop K-user networks.

CHAPTER 2

TWO-UNICAST WIRELESS NETWORKS

We start our efforts on degrees-of-freedom characterizations by considering perhaps the simplest class of multi-hop multi-flow wireless networks: two-unicast networks. We focus on layered networks, i.e., networks whose nodes can be partitioned into several sets (the layers), and links may only exist between nodes in adjacent layers. We show that, if the channel gains of the network are all chosen independently according to continuous distributions, then, with probability 1, two-unicast layered Gaussian networks can only have 1, 3/2 or 2 sum degrees-of-freedom and the five degrees-of-freedom regions shown in Fig. 1.5. We provide sufficient and necessary conditions for each case based on network connectivity and a new notion of source-destination paths with manageable interference.

2.1 Manageable Interference

In order to study the degrees of freedom of two-unicast layered wireless networks, it will be key to characterize when the interference between the two communicating pairs – (s_1, d_1) and (s_2, d_2) – will be *manageable*, in the sense that the interference from s_1 to d_2 and from s_2 to d_1 can be simultaneously neutralized while the source signals still reach their intended destinations. We start with a few definitions.

Definition 2.1 *A path between $v_1 \in V$ and $v_k \in V$ is an ordered set of nodes $\{v_1, v_2, ..., v_k\}$ such that $(v_i, v_{i+1}) \in E$ for $i = 1, ..., k-1$.*

We will commonly refer to a path between v_1 and v_k by P_{v_1,v_k}. We write $v_1 \rightsquigarrow v_k$, if there is a path between v_1 and v_k. Otherwise, we write $v_1 \not\rightsquigarrow v_k$. Notice that for any node $v \in V$, $v \rightsquigarrow v$. If the path between v_1 and v_2 contains only two nodes, i.e., $(v_1, v_2) \in E$, we may also write $v_1 \rightarrow v_2$. Moreover, for sets $A, B \subset V$, we will write $A \rightsquigarrow B$ (resp. $A \rightarrow B$) if $u \rightsquigarrow v$ (resp. $u \rightarrow v$) for some $u \in A$ and $v \in B$.

For simplicity, we will assume that any $v \in V$ belongs to at least one path P_{s_i,d_j} for $i \in \{1,2\}$ and $j \in \{1,2\}$. This is reasonable since a node that does not belong to any source-destination path does not alter the achievable rates in the network and can be removed. Moreover, we will always assume that $s_i \rightsquigarrow d_i$ for $i = 1,2$, since $s_i \not\rightsquigarrow d_i$ implies that $R_i = 0$. In order to be able to "cut and paste" path segments we will also consider the following path operations. For a path $P_{v_a,v_b} = \{v_a, v_{a+1}, ..., v_b\}$, we will let $P_{v_a,v_b}[v_c, v_d] = \{v_c, v_{c+1}, ..., v_d\}$ if $a \leq c \leq d \leq b$. Moreover, if we have paths P_{v_e,v_f} and P_{v_f,v_g}, we will let $P_{v_e,v_f} \oplus P_{v_f,v_g}$ be the path which results from concatenating P_{v_e,v_f} and P_{v_f,v_g}.

Definition 2.2 *Paths P_{v_a,v_b} and P_{v_c,v_d} are disjoint if $P_{v_a,v_b} \cap P_{v_c,v_d} = \emptyset$.*

Definition 2.3 *For a subset of the vertices $S \subset V$, we say that $G[S]$ is the graph induced by S on G, if $G[S] = (S, E_s)$, where $E_s = \{(v_i, v_j) \in E : v_i, v_j \in S\}$.*

Definition 2.4 *We say that $\mathcal{N}' = (G', L')$ is a subnetwork of $\mathcal{N} = (G, L)$, if $G' = G[S]$, for some $S \subset V$ such that $L \subset S \times S$, and $L' = L$.*

Definition 2.5 *A set of nodes A is a (B, C) cut in $G = (V, E)$ if $B \not\rightsquigarrow C$ in $G[V \setminus A]$. We let $\mathcal{K}(B, C) \triangleq \min\{|A| : A \text{ is a } (B, C) \text{ cut}\}$.*

In order to characterize when the interference in a given network $\mathcal{N}$ is manageable, we start by extending a concept from the study of two-unicast *wireline* networks. In [53], it is shown that, if a two-unicast wireline (acyclic) network contains an edge $e = (u, v)$ whose removal disconnects s_1 from d_1, s_2 from d_2, and either s_1 from d_2 or s_2 from d_1, then its rate region is given by all rate pairs (R_1, R_2) with $R_1 + R_2 \leq 1$. Intuitively, the node v at the head of e can be thought of as a node that can decode both messages W_1 and W_2 for any possible coding scheme, or an *omniscient* node. The generalization of this concept to the wireless setting, discovered simultaneously in [55] and [68], is as follows:

Definition 2.6 *If node v is a $(\{s_1, s_2\}, d_i)$ cut and some node $u \in \mathcal{I}(v) \cup \{v\}$ is a $(s_{\bar{i}}, \{d_1, d_2\})$ cut, for $i = 1$ or 2, v is an omniscient node.*

An example of a network with an omniscient node is shown in Fig. 2.1. Analogous to the case in [53], an omniscient node is the graph-theoretic struc-

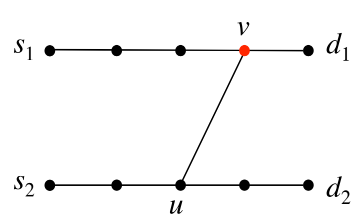


Figure 2.1: Two-unicast network containing omniscient node v. Notice that, according to Definition 2.6, there is a node $u \in \mathcal{I}(v)$ that is an $(\{s_2\}, \{d_1, d_2\})$ cut.

ture that captures when a network only has one degree of freedom. More precisely, as we will show in Section 2.3, we have the following result for wireless networks:

Proposition 2.1 *A two-unicast layered network has* 1 *degree of freedom for almost all values of channel gains if and only if it contains an omniscient node.*

As implied by the above proposition, the absence of an omniscient node guarantees that we can do strictly better than 1 degree of freedom. However, as we shall see, it does not guarantee that we can achieve the cut-set upper bound of 2 degrees of freedom. To characterize the achievability of 2 degrees of freedom, we need to define the notion of a *key node*, which will allow us to define *manageable interference*.

For the next definitions, we assume we have two paths P_{s_1,d_1} and P_{s_2,d_2}. Since we will often make statements which work for both P_{s_1,d_1} and P_{s_2,d_2}, we will let $\bar{i} = 2$ if $i = 1$ and $\bar{i} = 1$ if $i = 2$. Furthermore, we will let $\ell(v)$ be the index corresponding to the layer containing v, i.e., $v \in V_{\ell(v)}$. Notice that ℓ induces a partial ordering on the nodes.

Definition 2.7 *For $i = 1, 2$ and $S \supset P_{s_1,d_1} \cup P_{s_2,d_2}$, the* key node *$w_i(P_{s_i,d_i}, S)$ is the first node in P_{s_i,d_i} that is an $(s_{\bar{i}}, d_i)$ cut in $G[S]$ (i.e., the node $v \in P_{s_i,d_i}$ with minimum $\ell(v)$ that is a $(s_{\bar{i}}, d_i)$ cut).*

Notice that the above definition is well-posed. The existence of $w_i(P_{s_i,d_i}, S)$ is guaranteed by the fact that $d_i \in P_{s_i,d_i}$ and d_i is trivially a $(s_{\bar{i}}, d_i)$ cut. The uniqueness of $w_i(P_{s_i,d_i}, S)$ comes from the fact that no two nodes in P_{s_i,d_i} can be on the same layer. To simplify the notation, whenever the choice of paths P_{s_i,d_i} for $i = 1, 2$ is clear from context, we will write $w_i(S)$, and when S is also clear, we will simply write w_i.

Consider a subnetwork $(G[S], \{(s_1, d_1), (s_2, d_2)\})$ for some $S \supset (P_{s_1,d_1} \cup P_{s_2,d_2})$, for two paths P_{s_1,d_1} and P_{s_2,d_2}. We will define

$$n_i(S) \triangleq |\{v \in S \cap \mathcal{I}(w_i(S)) : s_{\bar{i}} \rightsquigarrow v \text{ in } G[S]\}|$$

for $i = 1, 2$, i.e., n_i counts the number of parent nodes of w_i that are reachable from $s_{\bar{i}}$ in $G[S]$. Moreover, we define $n_i^0 = n_i(P_{s_1,d_1} \cup P_{s_2,d_2})$, when there is no ambiguity in the choice of our paths P_{s_1,d_1} and P_{s_2,d_2}. The following observation relates these two quantities:

Lemma 2.1 *Given two paths P_{s_1,d_1} and P_{s_2,d_2} and any $S \supset P_{s_1,d_1} \cup P_{s_2,d_2}$, we have $n_i(S) \geq n_i^0$, for $i = 1, 2$.*

Proof: Since $G[P_{s_1,d_1} \cup P_{s_2,d_2}]$ has at most two nodes per layer, $|\mathcal{I}(w_i)| \leq 2$, which implies $n_i^0 \in \{0, 1, 2\}$ for $i = 1, 2$. If $n_i^0 = 0$ there is nothing to prove. If $n_i^0 = 1$, we have $s_{\bar{i}} \rightsquigarrow d_i$ and then $n_i(S) \geq 1$. Finally, suppose $n_i^0 = 2$. Clearly, $n_i(S) \geq 1$. Since $w_i(S)$ is the first $(\{s_{\bar{i}}\}, \{d_i\})$ cut on P_{s_i,d_i}, $\ell(w_i(P_{s_1,d_1} \cup P_{s_2,d_2})) \leq \ell(w_i(S))$. Therefore, if we had $n_i(S) = 1$, the unique node in $\mathcal{I}(w_i(S))$ reachable from $s_{\bar{i}}$ would be on P_{s_i,d_i} and would thus be also an $(s_{\bar{i}}, d_i)$ cut on P_{s_i,d_i}, which is a contradiction. ■

The following notion is the central ingredient in our degrees-of-freedom characterization.

Definition 2.8 *Two paths P_{s_1,d_1} and P_{s_2,d_2} have manageable interference if we can find $S \supseteq P_{s_1,d_1} \cup P_{s_2,d_2}$, such that $w_1(S) \notin P_{s_2,d_2}$, $w_2(S) \notin P_{s_1,d_1}$, $n_1(S) \neq 1$ and $n_2(S) \neq 1$.*

The following example illustrates the definitions above.

Example 2.1. Consider the network depicted in Figure 2.2. First suppose we

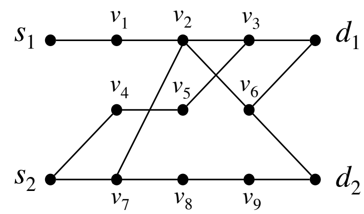


Figure 2.2: Two-unicast layered network considered in Example 1.

choose paths $P_{s_1,d_1} = \{s_1, v_1, v_2, v_3, d_1\}$ and $P_{s_2,d_2} = \{s_2, v_7, v_2, v_6, d_2\}$ and $S = P_{s_1,d_1} \cup P_{s_2,d_2}$. Then $w_1(S) = w_2(S) = v_2$ is an omniscient node since it is an $(\{s_1, s_2\}, \{d_1, d_2\})$ cut in $G[S]$. Next suppose we pick as paths $P_{s_1,d_1} = \{s_1, v_1, v_2, v_3, d_1\}$ and $P'_{s_2,d_2} = \{s_2, v_7, v_8, v_9, d_2\}$. For $S = P_{s_1,d_1} \cup P'_{s_2,d_2}$, $w_1(S) = v_3$ and $w_2(S) = s_2$, and $n_1^0 = 1$, $n_2^0 = 0$. If we consider the entire network, i.e., $S = V$, $w_1(S) = d_1$ and $w_2(S) = d_2$, and $n_1(V) = 2$ and $n_2(V) = 1$. Instead, if we consider the subnetwork $\mathcal{N} = (G[S'], L)$, where $S' = V \setminus \{v_6\}$, $w_1(S) = v_3$ and $w_2(S) = s_2$ are not omniscient and we have $n_1(S') = 2$ and $n_2(S') = 0$. We conclude that P_{s_1,d_1} and P'_{s_2,d_2} have manageable interference.

As it will turn out, paths with manageable interference is precisely the graph-theoretic structure that characterizes when 2 degrees of freedom are achievable.

Proposition 2.2 *A two-unicast layered network has 2 degrees of freedom for almost all values of channel gains if and only if it has paths with manageable interference.*

2.2 Characterizing the Degrees of Freedom

In this section, we describe how the sum degrees of freedom and the degrees-of-freedom region can be characterized solely as a function of the network graph. As mentioned in the previous section, the notions of omniscient nodes and paths with manageable interference will suffice to characterize the two extreme cases, i.e., when 1 and 2 degrees of freedom are achievable. Hence, the main question we need to answer is what happens in the remaining cases, when the network does not contain an omniscient node, but its interference is not manageable. Our main result, stated below, surprisingly establishes that, besides the cases $D_\Sigma = 1$ and $D_\Sigma = 2$, there is only on more case.

Theorem 2.1 *The sum degrees of freedom D_Σ of a two-unicast layered Gaussian network $\mathcal{N} = ((V, E), \{(s_1, d_1), (s_2, d_2)\})$ are given by*

A) $D_\Sigma = 1$ *if $\mathcal{N}$ contains an omniscient node,*

B) $D_\Sigma = 2$ *if $\mathcal{N}$ has paths with manageable interference,*

C) $D_\Sigma = \frac{3}{2}$ *in all other cases,*

for almost all values of channel gains.

Each of the three cases in Theorem 2.1 is illustrated in Fig. 2.3. Our characterization can in fact be extended to yield the full degrees-of-freedom region of two-unicast layered wireless networks. First we consider the following definition.

Definition 2.9 *Two disjoint paths P_{s_1,d_1} and P_{s_2,d_2} have (s_i, d_i)-manageable interference if we can find $S \subset V$ such that $P_{s_1,d_1}, P_{s_2,d_2} \subset S$, $n_i(S) \neq 1$, for $i = 1$ or 2.*

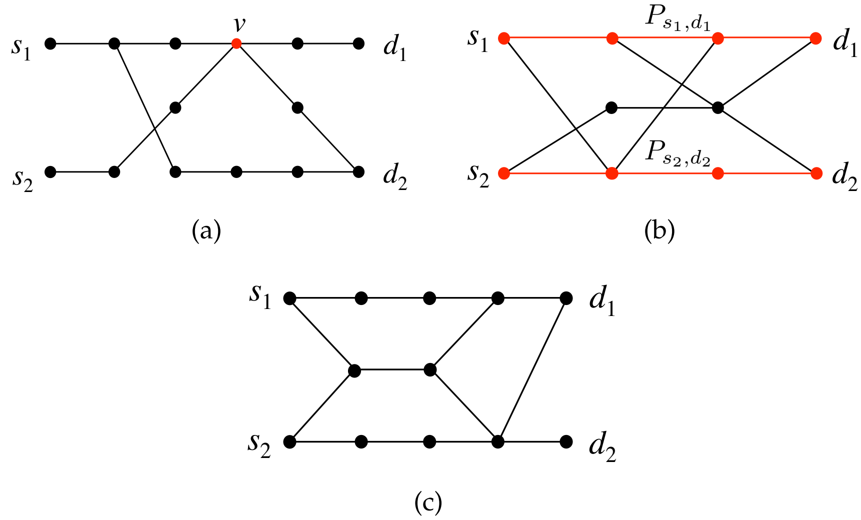


Figure 2.3: (a) Network with an omniscient node v; (b) Network with paths P_{s_1,d_1} and P_{s_2,d_2} with manageable interference; (c) Network that does not contain an omniscient node nor two paths with manageable interference

The following result shows that any two-unicast layered wireless network has one of the five degrees-of-freedom regions shown in Fig. 2.4.

Theorem 2.2 *The degrees-of-freedom region* $\mathbf{D}$ *of a two-unicast layered Gaussian network* $\mathcal{N} = ((V, E), \{(s_1, d_1), (s_2, d_2)\})$ *is given by*

I. $\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 + D_2 \le 1\right\}$ *if* $\mathcal{N}$ *contains an omniscient node,*

II. $\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \le 1, D_2 \le 1\right\}$ *if* $\mathcal{N}$ *has paths with manageable interference,*

III. $\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \le 1, D_2 \le 1, D_1 + D_2 \le \frac{3}{2}\right\}$ *if* $\mathcal{N}$ *is not in cases I, II and contains disjoint paths* P_{s_1,d_1} *and* P_{s_2,d_2} *whose interference is* (s_1, d_1) *and* (s_2, d_2)*-manageable,*

IV. $\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \le 1, D_1 + 2D_2 \le 2\right\}$ *if* $\mathcal{N}$ *is not in cases I, II and III and contains paths* Q_{s_1,d_1}*,* Z_{s_1,d_1} *and* P_{s_2,d_2}*, such that* Q_{s_1,d_1} *and* P_{s_2,d_2} *are disjoint and* *have* (s_1, d_1)*-manageable interference, and* Z_{s_1,d_1} *and* P_{s_2,d_2} *are disjoint and have* (s_2, d_2)*-manageable interference,*

V. $\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_2 \leq 1, 2D_1 + D_2 \leq 2\right\}$ *if* $\mathcal{N}$ *is not in cases I, II and III and contains paths* P_{s_1,d_1}*,* Q_{s_2,d_2} *and* Z_{s_2,d_2}*, such that* Q_{s_2,d_2} *and* P_{s_1,d_1} *are disjoint and have* (s_1, d_1)*-manageable interference, and* Z_{s_2,d_2} *and* P_{s_1,d_1} *are disjoint and have* (s_2, d_2)*-manageable interference,*

for almost all values of channel gains. Moreover, any two-unicast layered Gaussian network $\mathcal{N}$ *falls in one of these cases.*

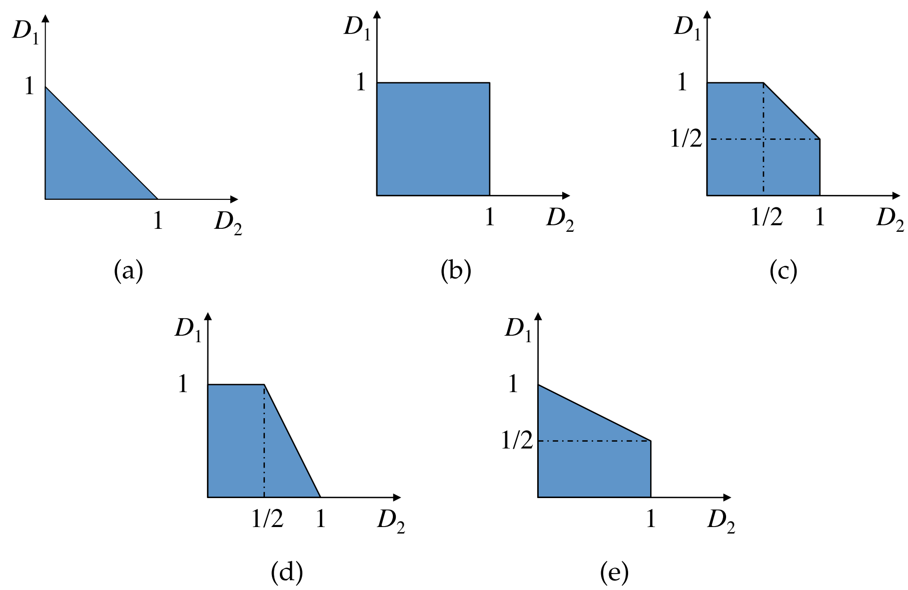


Figure 2.4: Degrees-of-freedom region for networks in (a) case I; (b) case II; (c) case III; (d) case IV; (e) case V.

In the following sections, we consider the three cases of sum degrees of freedom in Theorem 2.1 separately. We present essential parts of the proofs and defer the remaining details to [55].

2.3 Networks with an Omniscient Node

In this section, we provide converse results for networks that contain an omniscient node and, according to Theorem 2.1, have one degree of freedom. We will derive information inequalities which allow us to bound the achievable sum rates, and thus the degrees of freedom.

Suppose a given node v is omniscient, according to Definition 2.6. Then we can assume without loss of generality that the removal of v disconnects d_1 from $\{s_1, s_2\}$ and there is a node $u \in \mathcal{I}(v) \cup \{v\}$ whose removal disconnects s_2 from $\{d_1, d_2\}$. We will focus on the case where $u \in \mathcal{I}(v)$, and show that $D_\Sigma \leq 1$. The case $u = v$ is simpler and follows similar steps.

In order to simplify the converse proofs in this section and in Section 2.5, we will consider a decomposition of the additive Gaussian noise Z_v associated with each node v. More specifically, if $m = |\mathcal{I}(v)|$, we break the noise at node v into m independent noise components, each with variance $1/m$. Then we associate each of these components with one of the incoming edges, and we can define, for $u \in \mathcal{I}(v)$,

$$\tilde{X}_{u,v} \triangleq h_{u,v} X_u + Z_{u,v},$$

where $Z_{u,v}$ is the noise term associated with the edge (u, v). Clearly, we have $Z_v = \sum_{u\in\mathcal{I}(v)} Z_{u,v}$, and Z_v has unit variance. Notice that we can now write, for a node v, $Y_v = \sum_{u\in\mathcal{I}(v)} \tilde{X}_{u,v}$. Moreover, we will define

$$\tilde{X}_u \triangleq \{\tilde{X}_{u,v} : v \in \mathcal{O}(u)\}.$$

As before, we let $\tilde{X}_S$ be the set of all $\tilde{X}_v$'s, for $v \in S$, and $\tilde{X}_v^n$ be a length n vector with all the $\tilde{X}_v[t]$'s, for $t = 1, ..., n$.

In order to find upper bounds to the rates, we will often be interested in showing that certain conditional mutual information terms can be upper bounded by a constant. In particular, if we have a Z structure across two layers in the network, such as the one shown in Fig. 2.5(a), we would like to say that $I(X_c^n; \tilde{X}_c^n | Y_b^n, \tilde{X}_a^n)$ can be upper bounded by a constant that does not depend on P. Intuitively, the reason is that, given $\tilde{X}_a^n$ and Y_b^n, one can subtract $\tilde{X}_{a,b}^n$ from Y_b^n

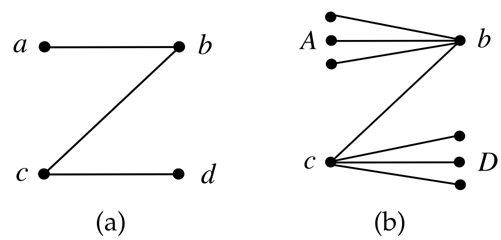


Figure 2.5: The Z structure.

and obtain $\tilde{X}_{c,b}^n$. This means that "almost all" information in $\tilde{X}_c^n$ can be deduced from $(Y_b^n, \tilde{X}_a^n)$, and thus the conditional mutual information cannot be very large. This reasoning is formalized in the following lemma, where we generalize the Z structure to one where $|\mathcal{I}(b)| \geq 2$ and $|\mathcal{O}(c)| \geq 2$, as shown in Fig. 2.5(b). Moreover, we generalize this notion to the case where the mutual information may be conditioned on other signals as well, provided that these signals do not contain information about $Z_{c,d}^n$, for some $d \in D$.

Lemma 2.2 *Suppose we have nodes b and c such that $(c, b) \in E$, and let $A = \mathcal{I}(b) \setminus \{c\}$ and $D = \mathcal{O}(c) \setminus \{b\}$. Suppose, in addition, that we have a set of nodes S such that, if $u \in \mathcal{O}(c)$ and $w \in S$, we have $u \not\rightsquigarrow w$, and a set of nodes T with the property that, if $u \in D$ and $w \in T$, then $u \not\rightsquigarrow w$. Then, we have*

$$I(X_S^n; \tilde{X}_c^n | Y_b^n, \tilde{X}_A^n, X_T^n) \leq n\kappa,$$

where κ is only a function of the channel gains in the network.

Remarks: If, in the statement of Lemma 2.2, we condition the mutual information on $\tilde{X}_T^n$ instead of X_T^n the same result holds. Also, if instead of conditioning on $\tilde{X}_A^n$ and Y_b^n we condition on $\tilde{X}_{c,b}^n$, the same result holds, since, in the proof, we use $\tilde{X}_A^n$ and Y_b^n to construct $\tilde{X}_{c,b}^n$. We will consider these cases to be covered by Lemma 2.2 as well.

Proof:

$$
\begin{aligned}
I(X_S^n; \tilde{X}_c^n | Y_b^n, \tilde{X}_A^n, X_T^n) &= I(X_S^n; \{\tilde{X}_{c,j}^n : j \in \mathcal{O}(v_c)\} | Y_b^n, \tilde{X}_A^n, X_T^n) \\
&\overset{(i)}{=} I(X_S^n; \{\tilde{X}_{c,j}^n - \tfrac{h_{c,j}}{h_{c,b}} \tilde{X}_{c,b}^n : j \in \mathcal{O}(v_c)\} | Y_b^n, \tilde{X}_A^n, X_T^n) \\
&\overset{(ii)}{=} I(X_S^n; \{Z_{c,j}^n - \tfrac{h_{c,j}}{h_{c,b}} Z_{c,b}^n : j \in D\} | Y_b^n, \tilde{X}_A^n, X_T^n) \\
&\leq h(\{Z_{c,j}^n - \tfrac{h_{c,j}}{h_{c,b}} Z_{c,b}^n : j \in D\}) \\
&\quad - h(\{Z_{c,j}^n - \tfrac{h_{c,j}}{h_{c,b}} Z_{c,b}^n : j \in D\} | Y_b^n, \tilde{X}_A^n, X_T^n, X_S^n) \\
&\overset{(iii)}{\leq} \frac{n|D|}{2} \log(2\pi e \kappa) - h(\{Z_{c,j}^n - \tfrac{h_{c,j}}{h_{c,b}} Z_{c,b}^n : j \in D\} | Y_b^n, \tilde{X}_A^n, X_T^n, X_S^n) \\
&\overset{(iv)}{\leq} \frac{n|D|}{2} \log(2\pi e \kappa) - h(\{Z_{c,j}^n : j \in D\} | Z_{c,b}^n, Y_b^n, \tilde{X}_A^n, X_T^n, X_S^n) \\
&\overset{(v)}{=} \frac{n|D|}{2} \log(2\pi e \kappa) - h(\{Z_{c,j}^n : j \in D\}) \\
&= n \left(\frac{|D|}{2} \log(2\pi e \kappa) - \sum_{j \in D} \frac{1}{2} \log \left(\frac{2\pi e}{|\mathcal{I}(j)|} \right) \right),
\end{aligned}
$$

where (i) follows from the fact that $Y_b^n - \sum_{a \in A} \tilde{X}_{a,b}^n = \tilde{X}_{c,b}^n$; (ii) follows since, for $j = b$, $Z_{c,j}^n - \frac{h_{c,j}}{h_{c,b}} Z_{c,b}^n = 0$; (iii) follows by letting $\kappa \triangleq 1 + (\max_{e,f \in E} h_e/h_f)^2$; (iv) follows because conditioning reduces entropy and thus we can condition on $Z_{c,b}^n$; (iv) follows from the fact that, since for $u \in D$ and $w \in T$, $u \not\rightsquigarrow w$, $Z_{c,u}^n$ is independent of all the random variables conditioned on. ■

We can now proceed to the proof of case (A) in Theorem 2.1. We assume wlog that we have an edge $(u, v) \in E$ such that the removal of v disconnects d_1 from both sources and the removal of u disconnects s_2 from both destinations. We let $A = \{v \in V : s_2 \not\rightsquigarrow v\}$, and we notice that $\mathcal{I}(v) \setminus \{u\} \subset A$, since the removal of

u disconnects s_2 from d_1. Moreover, $u \notin A$, because all paths from s_2 to d_2 contain u and we must have at least one such path. Using Fano's inequality, we have

$$\begin{aligned} nR_1 &\leq I(W_1; Y_{d_1}^n) + n\epsilon_n \overset{(i)}{\leq} I(\tilde{X}_A^n; Y_v^n) + n\epsilon_n \\ &= I(\tilde{X}_A^n, X_u^n; Y_v^n) - I(X_u^n; Y_v^n | \tilde{X}_A^n) + n\epsilon_n \\ &\overset{(ii)}{\leq} \frac{n}{2} \log P + n\kappa_1 - I(X_u^n; Y_v^n | \tilde{X}_A^n) + n\epsilon_n, \end{aligned} \tag{2.1}$$

where (i) follows because v disconnects d_1 from both sources and $s_1 \in A$, and thus we have $W_1 \leftrightarrow \tilde{X}_A^n \leftrightarrow Y_v^n \leftrightarrow Y_{d_1}^n$; and (ii) follows because $\mathcal{I}(v) \setminus \{u\} \subset A$ and $u \notin A$, and hence

$$\begin{aligned} I(\tilde{X}_A^n, X_u^n; Y_v^n) &= h(Y_v^n) - h(Z_{u,v}^n) \\ &\leq \frac{n}{2} \log \left(\frac{1 + \left(\sum_{w \in \mathcal{I}(v)} |h_{w,v}| \right)^2 P}{1/|\mathcal{I}(v)|} \right) \\ &\leq \frac{n}{2} \log P + n\kappa_1, \end{aligned} \tag{2.2}$$

where κ_1 is a constant, independent of P for P sufficiently large. Next we bound the second rate as

$$\begin{aligned} nR_2 &\leq I(W_2; Y_{d_2}^n) + n\epsilon_n \overset{(i)}{\leq} I(W_2; \tilde{X}_u^n | \tilde{X}_A^n) + n\epsilon_n \\ &\overset{(ii)}{\leq} I(X_u^n; \tilde{X}_u^n | \tilde{X}_A^n) + n\epsilon_n \leq I(X_u^n; \tilde{X}_u^n, Y_v^n | \tilde{X}_A^n) + n\epsilon_n \\ &= I(X_u^n; Y_v^n | \tilde{X}_A^n) + I(X_u^n; \tilde{X}_u^n | \tilde{X}_A^n, Y_v^n) + n\epsilon_n \\ &\overset{(iii)}{\leq} I(X_u^n; Y_v^n | \tilde{X}_A^n) + n\kappa_2 + n\epsilon_n, \end{aligned} \tag{2.3}$$

where (i) follows from the fact that the removal of u disconnects s_2 and d_2 and $s_1 \in A$; (ii) follows from the fact that we have $W_2 \leftrightarrow X_u^n \leftrightarrow \tilde{X}_u^n$ given $\tilde{X}_A^n$; (iii) follows from the application of Lemma 2.2 to $I(X_u^n; \tilde{X}_u^n | \tilde{X}_A^n, Y_v^n)$, since $\mathcal{I}(v) \setminus \{u\} \subset A$. Finally, by adding (2.1) and (2.3) we obtain

$$n(R_1 + R_2) \leq \frac{n}{2} \log P + n(\kappa_1 + \kappa_2) + n\epsilon_n,$$

which implies that $D_\Sigma \leq 1$. Since one degree of freedom is trivially achievable, this concludes the proof of case (A) in Theorem 2.1.

2.4 Achieving Two Degrees of Freedom

In this section, we will discuss how to achieve two degrees of freedom in networks with manageable interference. We will assume throughout that we have two paths P_{s_1,d_1} and P_{s_2,d_2} and a set $S \supseteq P_{s_1,d_1} \cup P_{s_2,d_2}$, such that $w_1(S) \notin P_{s_2,d_2}$, $w_2(S) \notin P_{s_1,d_1}$, $n_1(S) \neq 1$ and $n_2(S) \neq 1$. To simplify the exposition, we will assume that the nodes in $V \setminus S$ have been removed, and drop S from the notation.

While for most specific network topologies in this class, devising a scheme to achieve 2 degrees of freedom is not difficult, describing general procedures that apply to a large number of networks is challenging. The main tool we will use to handle this issue will be the notion of *network condensation*, which allows us to "convert" a network with an arbitrary number of layers into a network with at most four layers, while *preserving paths with manageable interference*.

2.4.1 Network Condensation

The main idea behind network condensation is to identify (at most) two *key layers* (other than the source and the destination layers) whose nodes will be responsible for performing non-trivial relaying operations. All the nodes which do not belong to the key layers will simply forward their received signal. This will allow us to build a *condensed* version of the network. The condensed network only contains the nodes in the key layers, V_1, and V_r. The edges and respective channel gains are determined according to the effective transfer matrices between two consecutive layers of the condensed network, which are obtained by assuming that all intermediate nodes that are not in the key layers, V_1 or V_r are simply forwarding their received signals. An example is shown in Figure 2.6.

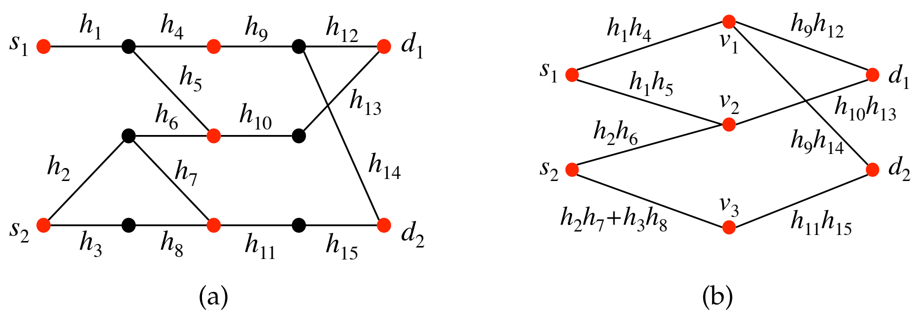


Figure 2.6: A 5-layer network (a) and its 3-layer condensed version (b).

We will refer to the effective channel gains of the edges in the condensed network by $\hat{h}(v, u)$, where v is the starting node and u is the ending node. For example, in Figure 2.6, we have $\hat{h}(s_2, v_3) = h_2h_7 + h_3h_8$ and $\hat{h}(v_2, d_2) = 0$. Notice that, in the condensed network, the effective additive noises at the nodes are not necessarily independent and identically distributed.

We will describe schemes to achieve $D_\Sigma = 2$ in essentially two ways, according to the structure of the condensed network. If the resulting condensed network contains a $2 \times 2 \times 2$ interference channel, then we will use the scheme described in [24] to achieve $D_\Sigma = 2$. Otherwise, we will describe simple scalar linear operations for the nodes in the key layers that convert the network's end-to-end transfer matrix into the form $\left[\begin{smallmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{smallmatrix}\right]$, for $\beta_1, \beta_2 \neq 0$. Thus we have $Y_{d_i} = \beta_i X_{s_i} + \tilde{Z}_{d_i}$, for $i = 1, 2$, where $\tilde{Z}_{d_i}$ is the effective additive noise at d_i. In order to make sure that the output power constraint is satisfied at all nodes, we will restrict the sources to using power αP, for some $\alpha \in (0, 1)$. Since the scaling factors used at the key layers will be functions of the channel gains only, it is not difficult to see that β_1, β_2 and the effective noise variances, σ_i^2, are independent of P. Thus, we can we can choose $\alpha > 0$ small enough so that, if the source restricts its output power to αP, the output power constraint is satisfied at all nodes, and each source-destination pair (s_i, d_i), for $i = 1, 2$, can achieve rate

$$R_i = \frac{1}{2} \log\left(1 + \frac{\alpha \beta_i^2 P}{\sigma_i^2}\right),$$

and hence one degree of freedom.

For networks with paths with manageable interference, we will choose the two key layers to be $V_{\ell(w_1)-1}$ and $V_{\ell(w_2)-1}$, i.e., the two layers preceding the two key nodes w_1 and w_2. Notice that we may have less than two key layers if if $w_i = s_i$ (so that layer $V_{\ell(w_i)-1}$ does not exist) and if $\ell(w_1) = \ell(w_2)$ (so that $V_{\ell(w_1)-1} = V_{\ell(w_2)-1}$). We let the resulting network be $\tilde{\mathcal{N}}$. This condensation process is useful because of the following observation:

Proposition 2.3 *Suppose paths P_{s_1,d_1} and P_{s_2,d_2} have manageable interference in $\mathcal{N}$. Then, the resulting condensed paths $\tilde{P}_{s_1,d_1}$ and $\tilde{P}_{s_2,d_2}$ have manageable interference in $\tilde{\mathcal{N}}$.*

It is important to notice, however, that the resulting channel gains of $\tilde{\mathcal{N}}$ are polynomials on the channel gains of $\mathcal{N}$ and, thus, proving that 2 degrees of freedom are achievable in $\tilde{\mathcal{N}}$ for almost all values of channel gains does not imply that 2 degrees of freedom are achievable in $\mathcal{N}$ for almost all values of channel gains. Hence we will need a result that allows us to infer properties about the channel gains of the condensed network. This is accomplished with the following lemma:

Lemma 2.3 *If $n_i \geq 2$, we must have $\mathcal{K}(\{s_1, s_2\}, \mathcal{I}(w_i)) = 2$.*

Proof: Clearly $1 \leq \mathcal{K}(\{s_1, s_2\}, \mathcal{I}(w_i)) \leq 2$. If $\mathcal{K}(\{s_1, s_2\}, \mathcal{I}(w_i(S))) = 1$, then there must be a node v that is a $(\{s_1, s_2\}, \mathcal{I}(w_i))$ cut. Thus, we must have $v \in P_{s_i,d_i}$. Moreover, since all paths from $\{s_1, s_2\}$ to d_i must contain w_i, v must also be a $(\{s_1, s_2\}, d_i)$ cut. But this is a contradiction since w_i was the first single-node $(s_{\bar{i}}, d_i)$ cut in the path P_{s_i,d_i}. ■

As we will see, the existence of two disjoint paths from $\{s_1, s_2\}$ to $\mathcal{I}(w_i)$ in $\mathcal{N}$ will guarantee that, in the condensed network $\tilde{\mathcal{N}}$, the channel gains along these two paths are in some sense "independent".

2.4.2 Scheme to achieve $D_\Sigma = 2$

In order to describe the achievability schemes, we consider three cases of the resulting condensed network. We first describe the achievability scheme for the case where $n_1(G[S]) \geq 2$ and $n_2(G[S]) = 0$ in detail. Then we divide the case $n_i \geq 2$ for $i = 1, 2$ in two subcases. The subcase $\ell(w_1) \neq \ell(w_2)$ follows by first choosing the coefficients of the first key layer in order to obtain a new condensed network with $n_1(G[S]) \geq 2$ and $n_2(G[S]) = 0$, where the steps from the previous case can be used. The case where $\ell(w_1) = \ell(w_2)$ also follows similar steps, except that when the network reduces to a $2 \times 2 \times 2$ wireless network, we must resort to alignment-based techniques, as done in [24] (also in Chapter 3, for the case of general $K \times K \times K$ wireless networks).

Case I: $n_1(G[S]) \geq 2$ **and** $n_2(G[S]) = 0$

In this case, $w_2 = s_2$, and the condensed network will be formed by layers $V_1, V_{\ell(w_1)-1}$ and V_r. Notice that, in this case, P_{s_1,d_1} and P_{s_2,d_2} must be disjoint paths. As shown in Fig. 2.7, we will let $V_{\ell(w_1)-1} = \{v_1, ..., v_m\}$, where $v_1 \in P_{s_1,d_1}$ and $v_m \in P_{s_2,d_2}$. To each $v_i \in V_{\ell(w_1)-1}$, $i = 1, ..., m$, we associate a scaling factor x_i. We must

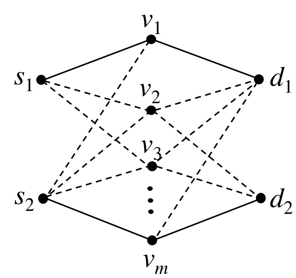


Figure 2.7: Illustration of a condensed network with $n_1(G[S])$ and $n_2(G[S]) = 0$. Solid lines represent edges that must exist in the condensed network, while the dashed lines represent edges that may not exist.

show that the end-to-end transfer matrix T,

$$\begin{bmatrix} T_{1,1} & T_{1,2} \\ T_{2,1} & T_{2,2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m \hat{h}(s_1, v_i)\hat{h}(v_i, d_1)x_i & \sum_{i=1}^m \hat{h}(s_2, v_i)\hat{h}(v_i, d_1)x_i \\ \sum_{i=1}^m \hat{h}(s_1, v_i)\hat{h}(v_i, d_2)x_i & \sum_{i=1}^m \hat{h}(s_2, v_i)\hat{h}(v_i, d_2)x_i \end{bmatrix},$$

can be made diagonal with non-zero diagonal entries by an appropriate choice of $x_1, ..., x_m$. Since, in this case, $n_2(G[S]) = 0$, there is no path from s_1 to d_2, and therefore we must have $\hat{h}(s_1, v_i)\hat{h}(v_i, d_2) = 0$ for $i = 1, ..., m$ and $T_{2,1}$ is always 0. From Lemma 2.3, we can find two nodes $v_a, v_b \in \mathcal{I}(w_1) \subset V_{\ell(w_1)-1}$ with associated variables x_a and x_b, and two disjoint paths P_{s_1,v_a} and P_{s_2,v_b}. Moreover, it is clear from the definition of w_1 and from the fact that $n_1 \geq 2$ that there are at least two nodes $v_c, v_d \in \mathcal{I}(w_1)$ reachable from s_2. So we can assume wlog that $v_c \neq v_m$. We claim that if the matrices

$$M_1 = \begin{bmatrix} \hat{h}(s_1, v_a)\hat{h}(v_a, d_1) & \hat{h}(s_1, v_b)\hat{h}(v_b, d_1) \\ \hat{h}(s_2, v_a)\hat{h}(v_a, d_1) & \hat{h}(s_2, v_b)\hat{h}(v_b, d_1) \end{bmatrix} \text{ and}$$

$$M_2 = \begin{bmatrix} \hat{h}(s_2, v_c)\hat{h}(v_c, d_1) & \hat{h}(s_2, v_m)\hat{h}(v_m, d_1) \\ \hat{h}(s_2, v_c)\hat{h}(v_c, d_2) & \hat{h}(s_2, v_m)\hat{h}(v_m, d_2) \end{bmatrix}$$

are both invertible, then we can choose $x_1, ..., x_m$ so that T is as desired. To see this, consider $\mathbf{x}' = [x'_1 \ ... \ x'_m]$, where $x'_j = 0$ for $j \neq a, b$, and $[x'_a \ x'_b]^T = M_1^{-1}[1 \ 0]^T$. This choice of scaling factors results in $T_{1,1} = 1$ and $T_{1,2} = 0$. If $T_{2,2} \neq 0$ we are done. Otherwise, if $T_{2,2} = 0$, we let $\mathbf{x}'' = [x''_1 \ ... \ x''_m]$, where $x''_j = 0$ for $j \neq c, m$ and $[x''_c \ x''_m]^T = M_2^{-1}[0 \ 1]^T$. This choice results in $T_{1,2} = 0$ and $T_{2,2} = 1$. If we have $T_{1,1} \neq 0$, we are done. Otherwise, we set $\mathbf{x}''' = \mathbf{x}' + \mathbf{x}''$. By linearity, this choice will guarantee that T is the identity matrix.

Next we show that, for almost all choices of the channel gains h_e's of the original network $\mathcal{N}$, M_1 and M_2 are full-rank. First we consider the transfer matrix between (s_1, s_2) and (v_a, v_b), given by

$$L_1 = \begin{bmatrix} \hat{h}(s_1, v_a) & \hat{h}(s_2, v_a) \\ \hat{h}(s_1, v_b) & \hat{h}(s_2, v_b) \end{bmatrix}.$$

The determinant of L_1 can be seen as a polynomial in the channel gains h_e. If $\det L_1$ is not identically zero, then it is clear that for almost all choices of the h_e's, $\det L_1$ will be non-zero. To see that $\det L_1$ is not identically zero, notice that the existence of disjoint paths P_{s_1,v_a} and P_{s_2,v_b} in the original (non-condensed) network guarantees that, if we set $h_e = 1$ if e connects adjacent nodes of P_{s_1,v_a} or P_{s_2,v_b} and $h_e = 0$ otherwise, L_1 will be the identity matrix. Therefore, L_1 will be invertible, and $\det L_1$ cannot be identically zero. Now, we notice that $\det M_1 = \hat{h}(v_a, d_1)\hat{h}(v_b, d_1) \det L_1$. Since $v_a \rightsquigarrow d_1$ and $v_b \rightsquigarrow d_1$, we have that $\hat{h}(v_a, d_1)\hat{h}(v_b, d_1)$ is also nonzero for almost all values of h_e's, and we conclude that M_1 is invertible for almost all values of h_e's. To show that M_2 is invertible, we follow very similar steps. First we notice that we have $\mathcal{K}(\{v_c, v_m\}; \{d_1, d_2\}) = 2$ in the original network since we have disjoint paths $(v_c, w_1) \oplus P_{s_1,d_1}[w_1 : d_1]$ and $P_{s_2,d_2}[v_m : d_2]$. Thus, for almost all values of channel gains, L_2, the transfer matrix from $\{v_c, v_m\}$ to $\{d_1, d_2\}$ is invertible. Since $\det M_2 = \hat{h}(s_2, v_c)\hat{h}(s_2, v_m) \det L_2$, M_2 is invertible for almost all choices of h_e's. By the previous discussion, we conclude that the linear scheme that uses scaling coefficients $x_1, ..., x_m$ at the key layer $V_{\ell(w_1)-1}$ can be used to achieve 2 degrees of freedom.

Before moving on to the next case, we make the following observation:

Lemma 2.4 *Suppose paths P_{s_1,d_1} and P_{s_2,d_2} have manageable interference and the set S satisfies the conditions in Definition 2.8. Then, paths $P_{s_1,d_1}[w_1(S), d_1]$ and $P_{s_2,d_2}[w_2(S), d_2]$ are disjoint.*

Proof: Assume wlog that $\ell(w_1(S)) \leq \ell(w_2(S))$. Suppose by contradiction that $P_{s_1,d_1}[w_1(S), d_1]$ and $P_{s_2,d_2}[w_2(S), d_2]$ are not disjoint, and let $u \in P_{s_1,d_1}[w_1(S), d_1] \cap P_{s_2,d_2}[w_2(S), d_2]$. Since $w_1(S) \notin P_{s_2,d_2}$, the path $P_{s_2,d_2}[s_2, u] \oplus P_{s_1,d_1}[u, d_1]$ does not contain $w_1(S)$. But this is a contradiction to the fact that $w_1(S)$ is an (s_2, d_1) cut. ■

Case II: $n_1(G[S]) \geq 2$**,** $n_2(G[S]) \geq 2$ **and** $\ell(w_1) \neq \ell(w_2)$

We will assume wlog that $\ell(w_2) < \ell(w_1)$ and let $V_{\ell(w_2)-1} = \{u_1, ..., u_n\}$ and $V_{\ell(w_1)-1} = \{v_1, ..., v_m\}$, as illustrated in Fig. 2.8 for $m = n = 4$. We assume $v_1 \in P_{s_1,d_1}$ and $v_m \in P_{s_2,d_2}$. Notice that Lemma 2.4 implies that we must have $v_1 \neq v_m$.

To each of the nodes v_i, $i = 1, ..., m$, we associate a variable x_i which will be the scaling factor used by node v_i, and to each of the nodes u_i, $i = 1, ..., n$ we associate a variable y_i which will be the scaling factor used by node u_i.

We will show that, for almost all values of channel gains, there is a choice of $x_1, ..., x_m$ and $y_1, ..., y_n$ such that the effective end-to-end transfer matrix is diago-

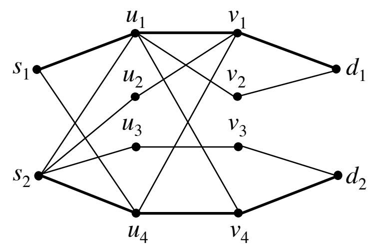


Figure 2.8: Example of a condensed network in the case where $n_1(G[S]) \geq 2$, $n_2(G[S]) \geq 2$ and $\ell(w_2) < \ell(w_1)$.

nal with non-zero diagonal entries. We will proceed in two steps. First we will show that, we can choose $y_1, ..., y_n$, all nonzero, such that, for some $v_a, v_b \in \mathcal{I}(w_1)$, the transfer matrix between (s_1, s_2) and (v_a, v_b) is invertible and the transfer matrix between (s_1, s_2) and v_m is of the form $[0\ \beta]$ for $\beta \neq 0$. Then, by "supressing" the key layer $V_{\ell(w_2)-1}$, we will essentially be in the case $n_1(S) \geq 2$, $n_2(S) = 0$, described previously.

We will show that, for almost all values of channel gains, it is possible to choose $y_1, ..., y_n$ all nonzero, such that the transfer matrix F between (s_1, s_2) and w_2 is of the form $[0\ \alpha]$ for $\alpha \neq 0$. We first notice that F is given by

$$[F_1\ F_2] = \left[\sum_{i=1}^n \hat{h}(s_1, u_i) h_{(u_i, w_2)} y_i \quad \sum_{i=1}^n \hat{h}(s_2, u_i) h_{(u_i, w_2)} y_i \right].$$

Since $n_2(S) \geq 2$, we know that there are at least two nodes $u_c, u_d \in \mathcal{I}(w_2)$ such that $s_1 \rightsquigarrow u_c$ and $s_1 \rightsquigarrow u_d$ (e.g., in Fig. 2.8 they would be u_1 and u_4). This implies that $\hat{h}(s_1, u_c) h_{(u_c, w_2)}$ and $\hat{h}(s_1, u_d) h_{(u_d, w_2)}$, if viewed as polynomials on the channel gains, are not identically zero. Thus, for almost all channel gain values, $F_1 = \sum_{i=1}^n \hat{h}(s_1, u_i) h_{(u_i, w_2)} y_i$ will have nonzero coefficients in front of y_c and y_d. This means that we can choose $\mathbf{y}' = (y'_1, ..., y'_n)$, with $y'_1, ..., y'_n$ all nonzero, so that $F_1 = \sum_{i=1}^n \hat{h}(s_1, u_i) h_{(u_i, w_2)} y'_i = 0$. If we have $F_2 = \sum_{i=1}^n \hat{h}(s_2, u_i) h_{(u_i, w_2)} y'_i \neq 0$, then we are done. Otherwise, if $F_2 = 0$, we proceed as follows. From Lemma 2.3, we know that we can choose $u_a, u_b \in \mathcal{I}(w_2) \subset V_{\ell(w_2)-1}$ so that we have two disjoint paths P_{s_1,u_a} and P_{s_2,u_b}. Therefore, for almost all channel gain values, the transfer matrix between (s_1, s_2) and (u_a, u_b), given by

$$L = \begin{bmatrix} \hat{h}(s_1, u_a) & \hat{h}(s_1, u_b) \\ \hat{h}(s_2, u_a) & \hat{h}(s_2, u_b) \end{bmatrix},$$

is full-rank. This also implies that the matrix

$$M = \begin{bmatrix} \hat{h}(s_1, u_a)h_{(u_a,w_2)} & \hat{h}(s_1, u_b)h_{(u_b,w_2)} \\ \hat{h}(s_2, u_a)h_{(u_a,w_2)} & \hat{h}(s_2, u_b)h_{(u_b,w_2)} \end{bmatrix}$$

is full-rank, because we have $\det M = h_{(u_a,w_2)}h_{(u_b,w_2)} \det L$, and, since $u_a, u_b \in \mathcal{I}(w_2)$, we have that $h_{(u_a,w_2)}h_{(u_b,w_2)}$ is nonzero for almost all channel gain values. The matrix M allows us to build $\mathbf{y}'' = (y_1'', ..., y_n'')$ by setting $y_i'' = 0$, for $i \neq a, b$, and $[y_a'' \;\; y_b'']^T = M^{-1}[0 \;\; 1]^T$. This choice guarantees that $F = [0 \;\; 1]$ as desired, but we do not have $y_1'', ..., y_n''$ all nonzero. However, it is easy to see that if we set $\mathbf{y}''' = \mathbf{y}' + \alpha\mathbf{y}''$, for some $\alpha \neq 0$, we will have $y_1''', ..., y_n'''$ all nonzero and $F = [0 \; \alpha]$.

We conclude that we can choose $y_1, ..., y_n$ all nonzero and have $F = [0 \; \alpha]$ with $\alpha \neq 0$. Moreover, since there exists a path from w_2 to v_m, and there exists no path from s_1 to v_m which does not contain w_2, we conclude that, with probability 1, our choice of $y_1, ..., y_m$ will make the transfer matrix from (s_1, s_2) to v_m be of the form $[0 \; \beta]$ for $\beta \neq 0$.

Next, by absorbing the operations performed by the nodes in $V_{\ell(w_2)-1}$ into the network, we obtain a new condensed network where $n_1(S) \geq 2$ and $n_2(S) = 0$. However, we cannot proceed exactly as in the previous case because the scaling coefficients $y_1, ..., y_m$ were chosen as functions of the channel gains. Nonetheless, if we let $\tilde{H}$ be the set of all $h_{(u_j,w_2)}$ for $j = 1, ..., n$ and all the channel gains that appear in $\hat{h}(s_i, u_j)$, for $i = 1, 2$ and $j = 1, ..., n$, we notice that our choice of $y_1, ..., y_n$

only depends on $\tilde{H}$. Therefore, we may assume that all the channel gains in $\tilde{H}$ are drawn according to their distributions, and are from now on viewed as constants. Then, we can also fix $y_1, ..., y_n$ as functions of $\tilde{H}$, following the steps described previously, and view them as constants. In [55], we prove the following result:

Claim 2.1 *If $y_1, ..., y_m$ are all nonzero and chosen only as functions of $\tilde{H}$, for almost all values of the channel gains, there are two nodes $v_a, v_b \in \mathcal{I}(w_1) \subset V_{\ell(w_1)-1}$ such that the transfer matrix from $\{s_1, s_2\}$ to $\{v_a, v_b\}$ is invertible.*

Because of this result, once we absorb the layer $V_{\ell(w_2)-1}$ into the condensed network, we obtain a network with $n_1 \geq 2$ and $n_2 = 0$, and where the steps presented for Case I can be followed to choose the scaling coefficients $x_1, ..., x_n$ such that the resulting end-to-end transfer matrix will be diagonal with nonzero diagonal entries. This in turn implies that 2 degrees of freedom can be achieved.

Case III: $n_1(S) \geq 2$**,** $n_2(S) \geq 2$ **and** $\ell(w_1) = \ell(w_2)$

When $\ell(w_1) = \ell(w_2)$, our condensed network will only contain three layers, V_1, $V_{\ell(w_1)-1} = V_{\ell(w_2)-1}$ and V_r. We will use two different approaches, depending on the size of $V_{\ell(w_1)-1}$. If $|V_{\ell(w_1)-1}| = 2$, it is not difficult to see that $n_1 \geq 2$ and $n_2 \geq 2$ imply that our condensed network should look like the $2 \times 2 \times 2$ network in Fig. 2.9. In this case, for almost all values of channel gains, the transfer matrix between (s_1, s_2) and (v_1, v_2) and the transfer matrix between (v_1, v_2) and (d_1, d_2), given respectively by

$$\begin{bmatrix} \hat{h}(s_1, v_1) & \hat{h}(s_2, v_1) \\ \hat{h}(s_1, v_2) & \hat{h}(s_2, v_2) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \hat{h}(v_1, d_1) & \hat{h}(v_2, d_1) \\ \hat{h}(v_1, d_2) & \hat{h}(v_2, d_2) \end{bmatrix},$$

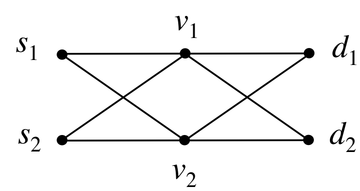


Figure 2.9: Illustration of the condensed network for the case where $n_1(G[S]) \geq 2$, $n_2(G[S]) \geq 2$, $\ell(w_2) = \ell(w_1)$ and $|V_{\ell(w_1)-1}| = 2$.

have only nonzero entries. Furthermore, Lemmas 2.3 and 2.4 imply that, for almost all values of channel gains, the transfer matrices between (s_1, s_2) and (v_1, v_2) and between (v_1, v_2) and (d_1, d_2) are full-rank. Therefore, we essentially have the $2 \times 2 \times 2$ interference channel described in [24]. The only difference is that additive noises at v_1, v_2, d_1 and d_2 are not independent and Gaussian. However, they still have a variance which does not depend on the power P (only on the channel gains), and thus the same scheme described in [24] will achieve $D_\Sigma = 2$.

In the case where $|V_{\ell(w_1)-1}| = m > 2$, we can follow similar steps to those in Case I in order to show that we can find scaling coefficients $x_1, ..., x_m$ such that the end-to-end transfer matrix is made diagonal with nonzero diagonal entries, and two degrees of freedom can be achieved.

2.5 Networks with 3/2 degrees of freedom

In this section, we consider the networks that do not contain paths with manageable interference nor an omniscient node, and thus fall in case (C) of Theorem 2.1. In order to characterize the networks in this class, we resort to a result from [15, 53], used to characterize which *wireline* two-unicast networks allow rate $(1, 1)$ to be achieved. In [53, Theorem 1], the authors show that rate pair

$(1, 1)$ is achievable in a two-unicast acyclic wireline network if and only if there is no edge e whose removal disconnects s_1 from d_1, s_2 from d_2 and either s_1 from d_2 or s_2 from d_1. In analogy to Definition 2.6, we will call such an edge an omniscient edge. The following alternative characterization for when rate pair $(1, 1)$ is achievable is found in [15]:

Theorem 2.3 ([15, Theorem 4.3]) *Rate pair* $(1, 1)$ *is achievable on a two-unicast acyclic network if and only if the network contains two disjoint paths* P_{s_1,d_1} *and* P_{s_2,d_2}*, a butterfly structure or a grail structure (as depicted in Fig. 2.10).*

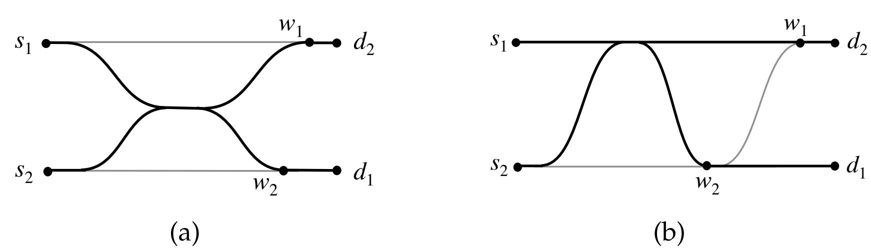


Figure 2.10: (a) Butterfly structure and (b) Grail Structure. The (non-disjoint) paths with manageable interference P_{s_1,d_1} and P_{s_2,d_2} and the key nodes w_1 and w_2 are marked.

Combining the two results, we conclude that if a two-unicast acyclic network does not contain an omniscient edge, then it must contain two disjoint paths P_{s_1,d_1}, P_{s_2,d_2}, a butterfly structure, or a grail structure. If an edge $e = (u, v)$ is omniscient, then it is clear that both u and v are omniscient nodes, in the sense of Definition 2.6. Therefore, for layered two-unicast wireless networks, we have:

Claim 2.2 *If a layered two-unicast (wireless) network* $\mathcal{N}$ *does not contain an omniscient node, it contains two disjoint paths* P_{s_1,d_1} *and* P_{s_2,d_2}*, a butterfly structure or a grail structure.*

Furthermore, it is not difficult to check that if a network contains a butterfly structure or a grail structure, then it contains two non-disjoint paths P_{s_1,d_1} and P_{s_2,d_2} with manageable interference, as illustrated in Fig. 2.10. Thus, such networks fall in case (B) of Theorem 2.1, and we conclude the following:

Claim 2.3 *All networks in case (C) contain two disjoint paths P_{s_1,d_1} and P_{s_2,d_2}, but no two paths P'_{s_1,d_1} and P'_{s_2,d_2} with manageable interference.*

We will thus assume that we have two disjoint paths P_{s_1,d_1} and P_{s_2,d_2} and we will first show that Claim 2.3 implies that wlog our network $\mathcal{N}$ falls into one of two cases:

C1. $n_1(V) \geq 2$, $n_1^0 = 1$, $n_2(V) = 1$ and $n_2^0 = 0$.

C2. $n_1(V) = n_1^0 = 1$

To see this, we start by observing that for two *disjoint* paths P_{s_1,d_1} and P_{s_2,d_2}, and $S \supset P_{s_1,d_1} \cup P_{s_2,d_2}$, we trivially have $w_1(S) \notin P_{s_2,d_2}$ and $w_2(S) \notin P_{s_1,d_1}$ and for manageable interference we only need to check whether $n_i(S) \neq 1$ for $i = 1, 2$. Thus, for a network in case (C), since the interference on the disjoint paths P_{s_1,d_1} and P_{s_2,d_2} is not manageable, we have that either $n_1(V) = 1$ or $n_2(V) = 1$. Moreover, we must also have either $n_1^0 = 1$ or $n_2^0 = 1$. So we assume wlog that $n_1^0 = 1$. From Lemma 2.1, this implies that $n_1(V) \geq 1$. If $n_1(V) = 1$, we are in case C2. If instead $n_1(V) \geq 2$, we must have $n_2(V) = 1$. Lemma 2.1 now implies that $n_2^0 \leq 1$. If $n_2^0 = 1$, we are again in case C2 by swapping the names of (s_1, d_1) and (s_2, d_2). Otherwise, if $n_2^0 = 0$, we are in C1.

These two cases are depicted in Fig. 2.11. We will provide an achievability and a converse for $D_\Sigma = \frac{3}{2}$ in case C1. Case C2 follows similarly, and we refer to

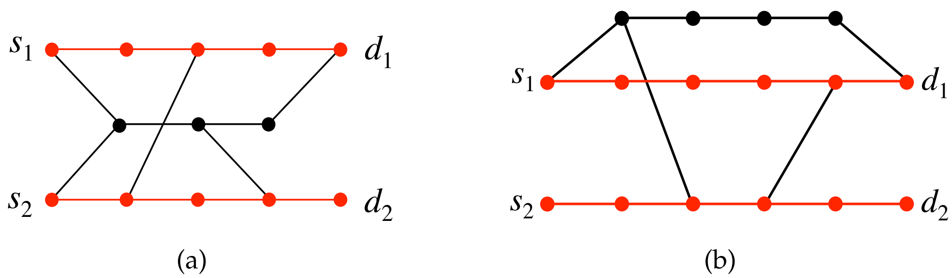


Figure 2.11: Example of a network in (a) case C1 and (b) case C2. Paths P_{s_1,d_1} and P_{s_2,d_2} are shown in red.

[55] for details.

2.5.1 Achievability: Two modes of Operation

In order to achieve $D_\Sigma = \frac{3}{2}$, the main idea will be to consider two modes of operation for network $\mathcal{N}$. In all cases, we will identify a single node u_0 that will operate as a "virtual destination" during the first mode and as a "virtual source" during the second mode.

We point out that even though the definitions in Section 2.2 assume that in a layered network, both sources are in the first layer and both destinations are in the last layer, they extend directly to a more general case where sources and destinations can be in arbitrary layers but sources have $\mathcal{I}(s_i) = \emptyset$ and destinations have $O(d_i) = \emptyset$. Under these new assumptions, for a given node u, we will let $\mathcal{N}_{d_i=u} = (G[S], \{(s_i, u), (s_{\bar{i}}, d_{\bar{i}})\})$ be the network obtained from $\mathcal{N}$ be restricting the node set to $S = \{v : v \rightsquigarrow \{u, d_{\bar{i}}\}\}$ and replacing destination d_i with u. Similarly, we can define $\mathcal{N}_{s_i=u}$. Our achievability result relies on the following lemma:

Lemma 2.5 *Any network $\mathcal{N}$ in case C1 contains a node u such that $\mathcal{N}_{d_i=u}$ and $\mathcal{N}_{s_i=u}$ have paths with manageable interference for $i \in \{1, 2\}$.*

Proof Sketch: Since $n_1(V) \geq 2$ and $n_1^0 = 1$, we must have a path P_{s_2,v_1} that is disjoint of P_{s_1,d_1} and such that $v_1 \rightarrow P_{s_1,d_1}$ and $v_1 \notin P_{s_2,d_2}$. We let z be the last node in $P_{s_2,d_2} \cap P_{s_2,v_1}$, and we have the path $P_{z,v_1} = P_{s_2,v_1}[z, v_1]$. Next we consider letting $S^* = P_{s_1,d_1} \cup P_{s_2,d_2} \cup P_{z,v_1}$ and we must have $n_1(G[S^*]) \geq 2$. Since P_{s_1,d_1} and P_{s_2,d_2} do not have manageable interference, we must have $n_2(G[S^*]) = 1$. Moreover, since $n_2^0 = 0$, we conclude that we must have a node $v_2 \in P_{z,v_1} - \{z\}$ such that $v_2 \rightarrow P_{s_2,d_2}$, and we must have a path $P_{s_1,v_2} \subset S^*$. It can then be seen that our network contains the network shown in Figure 2.12 up to a change in the position of the edge (v_3, v_4) and changes in the path lengths. Notice that we may also have $v_1 = v_2$. We define node u according to the position of v_3 with

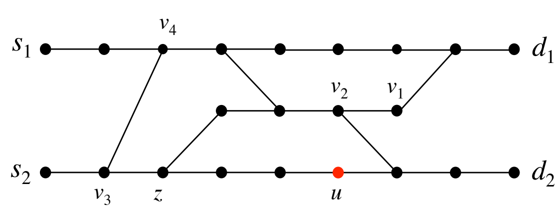


Figure 2.12: Illustration of a network in case C1. The choice of node u according to Lemma 2.5 is shown in red.

respect to v_2. If $\ell(v_3) \geq \ell(v_2)$, we let $u = P_{s_1,d_1} \cap V_{\ell(v_2)}$ and if $\ell(v_3) < \ell(v_2)$, we let $u = P_{s_2,d_2} \cap V_{\ell(v_2)}$. A careful analysis of $\mathcal{N}_{d_1=u}$ and $\mathcal{N}_{s_1=u}$ in the first case, or $\mathcal{N}_{d_2=u}$ and $\mathcal{N}_{s_2=u}$ in the second case, shows that they contain paths with manageable interference. ■

Once we find a node u satisfying Lemma 2.5, achieving $D_\Sigma = \frac{3}{2}$ is relatively simple. The scheme is illustrated in Fig. 2.13. During the first mode of operation, we use a scheme to achieve two degrees of freedom on $\mathcal{N}_{d_i=u}$. Node u decodes the message from s_i and stores it. In the second mode of operation, node u becomes source s_i. Since we have paths with manageable interference in $\mathcal{N}_{s_i=u}$, we can again use a scheme to achieve two degrees of freedom, where node u encodes the message decoded during the first mode of operation. Over both modes, source-destination pair (s_i, d_i) achieves 1/2 a degree of freedom, and $(s_{\bar{i}}, d_{\bar{i}})$ achieves one degree of freedom, for a total of $D_\Sigma = 3/2$.

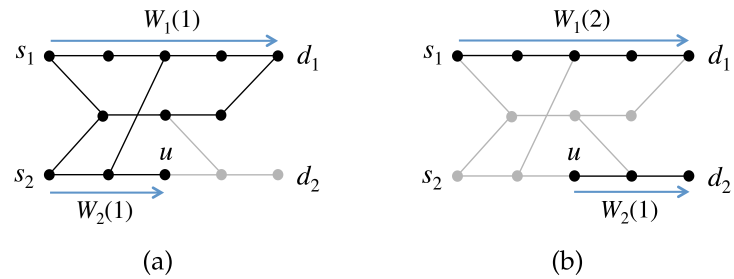


Figure 2.13: Illustration of scheme to achieve 3/2 degrees of freedom. In the first mode (a), u operates as a virtual destination, while in the second mode (b), u operates as a virtual source, forwarding the message decoded at the end of the first mode.

2.5.2 Converse for case C1

In this section, we will show that if a network falls in C1 but does not contain two disjoint paths with manageable interference, then $D_\Sigma \leq \frac{3}{2}$. We will start by naming some extra nodes that will be important to us, as shown in Figure 2.14. We will let v_0 be the node on P_{s_2,d_2} such that $(v_2, v_0) \in E$. From the proof sketch of Lemma 2.5, we know that we have a path P_{s_1,v_2}, which must be entirely contained in $S^* = P_{s_1,d_1} \cup P_{s_2,d_2} \cup P_{z,v_1}$. Thus, we let v_5 be the last node in $P_{s_1,d_1} \cap P_{s_1,v_2}$, and we let v_6 be its consecutive node on P_{s_1,v_2} (which must be part of P_{z,v_1} as well). We will also let $A \triangleq \{v \in V : s_2 \not\rightsquigarrow v\}$ and $B \triangleq \{v \in V : s_1 \not\rightsquigarrow v\}$.

Before we formally derive the inequalities, we will describe some of the intuition that leads to them, for the specific network example from Fig. 2.14.

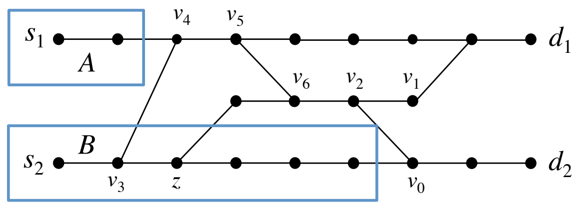


Figure 2.14: Illustration of a network in case C1.

We will let W_1 and W_2 be independent random variables corresponding to a uniform choice over the messages on sources s_1 and s_2 respectively. Also, to simplify the expressions, we will write X_a (or Y_a) instead of X_{v_a} (or Y_{v_a}). Suppose that for a given block length n, we have coding schemes for a network $\mathcal{N}$ for any value of the transmit power constraint $P > 0$. We will consider the quantities

$$\alpha \triangleq \liminf_{P\to\infty} \frac{I(Y_4^n, X_B^n; \tilde{X}_2^n)}{\frac{n}{2}\log P} \text{ and } \beta \triangleq \liminf_{P\to\infty} \frac{I(X_5^n; Y_6^n|\tilde{X}_B^n)}{\frac{n}{2}\log P}.$$

It is easy to see that $0 \leq \alpha, \beta \leq 1$. Intuitively, since all the information from the sources must go through either v_4 or the nodes in B to reach v_2, α can be thought of as the number of useful degrees of freedom (i.e., carrying information about the sources) transmitted by v_2. Similarly, β can be thought of as the number of degrees of freedom transmitted by v_5, but only counting the degrees of freedom with information about message W_1 (since we condition on $\tilde{X}_B^n$). Based on these quantities we will state three inequalities related to the degrees of freedom that can be achieved, and for each one we will provide an intuitive explanation. The formal proof is omitted, but it follows from the information inequalities we will derive later for a more general setting. In the sense of Definition 1.6, we let D_i be the degrees of freedom assigned to (s_i, d_i), for $i = 1, 2$. First, we have

$$D_1 \leq \beta, \tag{2.4}$$

since all information from W_1 must flow through v_5. Next, we claim that both

W_1 and W_2 can be decoded from Y_4^n and $\tilde{X}_2^n$, and thus

$$D_1 + D_2 \leq 1 + \alpha. \tag{2.5}$$

To see this, we first notice that, since the removal of v_4 and v_2 disconnects d_1 from $\{s_1, s_2\}$, from Y_4^n and $\tilde{X}_2^n$, W_1 can be decoded. Then, W_1 can be used to approximately obtain X_A^n (since the nodes in A cannot be influenced by W_2), and, by removing its contribution from Y_4^n, we can obtain a noisy version of the transmit signal from v_3. But since all the information about W_2 must flow through v_3, this allows one to use Y_4^n and $\tilde{X}_2^n$ to decode W_2 as well. For the third inequality we claim that, from Y_0^n, we can decode W_2 completely and $(\alpha + \beta - 1)$ degrees of freedom of W_1, and thus

$$D_2 + (\alpha + \beta - 1) \leq 1. \tag{2.6}$$

To see this, we first notice that, since the removal of v_0 disconnects d_2 from $\{s_1, s_2\}$, from Y_0^n, we can decode W_2, and thus obtain X_B^n approximately. By removing its contribution from Y_0^n, we obtain a noisy version of the transmit signal from node v_2, which allows us to decode the α degrees of freedom transmitted by it. Now we ask ourselves how many of the α degrees of freedom transmitted by v_2 must be carrying information about W_1. To answer this question, we notice that all the degrees of freedom transmitted by v_2 must have come through node v_6. Since node v_6 receives β degrees of freedom with information about W_1 from v_5, at most $1-\beta$ of its degrees of freedom can be not about W_1. Thus, any number of degrees of freedom above $1 - \beta$ that v_2 transmits, i.e., $\alpha - (1 - \beta) = \alpha + \beta - 1$, must contain information about W_1. Finally, by adding inequalities (2.4), (2.5) and (2.6), we obtain $2(D_1 + D_2) \leq 3$, and therefore $D_\Sigma \leq 3/2$.

Next, we formally derive information inequalities that can be used to show that $D_\Sigma \leq 3/2$ for all networks in case C1. In order to establish inequalities that hold more generally than (2.4), (2.5) and (2.6), we will need to use the assumption that there are no two disjoint paths with manageable interference to infer general connectivity properties about networks in case C1, illustrated in Figure 2.14. Next, we state and prove these properties.

P1. All paths from s_1 to d_2 contain nodes v_2 and v_0.

It is easy to see that if we have a path P_{s_1,d_2} not containing $\{v_2, v_0\}$, then we would have $n_2(V) \geq 2$ and not be in case C1.

P2. All paths from s_1 to d_2 contain nodes v_5 and v_6.

First consider the path $Q_{s_2,d_2} = P_{s_2,d_2}[s_2, z] \oplus P_{z,v_1}[z, v_2] \oplus (v_2, v_0) \oplus P_{s_2,d_2}[v_0, d_2]$. Clearly, $Q_{s_2,d_2} \cap P_{s_1,d_1} = \emptyset$ and $v_5 \xrightarrow{I} Q_{s_2,d_2}$. If we have a path P_{s_1,d_2} not containing $\{v_5, v_6\}$ we conclude that $n_2(V, Q_{s_2,d_2}) \geq 2$. But since $n_1(V, P_{s_1,d_1}) \geq 2$ we contradict the fact that there are no paths with manageable interference.

P3. All paths from s_2 to d_1 contain $\{v_6, v_2\}$ or $\{v_3, v_4\}$.

Suppose there is a path P_{s_2,d_1} not containing $\{v_6, v_2\}$ nor $\{v_3, v_4\}$. Then we let $S = P_{s_1,d_1} \cup P_{s_2,d_2} \cup P_{s_2,d_1}$ and we have $n_1(S, P_{s_1,d_1}) \geq 2$. But since P1 and P2 imply that any path from s_1 to d_2 must contain $\{v_6, v_2\}$, and $\{v_6, v_2\} \not\subset S$, we must have $n_2(S, P_{s_2,d_2}) = 0$, contradicting the fact that there are no paths with manageable interference.

P4. The removal of v_0 disconnects d_2 from both sources.

From P1, the removal of v_0 disconnects d_2 from s_1. So suppose the removal of v_0 does not disconnect d_2 from s_2 and we have a path Q_{s_2,d_2} not containing v_0. We know that Q_{s_2,d_2} must be disjoint from P_{s_1,d_1}, since otherwise we would contradict the fact that the removal of v_0 disconnects d_2 from s_1 (P1). Moreover, if we let $S = P_{s_1,d_1} \cup Q_{s_2,d_2}$, since $v_0 \notin S$, we must have

$n_2(S, Q_{s_2,d_2}) = 0$. If $n_1(S, P_{s_1,d_1}) \neq 1$, we contradict the assumption of no two disjoint paths with manageable interference. However, if $n_1(S, P_{s_1,d_1}) = 1$, we must have $Q_{s_2,d_2} \xrightarrow{I} P_{s_1,d_1}$, and we will have $n_1(V \setminus \{v_0\}, P_{s_1,d_1}) \geq 2$ and $n_2(V \setminus \{v_0\}, Q_{s_2,d_2}) = 0$, and we again reach a contradiction.

P5. The removal of v_5 disconnects s_1 from both destinations.

From P2, the removal of v_5 disconnects s_1 from d_2. So we suppose the removal of v_5 does not disconnect s_1 from d_1 and we have a path Q_{s_1,d_1} not containing v_5. The path Q_{s_1,d_1} must be disjoint from P_{s_2,d_2}, or else we would contradict the fact that the removal of v_5 disconnects s_1 from d_2 (P2). So first we let $S = Q_{s_1,d_1} \cup P_{s_2,d_2}$, and, since $v_5 \notin S$, we have $n_2(S, P_{s_2,d_2}) = 0$. If we have $n_1(S, Q_{s_1,d_1}) \neq 1$, we contradict the assumption of no two disjoint paths with manageable interference. However, if $n_1(S, Q_{s_1,d_1}) = 1$, we must have $P_{s_2,d_2} \xrightarrow{I} Q_{s_1,d_1}$, and we will have $n_1(V \setminus \{v_5\}, Q_{s_1,d_1}) \geq 2$ and $n_2(V \setminus \{v_5\}, P_{s_2,d_2}) = 0$, and we again reach a contradiction.

P6. The removal of v_2 and v_3 disconnects d_2 from both sources.

From P1, the removal of v_2 disconnects d_2 from s_1. So suppose the removal of v_2 and v_3 does not disconnect d_2 from s_2 and we have a path Q_{s_2,d_2} not containing v_2 nor v_3. We know that Q_{s_2,d_2} is disjoint form P_{s_1,d_1}, or else we would contradict the fact that the removal of v_2 disconnects s_1 from d_2 (P1). Then, we set $S = P_{s_1,d_1} \cup Q_{s_2,d_2}$. Since $v_2, v_3 \notin S$, from P1, we must have $n_2(S, Q_{s_2,d_2}) = 0$, and from P3, we must have $n_1(S, P_{s_1,d_1}) = 0$, which is again a contradiction.

P7. The removal of v_2 and v_4 disconnects d_1 from both sources.

From P3, the removal of v_2 and v_4 disconnects d_1 from s_2. Thus, we assume that we have a path Q_{s_1,d_1} not containing v_2 nor v_4. The path Q_{s_1,d_1} must be disjoint of P_{s_2,d_2}, or else we contradict P3. Thus we set $S = Q_{s_1,d_1} \cup P_{s_2,d_2}$.

Since $v_2, v_4 \notin S$, from P1, we must have $n_2(S, P_{s_2,d_2}) = 0$, and from P3, we must have $n_1(S, Q_{s_1,d_1}) = 0.$, which is a contradiction.

P8. All paths from s_1 or s_2 to v_2 contain v_6.

This follows from P1, P2 and P3, since $v_2 \rightsquigarrow d_1$ and $v_2 \rightsquigarrow d_2$.

These properties allow us to infer the information inequalities that will build the converse proof. The intuition is similar to that of inequalities (2.4), (2.5) and (2.6), but the inequalities are somewhat different since they need to hold for any network in case C1. First we have

$$\begin{aligned} nR_2 &\leq I(W_2; Y_{d_2}^n) + n\epsilon_n \overset{(i)}{\leq} I(\tilde{X}_B^n; Y_0^n) + n\epsilon_n \\ &= I(X_2^n, \tilde{X}_B^n; Y_0^n) - I(X_2^n; Y_0^n|\tilde{X}_B^n) + n\epsilon_n \\ &\overset{(ii)}{\leq} \frac{n}{2}\log P + n\kappa_3 - I(X_2^n; Y_0^n|\tilde{X}_B^n) + n\epsilon_n, \end{aligned} \tag{2.7}$$

where (i) follows from the Markov chain $W_2 \leftrightarrow \tilde{X}_B^n \leftrightarrow Y_0^n \leftrightarrow Y_{d_2}^n$, which is implied by P4 and the fact that $s_2 \in B$; (ii) follows from the fact that $I(X_2^n, \tilde{X}_B^n; Y_0^n)$ can be upper bounded by $h(Y_0^n) - h(Z_{2,0}^n)$ by following the steps in (2.2), where κ_3 is a positive constant, independent of P, for P sufficiently large. We also have that

$$\begin{aligned} nR_1 &\leq I(W_1; Y_{d_1}^n) + n\epsilon_n \overset{(i)}{\leq} I(W_1; \tilde{X}_5^n, \tilde{X}_B^n) + n\epsilon_n \\ &\overset{(ii)}{=} I(W_1; \tilde{X}_5^n|\tilde{X}_B^n) + n\epsilon_n \overset{(iii)}{\leq} I(X_5^n; \tilde{X}_5^n|\tilde{X}_B^n) + n\epsilon_n \\ &\leq I(X_5^n; Y_6^n|\tilde{X}_B^n) + I(X_5^n; \tilde{X}_5^n|\tilde{X}_B^n, Y_6^n) + n\epsilon_n \\ &\overset{(iv)}{=} I(X_5^n; Y_6^n|\tilde{X}_B^n) + n\kappa_4 + n\epsilon_n, \end{aligned} \tag{2.8}$$

where (i) follows because P5 and the fact that $s_2 \in B$ imply that the removal of v_5 and B disconnects d_1 from both sources and thus $W_1 \leftrightarrow (\tilde{X}_5^n, \tilde{X}_B^n) \leftrightarrow Y_{d_1}^n$; (ii) follows from the fact that $\tilde{X}_B$ is independent of W_1; (iii) follows from the fact that, given $\tilde{X}_B^n$, we have $W_1 \leftrightarrow X_5^n \leftrightarrow \tilde{X}_5^n$; (iv) follows from Lemma 2.2, since P2 implies that $\mathcal{I}(v_6) \setminus \{v_5\} \subset B$. To obtain the next inequalities, we consider two cases, according to the position of v_4 and v_5.

I) $\ell(v_4) \leq \ell(v_5)$: In this case, we have

$$
\begin{aligned}
nR_2 &\leq I(W_2; Y_{d_2}^n) + n\epsilon_n \\
&\overset{(i)}{\leq} I(X_{s_2}^n; \tilde{X}_2^n, \tilde{X}_3^n) + n\epsilon_n \\
&\overset{(ii)}{\leq} I(X_{s_2}^n; \tilde{X}_2^n, \tilde{X}_3^n | \tilde{X}_A^n) + n\epsilon_n \\
&\leq I(X_{s_2}^n; \tilde{X}_2^n, \tilde{X}_3^n, Y_4^n | \tilde{X}_A^n) + n\epsilon_n \\
&= I(X_{s_2}^n; \tilde{X}_3^n, Y_4^n | \tilde{X}_A^n) + I(X_{s_2}^n; \tilde{X}_2^n | \tilde{X}_A^n, \tilde{X}_3^n, Y_4^n) + n\epsilon_n \\
&\leq I(X_{s_2}^n; Y_4^n | \tilde{X}_A^n) + I(X_{s_2}^n; \tilde{X}_3^n | \tilde{X}_A^n, Y_4^n) + I(X_{s_2}^n, \tilde{X}_3^n; \tilde{X}_2^n | \tilde{X}_A^n, Y_4^n) + n\epsilon_n \\
&\overset{(iii)}{\leq} I(X_{s_2}^n; Y_4^n | \tilde{X}_A^n) + n\kappa_5 + I(X_{s_2}^n, \tilde{X}_3^n; \tilde{X}_2^n | \tilde{X}_A^n, Y_4^n) + n\epsilon_n \\
&\overset{(iv)}{\leq} I(X_B^n; Y_4^n | \tilde{X}_A^n) + I(X_{s_2}^n, \tilde{X}_3^n; \tilde{X}_2^n | \tilde{X}_A^n, Y_4^n) + n\kappa_5 + n\epsilon_n \\
&\leq I(X_B^n; Y_4^n | \tilde{X}_A^n) + I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n, Y_4^n) + I(X_{s_2}^n, \tilde{X}_3^n; \tilde{X}_2^n | \tilde{X}_A^n, Y_4^n, X_B^n) + n\kappa_5 + n\epsilon_n \\
&\overset{(v)}{\leq} I(X_B^n; Y_4^n | \tilde{X}_A^n) + I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n, Y_4^n) + n\kappa_5 + n\epsilon_n \\
&\leq I(X_B^n; Y_4^n, \tilde{X}_2^n | \tilde{X}_A^n) + n\kappa_5 + n\epsilon_n,
\end{aligned} \tag{2.9}
$$

where (i) follows because P6 implies the Markov chain $W_2 \leftrightarrow X_{s_2}^n \leftrightarrow (\tilde{X}_2^n, \tilde{X}_3^n) \leftrightarrow Y_{d_2}^n$; (ii) follows from the fact that $\tilde{X}_A^n$ is independent of $X_{s_2}^n$; (iii) follows by applying Lemma 2.2 to the second term, since $\ell(v_4) \leq \ell(v_5)$ implies that $\mathcal{I}(v_4) \setminus \{v_3\} \subset A$, or else we contradict P3; (iv) follows from the fact that $s_2 \in B$; and (v) follows because we have $(X_{s_2}^n, \tilde{X}_3^n) \leftrightarrow (\tilde{X}_A^n, Y_4^n, X_B^n) \leftrightarrow \tilde{X}_2^n$, since the removal of A, v_4 and B disconnects s_2 and $\mathcal{O}(v_3)$ from v_2. This can be seen as follows. From P8, all paths from s_2 or v_3 to v_2 must contain a node in $\mathcal{I}(v_6)$. From P2, we know that $\mathcal{I}(v_6) \setminus \{v_5\} \subset B$. From P3, we know that any path from v_3 or s_2 to v_5 must contain v_4. Finally, since $\ell(v_4) < \ell(v_6)$, we have that $v_3 \notin \mathcal{I}(v_6)$, and, therefore, any path from s_2 or $\mathcal{O}(v_3)$ to v_2 must either contain v_4 or a node in B. Notice that we had to consider $\mathcal{O}(v_3)$ instead of simply v_3, because we have $\tilde{X}_3^n$, and not X_3^n. Next, we have that

$$
\begin{aligned}
nR_1 &\leq I(W_1; Y_{d_1}^n) + n\epsilon_n \overset{(i)}{\leq} I(W_1; Y_4^n, \tilde{X}_2^n) + n\epsilon_n \\
&\overset{(ii)}{\leq} I(\tilde{X}_A^n; Y_4^n, \tilde{X}_2^n) + n\epsilon_n \\
&= I(\tilde{X}_A^n, X_B^n; Y_4^n, \tilde{X}_2^n) - I(X_B^n; Y_4^n, \tilde{X}_2^n | \tilde{X}_A^n) + n\epsilon_n \\
&= I(\tilde{X}_A^n, X_B^n; Y_4^n) + I(\tilde{X}_A^n, X_B^n; \tilde{X}_2^n | Y_4^n) - I(X_B^n; Y_4^n, \tilde{X}_2^n | \tilde{X}_A^n) + n\epsilon_n \\
&\overset{(iii)}{\leq} \frac{n}{2}\log P + n\kappa_6 + I(\tilde{X}_A^n, X_B^n, Y_4^n; \tilde{X}_2^n) - I(X_B^n; Y_4^n, \tilde{X}_2^n | \tilde{X}_A^n) + n\epsilon_n,
\end{aligned}
$$

where (i) follows because P7 implies the Markov chain $W_1 \leftrightarrow (Y_4^n, \tilde{X}_2^n) \leftrightarrow Y_{d_1}^n$; (ii) follows since $s_1 \in A$; (iii) follows from the fact that $I(\tilde{X}_A^n, X_B^n; Y_4^n)$ can be upper bounded by $h(Y_4^n) - h(Z_{3,4}^n)$ by following the steps in (2.2), where κ_6 is a positive constant, independent of P, for P sufficiently large. The second term in the inequality above can be bounded as

$$
\begin{aligned}
I(\tilde{X}_A^n, X_B^n, Y_4^n; \tilde{X}_2^n) &\overset{(i)}{\leq} I(\tilde{X}_A^n, \tilde{X}_B^n, Y_4^n; \tilde{X}_2^n) \\
&= I(\tilde{X}_B^n; \tilde{X}_2^n) + I(\tilde{X}_A^n, Y_4^n; \tilde{X}_2^n | \tilde{X}_B^n) \\
&\overset{(ii)}{\leq} I(\tilde{X}_B^n; Y_6^n) + I(\tilde{X}_A^n, Y_4^n; \tilde{X}_2^n | \tilde{X}_B^n) \\
&\overset{(iii)}{\leq} I(\tilde{X}_B^n; Y_6^n) + I(X_2^n; \tilde{X}_2^n | \tilde{X}_B^n) \\
&\leq I(X_5^n, \tilde{X}_B^n; Y_6^n) - I(X_5^n; Y_6^n | \tilde{X}_B^n) + I(X_2^n; Y_0^n | \tilde{X}_B^n) + I(X_2^n; \tilde{X}_2^n | \tilde{X}_B^n, Y_0^n) \\
&\overset{(iv)}{\leq} I(X_5^n, \tilde{X}_B^n; Y_6^n) - I(X_5^n; Y_6^n | \tilde{X}_B^n) + I(X_2^n; Y_0^n | \tilde{X}_B^n) + n\kappa_7 \\
&\overset{(v)}{\leq} \frac{n}{2}\log P - I(X_5^n; Y_6^n | \tilde{X}_B^n) + I(X_2^n; Y_0^n | \tilde{X}_B^n) + n(\kappa_7 + \kappa_8) \qquad (2.10)
\end{aligned}
$$

where (i) follows because of the Markov chain $(\tilde{X}_A^n, X_B^n, Y_4^n) \leftrightarrow (\tilde{X}_A^n, \tilde{X}_B^n, Y_4^n) \leftrightarrow \tilde{X}_2^n$; (ii) follows because P8 implies $\tilde{X}_B^n \leftrightarrow Y_6^n \leftrightarrow \tilde{X}_2^n$; (iii) follows since, given X_B^n, we have $(\tilde{X}_A^n, Y_4^n) \leftrightarrow X_2^n \leftrightarrow \tilde{X}_2^n$; (iv) follows by applying Lemma 2.2 to $I(X_2^n; \tilde{X}_2^n | \tilde{X}_B^n, Y_0^n)$, since $\mathcal{I}(v_0) \setminus \{v_2\} \subset B$, or else we contradict P1; (v) follows from the fact that $I(X_5^n, \tilde{X}_B^n; Y_6^n)$ can be upper bounded by $h(Y_6^n) - h(Z_{5,6}^n)$ by following the steps in

(2.2), where κ_8 is a positive constant, independent of P, for P sufficiently large. Thus, we obtain

$$\begin{aligned} nR_1 \leq\ & n \log P - I(X_5^n; Y_6^n | \tilde{X}_B^n) + I(X_2^n; Y_0^n | \tilde{X}_B^n) \\ & - I(X_B^n; Y_4^n, \tilde{X}_2^n | \tilde{X}_A^n) + n(\kappa_6 + \kappa_7 + \kappa_8) + n\epsilon_n. \end{aligned} \tag{2.11}$$

II) $\ell(v_4) > \ell(v_5)$: We will obtain similar inequalities to the ones in case I. We will define $C \triangleq \mathcal{I}(v_4) \setminus \{v_2, v_3\}$ and $D \triangleq \mathcal{O}(v_3) \setminus \{v_6\}$. Then, we will let $Y_{C,4}^n = \sum_{v_c \in C} \tilde{X}_{c,4}^n$. We also let $\tilde{X}_{3,D}^n = \{\tilde{X}_{3,v_j}^n : v_j \in D\}$. Notice that $\tilde{X}_{3,D}^n = \tilde{X}_3^n$ if $v_6 \notin \mathcal{O}(v_3)$. Then we have

$$\begin{aligned} nR_2 &\leq I(W_2; Y_{d_2}^n) + n\epsilon_n \stackrel{(i)}{\leq} I(X_{s_2}^n; \tilde{X}_2^n, \tilde{X}_{3,D}^n) + n\epsilon_n \\ &\stackrel{(ii)}{\leq} I(X_{s_2}^n; \tilde{X}_2^n, \tilde{X}_{3,D}^n | \tilde{X}_A^n) + n\epsilon_n \\ &= I(X_{s_2}^n; \tilde{X}_2^n | \tilde{X}_A^n) + I(X_{s_2}^n; \tilde{X}_{3,D}^n | \tilde{X}_A^n, \tilde{X}_2^n) + n\epsilon_n \\ &\leq I(X_{s_2}^n; \tilde{X}_2^n | \tilde{X}_A^n) + I(X_{s_2}^n; \tilde{X}_{3,D}^n, \tilde{X}_{3,4}^n | \tilde{X}_A^n, \tilde{X}_2^n) + n\epsilon_n \\ &\stackrel{(iii)}{\leq} I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n) + I(X_{s_2}^n; \tilde{X}_{3,4}^n | \tilde{X}_A^n, \tilde{X}_2^n) + I(X_{s_2}^n; \tilde{X}_{3,D}^n | \tilde{X}_A^n, \tilde{X}_2^n, \tilde{X}_{3,4}^n) + n\epsilon_n \\ &\stackrel{(iv)}{\leq} I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n) + I(X_{s_2}^n; \tilde{X}_{3,4}^n | \tilde{X}_A^n, \tilde{X}_2^n) + n\kappa_9 + n\epsilon_n \\ &\stackrel{(v)}{\leq} I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n) + I(X_B^n; \tilde{X}_{3,4}^n | \tilde{X}_A^n, \tilde{X}_2^n) + n\kappa_9 + n\epsilon_n \\ &\stackrel{(vi)}{\leq} I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n) + I(X_B^n; \tilde{X}_{3,4}^n | \tilde{X}_A^n, \tilde{X}_2^n, \tilde{X}_C^n) + n\kappa_9 + n\epsilon_n \end{aligned} \tag{2.12}$$

where (i) follows because P6 implies that the removal of $\mathcal{O}(v_3)$ and v_2 disconnects d_2 from both sources. Then, since P1 implies that all paths from v_6 to d_2 contain v_2, we know that the removal of D and v_2 also disconnects d_2 from both sources, and we have the Markov chain $W_2 \leftrightarrow X_{s_2}^n \leftrightarrow (\tilde{X}_2^n, \tilde{X}_{3,D}^n) \leftrightarrow Y_{d_2}^n$; (ii) follows from the fact that $\tilde{X}_A^n$ is independent of $X_{s_2}^n$; (iii) follows since $s_2 \in B$; (iv) follows by applying Lemma 2.2 to $I(X_{s_2}^n; \tilde{X}_{3,D}^n | \tilde{X}_A^n, \tilde{X}_2^n, \tilde{X}_{3,4}^n)$, since, in case II, if $u \in D \setminus \{v_4\}$, then $u \not\rightsquigarrow v_2$, or else we contradict P8; (v) follows since $s_2 \in B$; and (vi) follows from the fact that, given $\tilde{X}_2^n$ and $\tilde{X}_A^n$, $\tilde{X}_C^n$ is independent of X_B^n. This is true because P3 implies that any path from a node in B to a node in C must contain v_2, and, thus, the removal of A and v_2 disconnects C from B and both sources. Notice that (vi) is only non-trivial in the cases where $C \not\subset A$ (when $\ell(v_4) > \ell(v_1) + 1$). Next, we have that

$$
\begin{aligned}
nR_1 &\leq I(W_1; Y_{d_1}^n) + n\epsilon_n \stackrel{(i)}{\leq} I(W_1; \tilde{X}_{3,4}^n + Y_{C,4}^n, \tilde{X}_2^n) + n\epsilon_n \\
&\stackrel{(ii)}{\leq} I(\tilde{X}_A^n; \tilde{X}_{3,4}^n + Y_{C,4}^n, \tilde{X}_2^n) + n\epsilon_n \\
&= I(\tilde{X}_A^n; \tilde{X}_2^n) + I(\tilde{X}_A^n; \tilde{X}_{3,4}^n + Y_{C,4}^n | \tilde{X}_2^n) + n\epsilon_n \\
&\leq I(\tilde{X}_A^n, \tilde{X}_C^n; \tilde{X}_{3,4}^n + Y_{C,4}^n | \tilde{X}_2^n) + I(\tilde{X}_A^n; \tilde{X}_2^n) + n\epsilon_n \\
&= I(\tilde{X}_A^n, \tilde{X}_C^n, X_B^n; \tilde{X}_{3,4}^n + Y_{C,4}^n | \tilde{X}_2^n) - I(X_B^n; \tilde{X}_{3,4}^n + \tilde{Y}_{C,4}^n | \tilde{X}_2^n, \tilde{X}_A^n, \tilde{X}_C^n) \\
&\qquad\qquad + I(\tilde{X}_A^n, X_B^n; \tilde{X}_2^n) - I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n) + n\epsilon_n \\
&\leq I(\tilde{X}_A^n, \tilde{X}_C^n, X_B^n, \tilde{X}_2^n; \tilde{X}_{3,4}^n + \tilde{Y}_{C,4}^n) - I(X_B^n; \tilde{X}_{3,4}^n | \tilde{X}_2^n, \tilde{X}_A^n, \tilde{X}_C^n) \\
&\qquad\qquad + I(\tilde{X}_A^n, X_B^n; \tilde{X}_2^n) - I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n) + n\epsilon_n \\
&\stackrel{(iii)}{\leq} \frac{n}{2} \log P + n\kappa_{10} + I(\tilde{X}_A^n, X_B^n; \tilde{X}_2^n) - I(X_B^n; \tilde{X}_{3,4}^n | \tilde{X}_2^n, \tilde{X}_A^n, \tilde{X}_C^n) - I(X_B^n; \tilde{X}_2^n | \tilde{X}_A^n) + n\epsilon_n,
\end{aligned}
$$

where (i) follows because P7 implies the Markov chain $W_1 \leftrightarrow (Y_4^n, \tilde{X}_2^n) \leftrightarrow Y_{d_1}^n$, and $(Y_4^n, \tilde{X}_2^n)$ can be constructed from $(\tilde{X}_{3,4} + Y_{C,4}^n, \tilde{X}_2)$ (notice that it may be the case that $Y_4^n = \tilde{X}_{3,4} + \tilde{Y}_{C,4}^n + \tilde{X}_{2,4}$, if $v_2 \in \mathcal{I}(v_4)$); (ii) follows since $s_1 \in A$; (iii) follows from the fact that $I(\tilde{X}_A^n, \tilde{X}_C^n, X_B^n, \tilde{X}_2^n; \tilde{X}_{3,4} + Y_{C,4}^n)$ can be upper bounded by $h(\tilde{X}_{3,4} + Y_{C,4}^n) - h(Z_{3,4}^n)$ by following the steps in (2.2), where κ_{10} is a positive constant, independent of P, for P sufficiently large. The second term in the inequality above can be bounded as

$$
\begin{aligned}
I(\tilde{X}_A^n, X_B^n; \tilde{X}_2^n) &\leq I(\tilde{X}_A^n, \tilde{X}_B^n; \tilde{X}_2^n) \\
&= I(\tilde{X}_B^n; \tilde{X}_2^n) + I(\tilde{X}_A^n; \tilde{X}_2^n | \tilde{X}_B^n) \\
&\leq I(\tilde{X}_B^n; Y_6^n) + I(X_2^n; \tilde{X}_2^n | \tilde{X}_B^n)
\end{aligned}
$$

$$\leq I(X_5^n, \tilde{X}_B^n; Y_6^n) - I(X_5^n; Y_6^n|\tilde{X}_B^n) + I(X_2^n; Y_0^n|\tilde{X}_B^n) + n\kappa_{11}$$
$$\leq \frac{n}{2}\log P - I(X_5^n; Y_6^n|\tilde{X}_B^n) + I(X_2^n; Y_0^n|\tilde{X}_B^n) + n(\kappa_{11} + \kappa_{12})$$

where the inequalities are justified as in (2.10). Therefore, we obtain

$$\begin{aligned} nR_1 \leq n\log P - I(X_5^n; Y_6^n|\tilde{X}_B^n) + I(X_2^n; Y_0^n|\tilde{X}_B^n) - I(X_B^n; \tilde{X}_{3,4}^n|\tilde{X}_2^n, \tilde{X}_A^n, \tilde{X}_C^n) \\ - I(X_B^n; \tilde{X}_2^n|\tilde{X}_A^n) + n(\kappa_{10} + \kappa_{11} + \kappa_{12}) + n\epsilon_n. \end{aligned} \quad (2.13)$$

Finally, by adding equations (2.7), (2.8), (2.9) and (2.11) for case I, and (2.7), (2.8), (2.12) and (2.13) for case II, we obtain

$$2n(R_1 + R_2) \leq \frac{3n}{2}\log P + 6n\kappa_{\max} + n\epsilon_n$$
$$\Rightarrow \frac{R_1 + R_2}{\frac{1}{2}\log P} \leq \frac{3}{2} + \frac{6\kappa_{\max} + \epsilon_n}{\log P},$$

where $\kappa_{\max} = \max_j \kappa_j$. Thus, as we let $n \to \infty$ and then $P \to \infty$, we obtain

$$D_\Sigma \leq \frac{3}{2}.$$

2.6 Characterizing the Full Degrees-of-Freedom Region

In this section, we extend the results from Theorem 2.1 and characterize the full degrees-of-freedom region of two-unicast layered Gaussian networks. The degrees-of-freedom region (see Definition 1.6) can be understood as a high-SNR approximation to the capacity region, scaled down by $\frac{1}{2}\log P$. Since the sum degrees-of-freedom can be expressed as

$$D_\Sigma = \lim_{P\to\infty}\left(\sup_{(R_1,R_2)\in C(P)} \frac{R_1 + R_2}{\frac{1}{2}\log P}\right),$$

we conclude that if $(D_1, D_2) \in \mathbf{D}$, we must have $D_1 + D_2 \leq D_\Sigma$. Thus, the results from Theorem 2.1 provide an outer bound to the degrees-of-freedom region with at least one achievable point. Moreover, it is always possible to bound each individual rate R_i as

$$\begin{aligned} nR_i &\leq I(W_i; Y_{d_i}^n) + n\epsilon_n = h(Y_{d_i}^n) - h(Y_{d_i}^n|W_i) + n\epsilon \\ &\leq h(Y_{d_i}^n) - h(Y_{d_i}^n|W_i, X_{\mathcal{I}(d_i)}^n) + n\epsilon \\ &= h(Y_{d_i}^n) - h(Z_{d_i}^n) + n\epsilon \leq \frac{n}{2}\log P + n\kappa + n\epsilon, \end{aligned}$$

where κ is independent of P, for $i = 1, 2$. Hence, we conclude that $\mathbf{D}$ is always a subset of $\left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \leq 1, D_2 \leq 1\right\}$. It is straightforward to show that $\mathbf{D}$ is convex. Therefore, for networks that belong to case (A) from Theorem 2.1, the fact that $D_\Sigma = 1$ guarantees that the degrees-of-freedom region is

$$\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 + D_2 \leq 1\right\}.$$

This region is depicted in Figure 2.4(a). The degrees-of-freedom region for networks in case (B) can also be easily obtained from the result in Theorem 2.1. Since for all networks in cases (B), we have $(1, 1) \in \mathbf{D}$, we conclude that the degrees-of-freedom region in these cases is given by

$$\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \leq 1, D_2 \leq 1\right\},$$

as depicted in Figure 2.4(b).

For networks in case (C), we will once again consider the division into cases C1 and C2, as described in Section 2.5. First consider networks in case C1. We will show that in these cases the degrees-of-freedom region will be given by

$$\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \leq 1, D_2 \leq 1, D_1 + D_2 \leq 3/2\right\}, \tag{2.14}$$

as shown in Figure 2.4(c). In order to do that, since our network $\mathcal{N}$ is in C1, we will assume that it contains disjoint paths P_{s_1,d_1} and P_{s_2,d_2} such that $n_1(V) \geq 2$,

$n_1^0 = 1$, $n_2(V) = 1$ and $n_2^0 = 0$, and that the nodes from $\mathcal{N}$ are named as depicted in Figure 2.15(a). We will then consider the condensed network formed by layers V_1, $V_{\ell(v_2)}$ and V_r, which is shown in Fig. 2.15(b). Notice that u_2 is the same node

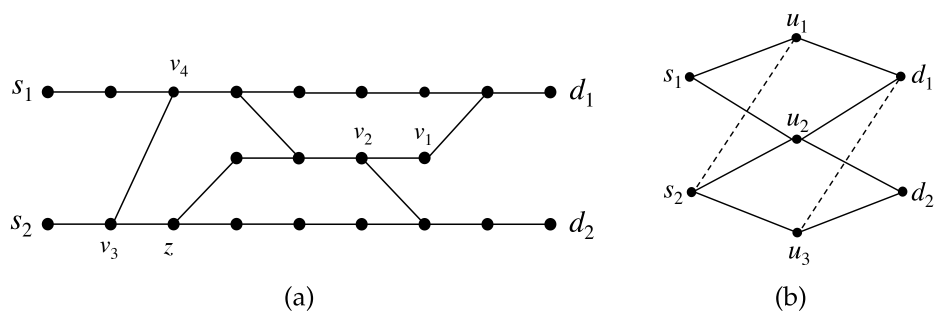


Figure 2.15: (a) Illustration of network from case C1, up to a change in the position of edge (v_3, v_4); (b) Condensed network for networks in case C1.

as v_2 in the original network. Therefore, edges $\hat{h}(s_1, u_3)$ and $\hat{h}(u_1, d_2)$ cannot exist due to property P1 in Section 2.5.2. Edges $\hat{h}(s_2, u_1)$ and $\hat{h}(u_3, d_1)$ may or may not exist, and that will depend on the position of the edge (v_3, v_4) in the original network. We will show that points $(1, 1/2)$ and $(1/2, 1)$ are included in $\mathbf{D}$ and, by convexity and the fact that $D_\Sigma \leq 3/2$, $\mathbf{D}$ must be as shown in Figure 2.4(c).

First note that we may assume that exactly one of the edges $\hat{h}(s_2, u_1)$ and $\hat{h}(u_3, d_1)$ exists. Otherwise, by removing u_2, P_{s_1,d_1} and P_{s_2,d_2} have manageable interference, and we are in case (B). Hence, we restrict ourselves to two cases: (1) $\hat{h}(u_3, d_1) \neq 0$ and (2) $\hat{h}(s_2, u_1) \neq 0$. We consider each one separately.

Case I: $\hat{h}(u_3, d_1) \neq 0$

This network is depicted in Figure 2.16(a). The achievability scheme described in Section 2.5.1 based on Lemma 2.5 shows that $(1/2, 1) \in \mathbf{D}$. In order to achieve the point $(1, 1/2)$, we need to use a scheme based on real interference alignment, similar to the ones described in [45] and [24].

At source s_1, the message W_1 will be split into two submessages $W_1^{(1)}$ and $W_1^{(2)}$, while s_2 will have a single message W_2. Each of these messages will be encoded using a single codebook with codewords of length n, which is obtained by uniform i.i.d. sampling of the set

$$\mathcal{U} = \mathbb{Z} \cap \left[-\gamma P^{\frac{1-\epsilon}{2(2+\epsilon)}}, \gamma P^{\frac{1-\epsilon}{2(2+\epsilon)}} \right], \tag{2.15}$$

for a small $\epsilon > 0$ and a constant γ. The rate of this code, i.e., the number of codewords, will be determined later. We will let $a_j[1], a_j[2], ..., a_j[n]$ be the n symbols of the codeword associated to message $W_1^{(j)}$, $j = 1, 2$, and $b[1], b[2], ..., b[n]$ be the n symbols of the codeword associated to message W_2. At time $t \in \{1, ..., n\}$, source s_1 will transmit

$$X_{s_1}[t] = G(a_1[t] + Ta_2[t]),$$

where T is an irrational number, and $G = \beta P^{\frac{1+2\epsilon}{2(2+\epsilon)}}$ is chosen to satisfy the power constraint for a constant β to be determined. Source s_2 will transmit

$$X_{s_2}[t] = G \frac{\hat{h}(s_1, u_2)}{\hat{h}(s_2, u_2)} b[t].$$

The maximum power of a transmit signal from s_1 is upper-bounded by

$$\beta^2 P^{\frac{1+2\epsilon}{2+\epsilon}} (1 + T^2) \gamma^2 P^{\frac{1-\epsilon}{2+\epsilon}} = \beta^2 (1 + T^2) \gamma^2 P,$$

and the maximum power of a transmit signal from s_2 is upper-bounded by

$$\beta^2 P^{\frac{1+2\epsilon}{2+\epsilon}} \frac{\hat{h}(s_1, u_2)^2}{\hat{h}(s_2, u_2)^2} \gamma^2 P^{\frac{1-\epsilon}{2+\epsilon}} = \beta^2 \frac{\hat{h}(s_1, u_2)^2}{\hat{h}(s_2, u_2)^2} \gamma^2 P.$$

Thus, for any choice of T and γ, parameter β can be chosen so that the maximum transmit power at the sources is less than P. Next we write the received signals at u_1, u_2 and u_3. We will drop the time t from the notation for simplicity.

$$Y_{u_1} = G\hat{h}(s_1, u_1)(a_1 + Ta_2) + Z_{u_1}$$

$$Y_{u_2} = G\hat{h}(s_1,u_2)(a_1+b+Ta_2)+Z_{u_2}$$
$$Y_{u_3} = \frac{G\hat{h}(s_2,u_3)\hat{h}(s_1,u_2)}{\hat{h}(s_2,u_2)}b+Z_{u_3}$$

Nodes u_1 and u_3 will simply perform amplify-and-forward. More precisely, their transmit signals will be given by

$$X_{u_1} = \alpha Y_{u_1} = \alpha G\hat{h}(s_1,u_1)(a_1+Ta_2)+\alpha Z_{u_1},$$

$$\begin{aligned} X_{u_3} &= -\alpha\frac{\hat{h}(s_2,u_2)}{\hat{h}(s_2,u_3)\hat{h}(s_1,u_2)}\frac{\hat{h}(s_1,u_1)\hat{h}(u_1,d_1)}{\hat{h}(u_3,d_1)}Y_{u_3} \\ &= -\alpha G\frac{\hat{h}(s_1,u_1)\hat{h}(u_1,d_1)}{\hat{h}(u_3,d_1)}b+\alpha\kappa_1 Z_{u_3}, \end{aligned}$$

where α is a constant and κ_1 is a function of the channel gains. We will choose α so that the power constraint at each relay is satisfied. For example, consider node u_1. The noise term at its received signal, Z_{u_1}, is a linear combination of noises in the original networks, whose coefficients are products of channel gains. Therefore, we may assume that $E[Z_{u_1}^2]$ is a constant $\sigma_{u_1}^2$. Thus, the transmitted power at u_1 is

$$\begin{aligned} E[\alpha^2 Y_{u_1}^2] &= \alpha^2\beta^2 P^{\frac{1+2\epsilon}{2+\epsilon}}\hat{h}^2(s_1,u_1)(1+T^2)\gamma^2 P^{\frac{1-\epsilon}{2+\epsilon}}+\alpha^2\sigma_{u_1}^2 \\ &= \alpha^2\beta^2\hat{h}^2(s_1,u_1)(1+T^2)\gamma^2 P+\alpha^2\sigma_{u_1}^2. \end{aligned} \tag{2.16}$$

It is now easy to see that α can be chosen independently of P to make sure that $E[\alpha^2 Y_{u_1}^2] \le P$, for P sufficiently large. The received signal at node u_2 can be seen as a noisy observation of a point in the set

$$\mathcal{U}_{u_2} = G\hat{h}(s_1,u_2)\{x_1+Tx_2 : x_1 \in \mathcal{U}+\mathcal{U}, x_2 \in \mathcal{U}\},$$

for $x_1 = a_1 + b$ and $x_2 = a_2$, where $\mathcal{U}+\mathcal{U} = \{u_1+u_2 : u_1,u_2 \in \mathcal{U}\} \subset \mathbb{Z} \cap \left[-2\gamma P^{\frac{1-\epsilon}{2(2+\epsilon)}}, 2\gamma P^{\frac{1-\epsilon}{2(2+\epsilon)}}\right]$. As explained in [45], the fact that T is irrational guarantees that there is a one-to-one map from the points in $\mathcal{U}_{u_2}$ to the points $(x_1,x_2) \in$

$(\mathcal{U}+\mathcal{U})\times\mathcal{U}$. Moreover, from Theorem 1 of [45] and subsequent remarks, we conclude that, for almost all choices of T, the minimum distance between two points in $\mathcal{U}_{u_2}$ satisfies

$$d_{\min} > G\hat{h}(s_1,u_2)\frac{\kappa}{(\max_{x\in\mathcal{U}} x)^{1+\epsilon}},$$

for some constant κ. Thus we have

$$d_{\min} > \hat{h}(s_1,u_2)\frac{\kappa\beta P^{\frac{1+2\epsilon}{2(2+\epsilon)}}}{\gamma^{1+\epsilon}P^{\frac{(1-\epsilon)(1+\epsilon)}{2(2+\epsilon)}}} = \frac{\hat{h}(s_1,u_2)\kappa\beta}{\gamma^{1+\epsilon}}P^{\epsilon/2}.$$

Node u_2 will map its received signal to the nearest point in $\mathcal{U}_{u_2}$, and then use the fact that this point uniquely determines (a_1+b) and a_2 to decode these two integers. We will refer to the output of this procedure as $\hat{a}_1+\hat{b}$ and $\hat{a}_2$. If the variance of Z_{u_2} is given by $\sigma_{u_2}^2$, then the probability of a wrong decoding at u_2 is given by

$$\begin{aligned}\Pr[\hat{a}_1+\hat{b}\neq a_1+b, \hat{a}_2\neq a_2] &\leq 2\,Q\left(\frac{d_{\min}}{2\sigma_{u_2}}\right)\\ &< \exp\left(-\frac{d_{\min}^2}{8\sigma_{u_2}^2}\right)\\ &= \exp\left(-\delta P^{\epsilon}\right),\end{aligned}$$

where δ is a positive constant, independent of P. The transmit signal at u_2 will then be

$$X_{u_2} = \alpha G\frac{\hat{h}(s_1,u_1)\hat{h}(u_1,d_1)}{\hat{h}(u_2,d_1)}(\hat{a}_1+\hat{b}).$$

We choose α independently of P, so that the power constraints at u_1, u_2 and u_3 are simultaneously satisfied, for P sufficiently large. The received signal at the destination d_1 is given by

$$\begin{aligned}Y_{d_1} &= \hat{h}(u_1,d_1)X_{u_1}+\hat{h}(u_2,d_1)X_{u_2}+\hat{h}(u_3,d_1)X_{u_3}+Z_{d_1}\\ &= \alpha G\hat{h}(s_1,u_1)\hat{h}(u_1,d_1)\left(a_1+Ta_2+\hat{a}_1+\hat{b}-b\right)+Z_{d_1}^{\text{eff}},\end{aligned}$$

where $Z_{d_1}^{\text{eff}} = \alpha\hat{h}(u_1, d_1)Z_{u_1} + \alpha\hat{h}(u_2, d_1)\kappa_1 Z_{u_3} + Z_{d_1}$. The received signal at d_2 is

$$\begin{aligned} Y_{d_2} &= \hat{h}(u_2, d_2)X_{u_2} + \hat{h}(u_3, d_2)X_{u_3} + Z_{d_2} \\ &= \alpha G \frac{\hat{h}(u_2, d_2)\hat{h}(s_1, u_1)\hat{h}(u_1, d_1)}{\hat{h}(u_2, d_1)}(\hat{a}_1 + \hat{b}) - \alpha \frac{\hat{h}(u_3, d_2)\hat{h}(s_1, u_1)\hat{h}(u_1, d_1)}{\hat{h}(u_3, d_1)} b + Z_{d_2}^{\text{eff}}, \end{aligned}$$

where $Z_{d_2}^{\text{eff}} = \alpha\hat{h}(u_3, d_2)\kappa_1 Z_{u_3} + Z_{d_2}$. Notice that with probability at least $1 - \exp(-\delta P^{\epsilon})$, Y_{d_1} and Y_{d_2} are given by

$$\begin{aligned} Y_{d_1} &= \alpha G\hat{h}(s_1, u_1)\hat{h}(u_1, d_1)(2a_1 + Ta_2) + Z_{d_1}^{\text{eff}}, \\ Y_{d_2} &= \alpha G\hat{h}(s_1, u_1)\hat{h}(u_1, d_1)\left(\frac{\hat{h}(u_2, d_2)}{\hat{h}(u_2, d_1)}(a_1 + b) - \frac{\hat{h}(u_3, d_2)}{\hat{h}(u_3, d_1)} b\right) + Z_{d_2}^{\text{eff}}. \end{aligned}$$

The destinations will first perform a hard-decoding, similar to the one performed by u_2. If we assume that the decoding at node u_2 was correct, the signal received by d_1 is a noisy version of a point in the set

$$\mathcal{U}_{d_1} = \alpha G\hat{h}(s_1, u_1)\hat{h}(u_1, d_1)\left\{2x_1 + Tx_2 : x_1, x_2 \in \mathcal{U}\right\},$$

for $x_1 = a_1$ and $x_2 = a_2$. Thus, by using the same argument used previously, it can be shown that d_1 can decode a_1 and a_2 with probability of error smaller than $\exp(-\delta_2 P^{\epsilon})$, for some positive constant δ_2. Assuming that the decoding at node u_2 was correct, the signal received by d_2 is a noisy version of a point in the set

$$\mathcal{U}_{d_2} = \alpha G\hat{h}(s_1, u_1)\hat{h}(u_1, d_1)\frac{\hat{h}(u_2, d_2)}{\hat{h}(u_2, d_1)}\left\{x_1 + T_2 x_2 : x_1 \in \mathcal{U} + \mathcal{U}, x_2 \in \mathcal{U}\right\},$$

for $x_1 = a_1 + b$ and $x_2 = b$, where $T_2 = -\frac{\hat{h}(u_2,d_1)\hat{h}(u_3,d_2)}{\hat{h}(u_2,d_2)\hat{h}(u_3,d_1)}$. Next we notice that $\hat{h}(u_2, d_1)$, $\hat{h}(u_3, d_2)$, $\hat{h}(u_2, d_2)$ and $\hat{h}(u_3, d_1)$ are each a polynomial on the channel gains h_e of the original network with only coefficients 1. Notice that for almost all choices of the channel gains h_e, since the polynomials $\hat{h}(u_2, d_1)\hat{h}(u_3, d_2)$ and $\hat{h}(u_2, d_2)\hat{h}(u_3, d_1)$ are not identical, T_2 is an irrational number. From the description of the original network, given in the proof of Lemma 2.5, we see that $u_2 = v_2$ is on a path P_{z,v_1}

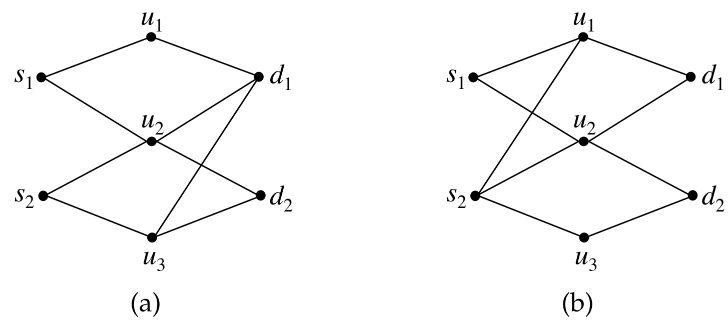


Figure 2.16: (a) Condensed network in case (1); (b) Condensed network in case (2).

such that $P_{z,v_1} \cap P_{s_2,d_2} = \{z\}$, $z \neq v_2$ and $v_1 \to P_{s_1,d_1}$. Therefore, there must exist two disjoint paths P_{u_2,d_1} and P_{u_3,d_2}. This implies that the determinant

$$\begin{vmatrix} \hat{h}(u_2, d_1) & \hat{h}(u_3, d_1) \\ \hat{h}(u_2, d_2) & \hat{h}(u_3, d_2) \end{vmatrix} = \hat{h}(u_2, d_1)\hat{h}(u_3, d_2) - \hat{h}(u_3, d_1)\hat{h}(u_2, d_2)$$

is non-zero for almost all channel gain values. This implies that $\hat{h}(u_2, d_1)\hat{h}(u_3, d_2)$ and $\hat{h}(u_3, d_1)\hat{h}(u_2, d_2)$ are not identical polynomials. Hence, we conclude that, for almost all channel gain values, T_2 is an irrational number and d_2 can decode b (and also $a_1 + b$, and thus a_1, even though d_2 does not require the message encoded by the a_1's) with probability at least $1 - \exp(-\delta_3 P^{\epsilon})$, for some $\delta_3 > 0$.

By applying these hard-decoders, destination d_1 obtains the estimates $\hat{a}_1[t]$ and $\hat{a}_2[t]$, and destination d_2 obtains the estimates $\hat{b}[t]$, for $t = 1, ..., n$. Then they can apply typicality-based decoders in order to decode the messages $W_1^{(1)}$, $W_1^{(2)}$ and W_2.

We now determine the rate of the codebook which is used to encode $W_1^{(1)}$, $W_1^{(2)}$ and W_2. We notice that for each of the messages $W_1^{(1)}$, $W_1^{(2)}$ and W_2, we effectively have a point-to-point discrete channel with input and output alphabet $\mathcal{U}$. Even though we do not calculate the actual transition probabilities, we know that the error probability is upper-bounded as

$$P_e \leq 1 - (1 - \exp(-\delta P^{\epsilon}))(1 - \exp(-\delta_2 P^{\epsilon}))(1 - \exp(-\delta_3 P^{\epsilon}))$$
$$\leq 1 - (1 - \exp(-\delta_{\min} P^{\epsilon}))^3 \leq 4 \exp(-\delta_{\min} P^{\epsilon}),$$

where $\delta_{\min} = \min(\delta, \delta_2, \delta_3)$. This allows us to lower bound the mutual information between the input U and the output $\hat{U}$ of this channel, for a uniform distribution over the input alphabet. Using Fano's inequality, we have

$$\begin{aligned}
I(U; \hat{U}) &\geq H(U) - H(U|\hat{U}) \\
&\geq \log |\mathcal{U}| - (1 + P_e \log |\mathcal{U}|) \\
&= (1 - P_e) \log |\mathcal{U}| - 1 \\
&\geq (1 - 4 \exp(-\delta_{\min} P^{\epsilon})) \left(\frac{1 - \epsilon}{2 + \epsilon} \frac{\log P}{2} + 1 \right) - 1 \\
&\geq (1 - 4 \exp(-\delta_{\min} P^{\epsilon})) \frac{1 - \epsilon}{2 + \epsilon} \frac{\log P}{2} - 4.
\end{aligned}$$

Therefore, since we constructed our code by taking independent samples uniformly at random from the set $\mathcal{U}$, it can achieve rate $R = (1 - 4 \exp(-\delta_{\min} P^{\epsilon})) \frac{1-\epsilon}{2+\epsilon} \frac{\log P}{2} - 4$, by having the codebook contain 2^{nR} codewords. Thus, each of the messages $W_1^{(1)}$, $W_1^{(2)}$ and W_2 possesses

$$\lim_{P \to \infty} \frac{R}{\frac{1}{2} \log P} = \frac{1 - \epsilon}{2 + \epsilon}$$

degrees-of-freedom. Since ϵ can be chosen arbitrarily small, we conclude that each of the messages may in fact achieve arbitrarily close to $1/2$ degrees-of-freedom. Therefore, we achieve the point $(1, 1/2)$ in the degrees-of-freedom region, and complete the proof in this case.

Case II: $\hat{h}(s_2, u_1) \neq 0$

This network is depicted in Figure 2.16(b). The achievability scheme described in Section 2.5.1 based on Lemma 2.5 shows that $(1, 1/2) \in \mathbf{D}$ Once more, we will use real interference alignment to achieve the other extreme point, i.e., $(1/2, 1)$. We will follow the steps in case I closely. This time, at source s_2, message W_2 will be split into two submessages $W_2^{(1)}$ and $W_2^{(2)}$, while s_1 will have a single message W_1. These three messages will be encoded using a single codebook with codewords of length n, which are obtained by uniform i.i.d. sampling of the set $\mathcal{U}$, defined in (A.1). The rate of this code will be determined later. We let $a[1], a[2], ..., a[n]$ be the n symbols of the codeword associated to message W_1, and $b_j[1], b_j[2], ..., b_j[n]$ be the n symbols of the codeword associated to message $W_2^{(j)}$, $j = 1, 2$. Sources s_1 and s_2 will respectively transmit

$$X_{s_1}[t] = Ga[t],$$

$$X_{s_2}[t] = G\left(\frac{\hat{h}(s_1, u_1)}{\hat{h}(s_2, u_1)}b_1[t] + \frac{\hat{h}(s_1, u_2)}{\hat{h}(s_2, u_2)}b_2[t]\right).$$

As in (1), G can be chosen as $G = \beta P^{\frac{1+2\epsilon}{2(2+\epsilon)}}$ to satisfy the sources power constraint, for some constant β. We again drop t from the notation for simplicity. The received signals at u_1, u_2 and u_3 are given by

$$Y_{u_1} = G\hat{h}(s_1, u_1)\left(a + b_1 + \frac{\hat{h}(s_1, u_2)\hat{h}(s_2, u_1)}{\hat{h}(s_2, u_2)\hat{h}(s_1, u_1)}b_2\right) + Z_{u_1},$$

$$Y_{u_2} = G\hat{h}(s_1, u_2)\left(\frac{\hat{h}(s_2, u_2)\hat{h}(s_1, u_1)}{\hat{h}(s_1, u_2)\hat{h}(s_2, u_1)}b_1 + b_2 + a\right) + Z_{u_2},$$

$$Y_{u_3} = G\hat{h}(s_2, u_3)\left(\frac{\hat{h}(s_1, u_1)}{\hat{h}(s_2, u_1)}b_1 + \frac{\hat{h}(s_1, u_2)}{\hat{h}(s_2, u_2)}b_2\right) + Z_{u_3}.$$

Nodes u_1 and u_3 will simply perform amplify-and-forward. More precisely, their transmit signals will be given by

$$\begin{aligned} X_{u_1} &= \frac{\alpha\hat{h}(u_2, d_1)}{\hat{h}(s_1, u_1)\hat{h}(u_1, d_1)}Y_{u_1} \\ &= \frac{\alpha G\hat{h}(u_2, d_1)}{\hat{h}(u_1, d_1)}\left(a + b_1 + \frac{\hat{h}(s_1, u_2)\hat{h}(s_2, u_1)}{\hat{h}(s_2, u_2)\hat{h}(s_1, u_1)}b_2\right) + \alpha\kappa_1 Z_{u_1}, \end{aligned}$$

$$X_{u_3} = \frac{-\alpha\hat{h}(s_2, u_1)\hat{h}(u_2, d_2)}{\hat{h}(s_1, u_1)\hat{h}(s_2, u_3)\hat{h}(u_3, d_2)}Y_{u_3}$$

$$= \frac{\alpha G\hat{h}(u_2,d_2)}{\hat{h}(u_3,d_2)}\left(-b_1 - \frac{\hat{h}(s_1,u_2)\hat{h}(s_2,u_1)}{\hat{h}(s_2,u_2)\hat{h}(s_1,u_1)}b_2\right) + \alpha\kappa_2 Z_{u_3},$$

for some constant α, where κ_1 and κ_2 are functions of the channel gains only. The received signal at node u_2 can be seen as a noisy observation of a point in the set

$$\mathcal{U}_{u_2} = G\hat{h}(s_1,u_2)\left\{Tx_1 + x_2 : x_1 \in \mathcal{U}, x_2 \in \mathcal{U} + \mathcal{U}\right\},$$

for $x_1 = b_1$ and $x_2 = a + b_2$, where $T = \frac{\hat{h}(s_2,u_2)\hat{h}(s_1,u_1)}{\hat{h}(s_1,u_2)\hat{h}(s_2,u_1)}$. Next we notice that $\hat{h}(u_2,d_1)$, $\hat{h}(u_3,d_2)$, $\hat{h}(u_2,d_2)$ and $\hat{h}(u_3,d_1)$ are each a polynomial on the channel gains h_e of the original network with only coefficients 1. From the description of the network in the proof of Lemma 2.5, we see that $u_2 = v_2$ is on a path P_{z,v_1} such that $P_{z,v_1} \cap P_{s_1,d_1} = \emptyset$ and $z \in P_{s_2,d_2}$. Therefore, there must exist two disjoint paths P_{s_1,u_1} and P_{s_2,u_2}. This implies that the determinant

$$\begin{vmatrix} \hat{h}(s_1,u_1) & \hat{h}(s_2,u_1) \\ \hat{h}(s_1,u_2) & \hat{h}(s_2,u_2) \end{vmatrix} = \hat{h}(s_2,u_2)\hat{h}(s_1,u_1) - \hat{h}(s_1,u_2)\hat{h}(s_2,u_1)$$

is non-zero for almost all channel gain values. As we argued before, this implies that, for almost all channel gain values, after mapping the received signal to the nearest point in $\mathcal{U}_{u_2}$, u_2 can decode b_1 and $a + b_2$ with probability at least $1 - \exp(-\delta_4 P^{\epsilon})$, for some positive constant δ_4. The transmit signal at u_2, will then be

$$X_{u_2} = -\alpha G\hat{b}_1,$$

where $\hat{b}_1$ is the output of the hard-decoding performed by u_2. We again notice that α can be chosen independently of P, for P sufficiently large, guaranteeing that the power constraints at u_1, u_2 and u_3 are simultaneously satisfied. The received signal at destination d_1 is given by

$$\begin{aligned} Y_{d_1} &= \hat{h}(u_1,d_1)X_{u_1} + \hat{h}(u_2,d_1)X_{u_2} + Z_{d_1} \\ &= \alpha G\hat{h}(u_2,d_1)\left(a + b_1 - \hat{b}_1 + \frac{\hat{h}(s_1,u_2)\hat{h}(s_2,u_1)}{\hat{h}(s_2,u_2)\hat{h}(s_1,u_1)}b_2\right) + Z_{d_1}^{\text{eff}}, \end{aligned}$$

where $Z_{d_1}^{\text{eff}} = \alpha\hat{h}(u_1, d_1)K_1 Z_{u_1} + Z_{d_1}$. The received signal at d_2 is given by

$$\begin{aligned} Y_{d_2} &= \hat{h}(u_2, d_2)X_{u_2} + \hat{h}(u_3, d_2)X_{u_3} + Z_{d_2} \\ &= \alpha G\hat{h}(u_2, d_2)\left(-b_1 - \hat{b}_1 - \frac{\hat{h}(s_1, u_2)\hat{h}(s_2, u_1)}{\hat{h}(s_2, u_2)\hat{h}(s_1, u_1)}b_2\right) + Z_{d_2}^{\text{eff}}, \end{aligned}$$

where $Z_{d_2}^{\text{eff}} = \alpha\hat{h}(u_3, d_2)K_2 Z_{u_3} + Z_{d_2}$. Notice that with probability at least $1 - \exp(-\delta_4 P^{\epsilon})$, Y_{d_1} and Y_{d_2} are given by

$$Y_{d_1} = \alpha G\hat{h}(u_2, d_1)\left(a + \frac{\hat{h}(s_1, u_2)\hat{h}(s_2, u_1)}{\hat{h}(s_2, u_2)\hat{h}(s_1, u_1)}b_2\right) + Z_{d_1}^{\text{eff}},$$

$$Y_{d_2} = \alpha G\hat{h}(u_2, d_2)\left(-2b_1 - \frac{\hat{h}(s_1, u_2)\hat{h}(s_2, u_1)}{\hat{h}(s_2, u_2)\hat{h}(s_1, u_1)}b_2\right) + Z_{d_2}^{\text{eff}}.$$

The destinations will first perform a hard-decoding, similar to the one performed by u_2. If we assume that the decoding at node u_2 was correct, the signal received by d_1 is a noisy version of a point in the set

$$\mathcal{U}_{d_1} = \alpha G\hat{h}(u_2, d_1)\left\{x_1 + T^{-1}x_2 : x_1, x_2 \in \mathcal{U}\right\},$$

for $x_1 = a$ and $x_2 = b_2$. Thus, it can be shown that, for almost all channel gain values, d_1 can decode a (and also b_2) with probability of error smaller than $\exp(-\delta_5 P^{\epsilon})$, for some positive constant δ_5. Again assuming that the decoding at node u_2 was correct, the signal received by d_2 is a noisy version of a point in the set

$$\mathcal{U}_{d_2} = \alpha G\hat{h}(u_2, d_2)\left\{x_1 + Tx_2 : x_1 \in 2\,\mathcal{U}, x_2 \in \mathcal{U}\right\},$$

for $x_1 = -2b_1$ and $x_2 = b_2$. For almost all channel gain values, d_2 can decode b_1 and b_2 with probability at least $\exp(-\delta_6 P^{\epsilon})$, for some positive constant δ_6 (if the decoding at u_2 was also correct). Therefore, destination d_1 obtains $a[t]$ and destination d_2 obtains both $b_1[t]$ and $b_2[t]$, for $t = 1, ..., n$, and, by applying typicality-based decoders, the messages W_1, $W_2^{(1)}$ and $W_2^{(2)}$ can be decoded by their intended destinations. By following the same steps as in case (1), our codebook can have rate

$$R = (1 - 4\exp(-\delta_{\min} P^{\epsilon}))\frac{1-\epsilon}{2+\epsilon}\frac{\log P}{2} - 4,$$

where $\delta_{\min} = \min(\delta_4, \delta_5, \delta_6)$. Thus, each of the messages carries $\frac{1-\epsilon}{2+\epsilon}$ degrees-of-freedom. Since ϵ can be chosen arbitrarily small, we conclude that $(1/2, 1) \in \mathbf{D}$, and $\mathbf{D}$ is as given in (2.14).

For networks in Case C2, as shown in [55], using mutual information inequalities similar to those in Section 2.5.2, one can always obtain either the bound $D_1 + 2D_2 \leq 1$ or the bound $2D_1 + D_2 \leq 1$. Therefore, achievability schemes based on two modes of operation, similar to those described in Section 2.5.1, are sufficient to show that the degrees-of-freedom region must be either

$$\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \leq 1, D_1 + 2D_2 \leq 2\right\} \tag{2.17}$$

$$\mathbf{D} = \left\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \leq 1, 2D_1 + D_2 \leq 2\right\}, \tag{2.18}$$

as depicted in Fig. 2.4(d) and Fig. 2.4(e). This concludes the derivation of the degrees-of-freedom region of all two-unicast layered Gaussian networks.

In order to state the result in a concise way, we will use the notion of disjoint paths with (s_i, d_i)-manageable interference (see Definition 2.9). Notice that if two disjoint paths P_{s_1,d_1} and P_{s_2,d_2} have interference that is both (s_1, d_1)-manageable and (s_2, d_2)-manageable, they do not necessarily have manageable interference, since the latter requires a single set S for which $n_1(G[S]) \neq 1$ and $n_2(G[S]) \neq 1$. We will describe case C1 in terms of (s_i, d_i)-manageable interference through the following claim.

Claim 2.4 *A network $\mathcal{N}$ is in case C1 and not in cases (A) and (B) if and only if it has disjoint paths P_{s_1,d_1} and P_{s_2,d_2} with interference that is not manageable, but is both (s_1, d_1)-manageable and (s_2, d_2)-manageable.*

Proof: By definition, if $\mathcal{N}$ is in case C1, we have wlog $n_1(V) \geq 2$, $n_1^0 = 1$, $n_2(V) = 1$ and $n_2^0 = 0$, which implies that P_{s_1,d_1} and P_{s_2,d_2} have (s_1, d_1)-manageable interference. Moreover, we see from Section 2.5.2 (and Figure 2.15(a)) that $P_{s_1,d_1} \cup P_{s_2,d_2} \subset V \setminus \{v_2\}$ and, from property P1, we have $n_2(V \setminus \{v_2\}) = 0$, which implies that the interference between P_{s_1,d_1} and P_{s_2,d_2} is also (s_2, d_2)-manageable. Next we argue that, conversely, if P_{s_1,d_1} and P_{s_2,d_2} do not have manageable interference but have interference that is both (s_1, d_1)-manageable and (s_2, d_2)-manageable, then we must be in case C1. Since P_{s_1,d_1} and P_{s_2,d_2} do not have manageable interference, we must have either $n_1^0 = 1$ or $n_2^0 = 1$. If $n_i^0 = 1$, for $i = 1$ or 2, then since P_{s_1,d_1} and P_{s_2,d_2} have (s_i, d_i)-manageable interference, we must have $n_i(V) \geq 2$. Therefore, we cannot have $n_1^0 = n_2^0 = 1$, or else we would have $n_1(V) \geq 2$ and $n_2(V) \geq 2$. We conclude that the only possible case is $n_i(V) \geq 2$, $n_i^0 = 1$, $n_{\bar{i}}(V) = 1$ and $n_{\bar{i}}^0 = 0$, and we are in case C1 wlog. ■

As shown in [55], we can similarly characterize the networks in case C2:

Claim 2.5 *A network $\mathcal{N}$ that is in case C2 and not in cases (A), (B) and C1, contains paths Z_{s_k,d_k}, Q_{s_k,d_k} and $P_{s_{\bar{k}},d_{\bar{k}}}$, for $k \in \{1, 2\}$ such that Q_{s_k,d_k} and $P_{s_{\bar{k}},d_{\bar{k}}}$ are disjoint and have (s_1, d_1)-manageable interference, Z_{s_k,d_k} and $P_{s_{\bar{k}},d_{\bar{k}}}$ are disjoint and have (s_2, d_2)-manageable interference. If $k = 1$, the degrees-of-freedom region is given by (2.17), and if $k = 2$, by (2.18).*

2.7 Extensions and Open Questions

In this chapter, we explored the degrees of freedom of two-unicast layered Gaussian networks. Our result shows that, in terms of the sum degrees of freedom, there are essentially three categories of such networks, which can be described based on the graph-theoretic notions of paths with manageable interference and omniscient nodes. This result can be extended in many directions and raises many questions for future work that currently remain open.

2.7.1 Non-Layered Two-Unicast Networks

One natural direction for further research is to relax the assumption of layered network topologies. We point out that, while this assumption is used mainly to simplify the problem, it is not entirely artificial. Since the layered topology simplifies the analysis and the implementation of coding schemes, it is desirable in practice, and it can actually be emulated in practical contexts by having the transmitters on each layer transmit on a different frequency band, which allows us to assume that the links only exist between consecutive layers. Moreover, a layered structure can also arise from the scheduling of the transmitting nodes in a wireless network with half-duplex nodes. In this context, each hop would capture which nodes are transmitting and which nodes are receiving at a given time slot, and the same node could appear in multiple layers, since they may be transmitting at multiple time-slots.

When non-layered networks are considered, a new issue that arises is that not all source destination paths have the same length; thus, interference may occur not only between signals originated at different sources, but also between signals originated at different times. Therefore, in order to perform interference cancelation, for example, one needs to make sure that the two canceling signals correspond to the same time-version of the source signal. For non-layered networks, other techniques such as signal delaying and backward decoding must be used to achieve the degrees of freedom, and the problem becomes significantly more difficult. Steps in that direction were taken in [26] and [25], where the authors considered the 2×2×2 topology first with interfering relays and then with arbitrary links between non-consecutive layers, which results in a class of non-layered two-unicast networks. Roughly speaking, their main result is that for this class, as long as there is no link directly connecting s_1 and d_2 or s_2 and d_1, the cut-set bound of 2 degrees of freedom are achievable. Otherwise, the network only admits one degree of freedom. The outer bound required in [25] to establish that only one degree of freedom when $(s_1, d_2) \in E$ or $(s_2, d_1) \in E$ can in fact be generalized, yielding a notion of omniscient node for (non-layered) acyclic two-unicast networks:

Lemma 2.6 *If a two-unicast (non-layered) acyclic Gaussian network contains a node v that is a $(\{s_1, s_2\}, d_i)$ cut and a node $u \in \mathcal{I}(v) \cup \{v\}$ that is a $(s_{\bar{i}}, \{d_1, d_2\})$ cut, for $i = 1$ or 2, then $D_\Sigma \leq 1$.*

Proof: We prove only the case $u \in \mathcal{I}(v)$. The case $u = v$ is easier to prove. We assume wlog that the removal of v disconnects d_1 from $\{s_1, s_2\}$ and the removal of u disconnects s_2 from $\{d_1, d_2\}$. We let $M = \{w : u \rightsquigarrow w, w \neq u\}$ and $A = \{w : s_2 \not\rightsquigarrow w\}$. Using Fano's inequality, we have

$$\begin{aligned}
n(R_1 + R_2 - \epsilon_n) = I(X^n_{s_1}; Y^n_{d_1}) + I(X^n_{s_2}; Y^n_{d_2}) &\overset{(i)}{\leq} I(X^n_{s_1}; Y^n_v) + I(X^n_{s_2}; Y^n_{d_2}) \\
&\overset{(ii)}{\leq} I(X^n_{s_1}; Y^n_v) + I(X^n_{s_2}; Y^n_{d_2} | X^n_A) \\
&\leq I(X^n_{s_1}; Y^n_v) + I(X^n_u; Y^n_{d_2} | X^n_A) + I(X^n_{s_2}; Y^n_{d_2} | X^n_A, X^n_u) \\
&\overset{(iii)}{=} I(X^n_{s_1}; Y^n_v) + I(X^n_u; Y^n_{d_2} | X^n_A) \\
&\overset{(iv)}{\leq} I(X^n_A; Y^n_v) + I(X^n_u; Y^n_M | X^n_A)
\end{aligned}$$

$$= h(Y_v^n) - h(Y_v^n|X_A^n) + h(Y_M^n|X_A^n) - h(Y_M^n|X_A^n, X_u^n)$$

$$= h(Y_v^n) + h(Y_{M\setminus\{v\}}^n|X_A^n, Y_v^n) - h(Y_M^n|X_A^n, X_u^n)$$

$$\stackrel{(v)}{\leq} \frac{n}{2}\log P + n\kappa_1 + h(Y_{M\setminus\{v\}}^n|X_A^n, Y_v^n) - h(Y_M^n|X_A^n, X_u^n) \tag{2.19}$$

where (i) follows because the removal of v disconnects d_1 from both sources, thus we have $X_{s_1}^n \leftrightarrow Y_v^n \leftrightarrow Y_{d_1}^n$; (ii) follows because X_A^n is indepencent of $X_{s_2}^n$; (iii) follows because, since the removal of u disconnects d_2 from s_2, the removal of A and u disconnects d_2 from both sources, and we have $X_{s_2}^n \leftrightarrow (X_u^n, X_A^n) \leftrightarrow Y_{d_2}^n$; (iv) follows since $s_1 \in A$ and $d_2 \in M$ (or else $s_2 \not\rightsquigarrow d_2$); and (v) follows by upper-bounding $h(Y_{d_1}^n)$, where κ_1 is a constant. Next, we show that the last two terms can also be upper-bounded by n times a constant. First we consider the following claims.

Claim 2.6 $(\mathcal{I}(v) \setminus \{u\}) \subset ((M \setminus \{v\}) \cup A)$

Proof of Claim 2.6: Suppose by contradiction that there exists a node $w \in \mathcal{I}(v)\setminus\{u\}$, such that $w \notin M \setminus \{v\} \cup A$. Since $w \neq v$, we must have $w \notin M \cup A$ which implies $u \not\rightsquigarrow w$ and $s_2 \rightsquigarrow w$. Since $w \in \mathcal{I}(v)$ implies $w \rightsquigarrow v$ and since $v \rightsquigarrow d_1$, we must have $w \rightsquigarrow d_1$. However, this implies the existence of a path P_{s_2,d_1} not containing u, which is a contradiction. ∎

Claim 2.7 *Let $\mathcal{I}(M) = \cup_{w\in M}\mathcal{I}(w)$. Then $\mathcal{I}(M) \subset (M \cup A \cup \{u\})$.*

Proof of Claim 2.7: Suppose by contradiction that there exists a node $z \in \mathcal{I}(M)$, such that $z \notin M\cup A\cup\{u\}$. Then we must have $u \not\rightsquigarrow z$ and $s_2 \rightsquigarrow z$, and, consequently, a path $P_{s_2,z}$ not containing u. Since $z \in \mathcal{I}(M)$, we must have a node $w \in M$ such that $(z, w) \in E$. Since any node is connected to at least one destination we must have a path P_{w,d_k}, for $k = 1$ or $k = 2$. Moreover, $u \notin P_{w,d_k}$, since $u \rightsquigarrow w$, and the network is acyclic. Thus, we have a path $P_{s_2,z} \oplus (z, w) \oplus P_{w,d_k}$ not containing u, which is a contradiction. ■

Claim 2.8 *For all $w \in M$ and $z \in A$ we have $w \not\rightsquigarrow z$*

Proof of Claim 2.8: Suppose by contradiction that there exist $w \in M$ and $z \in A$, such that $w \rightsquigarrow z$. Then, since $w \in M$, $u \rightsquigarrow w$. Moreover, since $s_2 \rightsquigarrow u$, we conclude that $s_2 \rightsquigarrow z$, which contradicts the fact that $z \in A$. ■

We can now upper bound the second term in (2.19) as

$$\begin{aligned}
&h(Y^n_{M\setminus\{v\}}|X^n_A, Y^n_v) = \sum_{j=1}^{n} h(Y_{M\setminus\{v\}}[j]|X^n_A, Y^n_v, Y^{j-1}_{M\setminus\{v\}}) \\
&\overset{(i)}{=} \sum_{j=1}^{n} h(Y_{M\setminus\{v\}}[j]|X^n_A, Y^n_v, Y^{j-1}_{M\setminus\{v\}}, X_{M\setminus\{v\}}[j]) \\
&\overset{(ii)}{=} \sum_{j=1}^{n} h(\{Y_w[j] : w \in M \setminus \{v\}\}|X^n_A, Y^n_v, Y^{j-1}_{M\setminus\{v\}}, X_{M\setminus\{v\}}[j], h_{u,v}X_u[j] + Z_v[j]) \\
&\overset{(iii)}{=} \sum_{j=1}^{n} h(\{Z_w[j] - \frac{h_{u,w}}{h_{u,v}} Z_v[j] : w \in M \setminus \{v\}\}|X^n_A, Y^n_v, Y^{j-1}_{M\setminus\{v\}}, X_{M\setminus\{v\}}[j], h_{u,v}X_u[j] + Z_v[j]) \\
&\leq \sum_{j=1}^{n} h(\{Z_w[j] - \frac{h_{u,w}}{h_{u,v}} Z_v[j] : w \in M \setminus \{v\}\}) \leq n\kappa_2
\end{aligned} \tag{2.20}$$

where (i) follows because $X_{M\setminus\{v\}}[j]$ is a deterministic function of $Y^{j-1}_{M\setminus\{v\}}$; (ii) follows because $\mathcal{I}(v)\setminus\{u\} \subset M\setminus\{v\} \cup A$ from Claim 1, therefore, by using X^n_A and $X_{M\setminus\{v\}}[j]$, it is possible to subtract $\sum_{w\in\mathcal{I}(v)\setminus\{u\}} h_{w,v}X_u[j]$ from $Y_v[j]$ to obtain $h_{u,v}X_u[j] + Z_v[j]$; (iii) follows since, from Claim 2.7, we can subtract $\sum_{z\in\mathcal{I}(w)\setminus\{u\}} h_{z,w}X_z[j] + \frac{h_{u,w}}{h_{u,v}}(h_{u,v}X_u[j] + Z_v[j])$ from $Y_u[j]$, for each $w \in M \setminus \{v\}$ (we assume that $h_{u,w} = 0$ if $(u, w) \notin E$).

For the third term in (2.19), we have

$$h(Y^n_M|X^n_A, X^n_u) = \sum_{j=1}^{n} h(Y_M[j]|X^n_A, X^n_u, Y^{j-1}_M)$$

$$\stackrel{(i)}{=} \sum_{j=1}^{n} h(\{Y_w[j] : w \in M\} | X_A^n, X_u^n, Y_M^{j-1}, X_M[j])$$
$$\stackrel{(ii)}{=} \sum_{j=1}^{n} h(\{Z_w[j] : w \in M\} | X_A^n, X_u^n, Y_M^{j-1}, X_M[j])$$
$$\stackrel{(iii)}{=} \sum_{j=1}^{n} h(\{Z_w[j] : w \in M\}) = n\kappa_3 \qquad (2.21)$$

where (i) follows because $X_M[j]$ is a deterministic function of Y_M^{j-1}; (ii) follows because $\mathcal{I}(v) \subset M \cup A \cup \{u\}$ from Claim 2.7, therefore, by using X_A^n, $X_M[j]$ and X_u^n, it is possible to subtract $\sum_{z \in \mathcal{I}(w)} h_{z,w} X_z[j]$ from $Y_w[j]$, for each $w \in M$; and (iii) follows since, for all $w \in M$ and all $z \in A$, we have $w \not\rightsquigarrow z$ (from Claim 2.8), and $w \not\rightsquigarrow u$ (because $u \rightsquigarrow w$, $w \neq u$ and the network is acyclic), $Z_w[j]$ is independent of X_A^n, X_u^n, Y_M^{j-1} and $X_M[j]$ (since $Z_u[j]$ occurs "after" Y_M^{j-1} and $X_M[j]$).

By combining (2.19), (2.20) and (2.21), we obtain

$$n(R_1 + R_2 - \epsilon_n) \leq \frac{n}{2} \log P + n(\kappa_1 + \kappa_2 - \kappa_3), \qquad (2.22)$$

and we conclude that $D_\Sigma \leq 1$. ■

Notice that, if we extend Definition 2.6 verbatim to the non-layered case, Lemma 2.6 implies that if a non-layered network contains an omniscient node, $D_\Sigma = 1$. The converse of this statement; i.e., $D_\Sigma > 1$ if no omniscient node exists, holds in the layered case, as we showed in this chatper. Whether it holds also in the non-layered acyclic case is an open question:

Open Question 2.1 *For non-layered acyclic two-unicast networks, is the existence of an omniscient node a necessary and sufficient condition for $D_\Sigma = 1$?*

More generally, one could define the notion of an omniscient node for K-unicast networks as follows:

Definition 2.10 *Let $A \subset \{1, ..., K\}$ and $A^c = \{1, ..., K\} - A$. In a K-unicast network, if node v is a $(\mathcal{S}, \{d_i : i \in A\})$ cut and some node $u \in \mathcal{I}(v) \cup \{v\}$ is a $(\{s_i : i \in A^c\}, \mathcal{D})$ cut, v is an omniscient node.*

By following similar steps to those in Lemma 2.6 after essentially grouping the sources (resp. destinations) with indices in A and in A^c as single sources (resp. destinations), one can show that the existence of an omniscient node in a K-unicast network implies that $D_\Sigma = 1$. Whether the converse direction is also true is currently unknown.

Open Question 2.2 *For non-layered acyclic K-unicast networks, is the existence of an omniscient node a necessary and sufficient condition for $D_\Sigma = 1$?*

2.7.2 Related Work

One can also consider extending the results presented in this chapter by relaxing the requirement that source s_i only has a message for destination d_i, for $i = 1, 2$. This was recently done in [65], where the authors considered networks with two sources and two destinations where each source has a message to each destination (for a total of four messages). Interestingly, it was shown that the sum degrees-of-freedom can also take values 4/3 and 5/3, in addition to the values 1, 3/2 and 2 that are possible in the setup considered in chapter.

Another research direction concerns modifying the assumptions on the availability of Channel State Information (CSI). If the channel gains in the network are varying fast (i.e., in a fast fading scenario), the assumption that the channel gain values are known instantaneously at all nodes in the network is unjustified. A more reasonable assumption would then be that a transmitter node obtains the CSI for its channel with some delay. In [63], $2 \times 2 \times 2$ wireless networks with time-varying channels were considered under the assumption that the sources obtain the CSI for the entire network within some finite delay while the relays have no CSI. It was shown that, under these assumptions, only 4/3 degrees of freedom are achievable. Later, in [67], layered two-unicast networks (as the ones considered in this chapter) were studied in the time-varying scenario under the assumption that CSI is available at each transmitter with a delay while it is available intantaneously at receivers. Interestingly, the main result is analogous to the results in this chapter: as long as there is no omniscient node in the network (for a modified notion of omniscient node, which captures the lack of instantaneous CSI), one can achieve more than 1 sum degree of freedom in the networks.

Two-unicast networks have also been studied in the presence of feedback channels. In [63], it is shown that in a $2 \times 2 \times 2$ wireless network where one relay obtains output feedback from one of the destinations has 4/3 sum degrees of freedom, even if all remaining nodes have no CSI nor feedback. In [66], two-unicast layered networks are considered under the linear deterministic setup from [68], but with destination-to-source feedback. It is shown that feedback can increase the capacity region in the cases where, without feedback, the capacity region is asymmetric; thus, feedback can be seen as a way to balance the network resources. Finally, security issues can also be brought into the problem. In [69], two-unicast layered wireless networks are considered under the additional constraint the source messages should not be decodable at the unintended receiver. By combining interference management techniques similar to those used in this chapter with cooperative jamming and neutralization techniques, it is shown that the sum *secure* degrees of freedom admit five values: 0, 2/3, 1, 3/2 and 2.

In general, extending results from two-unicast to general K-unicast networks is quite challenging, and the results along this research direction are scarcer. One effort along this direction is found in [59], where the authors focus on two-hop networks structured as $K \times K \times K$ wireless networks where K is very large (and edge effects can be neglected) and investigate communication strategies based on rate-splitting and successive interference cancellation at each hop. In [36], networks with K source-destination pairs and K hops with K nodes each were considered under the fast fading scenario. The authors show that, under some assumptions on the joint distribution of the channel gains, K degrees of freedom can be achieved. The main idea is to have the relays forward their received signals at carefully chosen times, so that the signals transmitted by the sources undergo an approximately diagonal end-to-end transformation. In the next chapter, we will focus on two-hop K-unicast networks, in particular the $K \times K \times K$ topology, and introduce new techniques to characterize the degrees of freedom.

CHAPTER 3

TWO-HOP WIRELESS NETWORKS

In the previous chapter, we began our study of multi-hop multi-flow wireless networks by considering two-unicast networks with an arbitrary number of hops. In this chapter, we approach multi-hop multi-flow networks from the opposite direction by considering two-hop K-unicast networks. In particular, we focus on the canonical example provided by the $K \times K \times K$ wireless network, a two-hop wireless network consisting of K sources, K relays, and K destinations.

Our main result is to prove that $K \times K \times K$ wireless networks with fully connected hops (see Fig. 1.2) have K degrees of freedom both in the case of time-varying channel coefficients and in the case of constant channel coefficients (in which case the result holds for almost all values of constant channel coefficients). We prove this result by introducing a new achievability scheme that we call Aligned Network Diagonalization (AND). This scheme allows the data streams transmitted by the sources to undergo a diagonal linear transformation from the sources to the destinations, thus being received free of interference by their intended destination. In addition, we extend our scheme to multi-hop networks with fully connected hops, and multi-hop networks with MIMO nodes, for which the degrees of freedom are also fully characterized.

3.1 Degrees of Freedom of $K \times K \times K$ wireless networks

The $K \times K \times K$ wireless network is made up of K sources $s_1, ..., s_K$, K relays $u_1, ..., u_K$, and K destinations $d_1, ..., d_K$, organized as a two-hop layered network, as shown in Fig. 1.2. We will consider two distinct scenarios.

- **Time-varying channels:** We let the channel gain between source s_i and relay u_j at time t be $h_{s_i,u_j}[t] \in \mathbb{R}$, and the channel gain between relay u_i and destination d_j at time t be $h_{u_i,d_j}[t] \in \mathbb{R}$, for $t = 0, 1, 2, ...$ We assume that $\{h_{s_i,u_j}[t]\}_{t=0}^{\infty}$ and $\{h_{u_i,d_j}[t]\}_{t=0}^{\infty}$ are mutually independent i.i.d random processes each obeying an absolutely continuous probability distribution with finite second moment.
- **Constant channels:** We assume that $h_{s_i,u_j}[t] = h_{s_i,u_j} \in \mathbb{R}$ and $h_{u_i,d_j}[t] = h_{u_i,d_j} \in \mathbb{R}$ remain the same throughout the entire communication period.

In this chapter, we characterize the number of degrees of freedom of a $K \times K \times K$ wireless network, in both the case of time-varying and constant channel coefficients. We have the following two results.

Theorem 3.1 *For a $K \times K \times K$ wireless network with time-varying channels, $D_\Sigma = K$.*

Theorem 3.2 *For a $K \times K \times K$ wireless network with constant channels, $D_\Sigma = K$ for almost all values of the channel gains.*

Since the cut-set outer bound trivially implies that, in both cases, $D_\Sigma \leq K$, we only need to show that K degrees of freedom are achievable. The achievability scheme we propose for both the time-varying channel case and the constant channel case are based on interference alignment techniques. Similar to the approach taken in [24], in the time-varying case our alignment is performed over *time dimensions*, while in the constant channel case, it is performed over *rational dimensions*. More precisely, when we have time-varying channels, the alignment is performed in the vector space created by multiple channel uses, using the framework introduced in [13]. In this case, our construction results in a *linear scheme*, i.e., where relaying functions are restricted to linear transformations. When the channels are constant, on the other hand, alignment over time dimensions is not feasible, and we instead use the real interference alignment framework introduced in [45].

In both cases, each of the K sources will transmit L data streams, each one along a different transmit dimension (be it time or rational). These data streams are aligned at the relays, which allows each relay to decode approximately L linear combinations of the data streams which can then be re-modulated using new transmit directions. These new transmit directions are chosen so that all the interference is cancelled at each destination, and the L data streams from each source arrive at their intended destination along independent directions, which allows perfect decoding with high probability. Since these operations guarantee that, with small probability of error, the LK data streams chosen at all K sources are mapped to LK received directions at the destinations by a diagonal linear transformation, we call the scheme Aligned Network Diagonalization.

The result in Theorems 3.1 and 3.2 has important consequences. Consider a two-hop K-unicast wireless network where, instead of K relays, we have A relays; i.e., a $K \times A \times K$ wireless network. It is easy to see that the cut-set bound states that no more than $\min(K, A)$ degrees of freedom can be achieved. Now, if $A \geq K$, we can ignore $A-K$ of the relays and use aligned network diagonalization to achieve K degrees of freedom. Similarly, if $K > A$, we ignore $K - A$ source-destination pairs to achieve A degrees of freedom. A similar idea can be used in a K-unicast multihop wireless network with J layers and A_j relays in the jth layer ($A_1 = A_J = K$). If we call such a network a $K \times A_2 \times ... \times A_{J-1} \times K$ wireless network, we have the following result.

Corollary 3.1 *For a $K \times A_2 \times ... \times A_{J-1} \times K$ wireless network, $D_\Sigma = \min_{1 \le j \le J} A_j$ in the time-varying case and for almost all channels in the constant-channel case.*

3.2 Aligned Network Diagonalization

In this section we describe the Aligned Network Diagonalization scheme, which achieves K degrees of freedom on the $K \times K \times K$ wireless network. First, in Section 3.2.1, we give a high-level overview and describe the intuition behind it. These ideas are then formalized in Section 3.2.2, where we focus on the time-varying case and describe in detail the operations performed by the sources, relays and destinations.

In the case of constant channel gains, a similar scheme based on asymptotic alignment can be proposed, with the main difference being that the alignment must be performed over rational dimensions, rather than over time. In the literature, there have been several examples of asymptotic alignment schemes that can be applied both over rational dimensions and over time (see, for instance, [13, 24, 45]). Converting from one of these frameworks to the other is relatively straightforward and, hence, we present a summary of the main ideas of AND for constant channels in Section 3.2.3.

3.2.1 Scheme Overview and Intuition

In order to understand the main idea behind AND, we start by considering a different but related problem. Suppose we have a two-hop network with K sources, K destinations, and a single MIMO relay with K (full-duplex) anten-

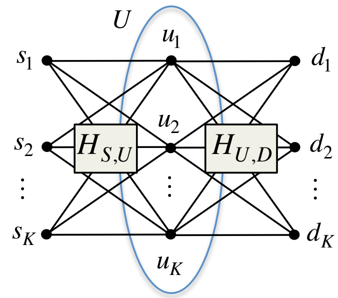


Figure 3.1: Network with a single MIMO relay node.

nas. Equivalently, this setup, illustrated in Fig. 3.1, can be seen as our $K \times K \times K$ wireless network where the K relay nodes are allowed to collaborate in the computation of their transmit signals. This new problem is clearly easier than our original problem, in the sense that any scheme for the $K \times K \times K$ wireless network can be used to achieve the same rates on the network with a single MIMO relay node.

Achieving K degrees of freedom in the setting from Fig. 3.1 is not difficult. As illustrated in Fig. 3.2, a simple linear scheme can be used to *diagonalize* the network. More precisely, if each source s_i transmits a signal $X_{s_i}[t]$ at time t, $i = 1, ..., K$, the received signal at the MIMO relay at time t is a length-K vector $\vec{Y}_U[t] = (Y_{u_1}[t], ..., Y_{u_K}[t])^\dagger$ given by $\vec{Y}_U[t] = H_{S,U}\vec{X}_S[t] + \vec{Z}[t]$, where $\vec{X}_S[t] = (X_{s_1}[t], ..., X_{s_K}[t])^\dagger$. Then, if we assume that the transfer matrices $H_{S,U}$ and $H_{U,D}$ are invertible (which is the case with probability 1 under the distribution assumptions in Section 3.1), the relay can build its transmit signal for time $t+1$ through the linear transformation $\vec{X}_U[t+1] = H_{U,D}^{-1} H_{S,U}^{-1} \vec{Y}_U[t]$. If we let $\vec{Y}_D[t+1] = (Y_{d_1}[t], ..., Y_{d_K}[t])^\dagger$ be the vector of the received signals at the destinations, it is clear that $\vec{Y}_D[t+1] = \vec{X}_S[t] + \vec{\tilde{Z}}[t+1]$, where $\vec{\tilde{Z}}[t+1]$ is the vector of effective noises at the destinations. Therefore, each destination receives its desired source signal plus a Gaussian noise term, meaning that the relay operations essentially diagonalized the end-to-end transfer matrix of the network, since $\vec{Y}_D[t+1] \approx I\vec{X}_S[t]$, where I is the identity matrix. It is easy to see that a slight modification of this scheme can guarantee that the power constraints are satisfied at the relays and can thus be used to show that K degrees of freedom are achievable in this setup.

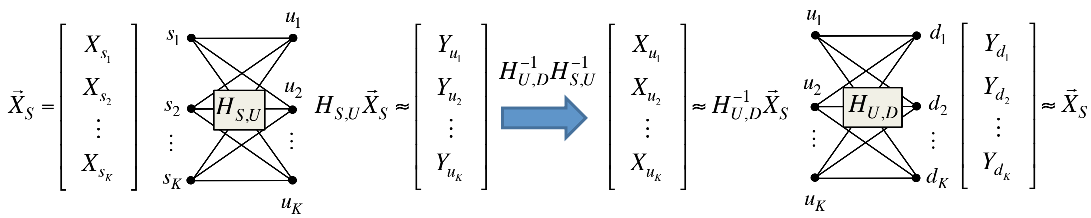


Figure 3.2: Achieving K degrees of freedom on the network with a single MIMO relay.

When we move back to our original problem with K single-antenna relay nodes, we notice that the same scheme cannot be implemented because the relays are not allowed to cooperate in order to compute $\vec{X}_U[t] = H_{U,D}^{-1}H_{S,U}^{-1}\vec{Y}_U[t]$. Therefore, a natural question is whether it is possible to apply the linear transformation $H_{U,D}^{-1}H_{S,U}^{-1}$ *distributedly*. More precisely, can we find functions $f_1, ..., f_K$ such that

$$\begin{bmatrix} f_1(y_1) \\ f_2(y_2) \\ \vdots \\ f_K(y_K) \end{bmatrix} = H_{U,D}^{-1}H_{S,U}^{-1} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix} \tag{3.1}$$

for all $(y_1, ..., y_K) \in \mathbb{R}^K$? In the case of general transfer matrices $H_{U,D}$ and $H_{S,U}$, the answer is no. In fact, if $H_{U,D}^{-1}H_{S,U}^{-1}$ is not diagonal, it is easy to see that at least one component of $H_{U,D}^{-1}H_{S,U}^{-1}(y_1, ..., y_K)^\dagger$ depends on multiple components of $(y_1, ..., y_K)$.

Therefore, in order to pursue our objective of diagonalizing the network with distributed relays, we must consider a more general question than the aforementioned one. In particular, we will reformulate the question of whether the network can be diagonalized by bringing in the channels' time variation, and by including linear transformations at each source and at each destination. Since our channels are time-varying, we notice that, if each hop of the network is used for d consecutive time steps, we can view both the transmit signals and the received signals of the network as length-d vectors. The transfer matrix of the first hop is now given by

$$H_{S,U} = \begin{bmatrix} H_{s_1,u_1} & H_{s_2,u_1} & \cdots & H_{s_K,u_1} \\ H_{s_1,u_2} & H_{s_2,u_2} & \cdots & H_{s_K,u_2} \\ \vdots & \vdots & \ddots & \vdots \\ H_{s_1,u_K} & H_{s_2,u_K} & \cdots & H_{s_K,u_K} \end{bmatrix},$$

$$\text{where } H_{s_i,u_j} = \begin{bmatrix} h_{s_i,u_j}[0] & 0 & \cdots & 0 \\ 0 & h_{s_i,u_j}[1] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{s_i,u_j}[d-1] \end{bmatrix},$$

for $1 \leq i, j \leq K$. The transfer matrix of the second hop, $H_{U,D}$ can be similarly built. In this new setting, we have transfer matrices constituted of diagonal blocks, and we could restate the goal in (3.1) by having each y_i be a length-d column vector. In this new setting, by assuming that each relay u_i applies a linear transformation to its vector of d received signals, the diagonalization problem becomes the problem of finding block diagonal matrices A_U (with $d \times d$ blocks A_{u_i}, for $i = 1, ..., K$), A_S (with $d \times d'$ blocks A_{s_i}, for $i = 1, ..., K$) and A_D (with $d' \times d$ blocks A_{d_i}, for $i = 1, ..., K$) such that

$$A_D H_{U,D} A_U H_{S,U} A_S = I, \tag{3.2}$$

where $A_S \in \mathbb{R}^{Kd\times Kd'}$, $A_U \in \mathbb{R}^{Kd\times Kd}$ and $A_D \in \mathbb{R}^{Kd'\times Kd}$ correspond to the linear transformations applied by the sources, relays and destinations. Notice that the identity matrix I is $Kd' \times Kd'$, and the parameter d' regulates how much information the sources are transmitting. Our goal is to solve the problem specified by (3.2) for $d' \leq d$ large enough so that $d'/d \approx 1$.

In this work, our main contribution is to show that the problem in (3.2), with probability 1 over the channel realizations, indeed admits a sequence of solutions parameterized by d, with the property that $d'/d \to 1$ as $d \to \infty$. The scheme that provides this solution, which we call Aligned Network Diagonalization, can be roughly described as follows. The source matrices A_{s_i}, $i = 1, ..., K$, are all chosen to be the same $d \times d'$ matrix A_{s_0}, whose columns are all of the form

$$T_{s_{1,1},s_{1,2},...,s_{K,K}} = \prod_{1\leq i,j\leq K} H_{s_i,u_j}^{s_{i,j}} \mathbb{1}, \tag{3.3}$$

for some nonnegative integers $s_{i,j}$, $1 \leq i, j \leq K$, where $\mathbb{1}$ is a column vector with all entries equal to 1. It is then not difficult to see that the result of

$$H_{S,U}A_S = \begin{bmatrix} H_{s_1,u_1} & \cdots & H_{s_K,u_1} \\ \vdots & \ddots & \vdots \\ H_{s_1,u_K} & \cdots & H_{s_K,u_K} \end{bmatrix} \begin{bmatrix} A_{s_0} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{s_0} \end{bmatrix}$$

is a $Kd \times Kd'$ matrix with $d \times d'$ blocks whose columns are again of the form in (3.3). The key idea in the AND scheme is in the design of the relaying matrices A_{u_i}. Once again, we will choose a single matrix A_{u_0} and let $A_{u_i} = A_{u_0} = \tilde{\mathbf{T}}\mathbf{T}^{-1}$ for $i = 1, ..., K$, where $\mathbf{T}$ is a matrix whose columns are the vectors of the form (3.3) that appear in any of the blocks in $H_{S,U}A_S$ and $\tilde{\mathbf{T}}$ is obtained from $\mathbf{T}$ by replacing each column $T_{s_{1,1},s_{1,2},...,s_{K,K}}$ as given in (3.3) with the column

$$\tilde{T}_{s_{1,1},s_{1,2},...,s_{K,K}} = \prod_{1\leq i,j\leq K} B_{i,j}^{s_{i,j}} \mathbb{1},$$

for diagonal matrices $B_{i,j}$ to be defined. The key observation is that the result of any vector $T_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$, as given in (3.3), undergoing the transformation A_{u_0} is

$$\begin{aligned} \tilde{\mathbf{T}}\,\mathbf{T}^{-1}T_{s_{1,1},s_{1,2},\ldots,s_{K,K}} &= \tilde{\mathbf{T}}\,\mathbf{T}^{-1}\,\mathbf{T}\,\mathbf{e}_{s_{1,1},s_{1,2},\ldots,s_{K,K}} \\ &= \tilde{T}_{s_{1,1},s_{1,2},\ldots,s_{K,K}}, \end{aligned} \tag{3.4}$$

where $\mathbf{e}_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$ is a standard basis vector with the 1 at the entry corresponding to the position of the column $T_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$ in $\mathbf{T}$. Therefore, the transformation A_{u_0} applied by each relay can be understood as replacing each "direction" $T_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$ with a new direction $\tilde{T}_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$. Each matrix $B_{i,j}$ is chosen as *what* $T_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$ *would have been if* $H_{S,U} = H_{U,D}^{-1}$. This essentially makes it look like the first hop of the network is $H_{U,D}^{-1}$, rather than $H_{S,U}$. More precisely, we have $A_U H_{S,U} A_S = H_{U,D}^{-1}\tilde{A}_S$, where $\tilde{A}_S$ is obtained by replacing each column $T_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$ in one of the blocks of A_S with $\tilde{T}_{s_{1,1},s_{1,2},\ldots,s_{K,K}}$. This reduces the end-to-end transformation in (3.2) to

$$A_D H_{U,D} A_U H_{S,U} A_S = A_D H_{U,D} H_{U,D}^{-1}\tilde{A}_S = A_D \tilde{A}_S.$$

Finally, since $\tilde{A}_S$ can be seen to admit a block diagonal left inverse, we can set A_D to be this matrix and obtain our desired end-to-end diagonalization. In the next section, we describe this scheme in more detail. In particular, several issues such as power constraints and invertibility of the matrices are properly addressed, and the fact that we can choose d' and d sufficiently large such that d'/d approaches 1 is proved.

The intuition behind AND can also be understood through the lenses of the asymptotic alignment schemes in [13, 24]. The column vector in (3.3) corresponds to distinct signal directions used by the transmitters. These transmit directions are chosen so that each relay receives, along each direction, a superposition of K symbols from distinct transmitters. Similar to [13], the specific superpositions of symbols received by the relays along each direction do not depend on the specific channel gain values, and they can be recovered by the relays with high probability. The relaying operation described in (3.4) can then be understood as taking each of these superpositions and transmitting it along a new direction, chosen according to the inverse of the second hop channel matrix. This essentially creates a mapping from the symbols transmitted at each direction at the relays to the symbols received at each direction at the destinations which is the inverse of the mapping over the first hop.

3.2.2 Aligned Network Diagonalization for the Case of Time-Varying Channels

In this section, we describe in detail the achievability scheme that achieves K degrees of freedom on the $K \times K \times K$ wireless network when the channels are time-varying. We first describe the encoding at the sources, followed by the relaying operations and the decoding operations.

Encoding at the sources:

Each source s_i starts by breaking its message W_i into L submessages. Each of the submessages will be encoded in a separate data stream, using Gaussian random codebooks with codewords of length n and entries drawn as $\mathcal{N}(0, P)$. We let

$$T_{s_{11},s_{12},\ldots,s_{KK}}[t] = \prod_{\substack{1\leq i\leq K\\ 1\leq j\leq K}} h_{s_i,u_j}[t]^{s_{ij}}, \tag{3.5}$$

and $\Delta_N = \{0, \ldots, N-1\}^{K^2}$, and we define the set of transmit directions for the sources at time t to be

$$\mathcal{T}_N[t] = \{T_{s_{11},s_{12},...,s_{KK}}[t] : (s_{11}, s_{12}, ..., s_{KK}) \in \Delta_N\}, \tag{3.6}$$

for some arbitrary N. This selection of directions is similar in flavor to the directions chosen in the Interference Alignment scheme introduced in [13]. Notice that the number of transmit directions (which is also the number of data streams) is $L = |\mathcal{T}_N[t]| = |\Delta_N| = N^{K^2}$. To simplify the notation we will let $\vec{s}$ be a vector of indices $(s_{11}, s_{12}, ..., s_{KK})$ and write $T_{\vec{s}}$.

Communication will take place over a block of nd time steps, where $d \triangleq (N+1)^{K^2}$. The $(m+1)$th symbol of the codeword associated to the submessage of stream $\vec{s} \in \Delta_N$ of source s_i will be written as $c_{i,\vec{s}}[m]$, for $0 \leq m \leq n-1$. At time $t = md + j$ for $m \in \{0, ..., n-1\}$ and $j \in \{0, ..., d-1\}$, source s_i will thus transmit

$$X_{s_i}[t] = \gamma \sum_{\vec{s} \in \Delta_N} T_{\vec{s}}[t]\ c_{i,\vec{s}}[m].$$

The constant γ is chosen so that the transmit power

$$E\left[X_{s_i}[t]^2\right] = \gamma^2 E\left[\left(\sum_{\vec{s} \in \Delta_N} T_{\vec{s}}[t]\ c_{i,\vec{s}}[m]\right)^2\right] = \gamma^2 P \sum_{\vec{s} \in \Delta_N} E\left[T_{\vec{s}}[t]^2\right] \tag{3.7}$$

does not exceed P. In (3.7), we used the fact that the $c_{i,\vec{s}}$ were independently generated. Notice that γ does not depend on P or t and can be chosen strictly positive, since the fact that the channel gains are independent and have finite variances implies $E\left[T_{\vec{s}}[t]^2\right] < \infty$ for all $\vec{s}$.

Relaying operations:

The received signal at relay u_j at time $t = md + j$ can be written as

$$Y_{u_j}[t] = \gamma \sum_{\vec{s} \in \Delta_N} T_{\vec{s}}[t] \left(\sum_{i=1}^{K} h_{s_i,u_j}[t] c_{i,\vec{s}}[m] \right) + Z_{u_j}[t]. \tag{3.8}$$

Even though writing the received signal as in (3.8) does not emphasize the alignment that occurs at the relays, it will still be a useful representation of the received signal. To capture the alignment, we consider rearranging the terms in the summation in (3.8) by viewing it as a polynomial on the variables $h_{s_i,u_j}[t]$, for $1 \leq i, j \leq K$, where the coefficients are given by sums of $c_{i,\vec{s}}$ terms. It can then be seen that the actual set of received directions at each relay is a subset of $\mathcal{T}_{N+1}[t]$, and the received signal at relay u_j at time t can be alternatively written as

$$Y_{u_j}[t] = \gamma \sum_{\vec{s} \in \Delta_{N+1}} T_{\vec{s}}[t]\ a_{j,\vec{s}}[m] + Z_{u_j}[t], \tag{3.9}$$

where $a_{j,(s_{11},\ldots,s_{KK})}[m] = \sum_{i=1}^{K} c_{i,(s_{11},\ldots,s_{ij}-1,\ldots,s_{KK})}[m]$ and we define $c_{i,\vec{s}}[m] = 0$ if any component of $\vec{s}$ is -1 or N. At the end of the $(m+1)$th block of d received signals (i.e., the block consisting of signals received at $t = md, md+1, \ldots, (m+1)d-1$), relay u_j can form a d-dimensional vector of received signals $\vec{Y}_{u_j}[m]$ as

$$\begin{bmatrix} Y_{u_j}[md] \\ Y_{u_j}[md+1] \\ \vdots \\ Y_{u_j}[md+d-1] \end{bmatrix} = \gamma \sum_{\vec{s} \in \Delta_{N+1}} \begin{bmatrix} T_{\vec{s}}[md] \\ T_{\vec{s}}[md+1] \\ \vdots \\ T_{\vec{s}}[md+d-1] \end{bmatrix} a_{j,\vec{s}}[m] + \begin{bmatrix} Z_{u_j}[md] \\ Z_{u_j}[md+1] \\ \vdots \\ Z_{u_j}[md+d-1] \end{bmatrix} \tag{3.10}$$

for $m \in \{0, \ldots, n-1\}$. Notice that, for each $\vec{s} \in \Delta_{N+1}$, $T_{\vec{s}}[t]$ is a distinct monomial on the variables $h_{s_i,u_j}[t]$ for $i, j \in \{1, \ldots, K\}$. The following lemma, whose proof is in Appendix A.1, will thus be useful.

Lemma 3.1 *Consider the vector* $\vec{p}(x_1, \ldots, x_m) = [p_1(x_1, \ldots, x_m), \ldots, p_\ell(x_1, \ldots, x_m)]^\dagger$, *where each* $p_i(x_1, \ldots, x_m)$ *is a distinct monomial on the variables* $x_1, \ldots, x_m$. *The determinant of the* $\ell \times \ell$ *matrix*

$$[\vec{p}(x_{1,1}, \ldots, x_{1,m}), \vec{p}(x_{2,1}, \ldots, x_{2,m}), \ldots, \vec{p}(x_{\ell,1}, \ldots, x_{\ell,m})]$$

is a non-identically zero polynomial on the variables $x_{1,1}, \ldots, x_{1,m}, \ldots, x_{\ell,1}, \ldots, x_{\ell,m}$.

Let $\mathbf{T}[m]$ be the $d \times d$ matrix whose columns are

$$\vec{T}_{\vec{s}}[m] = \begin{bmatrix} T_{\vec{s}}[md] \\ T_{\vec{s}}[md+1] \\ \vdots \\ T_{\vec{s}}[(m+1)d-1] \end{bmatrix}, \tag{3.11}$$

for $\vec{s} \in \Delta_{N+1}$. From Lemma 3.1, we see that $\det \mathbf{T}[m]^{\dagger} = \det \mathbf{T}[m]$, seen as a polynomial on the variables $h_{s_i,u_j}[t]$ for $i, j \in \{1, ..., K\}$ and $t = md, ..., (m+1)d-1$, is not identically zero. Thus, since $h_{s_i,u_j}[t]$ for $i, j \in \{1, ..., K\}$ and $t = 0, ..., nd-1$ are all indepedent and drawn from absolutely continuous distributions, $\mathbf{T}[m]$ is invertible with probability 1. Moreover, if we fix some arbitrary $\epsilon > 0$, we can find $\delta > 0$ such that $|\det \mathbf{T}[m]| > \delta$ with probability $1-\epsilon$. At time $t = (m+1)d-1$, the relays will verify whether this is satisfied. In case $|\det \mathbf{T}[m]| \leq \delta$, all the relays will simply remain silent at times $t = (m+1)d, ..., (m+2)d-1$. As we will see later, this is important to guarantee that the entries of $\mathbf{T}^{-1}[m]$ are not too large, which could lead to a violation of the transmit power constraints at the relays. Otherwise, if $|\det \mathbf{T}[m]| > \delta$, in order to build its transmit signals, each relay u_j will construct the vector of estimates of the $a_{j,\vec{s}}$s

$$\begin{aligned} \left[\hat{a}_{j,\vec{s}}[m]\right]_{\vec{s}\in\Delta_{N+1}} &= \gamma^{-1}\mathbf{T}[m]^{-1}\vec{Y}_{u_j}[m] \\ &= \left[a_{j,\vec{s}}[m]\right]_{\vec{s}\in\Delta_{N+1}} + \gamma^{-1}\mathbf{T}[m]^{-1} \begin{bmatrix} Z_{u_j}[md] \\ Z_{u_j}[md+1] \\ \vdots \\ Z_{u_j}[(m+1)d-1] \end{bmatrix}. \end{aligned} \tag{3.12}$$

In order to build the transmit signal for time $t = (m+1)d, ..., (m+2)d-1$, each relay will compute the determinant of

$$\mathbf{H}_{U,D}[t] = \begin{bmatrix} h_{u_1,d_1}[t] & \dots & h_{u_K,d_1}[t] \\ \vdots & \ddots & \vdots \\ h_{u_1,d_K}[t] & \dots & h_{u_K,d_K}[t] \end{bmatrix}.$$

It is obvious that $\det \mathbf{H}_{U,D}[t]$ is a non-identically zero polynomial on the variables $h_{u_i,d_j}[t]$, $i, j \in \{1, ..., K\}$. Since the event $\{|\det \mathbf{H}_{U,D}[t]| > \delta'\}$ is independent for each time t, we can choose $\delta' > 0$ small enough so that

$$\Pr\left[\left|\{t : md \le t \le (m+1)d - 1, |\det \mathbf{H}_{U,D}[t]| > \delta'\}\right| \le |\Delta_N|\right] < \epsilon, \tag{3.13}$$

where ϵ is the same previously chosen parameter. Now, if at time t, $|\det \mathbf{H}_{U,D}[t]| \le \delta'$, all relays will simply stay silent. Otherwise, using $\left[\hat{a}_{j,\vec{s}}[m]\right]_{\vec{s}\in\Delta_{N+1}}$ from (3.12), relay u_j will encode all these $d = |\Delta_{N+1}|$ symbols using new transmit directions. To describe the new set of transmit directions, we first define

$$\begin{bmatrix} b_{11}[t] & \dots & b_{K1}[t] \\ \vdots & \ddots & \vdots \\ b_{1K}[t] & \dots & b_{KK}[t] \end{bmatrix} = \mathbf{H}_{U,D}[t]^{-1}. \tag{3.14}$$

Next, we let

$$\tilde{T}_{s_{11},s_{12},...,s_{KK}}[t] = \prod_{\substack{1\le i\le K \\ 1\le j\le K}} b_{ij}[t]^{s_{ij}}, \tag{3.15}$$

and, similar to (3.6), we can define the set of transmit directions for the relays to be

$$\tilde{\mathcal{T}}_{N+1}[t] = \left\{\tilde{T}_{s_{11},s_{12},...,s_{KK}}[t] : (s_{11}, s_{12}, ..., s_{KK}) \in \Delta_{N+1}\right\}. \tag{3.16}$$

Relay u_j will encode the $\hat{a}_{j,\vec{s}}$s by transmitting, at time $t = (m+1)d, (m+1)d + 1, ..., (m+2)d - 1$,

$$X_{u_j}[t] = \gamma'\left(\sum_{\vec{s}\in\Delta_{N+1}} \tilde{T}_{\vec{s}}[t]\ \hat{a}_{j,\vec{s}}[m]\right) = \gamma'\left(\sum_{\vec{s}\in\Delta_{N+1}} \tilde{T}_{\vec{s}}[t]\ a_{j,\vec{s}}[m]\right) + \gamma'\tilde{Z}_{u_j}[t], \tag{3.17}$$

where $\tilde{Z}_{u_j}[t]$ is the effective noise term which results from the additive noise terms in the estimates $\hat{a}_{j,\vec{s}}$s. The constant γ' is chosen so that the transmit power

$$E\left[X_{u_j}[t]^2\right] = \gamma'^2 E\left[\left(\sum_{\vec{s}\in\Delta_{N+1}} \tilde{T}_{\vec{s}}[t]\ a_{j,\vec{s}}[m]\right)^2\right] + \gamma'^2 E\left[\tilde{Z}_{u_j}[t]^2\right]$$
$$\leq \gamma'^2 KP \sum_{\vec{s}\in\Delta_{N+1}} E\left[\tilde{T}_{\vec{s}}[t]^2\right] + \gamma'^2 E\left[\tilde{Z}_{u_j}[t]^2\right]$$

does not exceed P. By expressing the inverse in (3.14) in terms of the cofactor matrix, we see that each $b_{ij}[t]$ can be written as a ratio between a polynomial on the variables $h_{u_i,d_j}[t]$, $i, j \in \{1, ..., K\}$ and $\det \mathbf{H}_{U,D}[t]$. Thus, since $|\det \mathbf{H}_{U,D}[t]| > \delta'$, we see that $E\left[\tilde{T}_{\vec{s}}[t]^2\right] < \infty$ for all $\vec{s}$. Moreover, the fact that $E\left[h_{u_i,d_j}[t]^2\right] < \infty$, for each $i, j \in \{1, ..., K\}$, and $|\det \mathbf{T}[m]| > \delta$ guarantees that the variance of $\tilde{Z}_{u_j}[t]$ is finite and independent of P. Thus, for P sufficiently large, γ' can be chosen independent of P and t.

We then have the following claim.

Claim 3.1 *The transmit signal of relay u_j, given in (3.17), can be re-written as*

$$X_{u_j}[t] = \gamma' \sum_{\vec{s}\in\Delta_N} \tilde{T}_{\vec{s}}[t] \left(\sum_{i=1}^{K} b_{ij}[t]\, c_{i,\vec{s}}[m]\right) + \gamma' \tilde{Z}_{u_j}[t]. \tag{3.18}$$

Proof: The main idea is to notice that, just as (3.9) can be written as (3.8), (3.17) can be re-written as (3.18). This can be more easily understood if we think of the (noiseless version of the) received signal in (3.9) as a polynomial on the variables $h_{s_i,u_j}[t]$, $1 \leq i, j \leq K$. When relay u_j estimates each coefficient $a_{j,\vec{s}}[t]$ of this polynomial and then replaces each monomial $T_{\vec{s}}$ with $\tilde{T}_{\vec{s}}$, it is essentially re-building the same polynomial with each variable $h_{s_i,u_j}[t]$ replaced by $b_{ij}[t]$. Therefore, the same factorization used on the polynomial on the $h_{s_i,u_j}[t]$ variables in (3.8) can be used on the polynomial on the $b_{ij}[t]$ variables, as shown in (3.18).
■

Decoding at the destinations:

In order to compute the received signals at the destinations, we first notice that, from (3.18), the vector of the K relay transmit signals at time $t = (m+1)d, (m+1)d+1, ..., (m+2)d-1$, can be written as

$$\begin{bmatrix} X_{u_1}[t] \\ \vdots \\ X_{u_K}[t] \end{bmatrix} = \gamma' \sum_{\vec{s}\in\Delta_N} \tilde{T}_{\vec{s}}[t] \begin{bmatrix} b_{11}[t] & \dots & b_{K1}[t] \\ \vdots & \ddots & \vdots \\ b_{1K}[t] & \dots & b_{KK}[t] \end{bmatrix} \begin{bmatrix} c_{1,\vec{s}}[m] \\ \vdots \\ c_{K,\vec{s}}[m] \end{bmatrix} + \gamma' \begin{bmatrix} \tilde{Z}_{u_1}[t] \\ \vdots \\ \tilde{Z}_{u_K}[t] \end{bmatrix}. \tag{3.19}$$

We can then write the vector of the K received signals at the destinations as

$$\begin{aligned}
\begin{bmatrix} Y_{d_1}[t] \\ \vdots \\ Y_{d_K}[t] \end{bmatrix} &= \begin{bmatrix} h_{u_1,d_1}[t] & \dots & h_{u_K,d_1}[t] \\ \vdots & \ddots & \vdots \\ h_{u_1,d_K}[t] & \dots & h_{u_K,d_K}[t] \end{bmatrix} \begin{bmatrix} X_{u_1}[t] \\ \vdots \\ X_{u_K}[t] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t] \\ \vdots \\ Z_{d_K}[t] \end{bmatrix} \\
&= \begin{bmatrix} b_{11}[t] & \dots & b_{K1}[t] \\ \vdots & \ddots & \vdots \\ b_{1K}[t] & \dots & b_{KK}[t] \end{bmatrix}^{-1} \begin{bmatrix} X_{u_1}[t] \\ \vdots \\ X_{u_K}[t] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t] \\ \vdots \\ Z_{d_K}[t] \end{bmatrix} \\
&= \gamma' \sum_{\vec{s}\in\Delta_N} \tilde{T}_{\vec{s}}[t] \begin{bmatrix} c_{1,\vec{s}}[m] \\ \vdots \\ c_{K,\vec{s}}[m] \end{bmatrix} + \underbrace{\gamma' \begin{bmatrix} b_{11}[t] & \dots & b_{K1}[t] \\ \vdots & \ddots & \vdots \\ b_{1K}[t] & \dots & b_{KK}[t] \end{bmatrix}^{-1} \begin{bmatrix} \tilde{Z}_{u_1}[t] \\ \vdots \\ \tilde{Z}_{u_K}[t] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t] \\ \vdots \\ Z_{d_K}[t] \end{bmatrix}}_{\left[\tilde{Z}_{d_1}[t] \dots \tilde{Z}_{d_K}[t]\right]^{\dagger}}.
\end{aligned} \tag{3.20}$$

Thus, the received signal at destination d_j at time $t = (m+1)d, (m+1)d+1, ..., (m+2)d-1$ is simply given by

$$Y_{d_j}[t] = \gamma' \sum_{\vec{s}\in\Delta_N} \tilde{T}_{\vec{s}}[t]\; c_{j,\vec{s}}[m] + \tilde{Z}_{d_j}[t], \tag{3.21}$$

and we see that all the interference has been cancelled, and destination d_j receives only the data streams originated at source s_j. Moreover, the effective noise $\tilde{Z}_{d_j}[t]$ has a finite variance that is independent of P.

Destination d_j will use decoding operations similar to those used at the relays. The block of the d signals received at times $t = (m+1)d, (m+1)d+1, ..., (m+2)d-1$ can be used to form a length-d vector

$$\vec{Y}_{d_j}[m+1] = \begin{bmatrix} Y_{d_j}[(m+1)d] \\ Y_{d_j}[(m+1)d+1] \\ \vdots \\ Y_{d_j}[(m+2)d-1] \end{bmatrix}. \tag{3.22}$$

Notice that, in case $|\det \mathbf{T}[m]| \leq \delta$, these received signals will contain just noise, since the relays stayed silent in times $t = (m+1)d, (m+1)d+1, ..., (m+2)d-1$. Moreover, at any time $t \in \{(m+1)d, (m+1)d+1, ..., (m+2)d-1\}$ for which $|\det \mathbf{H}_{U,D}[t]| \leq \delta'$, the corresponding entry of $\vec{Y}_{d_j}[m+1]$ will contain only noise. Notice that, from (3.13), this will be the case of less than $d - |\Delta_N|$ of the entries, with probability at least $1-\epsilon$. Thus, with probability at least $1-\epsilon$, destination d_j can let $t_1, ..., t_{|\Delta_N|}$ be the first $|\Delta_N|$ values of $t \in \{(m+1)d, ..., (m+2)d-1\}$ for which $|\det \mathbf{H}_{U,D}[t]| > \delta'$, and, from (3.21), the resulting received signals satisfy

$$\begin{bmatrix} Y_{d_j}[t_1] \\ Y_{d_j}[t_2] \\ \vdots \\ Y_{d_j}[t_{|\Delta_N|}] \end{bmatrix} = \gamma' \sum_{\vec{s} \in \Delta_N} \begin{bmatrix} \tilde{T}_{\vec{s}}[t_1] \\ \tilde{T}_{\vec{s}}[t_2] \\ \vdots \\ \tilde{T}_{\vec{s}}[t_{|\Delta_N|}] \end{bmatrix} c_{j,\vec{s}}[m] + \begin{bmatrix} \tilde{Z}_{d_j}[t_1] \\ \tilde{Z}_{d_j}[t_2] \\ \vdots \\ \tilde{Z}_{d_j}[t_{|\Delta_N|}] \end{bmatrix}. \tag{3.23}$$

The remaining $d - |\Delta_N|$ received signals are simply discarded by the destinations. We will let $\vec{\tilde{T}}_{\vec{s}}[m+1] = \left[\tilde{T}_{\vec{s}}[t_1], \tilde{T}_{\vec{s}}[t_2], ..., \tilde{T}_{\vec{s}}[t_{|\Delta_N|}]\right]^\dagger$, for each $\vec{s} \in \Delta_N$. Notice that, for each $\vec{s} \in \Delta_N$, $\tilde{T}_{\vec{s}}[t]$ is a distinct monomial on the variables $b_{ij}[t]$ for $i, j \in \{1, ..., K\}$. We will then let $\tilde{\mathbf{T}}[m]$ be the $|\Delta_N| \times |\Delta_N|$ (i.e., $N^{K^2} \times N^{K^2}$) matrix whose columns are $\vec{\tilde{T}}_{\vec{s}}[m+1]$, for $\vec{s} \in \Delta_N$. Lemma 3.1 now implies that $\det \tilde{\mathbf{T}}[m]^\dagger = \det \tilde{\mathbf{T}}[m]$ is a non-identically zero polynomial on the variables $b_{ij}[t]$, for $i, j \in \{1, ..., K\}$ and $t \in \{t_1, ..., t_{|\Delta_N|}\}$. Since, by expressing the inverse in (3.14) in terms of the cofactor matrix, each $b_{ij}[t]$ can be written as a ratio between two polynomials on the variables $h_{u_i,d_j}[t]$, $i, j \in \{1, ..., K\}$, we see that $\det \tilde{\mathbf{T}}[m]$ is also a ratio of polynomials on the variables $h_{u_i,d_j}[t]$, $i, j \in \{1, ..., K\}$, $t \in \{t_1, ..., t_{|\Delta_N|}\}$. We claim that this ratio of polynomials is also non-identically zero. To see this we notice that, since $\det \tilde{\mathbf{T}}[m]$ is a non-identically zero polynomial on the variables $b_{ij}[t]$, we can pick values for each $b_{ij}[t]$ such that $\det \tilde{\mathbf{T}}[m] = \alpha$ for some $\alpha \neq 0$. Then, using the inverse relation in (3.14), we can find corresponding values for each $h_{u_i,d_j}[t]$, so that $\det \tilde{\mathbf{T}}[m] = \alpha$, which shows that $\det \tilde{\mathbf{T}}[m]$ cannot be an identically zero ratio of polynomials in the $h_{u_i,d_j}[t]$ variables. Therefore, one can find $\delta'' > 0$ such that, with probability $1 - \epsilon$, $|\det \tilde{\mathbf{T}}[m]| > \delta''$.

At time $t = (m + 2)d - 1$, destination d_j will construct a length-$|\Delta_N|$ vector of effective outputs as follows. If $|\det \tilde{\mathbf{T}}[m]| \leq \delta''$, $|\det \mathbf{T}[m]| \leq \delta$ or if there are more than $d - |\Delta_N|$ times $t \in \{(m + 1)d, (m + 1)d + 1, ..., (m + 2)d - 1\}$ for which $|\det \mathbf{H}_{U,D}[t]| \leq \delta'$, it simply outputs $[\varepsilon, ..., \varepsilon]$, where ε simbolizes an erasure. Since each of these three events occurs with probability at most ϵ, their union occurs with probability at most 3ϵ. If none of these events occurs, destination d_j will output the vector of estimates of the $c_{j,\vec{s}}$s

$$\left[\hat{c}_{j,\vec{s}}[m]\right]_{\vec{s}\in\Delta_N} = \frac{1}{\gamma'}\tilde{\mathbf{T}}[m]^{-1}\begin{bmatrix} Y_{d_j}[t_1] \\ Y_{d_j}[t_2] \\ \vdots \\ Y_{d_j}[t_{|\Delta_N|}] \end{bmatrix} = \left[c_{j,\vec{s}}[m]\right]_{\vec{s}\in\Delta_N} + \underbrace{\frac{1}{\gamma'}\tilde{\mathbf{T}}[m]^{-1}\begin{bmatrix} \tilde{Z}_{u_j}[t_1] \\ \tilde{Z}_{u_j}[t_2] \\ \vdots \\ \tilde{Z}_{u_j}[t_{|\Delta_N|}] \end{bmatrix}}_{\left[\tilde{\tilde{Z}}_{j,\vec{s}}[m]\right]_{\vec{s}\in\Delta_N}}. \tag{3.24}$$

Notice that $|\det \tilde{\mathbf{T}}[m]| > \delta''$ implies that the entries of $\tilde{\mathbf{T}}[m]^{-1}$ have a finite variance, which in turn implies that the resulting additive noise vector $\left[\tilde{\tilde{Z}}_{j,\vec{s}}[m]\right]_{\vec{s}\in\Delta_N}$ has a finite covariance matrix, with entries that are independent of P. Destination d_j will then view each entry of its output vector as the output of a separate channel, with input $c_{j,\vec{s}}[m]$ and output $c_{j,\vec{s}}[m] + \tilde{\tilde{Z}}_{j,\vec{s}}[m]$ with probability $1 - q$ and ε with probability q, where $q \leq 3\epsilon$. Therefore, we essentially create N^{K^2} parallel AWGN channels with erasure probability at most 3ϵ. The fact that the additive noises are correlated is irrelevant (in fact it can only improve the achievable rates), and it is clear that we can achieve $1 - 3\epsilon$ degrees of freedom in each of these effective channels. Since we need d time steps to transmit one symbol in each of these channels, we achieve a total of

$$(1 - 3\epsilon)\frac{N^{K^2}}{d} = \frac{(1 - 3\epsilon)N^{K^2}}{(N + 1)^{K^2}} = (1 - 3\epsilon)\left(\frac{N}{N + 1}\right)^{K^2}$$

degrees of freedom per user, for arbitrarily chosen N and ϵ. Thus, by choosing N large and $\epsilon > 0$ small, each user can achieve arbitrarily close to one degree of freedom.

3.2.3 Aligned Network Diagonalization for Constant Channels

In the case of constant channel gains, the AND scheme presented in Section 3.2.2 does not work. The lack of time diversity makes the entries in the vector $\vec{T}_{\vec{s}}[m]$, given in (3.11), be all equal, and $\mathbf{T}[m]$ is not invertible (as its rank is one). Therefore, in order to achieve K degrees of freedom with constant channels, we must perform the alignment operations of AND not over time dimensions, but over rational dimensions, in the spirit of [45]. In this section, we present the main ideas to convert the scheme from Section 3.2.2 to the constant-channel setting.

In the case of constant channel gains, once again each source s_i starts by breaking its message W_i into $L = |\Delta_N|$ submessages, and the transmit signals are of the form

$$X_{s_i}[t] = \gamma \sum_{\vec{s} \in \Delta_N} T_{\vec{s}}\, c_{i,\vec{s}}[t],$$

where γ is chosen so that the power constraint is satisfied. The directions $T_{\vec{s}}$ are again defined as in (3.5), but the code symbols $c_{i,\vec{s}}[t]$ are now integer-valued. Similar to (3.9), the received signal at relay u_j will be given by

$$Y_{u_j}[t] = \gamma \sum_{\vec{s} \in \Delta_{N+1}} T_{\vec{s}}\, a_{j,\vec{s}}[t] + Z_{u_j}[t],$$

where each $a_{j,\vec{s}}$ is a sum of $c_{i,\vec{s}}$s and thus also integer-valued. By following the rational dimensions framework of [45], with high probability, relay u_j can extract from this signal the integers $a_{j,\vec{s}}$, and transmit them along new directions, according to

$$X_{u_j}[t+1] = \gamma' \sum_{\vec{s} \in \Delta_{N+1}} \tilde{T}_{\vec{s}}\, a_{j,\vec{s}}[t],$$

where γ' is chosen so that the transmit power constraint is satisfied, and the new transmit directions $\tilde{T}_{\vec{s}}$ are defined as in (C.6). Next, by following steps as those in (C.8) and (3.20), the received signal at destination d_j can be expressed as

$$Y_{d_j}[t+1] = \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}}\, c_{j,\vec{s}}[t] + Z_{d_j}[t+1], \tag{3.25}$$

and we again have each destination receiving the symbols from its corresponding source free of interference. Finally, the rational dimensions framework is used once again to show that each destination d_j can in fact decode its desired symbols $c_{j,\vec{s}}$.

A detailed description of this construction is found in Appendix A.2, as well as the technical steps required to show the decodability of the integers at the relays and destinations, and a performance analysis showing that this scheme can indeed achieve close to one degree of freedom per user.

3.3 Two-Hop Networks with MIMO Nodes

In this section, we use the result from Theorem 3.1 in order to characterize the degrees of freedom of two-hop networks where we still have full connectivity at each hop, but each node (sources, relays and destinations) is allowed to have multiple antennas. In general, we want to focus on a $K \times A \times K$ network, where each node $u \in \{s_i, u_i, d_i : 1 \leq i \leq K\}$ has M_u (full-duplex) antennas. For simplicity of exposition, we will focus on the case of time-varying channels. However, it should be clear that the results in this section can also be obtained in the case of constant channels, by extending Theorem 3.2 instead.

It is obvious that, in this setting, for certain choices of the number of antennas at each node, it may not be optimal to assign the same number of degrees of freedom to each source-destination pair, as was the case when $M_u = 1$ for all u. Therefore, in this section, instead of focusing on the sum degrees of freedom, we will instead consider the degrees-of-freedom region, defined in Definition 1.6.

While the formal definition is technical, the degrees-of-freedom region can be intuitively understood as a high-SNR approximation to the capacity region, scaled down by $\frac{1}{2}\log P$. The sum degrees of freedom D_Σ from Definition 1.5 is simply the point in $\mathbf{D}$ that maximizes the (unweighted) sum of its components. For two-hop networks with MIMO nodes we then have the following result.

Theorem 3.3 *For a $K \times A \times K$ wireless network with time-varying channels where each node u has M_u antennas, the degrees-of-freedom region comprises all nonnegative K-tuples $(D_1, ..., D_K)$ satisfying*

$$\sum_{i=1}^{K} D_i \leq M_{relays} \tag{3.26}$$

$$D_i \leq \min\left[M_{s_i}, M_{d_i}\right], \quad \textit{for } i = 1, ..., K, \tag{3.27}$$

where $M_{relays} = \sum_{i=1}^{A} M_{u_i}$ *is total number of antennas at the relays.*

Once again, the converse part of this Theorem is obtained from the cut-set bound in a straigthforward manner. Moreover, given Theorem 3.1, the achievability is also easily obtained. More precisely, for any degrees-of-freedom tuple $(D_1, ..., D_K)$ satisfying (3.26), we first discard $M_{\text{relays}} - \sum_{i=1}^{K} d_i$ out of the total relay antennas. Moreover, since $d_i \leq \min\left[M_{s_i}, M_{d_i}\right]$ from (3.27), we can discard $M_{s_i} - d_i$ out of the source antennas and $M_{d_i} - D_i$ out of the destination antennas. Then, if we view all remaining antennas as separate nodes, we obtain a $K' \times K' \times K'$ wireless network with time-varying channels, where $K' = \sum_{i=1}^{K} D_i$. It is then clear that, by applying Theorem 3.1, we can achieve $\sum_{i=1}^{K} D_i$ sum degrees of freedom on this network, which corresponds to the degrees-of-freedom tuple $(D_1, ..., D_K)$ in the original network.

The most interesting aspect of this result is the fact that the cooperation that is allowed among the antennas due to the MIMO setting does not improve the degrees of freedom that can be achieved in the case that all the antennas are viewed as separate nodes. We point out that similar observations had already been made in the literature. In [6], for instance, this observation was made in the context of relay networks and constant-gap capacity approximations. In [36], this was also noted in a context similar to the one considered in this section.

3.4 Future Research Directions

While our results imply the tightness of the cut-set bound, it is important to point out that this is likely only the case for the degrees of freedom. In general, one would expect that more sophisticated outer bounds can be developed for the capacity region of $K \times K \times K$ wireless network. Towards this goal, a promising direction is to consider a deterministic model of the $K \times K \times K$ wireless network. Deterministic models of wireless networks have been proven useful in the study of the capacity of both multi-hop single-flow networks [6] and single-hop multi-flow networks [12, 22]. Not only do they usually provide new insights about the original stochastic problem, but they can in fact be shown, in several cases, to approximate well the capacity of their non-deterministic (usually AWGN) counterparts. A step towards studying $K \times K \times K$ wireless networks under deterministic models is taken in Chapter 6. By using the worst-case noise result from Chapter 4, it is shown that the capacity region of an AWGN $K \times K \times K$ wireless network is a subset of the capacity region of the same network under the truncated deterministic model [6] (where nodes are given slightly more power). This fact is particularly interesting because it allows us to look for outer bounds on the capacity region of the AWGN $K \times K \times K$ wireless network by focusing on the truncated deterministic channel model, which is expected to reveal combinatorial structures of the problem that are not apparent in the AWGN setting.

Since the degrees-of-freedom characterization only provides a capacity approximation at high SNR, an important direction for future work is to understand the $K \times K \times K$ wireless network at low and moderate SNRs. One effort along this direction is found in [37], where the $2 \times L \times 2$ wireless network was considered under the fast fading scenario. The capacity was characterized to within a constant number of bits for several channel fading distributions, with an achievability scheme based on ergodic interference neutralization [36], which is based on the ideas of ergodic interference alignment [46]. In the case of constant channel gains, however, such schemes cannot be used, and some research has focused on devising achievability schemes whose achieved rates can be explicitly computed at any given SNR (which is not the case for the AND scheme described in this chapter). Efforts along this direction are found in [32, 33], where a linear scheme for the $2 \times 2 \times 2$ wireless network based on linear relaying operations is introduced, and in [30, 31], where a scheme based on AND and lattices is described also for the $2\times2\times2$ wireless network. In both cases these coupled schemes outperform simple schemes (such as TDMA) at moderate and low SNRs.

Other directions for future work include studying the channel diversity required for our proposed scheme to be performed, and what can be done with limited channel diversity. In particular, we notice that the linear version of AND relies on the fast variation of the channel gains in the network and requires a large number of distinct channel realizations in order to achieve close to one degree of freedom per user. If we limit the available time (and space) diversity, as considered for instance in [10], it is not clear if the same gains achieved by AND can be obtained. Of particular interest are the achievable degrees of freedom once we restrict ourselves to linear schemes, but assume a finite amount of channel diversity, or simply constant channels. This is done in [32, 33], for the case of $2\times2\times2$ wireless networks. It is shown that, even if the channels are constant, linear schemes can still outperform simple decoupled schemes, and achieve $4/3$ degrees of freedom for real-valued channel gains and $5/3$ degrees of freedom for complex-valued channel gains. However, for any $K > 2$, this problem remains unsolved, and it is not even clear how the linearly-achievable degrees of freedom scale with K, or whether the gap between the linearly-achievable degrees of freedom and the cut-set bound of K is unbounded.

Another future direction would be to investigate the role of relays for interference management when the channel state information is obtained with some delay. Recently, in [1, 2], it has been shown that even under such constraint, the sum degrees of freedom can scale with the number of users (assuming that the number of hops of communication via relays also scales with the number of users). However, there is still a large gap between the state-of-the-art inner bounds and outer bounds on the degrees of freedom of multi-hop multi-flow wireless networks with delayed knowledge of the channel states.

Part II

Robustness of Gaussian Models

Stochastic modeling of the data source and the communication medium are essential in data compression and data communication problems. However, extracting these descriptions from a practical system is in general difficult and often leads to intractable problems from a theoretical point of view. As a result, Gaussian models for both the data sources and the noise in communication networks prevail.

The modeling of the noise in communication links as additive Gaussian is generally justified through the Central Limit Theorem, which suggests that the cumulative effect of many independent noise sources should be approximately Gaussian. The modeling of data sources as Gaussian, on the other hand, is less justifiable and done largely for the sake of analytical tractability.

From a theoretical standpoint, one way of supporting the Gaussian assumption is by establishing that it is worst-case, meaning that, within a given family of distributions (usually defined by a covariance constraint), the Gaussian assumption results in the smallest possible capacity or rate-distortion region. In fact, this has long been known to be the case in two classical single-user Information Theory scenarios. In the channel coding setting, it is known that, given a fixed variance of the noise, the Gaussian distribution minimizes the capacity of a memoryless additive-noise channel. The source coding counterpart of this result is that, for a fixed-variance i.i.d. random source, the Gaussian distribution minimizes the rate-distortion region (i.e., maximizes the rate-distortion function). Both of these assertions can be proved using the fact that, subject to a variance constraint, the Gaussian distribution maximizes the entropy. In the channel coding case, a more operational proof of the fact that Gaussian noise is the worst-case noise was provided in [41], where it was shown that random

Gaussian codebooks and nearest-neighbor decoding achieve the capacity of the corresponding AWGN channel on a non-Gaussian channel.

There are a few other worst-case characterizations in the literature. One example is [18], where the authors consider vector channels with additive noise subject to the constraint that the noise covariance matrix lies in a convex set. It is shown that, in this setting, the worst-case noise is vector Gaussian with a covariance matrix that depends on the transmit power constraints. In [51], a scalar additive-noise channel with binary input is considered. In this setting, the probability mass function of the (discrete) worst-case noise is characterized, and the worst-case capacity (i.e., the capacity under the worst-case noise) is found. Another example is the work in [64], which characterizes the rate-distortion region for the two-encoder source coding problem with quadratic distortion constraints and Gaussian sources, which in turn allows the characterization of the joint Gaussian source as the worst-case source for the two-encoder quadratic source coding problem.

Beyond the aforementioned examples, worst-case analyses of more general multi-user networks was, until recently, fairly limited. In this part of the dissertation, we introduce novel techniques that enable the characterization of the Gaussian distribution as worst-case in several different new scenarios. Of particular interest to us will be the characterization of the Gaussian noise as the worst-case additive noise in general multi-hop multi-flow wireless networks. As we will see, this result will allow us to establish a relationship between the capacity region of multi-hop multi-flow wireless networks under different channel models in Chapter 6.

CHAPTER 4

WORST-CASE ADDITIVE NOISE IN WIRELESS NETWORKS

In this chapter, we generalize the classical result that characterizes the Gaussian noise as the worst-case additive noise in point-to-point channels to the case of arbitrary multi-hop multi-flow wireless networks. We show that, if we fix the variance of the additive noise at all nodes in the network, the capacity region is minimized by choosing them to be normally distributed. More precisely, we prove the following result.

Theorem 4.1 *From a sequence of coding schemes that achieve rate tuple* $\mathbf{R}$ *on an AWGN* K*-unicast wireless network, it is possible to construct a sequence of coding schemes that achieves arbitrarily close to* $\mathbf{R}$ *on the same* K*-unicast wireless network, where, for each relay* v*, the distribution of* Z_v *is replaced with any distribution satisfying* $E[Z_v] = 0$ *and* $E\left[Z_v^2\right] = \sigma_v^2$*. Therefore, if* C_{AWGN} *is the capacity region of the AWGN* K*-unicast wireless network, and* $C_{\text{non-AWGN}}$ *is the capacity region of the same wireless network where, for each relay* v*, the distribution of* Z_v *is replaced with an arbitrary distribution satisfying* $E[Z_v] = 0$ *and* $E\left[Z_v^2\right] = \sigma_v^2$*, then*

$$C_{\text{AWGN}} \subseteq C_{\text{non-AWGN}}.$$

We prove Theorem 4.1 based on two main results. The first one is that, given a coding scheme with *finite reading precision* for an AWGN network, one can build a coding scheme that achieves the same rates on a non-Gaussian wireless network. A coding scheme is said to have finite reading precision if, for any node, its transmit signals only depend on its received signals read up to a finite number of digits after the decimal point. This result by first applying a transformation at the transmit signals and received signals of all nodes in the network in order to create an “approximately Gaussian” effective network. The technique resembles OFDM in that it uses the Discrete Fourier Transform in order to mix together multiple uses of the same channel. This mixing causes the additive noise terms from distinct network uses to be averaged over time and, by making use of Lindeberg’s Central Limit Theorem, it can be shown that the resulting effective noise is approximately Gaussian in the distribution sense. By combining this OFDM-like approach with an interleaving operation to take care of the dependency between the resulting noise realizations, we create an approximately Gaussian network. Then we show that, coding schemes with finite reading precision are robust in the sense that, when applied on an approximately Gaussian network, their performance does not deviate much from what it would be if the noises were truly Gaussian.

The second main result we need is that, for any wireless network, the capacity when we restrict ourselves to coding schemes with finite reading precision, and allow the precision to tend to infinity along the sequence of coding schemes, is the same as the unrestricted capacity. To prove this we show that, for any coding scheme with infinite precision, there exists a quantization scheme of the received signals which does not increase the error probability of the coding scheme too much. This is done by showing that a truncation of the bit expansion of the received signal followed by a random shift performs well; thus, there must exist a fixed shift for each node which guarantees the same performance. This quantization operation makes the coding scheme have finite reading precision, and the result follows.

We will first focus on coding schemes that have *finite reading precision*. Then we will show that coding schemes with infinite reading precision can be converted into coding schemes with finite reading precision without much loss in performance.

Definition 4.1 *For some $x \in \mathbb{R}$ and a positive integer ρ, let $\lfloor x \rfloor_\rho = 2^{-\rho}\lfloor 2^\rho x \rfloor$. A coding scheme C is said to have finite reading precision $\rho \in \mathbb{N}$ if its relaying functions satisfy*

$$r_v^{(t)}(y_1, ..., y_{t-1}) = r_v^{(t)}(\lfloor y_1 \rfloor_\rho, ..., \lfloor y_{t-1} \rfloor_\rho),$$

for any $(y_1, ..., y_{t-1}) \in \mathbb{R}^{t-1}$, any $v \in V - \{s_1, ..., s_K\}$, and any time t, and its decoding functions satisfy

$$g_i(y_1, ..., y_n) = g_i(\lfloor y_1 \rfloor_\rho, ..., \lfloor y_n \rfloor_\rho),$$

for any $(y_1, ..., y_n) \in \mathbb{R}^n$, and $i \in \{1, ..., K\}$.

Definition 4.2 *Rate tuple $\mathbf{R}$ is achievable by coding schemes with finite reading precision if we have a sequence of coding schemes C_n, where C_n has finite reading precision ρ_n, which achieves rate tuple $\mathbf{R}$ according to Definition 1.3.*

Remark 4.1 *Notice that we allow the precision ρ_n to vary arbitrarily along the sequence of codes, and it may be the case that $\rho_n \to \infty$ as $n \to \infty$.*

Our main result, presented in Theorem 4.1, shows that any rate tuple that is achievable on a network where each Z_v is Gaussian for each $v \in V$ is also achievable on a network where each Z_v instead has *any* distribution with the same mean and variance. In the special case of K-unicast wireless networks.

We will prove Theorem 4.1 using the following two auxiliary results.

Theorem 4.2 *Suppose a rate tuple* $\mathbf{R}$ *is achievable by coding schemes with finite reading precision on an AWGN* K*-unicast wireless network. Then it is possible to construct a single sequence of coding schemes that achieves arbitrarily close to* $\mathbf{R}$ *on the same* K*-unicast wireless network where, for each relay* v*, the distribution of* Z_v *is replaced with an arbitrary distribution satisfying* $E[Z_v] = 0$ *and* $E\left[Z_v^2\right] = \sigma_v^2$.

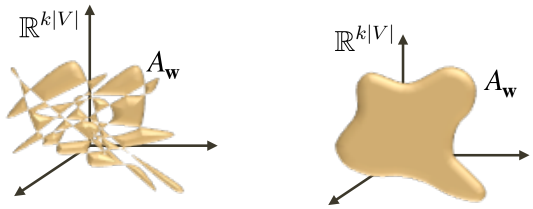
Theorem 4.3 *Suppose we have a sequence of coding schemes* C_n *achieving a rate tuple* $\mathbf{R}$ *on an AWGN network. Then it is possible to construct a sequence of coding schemes* $C_n^\star$ *with finite reading precision that also achieves* $\mathbf{R}$ *on the same AWGN network.*

It is clear that by combining Theorems 4.2 and 4.3, Theorem 4.1 will follow. The proof of Theorems 4.2 and 4.3 will be presented in Section 4.1. The result in Theorem 4.1 can be generalized to networks with arbitrary traffic demands [56].

4.1 Proof of Worst-Case Noise Result

In this Section, we will prove Theorems 4.2 and 4.3, from which Theorem 4.1 will follow. To prove Theorem 4.2, we start by assuming that we have a sequence of coding schemes with finite reading precision designed to achieve a rate tuple $\mathbf{R}$ on an AWGN network. Then, through a series of steps, we will use this sequence of coding schemes to construct another sequence of coding schemes that achieves arbitrarily close to the rate tuple $\mathbf{R}$ on the corresponding network where the additive noises are not Gaussian.

A diagram illustrating the proof steps of Theorem 4.2 is shown in Fig. 4.2. We start by describing an OFDM-like scheme that is applied to all nodes in the network. The main idea is that, by applying an Inverse Discrete Fourier Transform (IDFT) to the block of transmit signals of each node, and a Discrete Fourier Transform (DFT) to the block of received signals of each node, we create effective additive noise terms that are weighted averages of the additive noise realizations during that block. We describe this procedure in detail in Section 4.1.1. Then, in Section 4.1.2, we show that this mixture of noises converges in distribution to a Gaussian additive noise term. This is done by showing that the weighted average of the noise realizations satisfies Lindeberg's Central Limit Theorem Condition [8]. Therefore, the OFDM-like scheme effectively produces a network where the noises at each node are dependent across time and approximately Gaussian. The dependence across time is undesirable since our original coding scheme designed for the AWGN network assumed that the additive noise at each receiver is i.i.d. over time. To overcome this problem, in Section 4.1.3, we apply the OFDM-like scheme over multiple blocks, and then we interleave the effective network uses from distinct blocks. This effectively creates several blocks in which the network behaves as an Approximately AWGN network (with i.i.d. noises). Then our original code for the AWGN network can be applied to each approximately AWGN block. The fact that this code has finite reading precision guarantees that, when applied to the approximately AWGN block, its error probability is close to its error probability on the AWGN network. More formally, the error probability of a coding scheme with block length k, for a given choice of messages $\mathbf{w} \in \prod_{i=1}^{K}\{1,...,2^{kR_i}\}$, can be seen as the probability measure of the error set $A_{\mathbf{w}}$ (i.e., the set of noise realizations which causes an error to occur). As illustrated in Fig. 4.1, in general, this set could be arbitrarily ill shaped. However, if the coding scheme has finite reading precision, $A_{\mathbf{w}}$ can be shown to be a continuity set, which implies that its measure under similar probability measures cannot change much. Finally, we take care of the dependence



(a) (b)

Figure 4.1: (a) Illustration of arbitrarily shaped error set $A_{\mathbf{w}}$ and (b) continuity set $A_{\mathbf{w}}$.

between the noises of different blocks created in the interleaving operation by using a random outer code for each source-destination pair. This can be done if we view the coding scheme as creating a discrete channel between the message chosen at a given source and the decoded message at its corresponding destination. Then we can show via a mutual-information argument that we can use an outer code to achieve a rate tuple arbitrarily close to $\mathbf{R}$ on the non-Gaussian wireless network.

In Section 4.1.4, we prove Theorem 4.3. The main idea is to show that, given a coding scheme with infinite reading precision, there exists a set of quantization mappings, one for each node in the network, such that, if each node quantizes its received signal before applying the relaying or decoding function, the change in the error probability is arbitrarily small.

We point out that our results are not inconsistent with the intuition that, for a channel with a discrete output alphabet, the worst-case noise should be discrete. Theorems 4.2 and 4.3 do *not* imply that Gaussian noise is the worst-case noise if we restrict ourselves to coding schemes with finite precision, because, in Theorem 4.2, we may require coding schemes with *infinite precision* to achieve

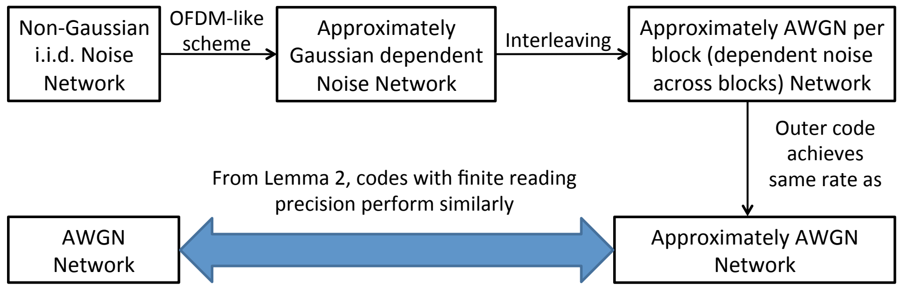


Figure 4.2: Diagram of proof steps of Theorem 4.2. Thin arrows relate to steps in the construction of our new coding scheme, while the thick arrow indicates a conceptual connection established through Lemma 4.2

the same point in the capacity region in the non-AWGN network (in fact we use coding schemes with infinite precision in our construction based on applying the OFDM-like scheme to the received signals first).

4.1.1 An OFDM-like scheme to mix the noises over time

We use an approach similar to OFDM in order to create an effective network with additive noises that are as close to normally-distributed as we wish. Essentially, each node in the network will apply transformations to its transmit signals and to its received signals, thus creating an effective network with new input-output relationships. If we focus on b uses of a single link of the network, then we convert the actual channel (i.e., a mapping from channel inputs $X[0], X[1], ..., X[b-1]$ to channel outputs $Y[0], Y[1], ..., Y[b-1]$) into an effective channel that maps inputs $d_0, d_1, ..., d_{b-1}$ into effective channel outputs $\tilde{Y}_0, \mathfrak{R}\left[\tilde{Y}_1\right], \mathfrak{I}\left[\tilde{Y}_1\right], ..., \mathfrak{R}\left[\tilde{Y}_{b/2-1}\right], \mathfrak{I}\left[\tilde{Y}_{b/2-1}\right], \tilde{Y}_{b/2}$, where $\mathfrak{R}[z]$ and $\mathfrak{I}[z]$ refer respectively to the real and imaginary parts of a complex number z. The overall transformation, depicted in Fig. 4.3, can be described as follows. Assume that a node

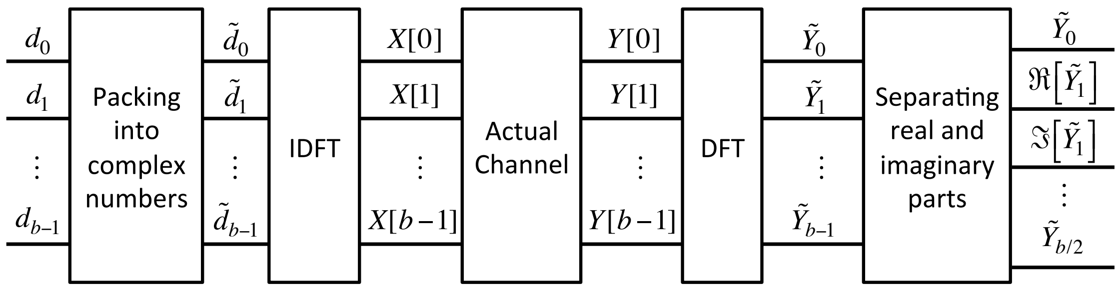


Figure 4.3: Diagram of the steps that create the effective channel.

$u \in V$ has b real numbers $d_0, d_1, ..., d_{b-1}$ which are the inputs to the effective channels we intend to create. We assume that b is even, to simplify the expressions. Then node u "packs" these signals into b complex numbers $\tilde{d}_0, ..., \tilde{d}_{b-1}$ as follows.

$$\begin{aligned}
&\tilde{d}_0 = d_0 \\
&\tilde{d}_i = d_{2i-1} + jd_{2i} \quad \text{for } i = 1, ..., \tfrac{b}{2} - 1 \\
&\tilde{d}_{b/2} = d_{b-1} \\
&\tilde{d}_i = \tilde{d}^*_{b-i} \qquad\qquad \text{for } i = \tfrac{b}{2} + 1, ..., b - 1
\end{aligned}$$

Next, node u takes the IDFT of the vector $\tilde{\mathbf{d}}_{\mathbf{u}} = (\tilde{d}_0, ..., \tilde{d}_{b-1})$ to obtain the vector $\mathbf{X}_u = \text{IDFT}(\tilde{\mathbf{d}}_{\mathbf{u}})$. Throughout the dissertation, we assume that DFT and IDFT refer to the *unitary* version of the DFT and IDFT. Since $\tilde{\mathbf{d}}_{\mathbf{u}}$ is conjugate symmetric, $\mathbf{X}_{\mathbf{u}}$ is a real vector (in $\mathbb{R}^b$). Moreover, we will require the original real-valued signals to satisfy

$$\text{avg}\left[d_0^2\right] \leq P, \tag{4.1}$$

$$\text{avg}\left[d_i^2\right] \leq P/2, \text{ for } i = 1, ..., b-2, \tag{4.2}$$

$$\text{avg}\left[d_{b-1}^2\right] \leq P, \tag{4.3}$$

where the avg operator refers to time average; i.e., if each d_i is seen as a stream of signals $d_i[0], ..., d_i[k-1]$, then $\text{avg}(d_i) = k^{-1} \sum_{t=0}^{k-1} d_i[t]$. Then we must have, by

Parseval's relationship,

$$\frac{1}{b}\operatorname{avg}\left[\|\mathbf{X}_u\|^2\right] = \frac{1}{b}\sum_{i=0}^{b-1}\operatorname{avg}\left[\left|\tilde{d}_i\right|^2\right]$$
$$= \frac{1}{b}\left\{\operatorname{avg}\left[d_0^2\right] + \operatorname{avg}\left[d_{b-1}^2\right] + 2\sum_{i=1}^{b/2-1}\operatorname{avg}\left[d_{2i-1}^2 + d_{2i}^2\right]\right\} \le P.$$

Therefore, u may transmit k vectors $\mathbf{X}_u$, each one over b time-slots, and the average power constraint of P over the block $n = kb$ will be satisfied. The parameter k can be understood as the number of blocks of length b to which we apply the OFDM-like scheme. A node v will receive, over each sequence of b time-slots,

$$\mathbf{Y}_v = \textstyle\sum_{u\in\mathcal{I}(v)} h_{u,v}\mathbf{X}_u + \mathbf{Z}_v.$$

By applying a DFT to each block of b received signals, node v will obtain

$$\tilde{\mathbf{Y}}_v = \mathrm{DFT}(\mathbf{Y}_v) = \textstyle\sum_{u\in\mathcal{I}(v)} h_{u,v}\tilde{\mathbf{d}}_u + \mathrm{DFT}(\mathbf{Z}_v).$$

The transformation induced by the the use of the IDFT on blocks of transmit signals and the DFT on blocks of received signals is illustrated in Fig. 4.4.

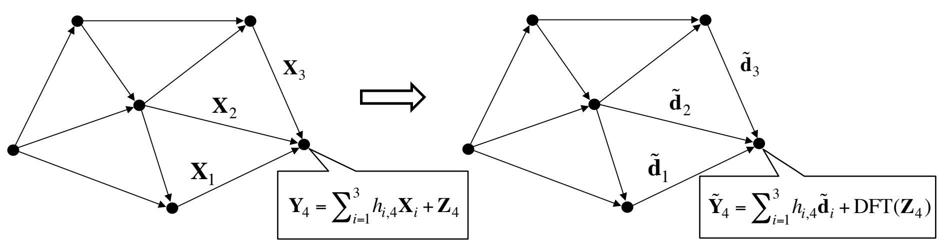


Figure 4.4: An illustration of the effect of taking the IDFT of blocks of transmit signals and the DFT of blocks of received signals.

Next, by looking at each component of $\tilde{\mathbf{Y}}_v$, we notice that we have effectively b complex-valued received signals. The additive noise on the ℓth received signal is given by

$$\mathrm{DFT}(\mathbf{Z}_v)_\ell = \frac{1}{\sqrt{b}}\sum_{i=0}^{b-1} Z_v[i]e^{-j2\pi\frac{i\ell}{b}}$$

$$= \frac{1}{\sqrt{b}} \sum_{i=0}^{b-1} Z_v[i] \cos\left(\frac{2\pi i \ell}{b}\right) - j \frac{1}{\sqrt{b}} \sum_{i=0}^{b-1} Z_v[i] \sin\left(\frac{2\pi i \ell}{b}\right). \tag{4.4}$$

By considering the real and imaginary parts of each component $\tilde{\mathbf{Y}}_{v,i}$ of $\tilde{\mathbf{Y}}_v$, for $i = 0, ..., b-1$, separately, we obtain the following $2b-2$ effective real-valued received signals:

(I) $\tilde{\mathbf{Y}}_{v,0} = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,0} + \mathrm{DFT}(\mathbf{Z}_v)_0$

(II) $\Re\left[\tilde{\mathbf{Y}}_{v,i}\right] = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2i-1} + \Re\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right]$ for $i = 1, ..., \frac{b}{2} - 1$

(III) $\Im\left[\tilde{\mathbf{Y}}_{v,i}\right] = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2i} + \Im\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right]$ for $i = 1, ..., \frac{b}{2} - 1$

(IV) $\tilde{\mathbf{Y}}_{v,b/2} = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,b-1} + \mathrm{DFT}(\mathbf{Z}_v)_{b/2}$

(V) $\Re\left[\tilde{\mathbf{Y}}_{v,i}\right] = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2(b-i)-1} + \Re\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right]$ for $i = \frac{b}{2} + 1, ..., b-1$

(VI) $\Im\left[\tilde{\mathbf{Y}}_{v,i}\right] = -\sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2(b-i)} + \Im\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right]$ for $i = \frac{b}{2} + 1, ..., b-1$

However, from the conjugate symmetry of $\mathrm{DFT}(\mathbf{Z}_v)$ (since $\mathbf{Z}_v$ is a real-valued vector), we have that $\Re\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right] = \Re\left[\mathrm{DFT}(\mathbf{Z}_v)_{b-i}\right]$ and $\Im\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right] = -\Im\left[\mathrm{DFT}(\mathbf{Z}_v)_{b-i}\right]$, for $i = 1, 2, ..., b-1$, and all the received signals in (V) and (VI) are repetitions (up to a change of sign) of the received signals in (II) and (III). Therefore, we conclude that we have effectively b distinct real-valued received signals with additive noise (i.e., the channels from (I), (II), (III) and (IV), which are the effective channel outputs shown in Fig. 4.3). It is important to notice that the additive noise terms are *dependent* across these b received signals. We also point out that the stricter power constraint in (4.2) will not constitute a problem. The reason is that the effective received signals during the network uses corresponding to (4.2), given by (II) and (III), will be shown in the next Section to be subject to a noise with variance $\sigma_v^2/2$ as opposed to σ_v^2. Thus, the effective SNR is still P/σ_v^2.

4.1.2 Noise mixture converges to Gaussian Noise

In this section, we show that the additive noise terms of the effective received signals we obtained in the previous Section approximate a Gaussian distribution as b gets large. Throughout this chapter, we will write $X_n \xrightarrow{d} X$ to denote that the random variables $X_1, X_2, ...$ converge in distribution to X, and $X_n \xrightarrow{p} X$ to denote that the random variables $X_1, X_2, ...$ converge in probability to X. We will use the following classical result.

Theorem 4.4 (Lindeberg's Central Limit Theorem [9]) *Suppose that for each $b = 1, 2, ...$, the random variables $Y_{b,1}, Y_{b,2}, ..., Y_{b,b}$ are independent. In addition, suppose that, for all b and $i \leq b$, $E[Y_{b,i}] = 0$, and let*

$$s_b^2 = \sum_{i=1}^{b} E\left[Y_{b,i}^2\right]. \tag{4.5}$$

Then, if for all $\varepsilon > 0$, Lindeberg's condition

$$\frac{1}{s_b^2} \sum_{i=1}^{b} E\left(Y_{b,i}^2 \, \mathbb{1}\left\{|Y_{b,i}| \geq \varepsilon s_b\right\}\right) \to 0 \text{ as } b \to \infty \tag{4.6}$$

holds, we have that

$$\frac{\sum_{i=1}^{b} Y_{b,i}}{s_b} \xrightarrow{d} \mathcal{N}(0, 1).$$

Lindeberg's CLT can be used to prove the following lemma.

Lemma 4.1 *Let $Z[0], Z[1], Z[2], ...$ be i.i.d. random variables that are zero-mean, have variance σ^2 and let*

$$N_b = \frac{1}{\sqrt{b}} \sum_{i=0}^{b-1} Z[i] \cos\left(\frac{2\pi i \ell_b}{b}\right), \tag{4.7}$$

for some $\ell_b \in \{1, ..., b-1\} \setminus \{b/2\}$. Then, N_b converges in distribution to $\mathcal{N}(0, \sigma^2/2)$ as $b \to \infty$.

Proof: We start by letting $Y_{b,i+1} = Z[i]\cos\left(\frac{2\pi i\ell_b}{b}\right)$, for $i = 0, 1, ..., b-1$. Then, by following (4.5), we have

$$\begin{aligned}
s_b^2 &= \sum_{i=1}^{b} E\left[Y_{b,i}^2\right] = \sum_{i=0}^{b-1} E\left[Z[i]^2\right]\cos^2\left(\frac{2\pi i\ell_b}{b}\right) \\
&= \frac{\sigma^2}{4}\sum_{i=0}^{b-1}\left(e^{j2\pi\ell_b\frac{i}{b}} + e^{-j2\pi\ell_b\frac{i}{b}}\right)^2 \\
&= \frac{\sigma^2}{4}\sum_{i=0}^{b-1}\left(e^{j4\pi\ell_b\frac{i}{b}} + e^{-j4\pi\ell_b\frac{i}{b}} + 2\right) \\
&= \frac{b\sigma^2}{2} + \frac{\sigma^2}{4}\sum_{i=0}^{b-1}\left(e^{j4\pi\ell_b\frac{i}{b}} + e^{-j4\pi\ell_b\frac{i}{b}}\right) \\
&= \frac{b\sigma^2}{2} + \frac{\sigma^2(1-e^{j4\pi\ell_b})}{4(1-e^{j4\pi\ell_b\frac{1}{b}})} + \frac{\sigma^2(1-e^{-j4\pi\ell_b})}{4(1-e^{-j4\pi\ell_b\frac{1}{b}})} = \frac{b\sigma^2}{2}.
\end{aligned}$$

The last equality follows because $e^{-j4\pi\ell_b} = 1$ and $e^{j4\pi\ell_b\frac{1}{b}} \neq 1$ for any $\ell_b \in \{1, ..., b-1\} \setminus \{b/2\}$. Consider any sequence i_b, for $b = 1, 2, ...,$ such that $i_b \in \{1, ..., b\}$, and any $\delta > 0$. Then we have that

$$\begin{aligned}
\Pr\left(U_{b,i_b} < \delta\right) &\geq \Pr\left(|Y_{b,i_b}| < \varepsilon\sigma\sqrt{b/2}\right) \\
&\geq \Pr\left(|Z[i_b-1]| < \varepsilon\sigma\sqrt{b/2}\right) \\
&= \Pr\left(|Z[1]| < \varepsilon\sigma\sqrt{b/2}\right) \to 1, \text{ as } b \to \infty,
\end{aligned}$$

which means that $U_{b,i_b} \xrightarrow{p} 0$ as $b \to \infty$. Moreover, we have that $|U_{b,i_b}| = U_{b,i_b} \leq Z[i_b-1]^2$ for all b, and $E\left[Z[i_b-1]^2\right] = \sigma^2 < \infty$. Next, we notice that $Z[i-1] \sim Z[1]$ for all $i \geq 1$, which implies that, for any $\tau > 0$,

$$\Pr\left[|U_{b,i_b}| \geq \tau\right] \leq \Pr\left[Z[i_b-1]^2 \geq \tau\right] = \Pr\left[Z[1]^2 \geq \tau\right].$$

Thus, we can apply the version of the Dominated Convergence Theorem described in [9, pp. 338-339], to conclude that $E[U_{b,i_b}] \to 0$ as $b \to \infty$. We conclude that

$$\frac{1}{s_b^2}\sum_{i=1}^{b} E\left(Y_{b,i}^2\,\mathbb{1}\left\{|Y_i| \geq \varepsilon s_b\right\}\right)$$

$$= \frac{2}{\sigma^2 b} \sum_{i=1}^{b} E\left[U_{b,i}\right] \leq \frac{2}{\sigma^2} \max_{1\leq i\leq b} E\left[U_{b,i}\right] \to 0 \text{ as } b \to \infty,$$

and Lindeberg's condition (4.6) is satisfied for any $\varepsilon > 0$. Hence, from Theorem 4.4, we have that

$$\frac{\sum_{i=1}^{b} Y_{b,i}}{\sigma\sqrt{b/2}} \xrightarrow{d} \mathcal{N}(0,1) \quad \Longrightarrow \quad N_b = \frac{\sigma}{\sqrt{2}} \frac{\sum_{i=1}^{b} Y_{b,i}}{\sigma\sqrt{b/2}} \xrightarrow{d} \mathcal{N}(0,\sigma^2/2).$$

■

Now consider the additive noise term in (II). It is the real part of (4.4), which, by Lemma 4.1, converges in distribution to $\mathcal{N}(0,\sigma_v^2/2)$, as $b \to \infty$. Moreover, it is easy to see that Lemma 4.1 can be restated with sines replacing the cosines, and the same result will hold. Thus, the additive noise in (III) also converges in distribution to $\mathcal{N}(0,\sigma_v^2/2)$. Finally, for the received signals in (I) and (IV), it is easy to see that the additive noise in (4.4) only has a real component, and by the usual Central Limit Theorem, it converges in distribution to $\mathcal{N}(0,\sigma_v^2)$.

Notice that, since in (4.2) we restricted the power used in the network uses corresponding to (II) and (III) to $P/2$, all of our effective channels have the same SNR they would have if the transmit signals had power P and the noise variance σ_v^2. Therefore, for the network uses corresponding to (II) and (III), we can instead assume that the power constraint is P, but all nodes divide their transmit signals by $\sqrt{2}$ prior to transmission, and multiply their received signals by $\sqrt{2}$. This yields the following b effective channels,

(I) $\quad \tilde{\mathbf{Y}}_{v,0} = \sum_{u\in\mathcal{I}(v)} h_{u,v}\mathbf{d}_{u,0} + \mathrm{DFT}(\mathbf{Z}_v)_0$

(II′) $\quad \sqrt{2}\cdot\Re\left[\tilde{\mathbf{Y}}_{v,i}\right] = \sum_{u\in\mathcal{I}(v)} h_{u,v}\mathbf{d}_{u,2i-1} + \sqrt{2}\cdot\Re\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right] \qquad$ for $i = 1, ..., \frac{b}{2}-1$

(III′) $\quad \sqrt{2}\cdot\Im\left[\tilde{\mathbf{Y}}_{v,i}\right] = \sum_{u\in\mathcal{I}(v)} h_{u,v}\mathbf{d}_{u,2i} + \sqrt{2}\cdot\Im\left[\mathrm{DFT}(\mathbf{Z}_v)_i\right] \qquad$ for $i = 1, ..., \frac{b}{2}-1$

(IV) $\quad \tilde{\mathbf{Y}}_{v,b/2} = \sum_{u\in\mathcal{I}(v)} h_{u,v}\mathbf{d}_{u,b-1} + \mathrm{DFT}(\mathbf{Z}_v)_{b/2}$

all of which have input power constraint P and additive noise with variance σ_v^2. The diagram describing the steps that create the effective channel from Fig. 4.3 can then be updated as shown in Fig. 4.5. We notice that the transformation

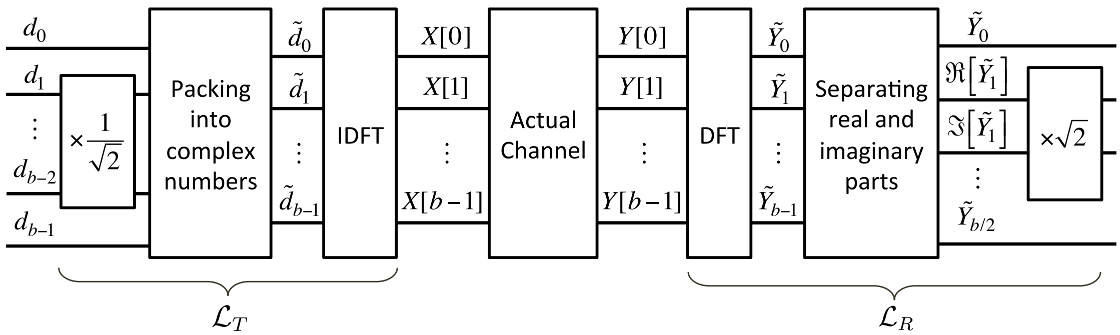


Figure 4.5: Diagram of the steps that create the effective channel. The overall transformation between the effective channel inputs and the actual channel inputs is represented by a linear transformation $\mathcal{L}_T$ and the overall transformation between the actual channel outputs and the effective channel outputs is represented by a linear transformation $\mathcal{L}_R$.

between the b inputs to the effective channels and the b inputs to the actual channel is in fact a 2-norm-preserving linear transformation, which we call $\mathcal{L}_T$. Similarly, the transformation between the b outputs of the actual channel and the b output of our effective channel is also a 2-norm-preserving linear transformation, which we call $\mathcal{L}_R$.

Now consider any sequence ℓ_b, $b = 1, 2, ...$, where $\ell_b \in \{0, ..., b-1\}$. Let N_{b,ℓ_b} now be the additive noise term of the ℓ_bth effective channel above. The sequence indices $b \in \{1, 2, ...\}$ can be partitioned into four sets J_1, J_2, J_3 and J_4, according to whether N_{b,ℓ_b} corresponds to the additive noise of an effective channel of type (I), (II′), (III′) or (IV). According to Lemma 4.1, if J_2 or J_3 are infinite sets, the subsequence that they define $\{N_{b,\ell_b}\}_{b\in J_2}$ or $\{N_{b,\ell_b}\}_{b\in J_3}$ converge in distribution to $\mathcal{N}(0, \sigma_v^2)$ (after the multiplication by $\sqrt{2}$). Moreover, as we noticed above, from the usual Central Limit Theorem, it follows that if J_1 or J_4 are infinite sets, the subsequences defined by $\{N_{b,\ell_b}\}_{b\in J_1}$ or $\{N_{b,\ell_b}\}_{b\in J_4}$ also converge in distribution to $\mathcal{N}(0, \sigma_v^2)$. Therefore, we conclude that, for any arbitrary sequence ℓ_b, $b = 1, 2, ...,$ where $\ell_b \in \{0, ..., b-1\}$, N_{b,ℓ_b} converges in distribution to $\mathcal{N}(0, \sigma_v^2)$.

4.1.3 Interleaving and Outer Code

In this Section, we address the fact that, as we mentioned before, the additive noise at node v in the b effective network uses are dependent of each other. In order to handle this dependence, we consider using the network for a total of bk times, performing the OFDM-like approach from Section 4.1.1 within each block of b time steps. Then, by interleaving the symbols, it is possible to view the result as b blocks of k network uses. This idea is illustrated in Fig. 4.6. Notice that,

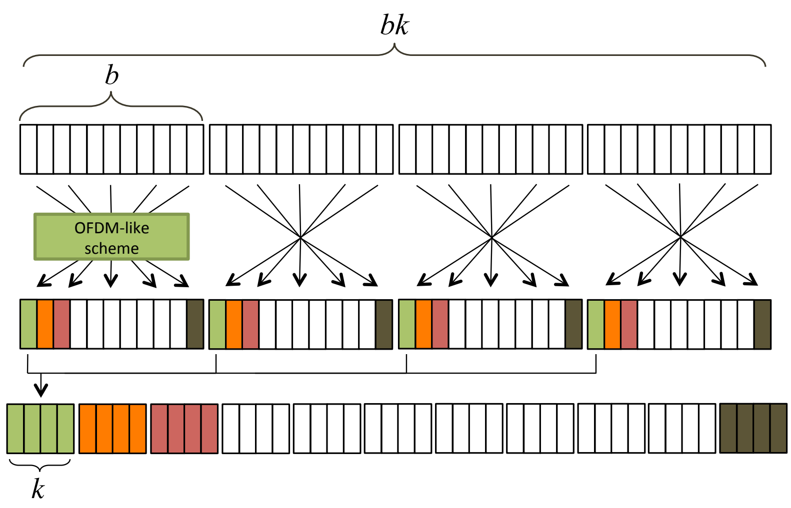


Figure 4.6: Interleaving the effective network uses obtained from the OFDM-like scheme.

within each block of k network uses, the additive noises are i.i.d., but they are dependent among distinct blocks. Intuitively, this makes each of these blocks of k network uses suitable for the application of a coding scheme C_k with block length k. The dependence between the noises of different blocks of length k will be handled at the end of this Section, through the application of a random outer code. Then, by considering a mutual-information argument, we will show that the performance of the resulting coding scheme on the wireless network with non-Gaussian noises is essentially the same as the performance of the original coding scheme C_k on the AWGN version of the network.

Example 4.2. Consider a simple relay channel, defined by a graph $G = (V, E)$, where $V = \{s, v, d\}$ and $E = \{(s, v), (s, d), (v, d)\}$. Suppose we have a coding scheme C_k of block length k and rate R for this network. The operations performed by the nodes under this scheme at time t can be illustrated as in Fig. 4.7. Now sup-

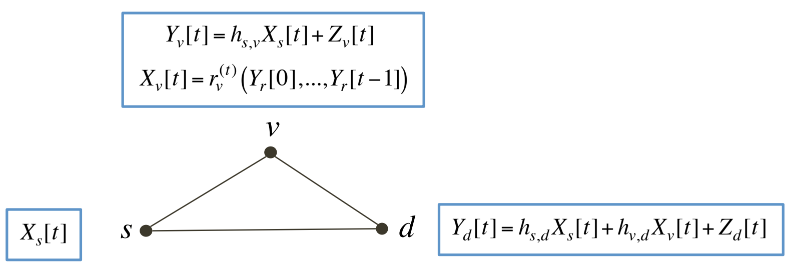


Figure 4.7: Illustration of a coding scheme C_k for a relay channel at time t. At times $t = 0, \ldots, k-1$, the source s transmits $X_s[t]$, which is the $(t+1)$th entry of the chosen codeword $f(w)$, for $w \in \{1, \ldots, 2^{kR}\}$. The relay v applies the relaying function $r_v^{(t)}$ to the signals it received up to time $t-1$, to obtain $X_v[t]$, which is then transmitted. The destination d waits until the end of the length-k block and applies the decoding function g to the block of received signals $(Y_d[0], \ldots, Y_d[k-1])$.

pose we want to apply the OFDM-like scheme and the interleaving procedure to this coding scheme C_k. In essence, b versions of this coding scheme will be simultaneously used. Encoding, relaying and decoding functions are applied "in parallel" for each of the b coding schemes, as shown in Fig. 4.8(a) in detail. First,

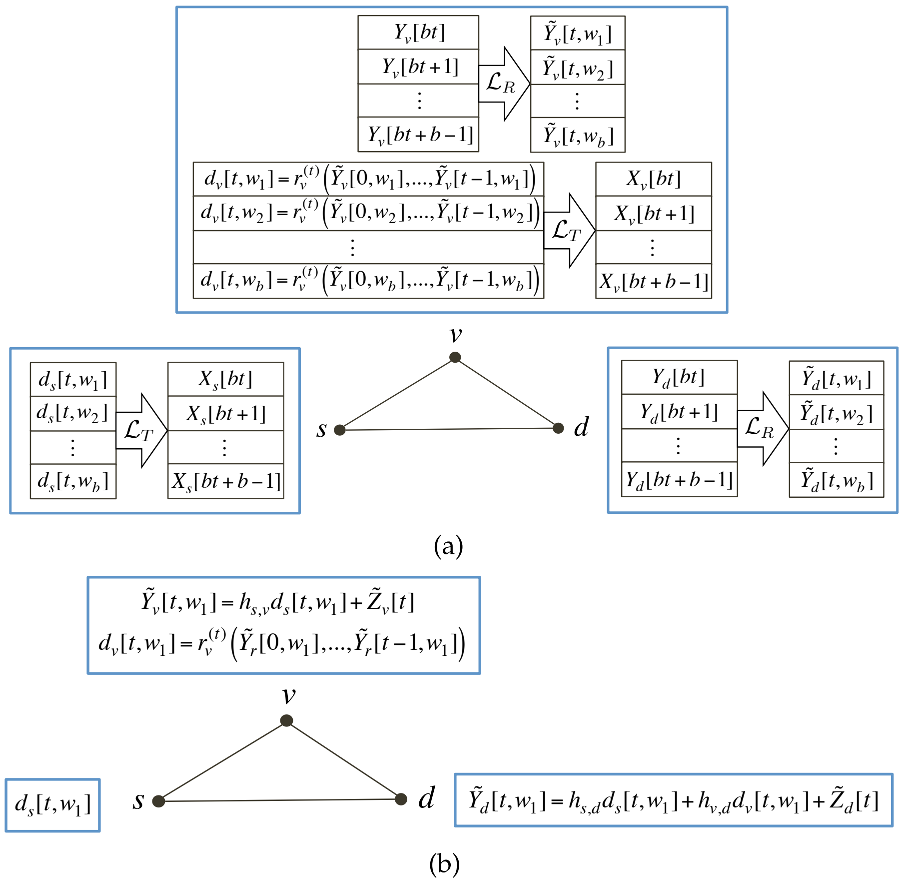


Figure 4.8: (a) Illustration of the source, relay and destination operations, after applying the OFDM-like scheme and the interleaving procedure to a coding scheme C_k. The source s chooses b messages $w_1, ..., w_b \in \{1, ..., 2^{kR}\}$. This yields b codewords $f(w_1), ..., f(w_b)$ which form the inputs $d_s(t, w_1), ..., d_s(t, w_b)$, for $t = 0, ..., k-1$, to the effective channel. At times $bt, bt+1, ..., bt+b-1$ for $t = 0, ..., k-1$, s transmits the b signals that result from applying $\mathcal{L}_T$ to the vector $(d_s[t, w_1], ..., d_s[t, w_b])$. At time $bt+b-1$, for $t = 0, ..., k-1$, the relay v finishes receiving the signals of a length-b block and can apply $\mathcal{L}_R$ to them. At time bt, for $t = 1, ..., k-1$, using all previously received effective signals, the relay can use relaying function $r_v^{(t)}$ b times to obtain $(d_r[t, w_1], ..., d_r[t, w_b])$. After applying $\mathcal{L}_T$ to this vector, the relay obtains the b signals to be transmitted at times $bt, bt+1, ..., bt+b-1$. The destination, at time $bt+b-1$, for $t = 0, ..., k-1$, finishes receiving the signals of a length-b block and can apply $\mathcal{L}_R$ to them. (b) Effective network experienced by the signals indexed by w_1.

b codewords $f(w_1), ..., f(w_b)$ are chosen at the source. At times $bt, bt+1, ..., bt+b-1$ for $t = 0, ..., k-1$, the source transmits the b signals obtained by applying $\mathcal{L}_T$ to the vector formed by the $(t+1)$th entries of these b codewords. Relay v, in turn, after applying $\mathcal{L}_R$ to the received signals at times $bt, bt+1, ..., bt+b-1$ for $t = 0, ..., k-1$, can use the relaying function $r_v^{(t+1)}$ a total of b times in order to obtain a length-b vector that goes through the transformation $\mathcal{L}_T$ to yield the b signals to be transmitted at times $b(t+1), b(t+1)+1, ..., b(t+1)+b-1$ for $t = 0, ..., k-2$. The destination, after applying $\mathcal{L}_R$ to each block of b received signals, obtains b sequences of n received signals, and can apply its decoding function to each of these sequences. As shown in Fig. 4.8(a), the application of the transformations $\mathcal{L}_T$ and $\mathcal{L}_R$ can be seen as creating b effective networks, where the transmit and received signals of the ith effective network are given by $d[t, w_i]$ and $\tilde{Y}[t, w_i]$ respectively.

The purpose of the interleaving procedure can be understood if we focus on what occurs to the signals in one of these effective networks, say the one indexed by w_1. By absorbing the transformations $\mathcal{L}_T$ and $\mathcal{L}_R$ into the network, and viewing the $d[t, w_1]$s and $\tilde{Y}[t, w_1]$s as inputs and outputs of the network, the network that is effectively experienced by the signals indexed by w_1 is shown in Fig. 4.8(b). Notice that the effective network in Fig. 4.8(b) is the same as the original network in Fig. 4.7 but with different additive noise terms $\tilde{N}_v[t]$ and $\tilde{N}_d[t]$. These effective noise terms are in fact i.i.d., since the operations $\mathcal{L}_T$ and $\mathcal{L}_R$ are applied to blocks of signals with different indices $w_1, w_2, ..., w_b$, and this cannot create dependence between effective noises $\tilde{N}_v[t]$ and $\tilde{N}_v[t']$ (or $\tilde{N}_d[t]$ and $\tilde{N}_d[t']$) for $t \neq t'$, since they both correspond to received signals indexed by w_1. Therefore, we are essentially applying coding scheme C_k in b parallel effective relay channels, each of which has i.i.d. noises at v and d. ■

Since from the statement of Theorem 4.1, the rate tuple $\mathbf{R}$ is achievable by coding schemes with finite reading precision, we may assume that we have a sequence of coding schemes C_k (with block length k and rate tuple $\mathbf{R}$) with finite reading precision ρ_k, whose error probability when used on the AWGN network is $\epsilon_k = P_{\text{error}}(C_k)$, and satisfies $\epsilon_k \to 0$ as $k \to \infty$. Now, consider applying this code over each of the b blocks of length-k that we obtained from the interleaving, as demonstrated in Example 1. Over each block of length k, the noises at all nodes are independent and i.i.d. over time, and, if b is chosen fairly large, they are very close to Gaussian in distribution, and, intuitively, the error probability we obtain should be close to ϵ_k. The actual distribution of the additive noise at each of these b length-k blocks is given by the noise terms in (I), (II'), (III') and (IV). For $\ell = 0, ..., b-1$, we let $\epsilon_{k,b}^{(\ell)}$ be the error probability of coding scheme C_k applied on the $(\ell+1)$th such block, for which the i.i.d. additive noise at node v is given by

$$N_{v,b}^{(\ell)} = \begin{cases} \mathrm{DFT}(\mathbf{Z}_v)_0 & \text{for } \ell = 0 \\ \sqrt{2} \cdot \Re\left[\mathrm{DFT}(\mathbf{Z}_v)_\ell\right] & \text{for } \ell = 1, ..., \frac{b}{2} - 1 \\ \sqrt{2} \cdot \Im\left[\mathrm{DFT}(\mathbf{Z}_v)_{(1+\ell-b/2)}\right] & \text{for } \ell = \frac{b}{2}, ..., b-2 \\ \mathrm{DFT}(\mathbf{Z}_v)_{b/2} & \text{for } \ell = b-1. \end{cases}$$

Then, for each value of b, we let $\epsilon_{k,b} = \max_{0\le\ell\le b-1} \epsilon_{k,b}^{(\ell)}$, and $\ell_b = \arg\max_{0\le\ell\le b-1} \epsilon_{k,b}^{(\ell)}$, which defines a sequence ℓ_b, $b = 1, 2, ...$ like the ones considered at the end of Section 4.1.2.

We let $\mathbf{N}_b \in \mathbb{R}^{k|V|}$ be the random vector associated with the effective additive noises at all nodes in V during the ℓ_bth length-k block assuming that we performed the OFDM-like scheme in blocks of size b; i.e.,

$$\mathbf{N}_b = \left(N_{v,b}^{(\ell_b)}[t]\right)_{v\in V, 0\le t\le k-1}.$$

Since each component of $\mathbf{N}_b$ is independent and they all converge in distribution to a zero-mean Gaussian random variable, we have that $\mathbf{N}_b$ converges in distribution to a Gaussian random vector. We let $\mathbf{N}$ be this limiting distribution, and we know that the component of $\mathbf{N}$ corresponding to node v and time t is distributed as $\mathcal{N}(0, \sigma_v^2)$, for any $t \in \{0, ..., k-1\}$. Now notice that, if we fix the messages chosen at the sources to be $\mathbf{w} = (w_1, w_2, ..., w_K) \in \prod_{i=1}^{K}\{1, ..., 2^{kR_i}\}$, then, whether C_k makes an error is only a deterministic function of $\mathbf{N}_b$. Therefore, for each $\mathbf{w} \in \prod_{i=1}^{K}\{1, ..., 2^{kR_i}\}$, we can define an error set $A_{\mathbf{w}}$, corresponding to all realizations of $\mathbf{N}_b$ that cause coding scheme C_k to make an error. It is important to notice that $A_{\mathbf{w}}$ is independent of the actual joint distribution of the noise terms; it only depends on the coding scheme C_k. Then we can write

$$\epsilon_{k,b} = 2^{-k\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \Pr\left[\mathbf{N}_b \in A_{\mathbf{w}}\right] \tag{4.8}$$

and also

$$\epsilon_k = 2^{-k\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \Pr\left[\mathbf{N} \in A_{\mathbf{w}}\right]. \tag{4.9}$$

Our first goal is to show that $\epsilon_{k,b} \to \epsilon_k$ as $b \to \infty$. Recall that a Borel set $A \subset \mathbb{R}^m$ is said to be a μ-continuity set for some probability measure μ on $\mathbb{R}^m$, if $\mu(\partial A) = 0$, where ∂A is the boundary of A (see, for example, [9]). Next, we state the following classical result, which provides an alternative characterization of convergence in distribution.

Theorem 4.5 (Portmanteau Theorem [8]) *Suppose we have a sequence of random vectors $\mathbf{N}_b \in \mathbb{R}^{k|V|}$ and another random vector $\mathbf{N} \in \mathbb{R}^{k|V|}$. Let μ_b and μ be the probability measures on $\mathbb{R}^{k|V|}$ associated to $\mathbf{N}_b$ and $\mathbf{N}$ respectively. Then $\mathbf{N}_b$ converges in distribution to $\mathbf{N}$ if and only if*

$$\lim_{b\to\infty} \mu_b(A) = \mu(A)$$

for all μ-continuity sets A.

Let μ be the probability measure on $\mathbb{R}^{k|V|}$ associated to $\mathbf{N}$. Then, if we show that $A_{\mathbf{w}}$ is a μ-continuity set for each choice of messages $\mathbf{w}$, from Theorem 4.5, the fact that $\mathbf{N}_b \xrightarrow{d} \mathbf{N}$ will imply that

$$\lim_{b\to\infty} \Pr\left[\mathbf{N}_b \in A_{\mathbf{w}}\right] = \Pr\left[\mathbf{N} \in A_{\mathbf{w}}\right] \tag{4.10}$$

for each $\mathbf{w}$, and from (4.8) and (4.9) we will conclude that $\epsilon_{k,b} \to \epsilon_k$ as $b \to \infty$. This is in fact what we do in the following Lemma.

Lemma 4.2 *Suppose we have a coding scheme C with block length k, rate tuple $\mathbf{R}$, and finite reading precision ρ. Then, for any choice of messages $\mathbf{w} \in \prod_{i=1}^{K}\{1,...,2^{kR_i}\}$, the error set $A_{\mathbf{w}}$ is a μ-continuity set.*

Proof: Fix some choice of messages $\mathbf{w}$. We will use the fact that C has finite reading precision ρ to show that our set $A_{\mathbf{w}}$ and its complement $A_{\mathbf{w}}^c = \mathbb{R}^{k|V|} \setminus A_{\mathbf{w}}$ can be represented as a countable union of disjoint convex sets, which will then imply the μ-continuity. Recall from Definition 4.1 that, in a coding scheme with finite reading precision ρ, a node v only has access to $\lfloor Y_v \rfloor_\rho$. Thus, we will call $\lfloor Y_v \rfloor_\rho$ the effective received signal at v. The set

$$\mathcal{Y} = \left\{(y_1,...,y_{k|V|}) \in \mathbb{R}^{k|V|} : y_i = \lfloor y_i \rfloor_\rho, i = 1,...,k|V|\right\}$$

can be understood as the set of all possible values of the effective received signals at all nodes in V during a length-k block. It is clear that $\mathcal{Y}$ is a countable set for any finite ρ.

Notice that, for our fixed choice of messages $\mathbf{w}$, the vector $\mathbf{y} \in \mathcal{Y}$ corresponding to the effective received signals at all nodes during the length-k block is a deterministic function of the value of all the noises in the network during the length-k block, $\mathbf{z} \in \mathbb{R}^{k|V|}$. Therefore, for each $\mathbf{y} \in \mathcal{Y}$, we define $Q(\mathbf{y}) \subset \mathbb{R}^{k|V|}$ to be the set of noise realizations $\mathbf{z}$ that will result in $\mathbf{y}$ being the effective received signals. In Lemma B.1 in the appendix, we prove that $Q(\mathbf{y})$ is a convex set. Furthermore, in Lemma B.2, we prove that, for any convex set S, $\lambda(\partial S) = 0$, where λ is the Lebesgue measure. Since our measure μ is absolutely continuous (as $\mathbf{N}$ is jointly Gaussian), it follows by definition [9] that

$$\lambda(S) = 0 \Rightarrow \mu(S) = 0,$$

for any Borel set S. Thus, since $\lambda(\partial Q(\mathbf{y})) = 0$, we have that $\mu(\partial Q(\mathbf{y})) = 0$. This, in turn, clearly implies that

$$\mu\left(Q(\mathbf{y})^\circ\right) = \mu\left(\overline{Q(\mathbf{y})}\right) = \mu\left(Q(\mathbf{y})\right), \tag{4.11}$$

where we use S° to represent the interior of a set S and $\overline{S}$ to represent its closure. Next, let $\mathcal{Y}_{A_\mathbf{w}} = \{\mathbf{y} \in \mathcal{Y} : A_\mathbf{w} \cap Q(\mathbf{y}) \neq \emptyset\}$. Notice that all noise realizations $\mathbf{z} \in Q(\mathbf{y})$ will cause all nodes and, in particular, the destination nodes to receive the exact same effective signals. Therefore, it must be the case that, if $A_\mathbf{w} \cap Q(\mathbf{y}) \neq \emptyset$, then $Q(\mathbf{y}) \subset A_\mathbf{w}$, which implies that

$$\bigcup_{\mathbf{y} \in \mathcal{Y}_{A_\mathbf{w}}} Q(\mathbf{y}) = A_\mathbf{w}.$$

Moreover, it is obvious that any noise realization must belong to exactly one set $Q(\mathbf{y})$, and we have

$$\bigcup_{\mathbf{y} \in \mathcal{Y} \setminus \mathcal{Y}_{A_\mathbf{w}}} Q(\mathbf{y}) = A_\mathbf{w}^c.$$

Finally, we obtain

$$\mu\left(A_\mathbf{w}^\circ\right) \overset{(i)}{\geq} \mu\left(\bigcup_{\mathbf{y} \in \mathcal{Y}_{A_\mathbf{w}}} Q(\mathbf{y})^\circ\right)$$

$$
\begin{aligned}
&\overset{(ii)}{=} \sum_{\mathbf{y}\in\mathcal{Y}_{A_{\mathbf{w}}}} \mu\left(Q(\mathbf{y})^\circ\right) \overset{(iii)}{=} \sum_{\mathbf{y}\in\mathcal{Y}_{A_{\mathbf{w}}}} \mu\left(Q(\mathbf{y})\right) \\
&= 1 - \sum_{\mathbf{y}\in\mathcal{Y}\setminus\mathcal{Y}_{A_{\mathbf{w}}}} \mu\left(Q(\mathbf{y})\right) = 1 - \sum_{\mathbf{y}\in\mathcal{Y}\setminus\mathcal{Y}_{A_{\mathbf{w}}}} \mu\left(Q(\mathbf{y})^\circ\right) \\
&= 1 - \mu\left(\bigcup_{\mathbf{y}\in\mathcal{Y}\setminus\mathcal{Y}_{A_{\mathbf{w}}}} Q(\mathbf{y})^\circ\right) \geq 1 - \mu\left((A_{\mathbf{w}}^c)^\circ\right) \\
&= \mu\left(((A_{\mathbf{w}}^c)^\circ)^c\right) = \mu\left(\overline{A_{\mathbf{w}}}\right),
\end{aligned}
$$

where (i) follows since, for sets $B_1, B_2, ...$, $(\cup_i B_i)^\circ \supseteq \cup_i B_i^\circ$, (ii) follows from the countability of $\mathcal{Y}_{A_{\mathbf{w}}}$ and the fact that $Q(\mathbf{y}_1) \cap Q(\mathbf{y}_2) = \emptyset$ for $\mathbf{y}_1 \neq \mathbf{y}_2$, and (iii) follows from (4.11). We conclude that $\mu(\partial A_{\mathbf{w}}) = \mu\left(\overline{A_{\mathbf{w}}}\right) - \mu\left(A_{\mathbf{w}}^\circ\right) = 0$; i.e., $A_{\mathbf{w}}$ is a μ-continuity set. ■

From our previous discussion, we conclude that $\epsilon_{k,b} \to \epsilon_k$ as $b \to \infty$. We then see that we can apply code C_k within each of the b blocks of length k and obtain a probability of error (within that block) that tends to ϵ_k as $b \to \infty$. However, since we have a total of b blocks of length k, we make an error if we make an error in any of the b blocks of length k. It turns out that a simple union bound does not work here, since the error probability would be of the form $b\epsilon_{k,b}$ and we would not be able to guarantee that it tends to 0 as b and k go to infinity. Instead we consider using an outer code for each source-destination pair.

The idea is to apply coding scheme C_k to each of the b length-k blocks, and then view this as creating a discrete channel for each source-destination pair. More specifically, for each length-bk block, source s_j chooses a *symbol* (rather than a message) from $\{1, ..., 2^{kR_j}\}^b$ and transmits the b corresponding codewords from C_k. Then destination d_j will apply the decoder from code C_k inside each length-k block and obtain an output symbol also from $\{1, ..., 2^{kR_j}\}^b$. Notice that, by viewing the input to bk network uses as a single input to this discrete channel, we make sure we have a discrete *memoryless* channel, and we can use the Channel Coding Theorem. We can view W_j^b and $\hat{W}_j^b$ as the discrete input and output of the channel between s_j and d_j. We will then construct a code (whose rate is to be determined) for this discrete channel between s_j and d_j by picking each entry uniformly at random from $\{1, ..., 2^{kR_j}\}^b$. Then, source-destination pair (s_j, d_j) can achieve rate

$$\begin{aligned}
\frac{1}{bk} I(W_j^b; \hat{W}_j^b) &= \frac{1}{bk}\left(H(W_j^b) - H(W_j^b|\hat{W}_j^b)\right) \\
&\geq R_j - \frac{1}{bk}\sum_{\ell=0}^{b-1} H(W_j[\ell]|\hat{W}_j[\ell]) \\
&\overset{(i)}{\geq} R_j - \frac{1}{k}(1 + \epsilon_{k,b}^{(\ell)} k R_j) \\
&\geq R_j - \frac{1}{k}(1 + \epsilon_{k,b} k R_j) \\
&= R_j(1 - \epsilon_{k,b}) - \frac{1}{k},
\end{aligned}$$

where (i) follows from Fano's Inequality, since, within the ℓth length-k block, we are applying code C_k and we have an average error probability of at most $\epsilon_{k,b}^{(\ell)}$ (it should in fact be less than $\epsilon_{k,b}^{(\ell)}$ since we are only considering the error event $W_j[\ell] \neq \hat{W}_j[\ell]$ and $\epsilon_{k,b}^{(\ell)}$ refers to the union of these events for all source-destination pairs).

We conclude that, by choosing b and k sufficiently large, it is possible for each source-destination pair to achieve arbitrarily close to rate R_j. Thus, our coding scheme can achieve arbitrarily close to the rate tuple $\mathbf{R}$. This concludes the proof of Theorem 4.2.

4.1.4 Optimality of Coding Schemes with Finite Reading Precision

In this Section, we prove Theorem 4.3. This theorem implies that, if we restrict ourselves to coding schemes with finite reading precision, and allow the reading precision to tend to infinity along the sequence of coding schemes, we can achieve any point in the capacity region of an AWGN wireless network, thus characterizing the optimality of coding schemes with finite reading precision for AWGN networks. We start by considering a sequence of coding schemes C_n (with infinite reading precision) that achieves rate tuple $\mathbf{R}$ on an AWGN K-unicast wireless network. We will build a sequence of coding schemes $C_n^\star$ with finite reading precision that also achieves rate tuple $\mathbf{R}$ on the same K-unicast wireless network.

Let ϵ_n be the error probability of coding scheme C_n, which achieves rate tuple $\mathbf{R}$ on the AWGN K-unicast wireless network. From Definition 1.3, we have that $\epsilon_n \to 0$ as $n \to \infty$. For any fixed n, we will first build a sequence of coding schemes with finite reading precision $C_{n,m}^\star$, $m = 1, 2, ...$, such that code $C_{n,m}^\star$ has error probability $\epsilon_{n,m}$, where $\epsilon_{n,m} \to \epsilon_n$ as $m \to \infty$. This will allow us to choose a finite m for which $\epsilon_{n,m}$ is arbitrarily close to ϵ_n.

Notice that, from Definition 1.1, relaying and decoding functions should be deterministic. However, in order to construct coding scheme $C_{n,m}^\star$, we will first assume that the relaying and decoding functions are allowed to be randomized, and later we will derandomize the constructed coding scheme. Recall that, from Definition 1.1, coding scheme C_n is comprised of encoding functions $\{f_i : 1 \leq i \leq K\}$, relaying functions $\left\{r_v^{(t)} : v \in V, 1 \leq t \leq n\right\}$ and decoding functions

$\{g_i : 1 \le i \le K\}$. We will build $C^\star_{n,m}$ from C_n by using the same encoding functions f_i, $i = 1, ..., K$, and replacing the relaying functions with

$$\tilde{r}_v^{(t)}\left(Y_v[1], ..., Y_v[t-1]\right) \triangleq r_v^{(t)}\left(\tilde{Y}_v^{(m)}[1], ..., \tilde{Y}_v^{(m)}[t-1]\right)$$

for $1 \le t \le n$ and $v \in V$, and replacing the decoding functions with

$$\tilde{g}_i\left(Y_v[1], ..., Y_v[n]\right) \triangleq g_i\left(\tilde{Y}_v^{(m)}[1], ..., \tilde{Y}_v^{(m)}[n]\right),$$

for $1 \le i \le K$, where we define

$$\tilde{Y}_v^{(m)}[t] = \lfloor Y_v[t] \rfloor_m + U_v^{(m)}[t], \tag{4.12}$$

for $v \in V$ and $1 \le t \le n$, where $U_v^{(m)}[1], ..., U_v^{(m)}[n]$ are independent uniform random variables drawn from $\left(-2^{-m-1}, 2^{-m-1}\right)$, independent from all signals and noises in the network. Notice that, since the relaying functions $r_v^{(t)}$ satisfy the power constraint in Definition 1.1, so will the new relaying functions $\tilde{r}_v^{(t)}$. In order to relate the error probability of $C^\star_{n,m}$ to the error probability of C_n, we will need the following lemma, whose proof is in Appendix B.2.

Lemma 4.3 *Suppose Y is a random variable with density f. Let $\tilde{Y}^{(m)} = \lfloor Y \rfloor_m + U^{(m)}$, where $U^{(m)}$ is uniformly distributed in $\left(-2^{-m-1}, 2^{-m-1}\right)$ and independent from Y. Then each $\tilde{Y}^{(m)}$ has a density $f^{(m)}$, and $f^{(m)}$ converges pointwise almost everywhere to f.*

This lemma will be used to show that, by picking m sufficiently large, we can make the error probability of code $C^\star_{n,m}$ arbitrarily close to ϵ_n. Suppose we fix the message vector $\mathbf{w} \in \prod_{i=1}^K \{1, ..., 2^{kR_i}\}$ and let $\mathbf{Y}$ be the random vector of length $n|V|$ corresponding to all the received signals at all nodes during the n time steps in the block if code C_n is used. More precisely, we write $\mathbf{Y} = (\mathbf{Y}[0], ..., \mathbf{Y}[n-1])$, where $\mathbf{Y}[t] = (Y_1[t], ..., Y_{|V|}[t])$ is the random vector of received signals at all $|V|$

nodes at time t, for $0 \leq t \leq n-1$. The received signal at node v at time t, $Y_v[t]$, is simply given by

$$Y_v[t] = \sum_{u \in \mathcal{I}(v)} h_{u,v} X_u[t] + Z_v[t].$$

Notice that here we assume that the set of nodes V can be written as $V = \{1, ..., |V|\}$, in order to simplify some expressions. We claim that the random vector $\mathbf{Y}$ conditioned on the choice of messages $\mathbf{W} = \mathbf{w}$ has a density. To see this, we first notice that, conditioned on the received signals received up to time $t-1$, i.e., on $(\mathbf{Y}[0], ..., \mathbf{Y}[t-1]) = (\mathbf{y}[0], ..., \mathbf{y}[t-1])$, and on $\mathbf{W} = \mathbf{w}$, the transmit signals at time t, $X_v[t]$ for $v \in V$, are all deterministic. Thus, the received signals $Y_v[t]$, for $v \in V$, are conditionally independent and each one is normally-distributed, conditioned on $(\mathbf{Y}[0], ..., \mathbf{Y}[t-1]) = (\mathbf{y}[0], ..., \mathbf{y}[t-1])$ and $\mathbf{W} = \mathbf{w}$. Therefore, the conditional pdf $f_{Y_v[t]|\mathbf{Y}[0],...,\mathbf{Y}[t-1],\mathbf{W}}(y_v[t]|\mathbf{y}[0], ..., \mathbf{y}[t-1], \mathbf{w})$ exists for each $v \in V$. We conclude that, conditioned on $\mathbf{W} = \mathbf{w}$, the random vector $\mathbf{Y}$ has a density given by

$$f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) = \prod_{v=1}^{|V|} f_{Y_v[0]|\mathbf{W}}\left(y_v[0]|\mathbf{w}\right) \prod_{t=1}^{n-1} \prod_{v=1}^{|V|} f_{Y_v[t]|\mathbf{Y}[0],...,\mathbf{Y}[t-1],\mathbf{W}}\left(y_v[t]|\,\mathbf{y}[0], ..., \mathbf{y}[t-1], \mathbf{w}\right). \tag{4.13}$$

Similarly, we let $\tilde{\mathbf{Y}}^{(m)}$ be the vector of $n|V|$ effective received signals (4.12) if code $C^{\star}_{n,m}$ is used instead, i.e., $\tilde{\mathbf{Y}}^{(m)} = \left(\tilde{\mathbf{Y}}^{(m)}[0], ..., \tilde{\mathbf{Y}}^{(m)}[n-1]\right)$, where $\tilde{\mathbf{Y}}[t] = \left(\tilde{Y}_1^{(m)}[t], ..., \tilde{Y}_{|V|}^{(m)}[t]\right)$. By using similar arguments to those that led to (4.13), we see that, when we condition on $(\tilde{\mathbf{Y}}^{(m)}[0], ..., \tilde{\mathbf{Y}}^{(m)}[t-1]) = (\mathbf{y}[0], ..., \mathbf{y}[t-1])$, and on $\mathbf{W} = \mathbf{w}$, the effective received signals $\tilde{Y}_v^{(m)}[t]$, for $v \in V$, are conditionally independent (although not normally-distributed). Then, using the fact that, from (4.12), $\tilde{Y}_v^{(m)}[t]$ is the sum of two independent random variables and $U_v^{(m)}[t]$ has a density (see page 266 in [9]), we conclude that, conditioned on $\mathbf{W}$, $\tilde{\mathbf{Y}}^{(m)}[t]$ has a conditional density given by

$$f_{\tilde{\mathbf{Y}}^{(m)}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) = \prod_{v=1}^{|V|} f_{\tilde{Y}_v^{(m)}[0]|\mathbf{W}}\left(y_v[0]|\mathbf{w}\right) \prod_{t=1}^{n-1} \prod_{v=1}^{|V|}$$

$$f_{\tilde{Y}_v^{(m)}[t]|\tilde{\mathbf{Y}}^{(m)}[0],...,\tilde{\mathbf{Y}}^{(m)}[t-1],\mathbf{W}}\left(y_v[t]|\,\mathbf{y}[0],...,\mathbf{y}[t-1],\mathbf{w}\right). \tag{4.14}$$

The random variables $\tilde{Y}_v^{(m)} = \lfloor Y_v[t] \rfloor_m + U_v^{(m)}[t]$, for $m = 1, 2, ...$, conditioned on $(\tilde{\mathbf{Y}}^{(m)}[0], ..., \tilde{\mathbf{Y}}^{(m)}[t-1]) = (\mathbf{y}[0], ..., \mathbf{y}[t-1])$ and $\mathbf{W} = \mathbf{w}$, satisfy the conditions of Lemma 4.3, and we have that

$$f_{\tilde{Y}_v^{(m)}[0]|\mathbf{W}}\left(y_v[0]|\mathbf{w}\right) \to f_{Y_v[0]|\mathbf{W}}\left(y_v[0]|\mathbf{w}\right) \quad \text{and}$$

$$f_{\tilde{Y}_v^{(m)}[t]|\tilde{\mathbf{Y}}^{(m)}[0],...,\tilde{\mathbf{Y}}^{(m)}[t-1],\mathbf{W}}\left(y_v[t]|\,\mathbf{y}[0],...,\mathbf{y}[t-1],\mathbf{w}\right) \to$$

$$f_{Y_v[t]|\mathbf{Y}[0],...,\mathbf{Y}[t-1],\mathbf{W}}\left(y_v[t]|\,\mathbf{y}[0],...,\mathbf{y}[t-1],\mathbf{w}\right),$$

as $m \to \infty$, for $t = 2, ..., n$ and $v \in V$, for almost all $\mathbf{y} \in \mathbb{R}^{n|V|}$. Therefore, we conclude that $f_{\tilde{\mathbf{Y}}^{(m)}|\mathbf{W}}\left(\mathbf{y}|\mathbf{w}\right) \to f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w})$ as $m \to \infty$ for almost all $\mathbf{y} \in \mathbb{R}^{n|V|}$ and any $\mathbf{w} \in \prod_{i=1}^{K}\{1, ..., 2^{kR_i}\}$.

Next we notice that, conditioned on the message vector $\mathbf{W} = \mathbf{w}$, whether we make an error or not is a function of the received signals at all nodes during the n time steps (it is in fact only a function of the received signals at the destinations). Thus, there exists a set $E_{\mathbf{w}} \subset \mathbb{R}^{n|V|}$ of received signals during the n time steps which cause a decoding error (at any of the decoders). We will let $\mu_{\mathbf{w}}^{(n)}$ be the probability measure on $\mathbb{R}^{n|V|}$ corresponding to $\mathbf{Y}$ (the received signals when using coding scheme C_n) conditioned on $\mathbf{W} = \mathbf{w}$ and $\mu_{\mathbf{w}}^{(m,n)}$ be the probability measure on $\mathbb{R}^{n|V|}$ corresponding to $\tilde{\mathbf{Y}}^{(m)}$ (the effective received signals when we use coding scheme $C_{n,m}^{\star}$) conditioned on $\mathbf{W} = \mathbf{w}$. By Scheffé's Theorem [9], we have that

$$\sup_{A \in \mathcal{B}} \left|\mu_{\mathbf{w}}^{(n)}(A) - \mu_{\mathbf{w}}^{(m,n)}(A)\right| \le \int_{\mathbb{R}^{n|V|}} \left|f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) - f_{\tilde{\mathbf{Y}}^{(m)}|\mathbf{W}}(\mathbf{y}|\mathbf{w})\right| d\lambda \to 0,$$

as $m \to \infty$, where $\mathcal{B}$ is the Borel σ-field on $\mathbb{R}^{n|V|}$, and λ is the Lebesgue measure. This, in turn, implies that for any choice of messages $\mathbf{w}$, we must have $\lim_{m\to\infty} \mu_{\mathbf{w}}^{(m,n)}(E_{\mathbf{w}}) = \mu_{\mathbf{w}}^{(n)}(E_{\mathbf{w}})$. We conclude that

$$\begin{aligned}\epsilon_{n,m} &= 2^{-n\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \Pr\left[\tilde{\mathbf{Y}}^{(m)} \in E_{\mathbf{w}} \,\middle|\, \mathbf{W} = \mathbf{w}\right] \\ &= 2^{-n\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \mu_{\mathbf{w}}^{(m,n)}\left(E_{\mathbf{w}}\right) \stackrel{m\to\infty}{\longrightarrow} 2^{-n\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \mu_{\mathbf{w}}^{(n)}\left(E_{\mathbf{w}}\right) = \epsilon_n. \end{aligned} \tag{4.15}$$

Therefore, we can choose, for each n, m_n sufficiently large such that the probability of error of code $C^{\star}_{m_n,n}$, $\epsilon_{m_n,n}$, is at most $2\epsilon_n$. Finally, we need to take care of the fact that $C^{\star}_{m_n,n}$ uses randomized relaying and decoding functions. First, we notice that if we let $\mathbf{U}_m$ be the random vector corresponding to the $n|V|$ samples from $(-2^{-(m+1)}, 2^{-(m+1)})$ drawn at the $|V|$ nodes during n time steps, then we can write

$$\begin{aligned}\epsilon_{m_n,n} &= 2^{-n\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \Pr\left[\tilde{\mathbf{Y}}^{(m_n)} \in E_{\mathbf{w}} \,\middle|\, \mathbf{W} = \mathbf{w}\right] \\ &= E\left[2^{-n\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \Pr\left[\tilde{\mathbf{Y}}^{(m_n)} \in E_{\mathbf{w}} \,\middle|\, \mathbf{W} = \mathbf{w}, \mathbf{U}_{m_n}\right]\right].\end{aligned}$$

Therefore, there must exist some $\mathbf{u} \in \mathbb{R}^{n|V|}$ for which

$$2^{-n\sum_{i=1}^{K} R_i} \sum_{\mathbf{w}} \Pr\left[\tilde{\mathbf{Y}}^{(m_n)} \in E_{\mathbf{w}} \,\middle|\, \mathbf{W} = \mathbf{w}, \mathbf{U}_{m_n} = \mathbf{u}\right] \leq \epsilon_{m_n,n}.$$

Thus, we define the coding scheme $C^{\star}_n$ by having each node v at time t quantize its received signal with resolution m_n, add to it $u_v[t]$ (i.e., the entry of $\mathbf{u}$ corresponding to node v and time t) and then apply the relaying/decoding function from code C_n. It is then clear that $C^{\star}_n$ has deterministic relaying/decoding functions, and its error probability is at most $\epsilon_{m_n,n} \leq 2\epsilon_n$. Therefore, the sequence of codes $C^{\star}_n$, $n = 1, 2, ...$, has finite reading precision and achieves the rate tuple $\mathbf{R}$.

4.2 Discussion and Extensions

In this chapter, we proved that the Gaussian noise is the worst-case noise in additive noise wireless networks. This extends the classical result that Gaussian noise is the worst-case noise for point-to-point additive noise channels, which is commonly used as a justification for the modeling of the noise in wireless systems as Gaussian noise. Thus, we provide formal evidence that this modeling is indeed justified beyond the point-to-point setting.

One simple extension of this work is to consider MIMO wireless networks; i.e., wireless networks where each node can have multiple antennas. It is not difficult to see that the same arguments will hold in this case, and the Gaussian noise can also be seen to be worst-case. But the tools we developed are in fact also useful for establishing several other worst-case results in different classes of problems. In particular, the same DFT-based linear transformation followed by an interleaving procedure was used in [58] in order to show that the Gaussian sources are worst-case data sources for distributed compression of correlated sources over rate-constrained, noiseless channels, with a quadratic distortion measure (i.e., in the context of the quadratic k-encoder source coding problem). A similar approach was also taken in [5], where the authors consider the problem of communicating a distributed correlated memoryless source over a memoryless network, under quadratic distortion constraints. In this setting they show that, (a) for an arbitrary memoryless network, among all distributed memoryless sources with a particular correlation, Gaussian sources are the worst compressible, that is, they admit the smallest set of achievable distortion tuples, and (b) for any arbitrarily distributed memoryless source to be communicated over a memoryless additive noise network, among all noise processes with a fixed correlation, Gaussian noise admits the smallest achievable set of distortion tuples. Result (a) is presented in the next chapter.

We observe that establishing the worst-case noise for wireless networks can also be a useful tool in determining the relationship between the capacity regions of the same network under different channel models. For example, in Chapter 6, an additive *uniform* noise network is used as a way to connect the capacity region of Gaussian networks with the capacity region of truncated deterministic networks (first introduced in [6]).

4.2.1 Correlated Noise

One other straightforward extension of the result in Theorem 4.1 is to allow correlation between the noise realization across different nodes. In fact, if we assume that the vector of noises at all nodes at time t, $(Z_v[t])_{v\in V}$, is an i.i.d. random process with an arbitrary distribution with covariance matrix $\mathbf{K}$, a strengthened version of Theorem 4.1 can be obtained. In order to do that, we need to replace the arguments in Section 4.1.2 with a result that shows that, after we apply the OFDM-like scheme at each node, we obtain an effective noise vector $(\tilde{Z}_v)_{v\in V}[t]$ that converges to a *jointly* Gaussian distribution as $b \to \infty$. First we observe that the compound effect of the operations $\mathcal{L}_R$ and $\mathcal{L}_T$ from Fig. 4.5 can be represented by a $b \times b$ linear transformation $\mathbf{Q}$ and its inverse, where the entry in the $(i+1)$th row and $(j+1)$th column of $\mathbf{Q}$ is

$$Q(i,j) = \begin{cases} 1/\sqrt{b} & \text{if } i = 0 \\ \sqrt{2/b}\cos\left(\frac{2\pi j i}{b}\right) & \text{if } i = 1, ..., \frac{b}{2} - 1 \\ (-1)^j/\sqrt{b} & \text{if } i = \frac{b}{2} \\ \sqrt{2/b}\sin\left(\frac{2\pi j (i-b/2)}{b}\right) & \text{if } i = \frac{b}{2} + 1, ..., b-1 \end{cases} \tag{4.16}$$

for $i, j \in \{0, ..., b-1\}$. It is straightforward to check that $\mathbf{Q}$ is a unitary transformation, i.e., that $\|\mathbf{Qx}\| = \|\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^b$. The fact that $\mathbf{Q}$ can make a random vector approximately Gaussian is expressed in the following lemma. The arguments in Section 4.1.2 can be understood as proving the special case in which $\mathbf{K}$ is diagonal.

Lemma 4.4 *Suppose $\{(Z_1[i], ..., Z_k[i])\}_{i=0}^{nb-1}$ is an i.i.d. sequence of length-k zero-mean random vectors with covariance matrix $\mathbf{K}$, and let $\mathbf{Q}$ be the $b \times b$ matrix defined in (4.16) and*

$$\begin{bmatrix} \tilde{Z}_1^{(0)}[t] & \cdots & \tilde{Z}_k^{(0)}[t] \\ \vdots & \ddots & \vdots \\ \tilde{Z}_1^{(b-1)}[t] & \cdots & \tilde{Z}_k^{(b-1)}[t] \end{bmatrix} = \mathbf{Q} \begin{bmatrix} Z_1[tb] & \cdots & Z_k[tb] \\ \vdots & \ddots & \vdots \\ Z_1[tb+b-1] & \cdots & Z_k[tb+b-1] \end{bmatrix} \tag{4.17}$$

for $t = 0, 1, ..., n-1$. Then, for any sequence ℓ_b such that, for $b = 1, 2, ...$, $\ell_b \in \{0, 1, ..., b-1\}$, and any $t \in \{0, 1, ..., n-1\}$,

$$\left(\tilde{Z}_1^{(\ell_b)}[t], ..., \tilde{Z}_k^{(\ell_b)}[t]\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{K}), \text{ as } b \to \infty.$$

The proof of this lemma is provided in Appendix B.3. Using this result, it is not difficult to show that the same techniques introduced in the previous sections (OFDM-like scheme, interleaving and outer-code) allow any rate tuple achievable over a network with jointly Gaussian noises to be achieved over the same network with non-Gaussian noises with the same covariance matrix $\mathbf{K}$. More precisely, we have the following generalization of Theorem 4.1:

Theorem 4.6 *Suppose we have a K-unicast additive noise wireless network where the received signal at each node at time t is given as in (1.3), and the vector of noises $(Z_v[t])_{v \in V}$ is i.i.d. over time and follows a distribution with fixed and finite covariance matrix $\mathbf{K}$. If C_{Gaussian} is the capacity region of this network when $(Z_v[t])_{v \in V}$ is jointly Gaussian, and $C_{\text{non-Gaussian}}$ is the capacity region of the same network when $(Z_v[t])_{v \in V}$ has some other arbitrary distribution, then*

$$C_{\text{Gaussian}} \subseteq C_{\text{non-Gaussian}}.$$

We point out that, if the noises are not i.i.d. across time and we try to apply the OFDM-like scheme to mix the noise realizations, the resulting effective noise terms are linear combinations of dependent random variables. It is then unclear whether a CLT-like phenomenon will cause them to converge in distribution to Gaussian. Therefore, it is unknown whether the techniques presented in this chapter can be adapted to handle these cases, and it remains an open question whether Gaussian noises are still worst-case in networks where the noises are dependent across time.

4.2.2 Other "Mixing" Matrices

Another question that can be raised regarding the methods introduced in this chapter for establishing the Gaussian distribution as worst-case is whether there is something fundamental about the DFT-based transformation $\mathbf{Q}$ or there are other linear transformations capable of achieving the same result. In essence, all we need is a $b \times b$ matrix $\mathbf{Q}$ for which Lemma 4.4 holds. Intuition suggests that any unitary matrix in which the entries are somewhat balanced (i.e., there are no entries whose magnitude is much larger than the magnitude of the other entries) will cause the noise realizations to mix well, allowing the CLT effect to kick in as $b \to \infty$. Thus, other constructions of the matrix $\mathbf{Q}$ should be possible.

As it turns out, one interesting class of matrices that satisfy these requirements are Hadamard matrices normalized by $\sqrt{b}$. A $b \times b$ normalized Hadamard matrix is unitary, since all entries are either $1/\sqrt{b}$ or $-1/\sqrt{b}$ and the rows are mutually orthogonal, and thus a good candidate for a matrix $\mathbf{Q}$ that satisfies Lemma 4.4. As a matter of fact, if we consider any sequence $\{b_i\}_{i=1}^{\infty}$ with $b_i \to \infty$ such that a $b_i \times b_i$ Hadamard matrix is known to exist for each i (e.g., $b_i = 2^i$), we can prove a lemma very similar to Lemma 4.4 with Q being Hadamard, by restricting the convergence to be along the sequence $\{b_i\}_{i=1}^{\infty}$. In fact, in this case, Lindeberg's CLT would not even be necessary: the resulting effective noise terms would be the sum of two i.i.d. subsequences (one corresponding to the $1/\sqrt{b}$ coefficients and one corresponding to the $-1/\sqrt{b}$ coefficients), and the classical Central Limit Theorem, applied to each of the two subsequences, would suffice.

4.2.3 Beyond K-unicast

The result in Theorem 4.1 can also be generalized to networks with arbitrary traffic demands. In this scenario, each node $v \in V$ may have a message $W_{v,D}$ intended for each set $D \subset V$. Notice that in this case we may have up to $|V|2^{|V|}$ messages in the network, which means that our rate tuple $\mathbf{R}$ is a vector in $\mathbb{R}_+^{|V|2^{|V|}}$. By generalizing Definition 1.1 for the case of general traffic demands, we can generalize the worst-case noise result as follows:

Theorem 4.7 *From a sequence of coding schemes that achieve rate tuple* $\mathbf{R}$ *on an*

AWGN network with arbitrary traffic demands, it is possible to construct a sequence of coding schemes that achieves arbitrarily close to $\mathbf{R}$ *on the same* K*-unicast wireless network, where, for each relay* v*, the distribution of* Z_v *is replaced with any distribution satisfying* $E[Z_v] = 0$ *and* $E\left[Z_v^2\right] = \sigma_v^2$*. Therefore, if* C_{AWGN} *is the capacity region of the AWGN* K*-unicast wireless network, and* $C_{\text{non-AWGN}}$ *is the capacity region of the same wireless network where, for each relay* v*, the distribution of* Z_v *is replaced with an arbitrary distribution satisfying* $E[Z_v] = 0$ *and* $E\left[Z_v^2\right] = \sigma_v^2$*, then*

$$C_{\text{AWGN}} \subseteq C_{\text{non-AWGN}}.$$

This result can be proved essentially by following the same steps described in this chapter for the special case of K-unicast networks. We refer the reader to [56] for details.

CHAPTER 5

WORST-CASE SOURCES IN NETWORK COMPRESSION PROBLEMS

As we see next, the framework introduced in the previous chapter can in fact be used to prove that, in network compression problems, the worst-case source is also Gaussian.

We consider a general setting in which k distinct nodes from a network observe distinct components of a k-component i.i.d. source with covariance matrix $\mathbf{K}$. These nodes wish to communicate their observations across a network to respective destination nodes, with the objective of minimizing the resulting quadratic distortion.

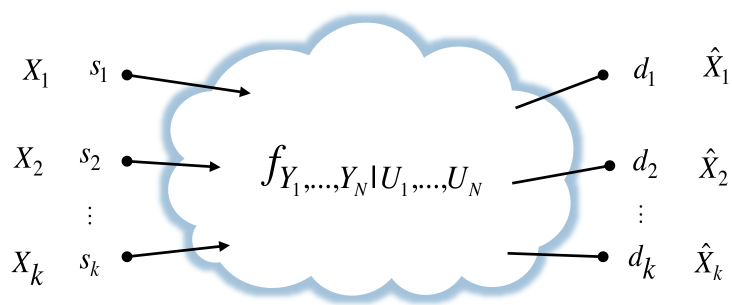


Figure 5.1: (k, N)-memoryless network.

A (k, N)-memoryless network, illustrated in Fig. 5.1, is characterized by the conditional density $f_{Y_1,...,Y_N|U_1,...,U_N}$, which relates the real valued network inputs $(U_1, \ldots, U_N)$ to real valued network outputs $(Y_1, \ldots, Y_N)$. The set of source nodes is denoted as $\mathcal{S} = \{s_1, s_2, \ldots, s_k\} \subseteq \{1, ..., N\}$, and the set of destination nodes is denoted as $\mathcal{D} = \{d_1, d_2, \ldots, d_k\} \subseteq \{1, ..., N\}$. The remaining nodes (we assume without loss of generality that the sets of source and destination nodes have empty intersection) are relays $\mathcal{R} = \{r_1, r_2, \ldots, r_{N-2k}\} \subseteq \{1, ..., N\}$. Source node $s_m \in$

$\mathcal{S}$ has access to the i.i.d. source $X_m[t]$, $t = 0, 1, ...$, which must be communicated to the corresponding destination node $d_m \in \mathcal{D}$. The i.i.d. vectors $(X_1[t], \ldots, X_k[t])$ have a joint distribution with covariance matrix $\mathbf{K}$.

Definition 5.1 *A coding scheme C with block length $n \in \mathcal{N}$ for distributed compression of a real-valued memoryless source $(X_1, X_2, \ldots, X_k)$ over a (k, N)-memoryless network consists of the following:*

1. Source Encoding Functions*: Source node $s_m \in \mathcal{S}$ encodes the source X_m as $U_{s_m}[t] = f_{s_m,t}(\mathbf{X}_m, Y_{s_m}^{t-1})$, $\forall\ t \in \{0, ..., n-1\}$, where $f_{s_m,t} : \mathbb{R}^n \times \mathbb{R}^{t-1} \to \mathbb{R}$, $\forall\ m \in \{1, ..., k\}$, $\forall\ t \in \{0, ..., n-1\}$ are the source encoding functions.*
2. *Relay Encoding Functions: Relay node $r_p \in \mathbb{R}$ receives the channel outputs from the network and encodes it as $U_{r_p}[t] = f_{r_p,t}(Y_{r_p}^{t-1})$, $\forall\ t \in \{0, ..., n-1\}$, where $f_{r_p,t} : \mathbb{R}^{t-1} \to \mathbb{R}$, $\forall\ p \in \{1, ..., N-2k\}$, $\forall\ t = \{0, ..., n-1\}$, are the relay encoding functions.*
3. *Destination Encoding Functions: Destination node $d_m \in \mathcal{D}$ receives the channel output from the network and encodes it as $U_{d_m}[t] = f_{d_m,t}(Y_{d_m}^{t-1})$, where $f_{d_m,t} : \mathbb{R}^{t-1} \to \mathbb{R}$, $\forall\ m \in \{1, ..., k\}$, $\forall\ t \in \{0, ..., n-1\}$, are the destination encoding functions.*
4. *Destination Decoding Functions: At the end of the block of communication, each destination $d_m \in \mathcal{D}$ constructs an estimate of the source as $\hat{\mathbf{X}}_m = g_{d_m}(\mathbf{Y}_{d_m})$, where $g_{d_m} : \mathbb{R}^n \to \mathbb{R}^n$, $\forall\ m \in \{1, ..., k\}$, are the destination decoding functions.*

Definition 5.2 *A distortion measure is a mapping $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$.*

Definition 5.3 *A distortion tuple $(D_1, D_2, \ldots, D_k)$ is said to be ϕ-achievable if for some block length n, there exists a coding scheme C, as described above, such that,*

$$\frac{1}{n} E\left[\sum_{t=1}^{n} \phi(X_m[t], \hat{X}_m[t])\right] \leq D_m,\ \forall\ m \in \{1, ..., k\}. \tag{5.1}$$

We focus on the quadratic distortion measure, i.e., where $\phi(x, y) = \xi(x, y) \triangleq (x-y)^2$. Notice that, in this case, the expression in (5.1) can be equivalently written as

$$\frac{1}{n}E\Big[\parallel \mathbf{X}_m - \hat{\mathbf{X}}_m \parallel^2\Big] \le D_m,\ \forall\, m \in \{1, ..., k\}.$$

Definition 5.4 *The ϕ-achievable distortion region $\mathcal{DR}$ of a (k, N)-memoryless network is the closure of the set of ϕ-achievable distortion tuples.*

Theorem 5.1 *For a (k, N) memoryless network, let $\mathcal{DR}_{NG}^{source}$ and $\mathcal{DR}_{G}^{source}$ stand for the ξ-achievable distortion regions for an arbitrary memoryless non-Gaussian source with covariance matrix $\mathbf{K}$ and for a memoryless Gaussian source with the same covariance matrix, respectively. Then*

$$\mathcal{DR}_{G}^{source} \subseteq \mathcal{DR}_{NG}^{source}. \tag{5.2}$$

Remark 5.1 *A special case of Theorem 5.1 is that of wireline networks where each link is a (noiseless) bit pipe. This gives us the result of Gaussian source being the worst case source for the k-encoder distributed compression problem studied in [58].*

5.1 Proof of Worst-Case Source Result

In this section we provide the essential steps to prove Theorem 5.1. We present the proofs of the lemmas in the appendix, and also refer the reader to [5] for more details. The main idea is to use a coding scheme C with block length n for Gaussian sources to construct a new coding scheme $\tilde{C}$ which achieves the same distortion tuple when the sources are non-Gaussian with the same covariance.

Notice that we may assume without loss of generality that the sources have zero mean, since otherwise, we can remove the mean at the source nodes and add it back at the destination nodes.

The first step in the construction of this new coding scheme is to use the DFT-based linear transformation introduced in Chapter 4 in order to transform blocks of i.i.d. non-Gaussian random variables into "approximately Gaussian" random variables. More specifically, to construct $\tilde{C}$, we take n blocks of b source symbols, apply the linear transformation $\mathbf{Q}$ from (4.16) to each of them and then interleave the resulting symbols, obtaining b blocks $\tilde{\mathbf{X}}_m^{(\ell)}$, for $\ell = 0, ..., b-1$, each of length n, for each source s_m, referred to as the effective source symbols (cf. Figure 5.2). It can be seen that each of the b resulting length-n blocks have

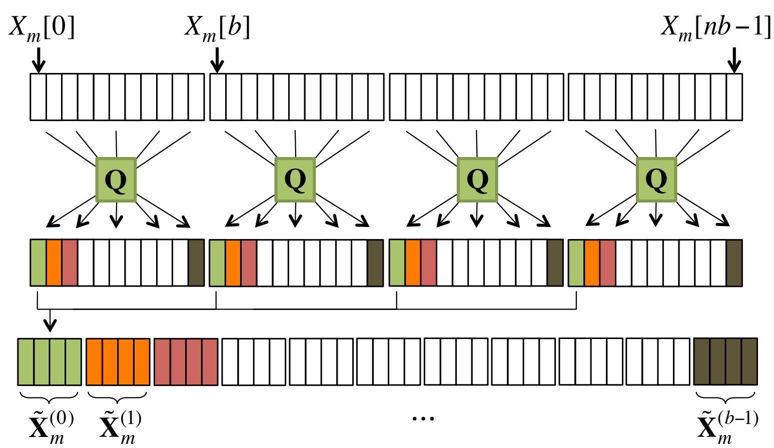


Figure 5.2: Construction of the effective source by source node s_m.

i.i.d. effective sources. We then apply the source encoding functions of C, designed to achieve a given distortion tuple $(D_1, ..., D_k)$ with Gaussian sources, to these resulting i.i.d. blocks. The relay and destination encoding functions of $\tilde{C}$ and C are the same, and the destination decoding functions of $\tilde{C}$ are the composition of the destination decoding functions of C with a transformation that inverts the construction of the effective sources in Fig. 5.2.

Our main goal is to show that, as $b \to \infty$, the distortion of the resulting coding scheme $\tilde{C}$ converges to $(D_1, ..., D_k)$. We begin with a lemma that allows us to just concentrate on *bounded output* coding schemes.

Lemma 5.1 *Suppose* $(X_1[t], ..., X_k[t])$ *has an arbitrary joint distribution with covariance matrix* $\mathbf{K}$ *and a coding scheme* C *with block length* n *achieves distortion vector* $(D_1, ..., D_k)$*. Then, for any* $\epsilon > 0$*, one can build another coding scheme* $\tilde{C}$ *of block length* n *with decoding functions* $\tilde{g}_{d_m}$ *such that*

$$\left\|\tilde{g}_{d_j}(y_1, ..., y_n)\right\|_\infty \leq M,$$

for any $(y_1, ..., y_n) \in \mathbb{R}^n$*,* $j = 1, ..., k$ *and a fixed* $M > 0$*, which achieves distortion vector* $(D_1 + \epsilon, ..., D_k + \epsilon)$*.*

Another important property that we need to assume for the original coding scheme designed for a Gaussian model is that of *finite precision*. For a real-valued vector $x^n = (x_1, ..., x_n)$ and a positive integer ρ, we let $\lfloor x^n \rfloor_\rho = 2^{-\rho}\left(\lfloor 2^\rho x_1 \rfloor, ..., \lfloor 2^\rho x_n \rfloor\right)$, and define the following:

Definition 5.5 *A coding scheme* C *of block length* n *is said to have finite encoding precision* $\rho = [\rho_1, \cdots, \rho_k] \in \mathbf{N}^k$ *if the encoding function at each source* $s_m \in \mathcal{S}$ *satisfies*

$$f_{s_m,t}(x_m^n, y^{t-1}) = f_{s_m,t}(\lfloor x_m^n \rfloor_{\rho_m}, y^{t-1}), \ \forall\, m \in \{1, ..., k\}$$

for any $x_m^n \in \mathbb{R}^n$*, any* $y^{t-1} \in \mathbb{R}^{t-1}$*, and any time* t*.*

Lemma 5.2 *Suppose the distortion tuple* $(D_1, \cdots, D_k)$ *is achievable over the* (k, N)*-memoryless network. Then for any* $\epsilon > 0$*, there exists a coding scheme with finite encoding precision that achieves distortion tuple* $(D_1 + \epsilon, \cdots, D_k + \epsilon)$*.*

From the proofs of Lemmas 5.1 and 5.2 [5], it can be seen that in fact there exists a single coding scheme that has both bounded outputs and finite encoding precision and achieves distortion tuple $(D_1+\epsilon, \cdots, D_k+\epsilon)$. Thus, we may assume initially that C has bounded outputs and finite encoding precision.

The importance of finite encoding precision is expressed in the following lemma.

Lemma 5.3 *If, for some $\rho \in \mathcal{N}$, $f : \mathbb{R}^a \to \mathbb{R}^b$ satisfies*

$$f(\mathbf{x}) = \mathbf{f}(\lfloor \mathbf{x} \rfloor_\rho)$$

for any $\mathbf{x} \in \mathbb{R}^a$, f is locally constant (and thus continuous) almost everywhere.

Lastly we will need a lemma that allows us to view our stochastic network as a collection of deterministic networks, which helps in bounding the resulting distortion.

Lemma 5.4 *For any two random vectors $\mathbf{Y}$ and $\mathbf{U}$, there exist a (deterministic, measurable) function F and a random vector $\mathbf{Z}$, independent of $\mathbf{U}$, for which the pair $(F(\mathbf{U}, \mathbf{Z}), \mathbf{U})$ has the same distribution as $(\mathbf{Y}, \mathbf{U})$.*

This lemma implies that there exist functions F_{d_m}, for $d_m \in \mathcal{D}$, and a random vector $\mathbf{Z}$, such that, if the length-n source sequences are $\mathbf{x}_1, ..., \mathbf{x}_k$, then the length-$n$ block of received signals at destination d_m is given by $F_{d_m}(\mathbf{x}_1, ..., \mathbf{x}_k, \mathbf{Z})$. Therefore, since $\mathbf{Q}$ is a unitary linear transformation, for each realization $\mathbf{z}$ of $\mathbf{Z}$, the distortion of $\tilde{C}$ can be written as

$$\frac{1}{b} \sum_{\ell=0}^{b-1} \frac{1}{n} \left\| \tilde{\mathbf{X}}_m^{(\ell)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell)}, ..., \tilde{\mathbf{X}}_k^{(\ell)}, \mathbf{z}) \right) \right\|^2 . \tag{5.3}$$

For each $b = 1, 2, ...$, we choose ℓ_b such that the ℓ_bth length-n block has the largest expected distortion, i.e.,

$$\ell_b = \arg \max_{0 \le \ell \le b-1} E \left\| \tilde{\mathbf{X}}_m^{(\ell)} - g_{d_m} \left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell)}, ..., \tilde{\mathbf{X}}_k^{(\ell)}, \mathbf{z}) \right) \right\|^2 .$$

Note that $\left\{ \left(\tilde{X}_1^{(\ell_b)}[i], ..., \tilde{X}_k^{(\ell_b)}[i] \right) \right\}_{i=0}^{n-1}$ is an i.i.d. sequence of length-k random vectors. From Lemma 4.4, we see that it converges in distribution to a sequence of i.i.d. jointly Gaussian random vectors with covariance matrix $\mathbf{K}$, as $b \to \infty$.

Each of the source encoding functions $f_{s_m,t}$ of the original coding scheme C is locally constant almost everywhere, since they were assumed to have finite encoding precision, by Lemma 5.3. In fact, this implies that the mapping

$$\left\{ \tilde{\mathbf{X}}_m^{(\ell_b)} \right\}_{m=1}^{k} \mapsto \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m} \left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2 ,$$

for $m = 1, ..., k$, is continuous almost everywhere. Hence,

$$\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m} \left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2 \xrightarrow{d} \left\| \mathbf{X}_m^G - g_{d_m} \left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{z}) \right) \right\|^2 ,$$

as $b \to \infty$, where $\mathbf{X}_m^G = (X_m^G[0], ..., X_m^G[n-1])$, for $m = 1, ..., k$, and $\left\{ \left(X_1^G[i], ..., X_k^G[i] \right) \right\}_{i=0}^{n-1}$ is an i.i.d. sequence such that $(X_1^G[0], ..., X_k^G[0])$ is jointly Gaussian with zero mean and covariance matrix $\mathbf{K}$. Moreover, we have that

$$\begin{aligned} &\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m} \left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2 \\ &\quad \le 2 \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} \right\|^2 + 2 \left\| g_{d_m} \left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2 \\ &\quad \le 2 \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} \right\|^2 + 2nM^2, \end{aligned} \tag{5.4}$$

and also that

$$\begin{aligned} E \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} \right\|^2 &= nE \left(\sum_{j=0}^{b-1} X_m[j] \; Q(\ell_b, j) \right)^2 \\ &= n\mathbf{K}_{m,m} \sum_{j=0}^{b-1} Q^2(\ell_b, j) = n\mathbf{K}_{m,m} < \infty. \end{aligned} \tag{5.5}$$

Thus, from a variation of the Dominated Convergence Theorem (see Appendix B.4), we conclude that, as $b \to \infty$,

$$E\left[\left\|\tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z})\right)\right\|^2 \middle| \mathbf{Z} = \mathbf{z}\right]$$
$$\to E\left[\left\|\mathbf{X}_m^G - g_{d_m}\left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{Z})\right)\right\|^2 \middle| \mathbf{Z} = \mathbf{z}\right],$$

for all $\mathbf{z}$. The Dominated Convergence Theorem can then be used once again yielding

$$E\left[E\left[\left\|\tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z})\right)\right\|^2 \middle| \mathbf{Z}\right]\right]$$
$$\to E\left[E\left[\left\|\mathbf{X}_m^G - g_{d_m}\left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{Z})\right)\right\|^2 \middle| \mathbf{Z}\right]\right],$$

which implies

$$E\left[\left\|\tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z})\right)\right\|^2\right] \to E\left[\left\|\mathbf{X}_m^G - g_{d_m}\left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{Z})\right)\right\|^2\right] \leq D_m.$$

Therefore, we can choose b sufficiently large so that

$$\frac{1}{n}E\left[\left\|\tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z})\right)\right\|^2\right] \leq D_m + \epsilon.$$

The expected distortion of code $\bar{C}$ (with block length nb) thus satisfies, for $m = 1, ..., k$,

$$\frac{1}{nb}\sum_{\ell=0}^{b-1} E\left[\left\|\tilde{\mathbf{X}}_m^{(\ell)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell)}, ..., \tilde{\mathbf{X}}_k^{(\ell)}, \mathbf{Z})\right)\right\|^2\right]$$
$$\leq \frac{1}{n}E\left[\left\|\tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z})\right)\right\|^2\right]$$
$$\leq D_m + \epsilon,$$

concluding the proof of Theorem 5.1.

Part III

Outer-Bounding Techniques

By analyzing the recent network information theory literature, many notable advances in the form of capacity inner bounds can be found. In many cases, these advances are in the form of general techniques that can be applied in a large class of networks (e.g., the Han-Kobayashi scheme [27], relaying techniques such as quantize-map-and-forward [6] and Noisy Network Coding [42], and interference-alignment-based schemes). In most cases, the proposed schemes have their performance evaluated against the classical cut-set bound. The cut-set bound is the standard outer-bounding technique due to its generality, as it applies to arbitrary memoryless networks. Moreover, applying it is relatively straightforward as it is a single-letter expression.

In the case of single-flow networks, the cut-set bound provides a fairly good way of evaluating the performance of proposed schemes: it is known to be tight in multicast wireline [3, 21] and linear deterministic networks and within a constant gap of capacity in AWGN relay networks [6]. For multi-flow networks, however, the cut-set bound is easily seen to be arbitrarily loose. Aside from the wireline scenario, where different improvements over the min-cut bound are known [28, 39, 62, 70], most outer-bound results are tied to specific network configurations (e.g., [20, 22, 55]), and few general techniques are known.

In this part of the dissertation, we look for new ways of deriving outer bounds for multi-hop multi-flow networks. First, in Chapter 6, we explore the possibility of deriving outer bounds through alternative channel models. In particular we look for a way to upper bound the capacity of networks under the AWGN channel model with the capacity of a deterministic counterpart. Then, in Chapter 7, we present a generalization of the cut-set bound for multi-flow deterministic networks. This new bound has several new applications. In particular, by combining it with the result from Chapter 6, we derive a new outer bound for the degrees of freedom of non-fully-connected $K \times K \times K$ wireless network, which turns out to provide necessary and sufficient conditions for K degrees of freedom to be achievable.

CHAPTER 6

OUTER BOUNDS VIA ALTERNATIVE CHANNEL MODELS

In this chapter, we use our worst-case noise result from Chapter 4 in order to establish a connection between the capacity region of AWGN and deterministic models of multi-hop multi-flow wireless networks. In particular, we are interested in finding a deterministic model that can serve as a tool to deriving outer bounds for the capacity region of these networks. We will show that the capacity region of the $K \times K \times K$ wireless network (see Fig. 1.2) under the usual AWGN channel model is a subset of the capacity region of the same network under the truncated deterministic channel model [6] provided that the nodes are given more power in the truncated channel model case. We start by describing three distinct channel models of interest:

1. *AWGN Channel Model:* The received signals at a node j are given by

$$Y_i[t] = \sum_{i:(i,j)\in E} h_{i,j} X_i[t] + Z_j[t] \tag{6.1}$$

   where $X_i[t]$ is the signal transmitted by node i at time t, and $Z_j[t]$ is a sequences of i.i.d. noise terms, disbributed as $\mathcal{N}(0, 1/12)$.

2. *Additive Uniform Noise Channel Model:* The received signals at a node j are also given by (6.1), but $Z_j[t]$ is instead a sequence of i.i.d. noise terms uniformly disbributed in $\left(-\frac{1}{2}, \frac{1}{2}\right)$.

3. *Truncated Channel Model:* The received signals at a node j are given by

$$Y_j[t] = \left\lfloor \sum_{i:(i,j)\in E} h_{i,j} X_i[t] \right\rfloor. \tag{6.2}$$

Definition 6.1 *We define $C_{\text{AWGN}}(P)$, $C_{\text{Uniform}}(P)$ and $C_{\text{Truncated}}(P)$ to be the capacity regions of the $K \times K \times K$ wireless network under the AWGN channel model, under the* *additive uniform noise channel model and under the truncated channel model respectively.*

Our main result will be to show that

$$C_{\text{AWGN}}(P) \subseteq C_{\text{Uniform}}(P) \subseteq C_{\text{Truncated}}(2P + \alpha).$$

where α is a constant that depends on the channel gains but not on P. The fact that the capacity region of the AWGN $K \times K \times K$ wireless network is a subset of the capacity region of the truncated $K \times K \times K$ wireless network (where the nodes have more power) is particularly interesting because it allows us to look for outer bounds on the capacity and on the degrees of freedom of the AWGN $K \times K \times K$ wireless network by focusing on the truncated deterministic channel model. This will be done in Chapter 7.

A key result that allows us to establish this connection between the AWGN channel model and the uniform noise channel model is the characterization of the Gaussian noise as the worst-case additive noise in wireless networks from Theorem 4.1. Since the $K \times K \times K$ wireless network is a K-unicast wireless network we have the following immediate corollary of Theorem 4.1.

Corollary 6.1 *The capacity regions of the $K \times K \times K$ wireless network under the AWGN channel model and under the additive uniform noise channel model satisfy*

$$C_{\text{AWGN}}(P) \subseteq C_{\text{Uniform}}(P).$$

The uniform noise channel model will serve as an intermediate step for us to connect the capacity regions of the $K \times K \times K$ wireless network under the AWGN channel model and under the truncated deterministic model.

6.1 Relating the Capacity of AWGN and Truncated Deterministic Networks

In this section we establish a relation between the capacity regions of the $K \times K \times K$ wireless network under the additive uniform noise model and under the truncated channel model. We let $H_{S,U}$ and $H_{U,D}$ be the transfer matrices of the first and second hops of the $K \times K \times K$ network respectively. We assume throughout this chapter that $H_{S,U}$ and $H_{U,D}$ are invertible.

Theorem 6.1 *The capacity regions of the $K \times K \times K$ wireless network under the additive uniform noise channel model and under the truncated channel model satisfy*

$$C_{\text{Uniform}}(P) \subseteq C_{\text{Truncated}}(2P + \alpha),$$

where α is a constant that depends on $H_{S,U}$ and $H_{U,D}$, but not on P.

It is clear from Theorem 6.1 that any outer bound on the capacity region of the truncated $K \times K \times K$ wireless network can be translated into an outer bound on the capacity of the uniform noise $K \times K \times K$ wireless network where the nodes have less power. It is important to point out, however, that in the high-SNR regime this extra power can provide at most 1/2 a bit per user. Thus, from a degrees-of-freedom point-of-view, this extra power is negligible, and we can relate the degrees-of-freedom regions under the two models in a simple way.

We will write $\mathbf{D}_{\text{Uniform}}$, $\mathbf{D}_{\text{Truncated}}$, and $\mathbf{D}_{\text{AWGN}}$ to represent the degrees-of-freedom region of the $K \times K \times K$ wireless network under the additive uniform noise channel model, under the truncated channel model and under the additive Gaussian noise channel model respectively. An immediate consequence of Corollary 6.1 and Theorem 6.1 is the following.

Corollary 6.2 *The degrees-of-freedom regions of the $K \times K \times K$ wireless network under the additive Gaussian noise channel model, under the additive uniform noise channel model and under the truncated channel model satisfy*

$$\mathbf{D}_{\text{AWGN}} \subseteq \mathbf{D}_{\text{Uniform}} \subseteq \mathbf{D}_{\text{Truncated}}.$$

The importance of Corollary 6.2 is that it guarantees that any outer bound on the degrees-of-freedom region of the $K \times K \times K$ wireless network under the truncated channel model is also an outer bound on the degrees-of-freedom region of the AWGN version of the same $K \times K \times K$ wireless network.

In order to prove Theorem 6.1, we need the following technical lemma:

Lemma 6.1 *Let X be an arbitrary random variable and U a random variable uniformly distributed in $\left(-\frac{1}{2}, \frac{1}{2}\right)$ and independent of X. Then*

$$\lfloor X + U \rfloor - U + \frac{1}{2} \sim X + U.$$

Proof: Let $Y = \lfloor X + U \rfloor - U + \frac{1}{2}$ and $Z = X + U$, and let F_X, F_Y and F_Z be the cdfs of X, Y and Z and $F_{Y|x}$ and $F_{Z|x}$ be the cdfs of Y and Z conditioned on $X = x$. We will show that $F_{Y|x}(y) = F_{Z|x}(y)$ for every x and y, and it will then follow that

$$F_Y(y) = E_X\left[F_{Y|X}(y)\right] = E_X\left[F_{Z|X}(y)\right] = F_Z(y),$$

establishing the result. We fix an arbitrary $x \in \mathbb{R}$ and it is straightforward to see that

$$F_{Z|x}(y) = \begin{cases} 0, & \text{if } y \leq x - \frac{1}{2} \\ y - x + \frac{1}{2}, & \text{if } x - \frac{1}{2} \leq y \leq x + \frac{1}{2} \\ 1, & \text{if } x + \frac{1}{2} \leq y. \end{cases}$$

Then we notice that

$$
\begin{aligned}
F_{Y|x}(y) &= \Pr\left(\lfloor x + U \rfloor - U \le y - \tfrac{1}{2}\right) \\
&= \Pr\left(\lfloor x + U \rfloor - y + \tfrac{1}{2} \le U\right) \\
&= \Pr\left(\lfloor x \rfloor - 1 - y + \tfrac{1}{2} \le U,\ U \le \lfloor x \rfloor - x\right) \\
&\quad + \Pr\left(\lfloor x \rfloor - y + \tfrac{1}{2} \le U,\ U \in (\lfloor x \rfloor - x, 1 + \lfloor x \rfloor - x)\right) \\
&\quad + \Pr\left(\lfloor x \rfloor + 1 - y + \tfrac{1}{2} \le U,\ U \ge 1 + \lfloor x \rfloor - x\right)
\end{aligned}
$$

Next we consider two separate cases. First, we suppose that $x - \lfloor x \rfloor \le \frac{1}{2}$. If we let $(\beta)^+ = \max(0, \beta)$, we will have

$$
\begin{aligned}
F_{Y|x}(y) &= \Pr\left(\lfloor x \rfloor - y - \tfrac{1}{2} \le U,\ U \le \lfloor x \rfloor - x\right) \\
&\quad + \Pr\left(\lfloor x \rfloor - y + \tfrac{1}{2} \le U,\ U \in (\lfloor x \rfloor - x, 1 + \lfloor x \rfloor - x)\right) \\
&= \left(\lfloor x \rfloor - x + \tfrac{1}{2} - (\lfloor x \rfloor - y)^+\right)^+ + \left(\tfrac{1}{2} - \lfloor x \rfloor - \max(-y + \tfrac{1}{2}, -x)\right)^+ \\
&= \begin{cases}
\left(y - x + \frac{1}{2}\right)^+ + (y - \lfloor x \rfloor)^+, & \text{if } y \le x - \frac{1}{2} \\
\left(y - x + \frac{1}{2}\right)^+ + (y - \lfloor x \rfloor)^+, & \text{if } x - \frac{1}{2} \le y \le \lfloor x \rfloor \\
\left(\lfloor x \rfloor - x + \frac{1}{2}\right) + (y - \lfloor x \rfloor)^+, & \text{if } \lfloor x \rfloor \le y \le x + \frac{1}{2} \\
\left(\lfloor x \rfloor - x + \frac{1}{2}\right) + \left(\frac{1}{2} - \lfloor x \rfloor + x\right)^+, & \text{if } x + \frac{1}{2} \le y
\end{cases} \\
&= \begin{cases}
0, & \text{if } y \le x - \frac{1}{2} \\
y - x + \frac{1}{2}, & \text{if } x - \frac{1}{2} \le y \le x + \frac{1}{2} \\
1, & \text{if } x + \frac{1}{2} \le y
\end{cases} \\
&= F_{Z|x}(y).
\end{aligned}
$$

Second, in the case where $x - \lfloor x \rfloor > \frac{1}{2}$, we have

$$
\begin{aligned}
F_{Y|x}(y) &= \Pr\left(\lfloor x \rfloor - y + \tfrac{1}{2} \le U, U \in (\lfloor x \rfloor - x, 1 + \lfloor x \rfloor - x)\right) \\
&\quad + \Pr\left(\lfloor x \rfloor - y + \tfrac{3}{2} \le U, U \ge 1 + \lfloor x \rfloor - x\right) \\
&= \left(1 + \lfloor x \rfloor - x - \max(\lfloor x \rfloor - y + \tfrac{1}{2}, -\tfrac{1}{2})\right)^+ + \left(\tfrac{1}{2} - \lfloor x \rfloor - 1 - \max(-y + \tfrac{1}{2}, -x)\right)^+
\end{aligned}
$$

$$
= \begin{cases} \left(y - x + \frac{1}{2}\right)^{+} + (y - \lfloor x \rfloor - 1)^{+}, & y \le x - \frac{1}{2} \\ \left(y - x + \frac{1}{2}\right)^{+} + (y - \lfloor x \rfloor - 1)^{+}, & \text{if } x - \frac{1}{2} \le y \le \lfloor x \rfloor + 1 \\ \left(\lfloor x \rfloor - x + \frac{3}{2}\right) + (y - \lfloor x \rfloor - 1)^{+}, & \text{if } \lfloor x \rfloor + 1 \le y \le x + \frac{1}{2} \\ \left(\lfloor x \rfloor - x + \frac{3}{2}\right) + \left(-\frac{1}{2} - \lfloor x \rfloor + x\right)^{+}, & \text{if } x + \frac{1}{2} \le y \end{cases}
$$

$$
= \begin{cases} 0, & \text{if } y \le x - \frac{1}{2} \\ y - x + \frac{1}{2}, & \text{if } x - \frac{1}{2} \le y \le x + \frac{1}{2} \\ 1, & \text{if } x + \frac{1}{2} \le y \end{cases} = F_{Z|x}(y).
$$

Therefore, $F_{Y|x}(y) = F_{Z|x}(y)$ for every x and y. ■

We can now proceed to the proof of our main result.

Proof of Theorem 6.1: Consider a sequence of coding schemes $\{C_n\}$ that achieves rate tuple $\mathbf{R}$ on the $K \times K \times K$ wireless network under the additive uniform noise channel model. We will use it to construct a new sequence of coding schemes that achieves the same rate tuple $\mathbf{R}$ on the $K \times K \times K$ wireless network under the truncated channel model. In order to do this, we will first assume that all the nodes in the network are allowed to share the outcome of a sequence of random vectors, drawn independently of the source message choices, prior to the beginning of communication. Later we will show that this assumption can be dropped.

Focus on coding scheme C_n with blocklength n, encoding functions f_i, relaying functions $r_i^{(t)}$, and decoding functions g_i as described in Definition 1.1. We assume that the random vectors $U^{(1)}[1], ..., U^{(1)}[n]$ and $U^{(2)}[1], ..., U^{(2)}[n]$ of length K and whose entries are i.i.d. and uniformly distributed in $\left(-\frac{1}{2}, \frac{1}{2}\right)$ are drawn prior to the beginning of communication and independently from the source messages, and shared among all nodes. Then we will construct a new coding scheme $\tilde{C}_n$ for the $K \times K \times K$ wireless network under the truncated channel model. The encoding functions $\tilde{f}_j(w_j)$, for $j = 1, ..., K$, of coding scheme $\tilde{C}_n$ will be defined as

$$\tilde{f}_j(w_j)[t] = f_j(w_j)[t] + \left(H_{S,U}^{-1} U^{(1)}[t]\right)_j,$$

where $f_j(w_j)[t]$ is the tth component of $f_j(w_j)$. Relay u_i will use relaying functions $\tilde{r}_i^{(t)}$, for $t = 1, ..., n$, defined as

$$\tilde{r}_i^{(t)}(y_i[1], ..., y_i[t-1]) = r_i^{(t)}\left(y_i[1] - U_i^{(1)}[1] + \tfrac{1}{2}, ..., y_i[t-1] - U_i^{(1)}[n] + \tfrac{1}{2}\right).$$

Notice that the operation applied by the relays in the new coding scheme $\tilde{C}_n$ is equivalent to simply applying the relaying functions of the original coding scheme C_n after modifying its received signal (thus creating an *effective* received signal) as

$$\tilde{Y}_{u_i}[t] = Y_{u_i}[t] - \left(U^{(1)}[t]\right)_i + \frac{1}{2}.$$

Notice that this effective received signal $\tilde{Y}_{u_i}[t]$ satisfies

$$\begin{aligned}
\tilde{Y}_{u_i}[t] &= \left\lfloor \sum_{j=1}^{K} h_{s_j,u_i} \tilde{f}_j(w_j)[t] \right\rfloor - \left(U^{(1)}[t]\right)_i + \frac{1}{2} \\
&= \left\lfloor \sum_{j=1}^{K} h_{s_j,u_i} \left(f_j(w_j)[t] + \left(H_{S,U}^{-1} U^{(1)}[t]\right)_j\right) \right\rfloor - \left(U^{(1)}[t]\right)_i + \frac{1}{2} \\
&= \left\lfloor \sum_{j=1}^{K} h_{s_j,u_i} f_j(w_j)[t] + \left(U^{(1)}[t]\right)_i \right\rfloor - \left(U^{(1)}[t]\right)_i + \frac{1}{2}.
\end{aligned}$$

From Lemma 1, this effective received signal is distributed as $\sum_{j=1}^{K} h_{s_j,u_i} f_j(w_j)[t] + Z_{u_i}[t]$, where $Z_{u_i}[t]$ is uniformly distributed in $\left(-\frac{1}{2}, \frac{1}{2}\right)$. Moreover, since the entries in $U^{(1)}[t]$ were independent, it is clear that, conditioned on the transmit signals at the sources, the received signals at the relays are independent. This implies that the joint distribution of the effective received signals at the relays when we use coding scheme $\tilde{C}_n$ on the truncated network is the same as the joint distribution of the actual received signals at the relays when we use coding scheme C_n on the uniform noise network.

The same procedure can then be applied to the second hop using the random vectors $U^{(2)}[t]$, and we make sure that the joint distribution of the effective received signals at the destinations is the same as what we would have if we used coding scheme C_n on the $K \times K \times K$ wireless network under the additive uniform noise channel model. Therefore, if we apply the same decoding functions from coding scheme C_n at the destinations, the probability of error of our newly created coding scheme $\tilde{C}_n$ on the $K \times K \times K$ wireless network under the truncated model is exactly the same as the error probability of C_n on the $K \times K \times K$ wireless network under the additive uniform noise channel model. In order to remove the necessity of shared randomness, we simply notice that, since the random vectors $U^{(1)}[t]$ and $U^{(2)}[t]$ for $t = 1, ..., n$ are drawn independently from the transmit signals during the communication block, there must be vectors $u^{(1)}[t]$ and $u^{(2)}[t]$ with entries in $\left(-\frac{1}{2}, \frac{1}{2}\right)$, for $t = 1, ..., n$, such that, conditioned on $U^{(1)}[t] = u^{(1)}[t]$ and $U^{(2)}[t] = u^{(2)}[t]$ for $t = 1, ..., n$, the error probability of coding scheme C'_n is no larger than its unconditional error probability. This eliminates the need for shared randomness.

Finally, we notice that the power used by source s_j in coding scheme $\tilde{C}_n$ satisfies

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} \tilde{f}_j^2(w_j)[t] &= \frac{1}{n}\sum_{t=1}^{n}\left(f_j(w_j)[t] + \left(H_{S,U}^{-1}u^{(1)}[t]\right)_j\right)^2 \\
&\le \frac{2}{n}\sum_{t=1}^{n} f_j^2(w_j)[t] + \frac{2}{n}\sum_{t=1}^{n}\left(H_{S,U}^{-1}u^{(1)}[t]\right)_j^2 \\
&\le \frac{2}{n}\sum_{t=1}^{n} f_j^2(w_j)[t] + \underbrace{2\left(\sum_{\ell=1}^{K}\frac{1}{2}\left|\left(H_{S,U}^{-1}\right)_{j,\ell}\right|\right)^2}_{\alpha(s_j)}
\end{aligned}
$$

$$\leq 2P + \alpha(s_j),$$

where the last inequality follows from the fact that each codeword in our original coding scheme C_n satisfies an average power constraint of P. We conclude that the average power of each codeword of source s_j in our new coding scheme $\tilde{C}_n$ is at most $2P + \alpha(s_j)$, for $j = 1, ..., K$. By repeating the same steps, it can be shown that the power used by relay u_j in this new coding scheme is at most $2P + \alpha(u_j)$, where $\alpha(u_j)$ is similarly defined, for $j = 1, ..., K$. Finally, we may let $\alpha = \max_v \alpha(v)$, where the maximum is taken over all transmitter nodes $v \in \{s_1, ..., s_K, u_1, ..., u_K\}$, and the theorem follows. ∎

CHAPTER 7

GENERALIZED CUT-SET BOUND FOR MULTI-FLOW DETERMINISTIC NETWORKS

In this chapter, we propose a new generalization to the cut-set bound for deterministic K-unicast networks. The intuition behind our bound comes from noticing that a coding scheme for a K-unicast network $\mathcal{N}$, when applied to a *concatenation* of multiple copies of $\mathcal{N}$, can be used to achieve the original rates while inducing essentially the same distribution on the transmit signals of each copy of $\mathcal{N}$. Hence, one should be able to apply the cut-set bound to the concatenated network with a restriction on the possible transmit signal distributions. As we show, one can in fact require the transmit signals distribution on each copy to be *the same*, which can significantly reduce the values that the mutual information terms attain.

In terms of applications, we first consider linear finite-field networks. These networks have recently received considerable attention as they allow the deterministic modeling of wireless networks and can provide insights about their AWGN counterparts. Similar to the cut-set bound in [6], we obtain a general outer-bound expression in terms of ranks of transfer matrices. We then focus on $K \times K \times K$ topologies. Besides being a canonical example of K-unicast multi-hop networks, as shown in Chapter 3, they reveal the significant role relays can play in interference management. For binary $K \times K \times K$ networks, our rank-based bound yields necessary and sufficient conditions for rate K to be achieved. Furthermore, using the result from the previous chapter, which relates the capacity of $K \times K \times K$ networks under the AWGN and the truncated deterministic models, we obtain a bound on the degrees of freedom of $K \times K \times K$ AWGN networks with general connectivity. This bound is tight in the case of the $K{\times}K{\times}K$ topology with "adjacent-cell interference" and allows us to establish graph-theoretic necessary and sufficient conditions for K degrees of freedom to be achievable in general topologies.

7.1 Generalizing the Cut-Set Bound

We consider a general K-unicast memoryless network $\mathcal{N}$, illustrated in Fig. 7.1. The network consists of a set of nodes V, out of which we have K sources $\mathcal{S} = \{s_1, ..., s_K\}$ and K corresponding destinations $\mathcal{D} = \{d_1, ..., d_K\}$. At each time $t = 1, 2, \ldots$, each node $v \in V$ transmits a symbol (or signal) $X_v[t] \in \mathcal{X}_v$ and each node $v \in V$ receives a signal $Y_v[t] \in \mathcal{Y}_v$, for arbitrary alphabets $\mathcal{X}_v$ and $\mathcal{Y}_v$. In general, for variables z_v indexed by $v \in V$, we will let $z_{\mathcal{A}} = (z_v : v \in \mathcal{A})$, and for $m \geq 1$, $z_v^m = (z_v[1], ..., z_v[m])$. Then, $Y_V[t]$, the signals received at time t, are determined by a function F as $Y_V[t] = F(X_V[t])$ if $\mathcal{N}$ is a deterministic network or by a conditional distribution $p(y_V|x_V)$ if $\mathcal{N}$ is a stochastic (memoryless) network. To simplify the exposition, we assume throughout that source nodes do not receive any signals (i.e., $\mathcal{Y}_{s_i} = \emptyset$) and destination nodes do not transmit any signal (i.e., $\mathcal{X}_{d_i} = \emptyset$).

For a K-unicast memoryless network, the classical cut-set bound states that,

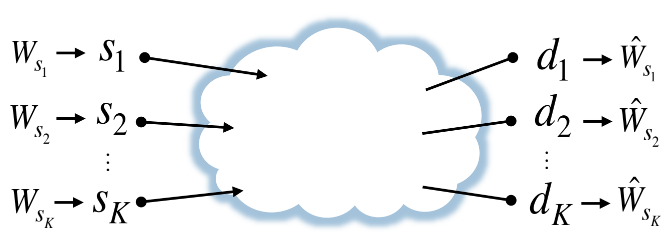


Figure 7.1: A general K-unicast network $\mathcal{N}$

if $(R_1, ..., R_K) \in C$, then there exists a distribution $p(x_V)$ on the transmit signals of all nodes in V (possibly with a power constraint in the case of AWGN networks) such that

$$\sum_{i=1}^{K} R_i \leq \min_{\substack{\Omega \subset V: \\ \mathcal{S} \subseteq \Omega \subseteq V - \mathcal{D}}} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}). \tag{7.1}$$

This outer bound is obtained by taking a coding scheme C_n out of a sequence that achieves a rate tuple $(R_1, ..., R_K)$ on $\mathcal{N}$ and showing that it induces a probability distribution $p(x_V)$ on the transmit signals such that, for any cut Ω, the sum rate $\sum_{i=1}^{K} R_i$ is upper-bounded by $I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c})$ plus the Fano error term. We generalize this bound in the case of deterministic networks as follows:

Theorem 7.1 *Consider a K-unicast deterministic network $\mathcal{N}$ with node set V. If a rate tuple $(R_1, ..., R_K)$ is achievable on $\mathcal{N}$, then there exists a joint distribution $p(x_V)$ on the transmit signals of the nodes in V, such that*

$$\sum_{i=1}^{K} R_i \leq \sum_{j=1}^{\ell} I(X_{\Omega_j}; Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_{\Omega_{j-1}^c}), \tag{7.2}$$

for all choices of ℓ node subsets $\Omega_1, ..., \Omega_\ell$ such that $V = \Omega_0 \supseteq \Omega_1 \supseteq \Omega_2 \supseteq ... \supseteq \Omega_\ell \supseteq \Omega_{\ell+1} = \emptyset$, and $d_i \in \Omega_j \Leftrightarrow s_i \in \Omega_{j+1}$ for $j = 0, 1, ..., \ell$, $i = 1, ..., K$ and any $\ell \geq 1$.

Remark 7.1 *If each $\mathcal{Y}_v$ is a discrete set, since the network is deterministic, the right-hand side of (7.2) reduces to $\sum_{j=1}^{\ell} H(Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_{\Omega_{j-1}^c})$.*

Remark 7.2 *The cut-set bound in (7.1) corresponds to $\ell = 1$.*

Remark 7.3 *If the network imposes a power constraint on the transmit signals, Theorem 7.1 holds for a distribution $p(x_V)$ whose covariance matrix satisfies such a constraint.*

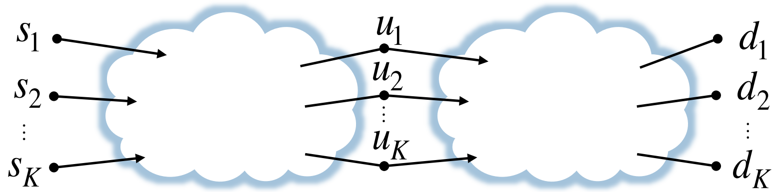


Figure 7.2: Concatenating two copies of $\mathcal{N}$ to obtain $\mathcal{N}^2$

Remark 7.4 *Both (7.1) and (7.2) can be used to bound the sum of a subset of the rates by treating the remaining sources and destinations as regular nodes.*

Remark 7.5 *In the case of wireline networks, the bound in Theorem 7.1 recovers and provides an alternative interpretation to the Generalized Network Sharing (GNS) bound [39, 70]. This is demonstrated in Section 7.2.4.*

The intuition behind this bound comes from noticing that a coding scheme C designed for a network $\mathcal{N}$ can also be applied on a *concatenation* of ℓ copies of $\mathcal{N}$, or $\mathcal{N}^\ell$, illustrated in Fig. 7.2 for $\ell = 2$, obtained by identifying each destination of copies $1, 2, ..., \ell - 1$ with the corresponding source on the next copy. More precisely, we have the following claim, whose proof, which is based on using coding scheme C on each copy of the network a repeated number of times, is presented in Appendix C.1.

Claim 7.1 *Let $C_\mathcal{N}$ and $C_{\mathcal{N}^\ell}$ be the capacity regions of a K-unicast memoryless network $\mathcal{N}$ and of the concatenation of ℓ copies of $\mathcal{N}$. Then $C_\mathcal{N} \subseteq C_{\mathcal{N}^\ell}$.*

Because of Claim 7.1, we can apply the cut-set bound to $\mathcal{N}^\ell$ in order to bound any sum rate achievable in $\mathcal{N}$. Hence, if we let V_1 and V_2 be the set of nodes of the first and second copies of the network respectively (and $V_1 \cap V_2 = \mathcal{U}$), we obtain

$$R_\Sigma \le \max_{p(x_{V_1 \cup V_2})} \min_{\Omega_1,\Omega_2} I(X_{\Omega_1}, X_{\Omega_2}; Y_{\Omega_1^c}, Y_{\Omega_2^c} | X_{\Omega_1^c}, X_{\Omega_2^c})$$

$$= \max_{p(x_{V_1 \cup V_2})} \min_{\Omega_1,\Omega_2} I(X_{\Omega_1}, X_{\Omega_2}; Y_{\Omega_1^c} | X_{\Omega_1^c}, X_{\Omega_2^c}) + I(X_{\Omega_1}, X_{\Omega_2}; Y_{\Omega_2^c} | X_{\Omega_1^c}, X_{\Omega_2^c}, Y_{\Omega_1^c}), \qquad (7.3)$$

where $R_\Sigma = \sum_{i=1}^K R_i$, $\Omega_i^c = V_i \setminus \Omega_i$ for $i = 1, 2$ and the minimization is over $\Omega_1 \subseteq V_1$ and $\Omega_2 \subseteq V_2 \setminus \mathcal{D}$ such that $\mathcal{S} \subseteq \Omega_1$ and $\Omega_1 \cap \mathcal{U} = \Omega_2 \cap \mathcal{U}$. We point out that this argument is tied to the multi-unicast nature of the network, which requires each u_i to be individually capable of decoding its message W_i. Since $\mathcal{Y}_{s_i} = \emptyset$ and $\mathcal{X}_{d_i} = \emptyset$ for $i = 1, ..., K$, we have the Markov chains $X_{V_2} \leftrightarrow X_{V_1} \leftrightarrow Y_{V_1}$ and $X_{V_1} \leftrightarrow X_{V_2} \leftrightarrow Y_{V_2 \setminus \mathcal{U}}$. Therefore, it is not difficult to see that the mutual information terms in (7.3) can be upper bounded as

$$I(X_{\Omega_1}, X_{\Omega_2}; Y_{\Omega_1^c} | X_{\Omega_1^c}, X_{\Omega_2^c}) + I(X_{\Omega_1}, X_{\Omega_2}; Y_{\Omega_2^c} | X_{\Omega_1^c}, X_{\Omega_2^c}, Y_{\Omega_1^c})$$

$$\le I(X_{\Omega_1}, X_{\Omega_2}, X_{\Omega_2^c}; Y_{\Omega_1^c} | X_{\Omega_1^c}) + I(X_{\Omega_1}, X_{\Omega_1^c}, X_{\Omega_2}; Y_{\Omega_2^c} | X_{\Omega_2^c}, Y_{\Omega_1^c})$$

$$= I(X_{\Omega_1}; Y_{\Omega_1^c} | X_{\Omega_1^c}) + I(X_{\Omega_2}, X_{\Omega_2^c}; Y_{\Omega_1^c} | X_{\Omega_1}, X_{\Omega_1^c})$$

$$+ I(X_{\Omega_2}; Y_{\Omega_2^c} | X_{\Omega_2^c}, Y_{\Omega_1^c}) + I(X_{\Omega_1}, X_{\Omega_1^c}; Y_{\Omega_2^c} | X_{\Omega_2}, X_{\Omega_2^c}, Y_{\Omega_1^c})$$

$$= I(X_{\Omega_1}; Y_{\Omega_1^c} | X_{\Omega_1^c}) + I(X_{\Omega_2}; Y_{\Omega_2^c} | X_{\Omega_2^c}, Y_{\Omega_1^c}),$$

and (7.3) can be written as

$$R_\Sigma \le \max_{p(x_{V_1 \cup V_2})} \min_{\Omega_1,\Omega_2} I(X_{\Omega_1}; Y_{\Omega_1^c} | X_{\Omega_1^c}) + I(X_{\Omega_2}; Y_{\Omega_2^c} | X_{\Omega_2^c}, Y_{\Omega_1^c}). \qquad (7.4)$$

For a general ℓ, by following the same argument, we conclude that, if a rate tuple $(R_1, ..., R_K)$ is achievable on $\mathcal{N}$, then there exists a joint distribution $p(x_{V_1 \cup ... \cup V_\ell})$ on the transmit signals of the nodes of the concatenated network $\mathcal{N}^\ell$, such that

$$R_\Sigma \le \min_{\Omega_1,...,\Omega_\ell} \sum_{j=1}^{\ell} I(X_{\Omega_j}; Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_1^c, ..., Y_{\Omega_{j-1}^c}), \qquad (7.5)$$

where the minimization is over subsets $\Omega_1, ..., \Omega_\ell$ such that $d_i \in \Omega_j \Leftrightarrow s_i \in \Omega_{j+1}$ for $j = 0, 1, ..., \ell$, $i = 1, ..., K$.

In order to obtain a bound on C_Σ from (7.5), one would need to maximize the right-hand side over all joint distributions $p(x_{V_1 \cup \ldots \cup V_\ell})$. However, as we shall see next, this maximization will result in uninteresting bounds. First we notice that, due to the Markov Chain $Y_{\Omega_1^c}, \ldots, Y_{\Omega_{j-1}^c} \leftrightarrow X_{V_j} \leftrightarrow Y_{\Omega_j^c \setminus \Omega_{j-1}^c}$, each term in (7.5) becomes

$$I(X_{\Omega_j}; Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_1^c, \ldots, Y_{\Omega_{j-1}^c}) = I(X_{\Omega_j}; Y_{\Omega_j^c \setminus \Omega_{j-1}^c} | X_{\Omega_j^c}) - I(Y_{\Omega_1^c}, \ldots, Y_{\Omega_{j-1}^c}; Y_{\Omega_j^c \setminus \Omega_{j-1}^c} | X_{\Omega_j^c}),$$

and it is not difficult to see that (7.5) is always maximized by product distributions $p(x_{V_1 \setminus V_2}) p(x_{V_2 \setminus V_3}) \ldots p(x_{V_\ell})$. In the case $\ell = 2$, for example, (7.5) implies that

$$R_\Sigma \leq \max_{p(x_{V_1 \cup V_2})} \min_{(\Omega_1, \Omega_2) \in \mathcal{K}} \left[I(X_{\Omega_1}; Y_{\Omega_1^c} | X_{\Omega_1^c}) + I(X_{\Omega_2}; Y_{\Omega_2^c \setminus \mathcal{U}} | X_{\Omega_2^c}) - I(Y_{\Omega_1^c}; Y_{\Omega_2^c} | X_{\Omega_2^c}) \right] \quad (7.6)$$

$$= \max_{p(x_{V_1 \setminus V_2}) p(x_{V_2})} \min_{(\Omega_1, \Omega_2) \in \mathcal{K}} \left[I(X_{\Omega_1}; Y_{\Omega_1^c} | X_{\Omega_1^c}) + I(X_{\Omega_2}; Y_{\Omega_2^c \setminus \mathcal{U}} | X_{\Omega_2^c}) \right] \quad (7.7)$$

which is similar to applying the cut-set bound first to the pairs $\{(s_i, d_i) : i \in \mathcal{I}\}$ where $\mathcal{I} = \{i : u_i \in \mathcal{U} \setminus \Omega_1\}$ and then to the pairs $\{(s_i, d_i) : i \notin \mathcal{I}\}$ (although not exactly the same).

In Theorem 7.1, we overcome this issue by, instead of taking cuts $\Omega_1 \subset V_1, \ldots, \Omega_\ell \subset V_\ell$ from concatenated copies of $\mathcal{N}$, taking multiple cuts from $\mathcal{N}$ itself; i.e., $\Omega_j \subset V$, for $j = 1, \ldots, \ell$ (with the additional restriction that $\Omega_1 \supseteq \ldots \supseteq \Omega_\ell$). Thus, for deterministic networks, this can be thought of as restricting the maximization in (7.7) to be over distributions where $X_{V_1} = X_{V_2}$ with probability 1. Intuitively, this choice makes the negative mutual information term in (7.7) as large as possible. The following example illustrates the gains of the bound in Theorem 7.1 over the traditional cut-set bound.

Example 7.3. Consider the binary Z-channel in Fig. 7.3(a). It is easy to see that $C_\Sigma = 1$, while the traditional cut-set bound only implies $C_\Sigma \leq 2$. Now consider concatenating two copies of this Z-channel and choosing Ω_1 and Ω_2 as shown in Fig. 7.3(b). By maximizing over all distributions $p(x_{V_1 \cup V_2})$, as in (7.6), we again

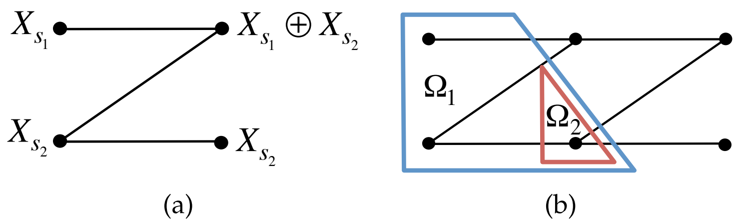


Figure 7.3: (a) A binary Z-channel, and (b) a possible choice of cuts for the concatenation of two binary Z-channels.

obtain $C_\Sigma \leq 2$. However, if we take the corresponding choices of Ω_1 and Ω_2 in Theorem 7.1 (i.e., $\Omega_1 = \{s_1, s_2, d_2\}$, $\Omega_2 = \{s_2\}$ in the original network), we obtain

$$\begin{aligned} C_\Sigma &\leq I(X_{\Omega_1}; Y_{\Omega_1^c} | X_{\Omega_1^c}) + I(X_{\Omega_2}; Y_{\Omega_2^c} | X_{\Omega_2^c}, Y_{\Omega_1^c}) \\ &= I(X_{s_1}, X_{s_2}; X_{s_1} \oplus X_{s_2}) + I(X_{s_2}; X_{s_1} \oplus X_{s_2}, X_{s_2} | X_{s_1}, X_{s_1} \oplus X_{s_2}) \leq 1 + 0. \end{aligned}$$

Next we prove Theorem 7.1. Even though the motivation behind the result is based on the concatenation of multiple copies of a network $\mathcal{N}$, the actual proof does not involve the notion of concatenation and follows by manipulating mutual-information inequalities on the original network $\mathcal{N}$.

Proof of Theorem 7.1: We first prove the case $\ell = 2$. We let $\Omega_1, \Omega_2 \subset V$ be such that $\mathcal{S} \subseteq \Omega_1$, $\Omega_2 \subseteq \Omega_1 \setminus \mathcal{D}$ and $d_i \in \Omega_1 \Leftrightarrow s_i \in \Omega_2$ and we let $\{C_n\}$ be a sequence of coding schemes that achieves sum rate R_Σ on $\mathcal{N}$. By applying coding scheme C_n of block length n on $\mathcal{N}$, we obtain

$$\begin{aligned} nR_\Sigma &= H(W_\mathcal{S}) = I(W_\mathcal{S}; Y_\mathcal{D}^n) + H(W_\mathcal{S} | Y_\mathcal{D}^n) \\ &\overset{(i)}{\leq} I(W_\mathcal{S}; Y_\mathcal{D}^n) + n\epsilon_n \leq I(W_\mathcal{S}; Y_{\Omega_2^c}^n) + n\epsilon_n \\ &= I(W_\mathcal{S}; Y_{\Omega_2^c \cap \Omega_1}^n, Y_{\Omega_1^c}^n) + n\epsilon_n \end{aligned}$$

$$= I(W_{\mathcal{S}}; Y^n_{\Omega_1^c}) + I(W_{\mathcal{S}\setminus\Omega_2}, W_{\mathcal{S}\cap\Omega_2}; Y^n_{\Omega_2^c\cap\Omega_1} | Y^n_{\Omega_1^c}) + n\epsilon_n$$

$$= \underbrace{I(W_{\mathcal{S}}; Y^n_{\Omega_1^c})}_{\mathrm{I}} + \underbrace{I(W_{\mathcal{S}\setminus\Omega_2}; Y^n_{\Omega_2^c\cap\Omega_1} | Y^n_{\Omega_1^c})}_{\mathrm{II}} + \underbrace{I(W_{\mathcal{S}\cap\Omega_2}; Y^n_{\Omega_2^c\cap\Omega_1} | Y^n_{\Omega_1^c}, W_{\mathcal{S}\setminus\Omega_2})}_{\mathrm{III}} + n\epsilon_n \tag{7.8}$$

where (i) follows from Fano's inequality. By following the steps in the usual cut-set bound proof (see [23, Theorem 18.1]), for term (I) we have

$$\begin{aligned} I(W_{\mathcal{S}}; Y^n_{\Omega_1^c}) &= \sum_{t=1}^{n} I(W_{\mathcal{S}}; Y_{\Omega_1^c}[t] | Y^{t-1}_{\Omega_1^c}) \\ &= \sum_{t=1}^{n} I(W^n_{\mathcal{S}}; Y_{\Omega_1^c}[t] | Y^{t-1}_{\Omega_1^c}, X_{\Omega_1^c}[t]) \\ &\le \sum_{t=1}^{n} I(W_{\mathcal{S}}, Y^{t-1}_{\Omega_1^c}; Y_{\Omega_1^c}[t] | X_{\Omega_1^c}[t]) \\ &\le \sum_{t=1}^{n} I(W_{\mathcal{S}}, Y^{t-1}_{\Omega_1^c}, X_{\Omega_1}[t]; Y_{\Omega_1^c}[t] | X_{\Omega_1^c}[t]) \\ &\le \sum_{t=1}^{n} I(X_{\Omega_1}[t]; Y_{\Omega_1^c}[t] | X_{\Omega_1^c}[t]). \end{aligned} \tag{7.9}$$

Term (II) can be upper-bounded by $H(W_{\mathcal{S}\setminus\Omega_2} | Y^n_{\Omega_1^c}) \le n\epsilon'_n$, where $\epsilon'_n \to 0$ from Fano's inequality, since $s_i \in \mathcal{S} \setminus \Omega_2 \Leftrightarrow d_i \in \Omega_1^c \cap \mathcal{D}$. Finally, for term (III), we obtain

$$\begin{aligned} &I(W_{\mathcal{S}\cap\Omega_2}; Y^n_{\Omega_2^c\cap\Omega_1} | Y^n_{\Omega_1^c}, W_{\mathcal{S}\setminus\Omega_2}) \\ &\quad = I(W_{\mathcal{S}\cap\Omega_2}; Y^n_{\Omega_2^c} | Y^n_{\Omega_1^c}, W_{\mathcal{S}\setminus\Omega_2}) \\ &\quad = \sum_{t=1}^{n} I(W_{\mathcal{S}\cap\Omega_2}; Y_{\Omega_2^c}[t] | Y^{t-1}_{\Omega_2^c}, Y^n_{\Omega_1^c}, W_{\mathcal{S}\setminus\Omega_2}) \\ &\quad \overset{(i)}{=} \sum_{t=1}^{n} I(W_{\mathcal{S}\cap\Omega_2}; Y_{\Omega_2^c}[t] | Y^{t-1}_{\Omega_2^c}, Y^n_{\Omega_1^c}, W_{\mathcal{S}\setminus\Omega_2}, X_{\Omega_2^c}[t]) \\ &\quad \le \sum_{t=1}^{n} I(W_{\mathcal{S}}, Y^{t-1}_{\Omega_2^c}, Y^n_{\Omega_1^c}, X_{\Omega_2}[t]; Y_{\Omega_2^c}[t] | Y_{\Omega_1^c}[t], X_{\Omega_2^c}[t]) \\ &\quad = \sum_{t=1}^{n} I(X_{\Omega_2}[t]; Y_{\Omega_2^c}[t] | Y_{\Omega_1^c}[t], X_{\Omega_2^c}[t]) + I(W_{\mathcal{S}}, Y^{t-1}_{\Omega_2^c}, Y^n_{\Omega_1^c}; Y_{\Omega_2^c}[t] | Y_{\Omega_1^c}[t], X_V[t]) \\ &\quad \overset{(ii)}{=} \sum_{t=1}^{n} I(X_{\Omega_2}[t]; Y_{\Omega_2^c}[t] | Y_{\Omega_1^c}[t], X_{\Omega_2^c}[t]) \end{aligned} \tag{7.10}$$

where (i) follows because from $Y^{t-1}_{\Omega_2^c}$ we can build $X_{\Omega_2^c\setminus\mathcal{S}}[t]$ and from $W_{\mathcal{S}\setminus\Omega_2}$ we can build $X_{\Omega_2^c\cap\mathcal{S}}[t]$ and (ii) because $Y_{\Omega_2^c}[t]$ is a function of $X_V[t]$. Therefore, (7.8) implies that

$$R_\Sigma \leq \frac{1}{n}\sum_{t=1}^{n}[I(X_{\Omega_1}[t];Y_{\Omega_1^c}[t]|X_{\Omega_1^c}[t]) + I(X_{\Omega_2}[t];Y_{\Omega_2^c}[t]|Y_{\Omega_1^c}[t],X_{\Omega_2^c}[t])] + (\epsilon_n + \epsilon_n').$$

Following [23], we let Q be a uniform r.v. on $\{1,...,n\}$ and we set $\tilde{X}_V = X_V[Q]$ so that $Q \leftrightarrow \tilde{X}_V \leftrightarrow \tilde{Y}_V$, and we obtain

$$\begin{aligned}
R_\Sigma &\leq I(X_{\Omega_1}[Q];Y_{\Omega_1^c}[Q]|X_{\Omega_1^c}[Q],Q) + I(X_{\Omega_2}[Q];Y_{\Omega_2^c}[Q]|Y_{\Omega_1^c}[Q],X_{\Omega_2^c}[Q],Q) + \epsilon_n + \epsilon_n' \\
&\leq I(X_{\Omega_1}[Q];Y_{\Omega_1^c}[Q]|X_{\Omega_1^c}[Q]) + I(X_{\Omega_2}[Q];Y_{\Omega_2^c}[Q]|Y_{\Omega_1^c}[Q],X_{\Omega_2^c}[Q]) + \epsilon_n + \epsilon_n' \\
&\leq I(\tilde{X}_{\Omega_1};\tilde{Y}_{\Omega_1^c}|\tilde{X}_{\Omega_1^c}) + I(\tilde{X}_{\Omega_2};\tilde{Y}_{\Omega_2^c}|\tilde{Y}_{\Omega_1^c},\tilde{X}_{\Omega_2^c}) + \epsilon_n'',
\end{aligned}$$

where we let $\epsilon_n'' = \epsilon_n + \epsilon_n'$, and $\epsilon_n'' \to 0$ as $n \to \infty$. This concludes the proof in the case $\ell = 2$.

Now consider the case $\ell = 3$. Similar to the expression obtained in (7.8), this time we upper bound the sum rate as

$$\begin{aligned}
nR_\Sigma &\leq I(W_\mathcal{S};Y_\mathcal{D}^n) + n\epsilon_n \leq I(W_\mathcal{S};Y_{\Omega_3^c}^n) + n\epsilon_n \\
&= I(W_{\mathcal{S}\setminus\Omega_2},W_{\mathcal{S}\cap\Omega_2};Y^n_{\Omega_3^c\cap\Omega_1},Y^n_{\Omega_1^c}) + n\epsilon_n \\
&= \underbrace{I(W_\mathcal{S};Y^n_{\Omega_1^c})}_{\mathrm{I}} + \underbrace{I(W_{\mathcal{S}\setminus\Omega_2};Y^n_{\Omega_3^c\cap\Omega_1}|Y^n_{\Omega_1^c})}_{\mathrm{II}} + \underbrace{I(W_{\mathcal{S}\cap\Omega_2};Y^n_{\Omega_3^c\cap\Omega_1}|Y^n_{\Omega_1^c},W_{\mathcal{S}\setminus\Omega_2})}_{\mathrm{III}} + n\epsilon_n
\end{aligned} \tag{7.11}$$

Term (I) can be bounded as in (7.9) and term (II) can be bounded with a Fano error term as we did for term (II) in (7.8). Term (III) can be rewritten as

$$\begin{aligned}
&I(W_{\mathcal{S}\cap\Omega_2};Y^n_{\Omega_3^c\cap\Omega_1}|Y^n_{\Omega_1^c},W_{\mathcal{S}\setminus\Omega_2}) \\
&\quad = I(W_{\mathcal{S}\cap\Omega_2};Y^n_{\Omega_2^c\cap\Omega_1},Y^n_{\Omega_3^c\cap\Omega_2}|Y^n_{\Omega_1^c},W_{\mathcal{S}\setminus\Omega_2}) \\
&\quad = \underbrace{I(W_{\mathcal{S}\cap\Omega_2};Y^n_{\Omega_2^c\cap\Omega_1}|Y^n_{\Omega_1^c},W_{\mathcal{S}\setminus\Omega_2})}_{\mathrm{IV}} + \underbrace{I(W_{\mathcal{S}\cap\Omega_2};Y^n_{\Omega_3^c\cap\Omega_2}|Y^n_{\Omega_1^c},W_{\mathcal{S}\setminus\Omega_2},Y^n_{\Omega_2^c\cap\Omega_1})}_{\mathrm{V}}.
\end{aligned}$$

Term (IV) is the same as term (III) in (7.8) and can be upper-bounded as in (7.10). Term (V) is further broken down as

$$I(W_{(\mathcal{S}\cap\Omega_2)\setminus\Omega_3},W_{\mathcal{S}\cap\Omega_3};Y^n_{\Omega_3^c\cap\Omega_2}|Y^n_{\Omega_2^c},W_{\mathcal{S}\setminus\Omega_2})$$

$$= I(W_{(\mathcal{S}\cap\Omega_2)\setminus\Omega_3}; Y^n_{\Omega_3^c\cap\Omega_2} | Y^n_{\Omega_2^c}, W_{\mathcal{S}\setminus\Omega_2}) + I(W_{\mathcal{S}\cap\Omega_3}; Y^n_{\Omega_3^c\cap\Omega_2} | Y^n_{\Omega_2^c}, W_{\mathcal{S}\setminus\Omega_3})$$

$$\leq \underbrace{I(W_{\mathcal{S}\setminus\Omega_3}; Y^n_{\Omega_3^c\cap\Omega_2} | Y^n_{\Omega_2^c}, W_{\mathcal{S}\setminus\Omega_2})}_{\text{VI}} + \underbrace{I(W_{\mathcal{S}\cap\Omega_3}; Y^n_{\Omega_3^c\cap\Omega_2} | Y^n_{\Omega_2^c}, W_{\mathcal{S}\setminus\Omega_3})}_{\text{VII}}.$$

As in the case of term (II), we can upper-bound term (VI) by $H(W_{\mathcal{S}\setminus\Omega_3} | Y^n_{\Omega_2^c}) \leq n\epsilon''_n$, where $\epsilon''_n \to 0$ from Fano's inequality, since $s_i \in \mathcal{S} \setminus \Omega_3 \Leftrightarrow d_i \in \Omega_2^c \cap \mathcal{D}$. Finally, we notice that term (VII) is exactly like term (IV) after increasing all indices by one, and can again be upper-bound as in (7.10). By combining all these facts, from (7.11), the sum-rate is upper-bounded as

$$R_\Sigma \leq \frac{1}{n} \sum_{t=1}^{n} [I(X_{\Omega_1}[t]; Y_{\Omega_1^c}[t] | X_{\Omega_1^c}[t]) + I(X_{\Omega_2}[t]; Y_{\Omega_2^c}[t] | Y_{\Omega_1^c}[t], X_{\Omega_2^c}[t])$$
$$+ I(X_{\Omega_3}[t]; Y_{\Omega_3^c}[t] | Y_{\Omega_2^c}[t], X_{\Omega_3^c}[t])] + \epsilon'''_n.$$

Using the same time-sharing variable Q as for the case $\ell = 2$, we conclude the proof for $\ell = 3$. It is straightforward to see that similar steps can be performed for any $\ell \geq 1$. ■

While in the wireline case, the bound in Theorem 7.1 recovers the GNS bound, its most interesting applications are in wireless settings. In the next section, we first consider several applications of the bound for wireless deterministic network models. We then show how, in the wireline case, the bound reduces to the GNS bound but, under certain restrictions on the allowed coding schemes (such as linear operations), it can be used to obtain tighter bounds.

7.2 Applications of the Bound

We will first consider finite-field deterministic networks and obtain a general outer-bound expression for the sum rate. In the case of binary $K \times K \times K$ networks, this bound provides necessary and sufficient conditions for sum rate K to be achievable. We then shift our focus to two-hop AWGN networks. Even though our outer bound only applies to deterministic networks, we will make use of a result from [54] that relates $K \times K \times K$ AWGN networks with a deterministic counterpart to obtain a bound for the degrees of freedom of $K \times K \times K$ networks with arbitrary connectivity. This bound, combined with a variation of the coding scheme introduced in [57] is then used to establish necessary and sufficient conditions for K degrees of freedom to be achievable on a $K \times K \times K$ AWGN network and to establish the degrees of freedom of the case of "adjacent-cell interference".

7.2.1 Linear Finite-Field Networks

A K-unicast linear finite-field network $\mathcal{N}$ is described by a directed graph $G = (V, E)$, where V is the node set and E is the edge set. If the network is layered, the node set V can be partitioned into r subsets $V_1, V_2, ..., V_r$ (the layers) in such a way that $E \subset \bigcup_{i=1}^{r-1} V_i \times V_{i+1}$, and $V_1 = \mathcal{S} = \{s_1, ..., s_K\}$, $V_r = \mathcal{D} = \{d_1, ..., d_K\}$. To each edge $(u, w) \in E$ we associate a nonzero channel gain $F(u, w)$ from a given finite field $\mathbb{F}$. For two sets of nodes $\mathcal{U} \subseteq V_i$ and $\mathcal{W} \subseteq V_{i+1}$, we let $F(\mathcal{U}, \mathcal{W})$ be the $|\mathcal{W}| \times |\mathcal{U}|$ transfer matrix from $\mathcal{U}$ to $\mathcal{W}$. The received signals at layer V_{j+1} are given by $Y_{V_{j+1}}[t] = F(V_j, V_{j+1}) X_{V_j}[t]$ for $t = 1, ..., n$. For conciseness, we let $\text{rank}(\mathcal{U}; \mathcal{W}) \triangleq \text{rank} F(\mathcal{U}, \mathcal{W})$, and for $\Omega \subset V$, we let $\Omega[j] = \Omega \cap V_j$. We also let $\bar{C}_\Sigma = C_\Sigma / \log |\mathbb{F}|$ be the normalized sum capacity. We have the following corollary of Theorem 7.1.

Corollary 7.1 *For a layered K-unicast linear finite-field network $\mathcal{N}$ as described above,* *if $R_\Sigma \le \bar{C}_\Sigma$, we must have*

$$R_\Sigma \le \sum_{j=1}^{r-1} \text{rank}(\Omega[j];\Omega^c[j+1]) + \text{rank}(\Theta[j];\Theta^c[j+1]) - \text{rank}(\Theta[j];\Omega^c[j+1]) \quad (7.12)$$

for any node subsets Ω and Θ such that $\Theta \subset \Omega \setminus \mathcal{D}$, $\mathcal{S} \subset \Omega$ and $d_i \in \Omega \Leftrightarrow s_i \in \Theta$.

Proof: We apply Theorem 7.1 with $\Omega_1 = \Omega$ and $\Omega_2 = \Theta$. For the first term in the sum in (7.2), we have

$$H(Y_{\Omega^c}|X_{\Omega^c}) \le \sum_{j=1}^{r-1} \text{rank}(\Omega[j];\Omega^c[j+1]) \cdot \log|\mathbb{F}|,$$

and for the second term we have

$$\begin{aligned} H(Y_{\Theta^c}|X_{\Theta^c}, Y_{\Omega^c}) &\le \sum_{j=1}^{r-1} H(Y_{\Theta^c[j+1]}|X_{\Theta^c[j]}, Y_{\Omega^c[j+1]}) \\ &\le \sum_{j=1}^{r-1} H\left(F(\Theta[j];\Theta^c[j+1])X_{\Theta[j]} \middle| F(\Theta[j];\Omega^c[j+1])X_{\Theta[j]}\right) \\ &\overset{(i)}{\le} \sum_{j=1}^{r-1} \Big(\text{rank}(\Theta[j];\Theta^c[j+1]) - \text{rank}(\Theta[j];\Omega^c[j+1])\Big) \cdot \log|\mathbb{F}|, \end{aligned}$$

where (i) follows since $H(A\mathbf{x}|B\mathbf{x})/\log|\mathbb{F}| \le \text{rank}\left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right] - \text{rank}B$ from Lemma C.1 in the appendix. ■

We point out that it is straightforward to generalize Corollary 7.1 to the wireless deterministic network model from [6] or to general finite-field networks with MIMO nodes.

We now shift our focus to $K \times K \times K$ networks, i.e., when $r = 3$ and $V_2 = \{u_1, ..., u_K\} \triangleq \mathcal{U}$. This network was recently studied in the AWGN case in [57], where K degrees of freedom were shown to be achievable. This result suggested that significant gains can be obtained from two-hop interference management, raising interest in the study of different two-hop network models. The following result provides necessary conditions for sum rate K to be achieved in the finite-field case.

Corollary 7.2 *For a $K \times K \times K$ finite-field network, if $\bar{C}_\Sigma = K$, then $F(\mathcal{U}, \mathcal{D})$ and $F(\mathcal{S}, \mathcal{U})$ must be invertible and*

(i) $F(s_i, u_j) = 0$ *if and only if* $\det F(\mathcal{U} \setminus \{u_j\}, \mathcal{D} \setminus \{d_i\}) = 0$, *for any* i, j

(ii) $F(u_j, d_i) = 0$ *if and only if* $\det F(\mathcal{S} \setminus \{s_i\}, \mathcal{U} \setminus \{u_j\}) = 0$, *for any* i, j.

Otherwise, $\bar{C}_\Sigma \le K - 1$.

Proof: Clearly, if $F(\mathcal{U}, \mathcal{D})$ or $F(\mathcal{S}, \mathcal{U})$ are not invertible, $\bar{C}_\Sigma \le K - 1$. We consider applying Corollary 7.1 with four different choices of Ω and Θ. For $\Omega = \mathcal{S} \cup \mathcal{U} \cup \{d_i\}$ and $\Theta = \{s_i\} \cup (\mathcal{U} \setminus \{u_j\})$, if $\bar{C}_\Sigma = K$, we obtain

$$\begin{aligned} K = \bar{C}_\Sigma &\le \operatorname{rank}(\mathcal{S}; \emptyset) + \operatorname{rank}(\{s_i\}; \{u_j\}) - \operatorname{rank}(\{s_i\}; \emptyset) \\ &\qquad + \operatorname{rank}(\mathcal{U}; \mathcal{D} \setminus \{d_i\}) + \operatorname{rank}(\mathcal{U} \setminus \{u_j\}; \mathcal{D}) \\ &\qquad - \operatorname{rank}(\mathcal{U} \setminus \{u_j\}; \mathcal{D} \setminus \{d_i\}) \\ &\le 2(K-1) + \operatorname{rank}(\{s_i\}; \{u_j\}) - \operatorname{rank}(\mathcal{U} \setminus \{u_j\}; \mathcal{D} \setminus \{d_i\}) \end{aligned}$$

which implies

$$\operatorname{rank}(\mathcal{U} \setminus \{u_j\}; \mathcal{D} \setminus \{d_i\}) - \operatorname{rank}(\{s_i\}; \{u_j\}) \le K - 2. \tag{7.13}$$

Next, by choosing $\Omega = \mathcal{S} \cup \mathcal{U} \setminus \{u_j\} \cup \{d_i\}$ and $\Theta = \{s_i\}$ $\Omega = \mathcal{S} \cup \{u_j\} \cup \mathcal{D} \setminus \{d_i\}$ and $\Theta = \mathcal{S} \setminus \{s_i\}$, and $\Omega = \mathcal{S} \cup \mathcal{U} \cup \mathcal{D} \setminus \{d_i\}$ and $\Theta = \mathcal{S} \setminus \{s_i\} \cup \{u_j\}$ we respectively obtain

$$K - 2 \le \operatorname{rank}(\mathcal{U} \setminus \{u_j\}; \mathcal{D} \setminus \{d_i\}) - \operatorname{rank}(\{s_i\}; \{u_j\}), \tag{7.14}$$

$$\operatorname{rank}(\mathcal{S} \setminus \{s_i\}; \mathcal{U} \setminus \{u_j\}) - \operatorname{rank}(\{u_j\}; \{d_i\}) \le K - 2, \tag{7.15}$$

$$K-2 \le \mathrm{rank}(\mathcal{S} \setminus \{s_i\}; \mathcal{U} \setminus \{u_j\}) - \mathrm{rank}(\{u_j\}; \{d_i\}). \tag{7.16}$$

Combining (7.13) and (7.14), we conclude that (7.14) holds with equality. Since $F(\mathcal{U}, \mathcal{D})$ is invertible, by Lemma C.3, $\mathrm{rank}(\mathcal{U} \setminus \{u_j\}; \mathcal{D} \setminus \{d_i\}) = K-2$ if $\det F(\mathcal{U} \setminus \{u_j\}, \mathcal{D} \setminus \{d_i\}) = 0$ and $\mathrm{rank}(\mathcal{U} \setminus \{u_j\}; \mathcal{D} \setminus \{d_i\}) = K-1$ otherwise, implying (*i*). Similarly, (7.15), (7.16) and the fact that $F(\mathcal{S}, \mathcal{U})$ is invertible imply (*ii*). ■

In the case $\mathbb{F} = GF(2)$, the conditions in Corollary 7.2 are in fact sufficient, and they imply the following:

Corollary 7.3 *For a $K \times K \times K$ finite-field network with $\mathbb{F} = GF(2)$, $C_\Sigma = K$ if and only if $F(\mathcal{U}, \mathcal{D})F(\mathcal{S}, \mathcal{U}) = I$.*

Proof: Clearly, if $C_\Sigma = K$, $F(\mathcal{S}, \mathcal{U})$ must be invertible. When $\mathbb{F} = GF(2)$, condition (*ii*) in Corollary 7.2 is equivalent to $\det F(\mathcal{S} \setminus \{s_i\}, \mathcal{U} \setminus \{u_j\}) = F(u_j, d_i)$. By definition, the (i, j)th entry of $F(\mathcal{S}, \mathcal{U})^{-1}$ can be written as the (j, i)th cofactor of $F(\mathcal{S}, \mathcal{U})$ divided by $\det F(\mathcal{S}, \mathcal{U}) = 1$, i.e.,

$$\begin{aligned} \left[F(\mathcal{S}, \mathcal{U})^{-1}\right]_{i,j} &= \frac{\det F(\mathcal{S} \setminus \{s_i\}, \mathcal{U} \setminus \{u_j\})}{\det F(\mathcal{S}, \mathcal{U})} \\ &= F(u_j, d_i) = [F(\mathcal{U}, \mathcal{D})]_{i,j}, \end{aligned}$$

and we conclude that $F(\mathcal{U}, \mathcal{D})F(\mathcal{S}, \mathcal{U}) = I$. Obviously, in this case, sum rate K can be achieved by having each relay forward its received signal. ■

7.2.2 Two-hop AWGN Networks

In this section we focus on $K \times K \times K$ wireless networks under an AWGN channel model. We follow the setup in Section 7.2.1, except that $\mathbb{F} = \mathbb{R}$,

$$Y_v[t] = \textstyle\sum_{u \in V} F(u, v) X_u[t] + Z_v[t] \tag{7.17}$$

is the received signal at node $v \in V \setminus \mathcal{S}$ at time t, where $Z_v[t]$ is the usual additive white Gaussian noise process, and there is a transmit power constraint $E[X_v^2] \leq P$ for $v \in V \setminus \mathcal{D}$. We will also consider the truncated deterministic channel model [6], where we still have a power constraint on X_v, but

$$Y_v[t] = \lfloor \textstyle\sum_{u \in V} F(u, v) X_u[t] \rfloor, \tag{7.18}$$

is the received signal. Based on the characterization of the Gaussian noise as the worst-case additive noise for wireless networks in Chapter 4, Corollary 6.2 was established in Chapter 6, relating the degrees-of-freedom region under these two models. We re-state it here in terms of sum degrees of freedom and with the notation of this chapter.

Lemma 7.1 *If $F(\mathcal{U}, \mathcal{D})$ and $F(\mathcal{S}, \mathcal{U})$ are invertible, the sum degrees of freedom of the $K \times K \times K$ wireless network under the AWGN channel model and under the truncated channel model satisfy $D_{\Sigma,\text{AWGN}} \leq D_{\Sigma,\text{Truncated}}$.*

Because of Lemma 7.1, any upper bound for the sum degrees of freedom of a $K \times K \times K$ network (with invertible transfer matrices) under the truncated model is also a bound for the sum degrees of freedom of the corresponding AWGN network. Since the $K \times K \times K$ network under the truncated channel model is a deterministic network, we can use Theorem 7.1 to upper-bound C_Σ and also D_Σ. Moreover, as implied by [6, Lemma 7.2], the degrees of freedom of a MIMO channel under the truncated deterministic model are given by the rank of the channel matrix. We obtain a version of Corollary 7.1 for truncated deterministic networks:

Corollary 7.1′. *For a layered K-unicast truncated deterministic network $\mathcal{N}$, we must have*

$$D_\Sigma \leq \sum_{j=1}^{r-1} \operatorname{rank}(\Omega[j]; \Omega^c[j+1]) + \operatorname{rank}(\Theta[j]; \Theta^c[j+1]) - \operatorname{rank}(\Theta[j]; \Omega^c[j+1])$$

for any node subsets Ω and Θ such that $\Theta \subset \Omega \setminus \mathcal{D}$, $\mathcal{S} \subset \Omega$ and $d_i \in \Omega \Leftrightarrow s_i \in \Theta$.

Proof: This result follows using the same steps as in the proof of Corollary 7.1, except that, instead of Lemma C.1, we use Lemma C.2, which is based on [6, Lemma 7.2]. ■

Since Corollary 7.2 follows directly from Corollary 7.1, we can also replace $\bar{C}_\Sigma$ with D_Σ in Corollary 7.2 and obtain necessary conditions for K degrees of freedom to be achievable in a truncated deterministic $K \times K \times K$ network. By Lemma 7.1, these conditions are also necessary in the case of AWGN networks, and interestingly, they turn out to also be sufficient. We will say that two node sets $\mathcal{A}, \mathcal{B} \subset V$ are matched if there is a perfect matching between $\mathcal{A}$ and $\mathcal{B}$ in E. Then we have:

Theorem 7.2 *For a $K \times K \times K$ AWGN network where $\mathcal{S}$ and $\mathcal{U}$ are matched and $\mathcal{U}$ and $\mathcal{D}$ are matched, if*

(i) $(s_i, u_j) \in E \Longleftrightarrow \mathcal{U} \setminus \{u_j\}$ *and* $\mathcal{D} \setminus \{d_i\}$ *are matched,*

(ii) $(u_j, d_i) \in E \Longleftrightarrow \mathcal{S} \setminus \{s_i\}$ *and* $\mathcal{U} \setminus \{u_j\}$ *are matched*

for any i, j, then, for almost all values of channel gains (of existing edges), $D_\Sigma = K$. Otherwise, $D_\Sigma \leq K - 1$ for almost all values of channel gains.

The necessary part follows by the previous discussion and by noticing that, for almost all choices of channel gains, (*i*) and (*ii*) are equivalent to (*i*) and (*ii*) in

Corollary 7.2. In order to prove the achievability part, we first need the following definition and lemma.

Definition 7.1 *A $K \times K \times K$ network with edge set E is diagonalizable if, for almost all assignments of real-valued channel gains to edges in E,*

- *$F(\mathcal{S}, \mathcal{U})^{-1}$ and $F(\mathcal{U}, \mathcal{D})$ have zeros at the same entries,*
- *$F(\mathcal{U}, \mathcal{D})^{-1}$ and $F(\mathcal{S}, \mathcal{U})$ have zeros at the same entries.*

Whereas the Aligned Network Diagonalization (AND) scheme was introduced in [57] for the case of $K \times K \times K$ networks with fully connected hops, it can be extended to the class of diagonalizable networks. This implies the following lemma, which we prove in Appendix C.3.

Lemma 7.2 *If a $K \times K \times K$ AWGN network is diagonalizable, then for almost all values of the channel gains, $D_\Sigma = K$.*

This lemma allows us to complete the proof of Theorem 7.2.

Proof of Achievability of Theorem 7.2: The (i, j)th entry of $F(\mathcal{S}, \mathcal{U})^{-1}$ can be written as

$$[F(\mathcal{S}, \mathcal{U})^{-1}]_{i,j} = \frac{\det(\mathcal{S} \setminus \{s_i\}, \mathcal{U} \setminus \{u_j\})}{\det F(\mathcal{S}, \mathcal{U})}.$$

Therefore, $[F(\mathcal{S}, \mathcal{U})^{-1}]_{i,j}$ is nonzero if and only if $\det(\mathcal{S} \setminus \{s_i\}, \mathcal{U} \setminus \{u_j\})$ is nonzero. The latter occurs for almost all values of channel gains if and only if $\mathcal{S} \setminus \{s_i\}$ and $\mathcal{U} \setminus \{u_j\}$ are matched, which by (*ii*) occurs if and only if $F(u_j, d_i) = [F(\mathcal{U}, \mathcal{D})]_{i,j} \neq 0$. Analogously we conclude that, for almost all values of channel gains,

$[F(\mathcal{U}, \mathcal{D})^{-1}]_{i,j}$ is nonzero if and only if $[F(\mathcal{S}, \mathcal{U})]_{i,j}$ is nonzero. Thus if a $K \times K \times K$ AWGN network satisfies the conditions in Theorem 7.2, it is diagonalizable and by Lemma 7.2, K degrees of freedom are achievable for almost all values of channel gains. ■

7.2.3 Two-Hop Networks with Adjacent-Cell Interference

The bound from Corollary 7.1, when applied to the degrees of freedom of $K \times K \times K$ AWGN networks, is also tight for the case of "adjacent-cell interference". As illustrated in Fig. 7.4, for this class of networks, $E = \{(s_i, u_j) : |i - j| \leq 1\} \cup \{(u_i, d_j) :$

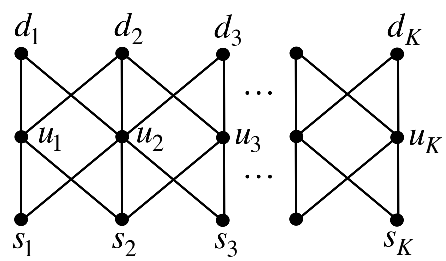


Figure 7.4: The $K \times K \times K$ Wireless Network with adjacent-cell interference.

$|i - j| \leq 1\}$. This configuration is motivated in the literature as the result of two-hop communication within each cell, when interference only occurs between adjacent cells [59].

Theorem 7.3 *The AWGN $K \times K \times K$ adjacent-cell interference network has $\left\lceil \frac{2K}{3} \right\rceil$ degrees of freedom for almost all values of channel gains.*

Proof: For the achievability, we consider several $2 \times 2 \times 2$ subnetworks formed by $\{s_i, s_{i+1}, u_i, u_{i+1}, d_i, d_{i+1}\}$ for $i = 1, 4, 7, ...$ In each one, we can use [24] to achieve

2 degrees of freedom, leaving the remaining nodes as "buffers" to prevent any interference between differen $2 \times 2 \times 2$ channels. If $K = 1 + 3m$ for some $m \in \mathbb{N}$, we utilize $\{s_K, u_K, d_K\}$ as a linear network where 1 degrees of freedom can be achieved. It is not difficult to see that this scheme achieves $\left\lceil \frac{2K}{3} \right\rceil$ degrees of freedom.

For the converse, we use the bound from Corollary 7.1, with $\Omega = \mathcal{S} \cup \mathcal{U} \cup d_{\mathcal{B}}$ and $\Theta = s_{\mathcal{B}} \cup u_{\mathcal{A}}$, where $\mathcal{A} = (\{1,2\} \cup \{5,6,7,8\} \cup \{11,12,13,14\}...) \cap \mathcal{K}$, $\mathcal{B} = (\{1\} \cup \{6,7\} \cup \{12,13\} \cup ...) \cap \mathcal{K}$ and $\mathcal{K} = \{1, ..., K\}$. First we notice that, since no index in $\mathcal{B}$ is adjacent to an index in $\mathcal{A}^c$, we have $\text{rank}(s_{\mathcal{B}}; u_{\mathcal{A}^c}) = 0$. Moreover, we have

$$
\begin{aligned}
&\text{rank}(\mathcal{U}; d_{\mathcal{B}^c}) + \text{rank}(u_{\mathcal{A}}; \mathcal{D}) \\
&\qquad \le |\mathcal{A}| + |\mathcal{B}^c| \\
&\qquad = |(\{1,2\} \cup \{5,6,7,8\} \cup \{11,12,13,14\} \cup ...) \cap \mathcal{K}| \\
&\qquad\quad + |(\{2,3,4,5\} \cup \{8,9,10,11\} \cup \{14,15,16,17\} \cup ...) \cap \mathcal{K}| \\
&\qquad = K + |\{2,5,8,11,...\} \cap \mathcal{K}| = K + \lfloor (K+1)/3 \rfloor .
\end{aligned}
$$

In order to compute $\text{rank}(u_{\mathcal{A}}, d_{\mathcal{B}^c})$, we notice that with the nodes of $u_{\mathcal{A}}$ and $d_{\mathcal{B}^c}$ we can build the matching

$$\{(1,2),(2,3),(5,4),(6,5),(7,8),(8,9),(11,10), ...\} \cap \mathcal{K} \times \mathcal{K},$$

which can be verified to have cardinality $\lceil 2(K-1)/3 \rceil$. Since either all the nodes in $u_{\mathcal{A}}$ or all the nodes in $d_{\mathcal{B}^c}$ are in this matching, we conclude that $\text{rank}(u_{\mathcal{A}}, d_{\mathcal{B}^c}) = \lceil 2(K-1)/3 \rceil$ for almost all values of channel gains, and the bound in (7.12) reduces to

$$K + \lfloor (K+1)/3 \rfloor - \lceil 2(K-1)/3 \rceil = \lceil 2K/3 \rceil ,$$

as we wanted to show. ■

7.2.4 Alternative Interpretation of the GNS Bound

A K-unicast wireline network $\mathcal{N}$ is characterized by a directed acyclic graph $G(V, E)$ where V is the node set and E the edge set. We let $\mathcal{I}(v) = \{u : (u, v) \in E\}$ and $\mathcal{O}(v) = \{u : (v, u) \in E\}$ and $\Delta = \max_v \max(|\mathcal{O}(v)|, |\mathcal{I}(v)|)$. At each time t, each $v \in V$ transmits a vector $X_v[t] \in \mathbb{F}^{|\mathcal{O}(v)|}$, for some finite field $\mathbb{F}$, where each component is called $X_{v,u}[t]$ for some $u \in \mathcal{O}(v)$. Each $v \in V$ receives a vector $Y_v[t] \in \mathbb{F}^{|\mathcal{I}(v)|}$, whose components are $X_{u,v}[t]$ for $u \in \mathcal{I}(v)$.

For K-unicast wireline networks, we consider a special case of the bound in Theorem 7.1 that can be seen to be equivalent to the Generalized Network Sharing bound [39, 70] and presents an alternative interpretation of this bound.

Corollary 7.4 (GNS Bound) *Let $\mathcal{N}^\ell$ be the concatenation of ℓ copies of a K-unicast wireline network $\mathcal{N}$. Suppose there is a set of edges $\mathcal{M} \subset E$ such that, by removing $\mathcal{M}$ from each of the ℓ copies of $\mathcal{N}$ in $\mathcal{N}^\ell$, all sources and destinations in $\mathcal{N}^\ell$ are disconnected. Then any rate tuple $(R_1, ..., R_K) \cdot \log |\mathbb{F}|$ achievable on $\mathcal{N}$ must satisfy*

$$\sum_{i=1}^{K} R_i \leq |\mathcal{M}|. \tag{7.19}$$

Proof: Let Ω be the set of nodes in $\mathcal{N}^\ell$ that are reachable from a source through a path that does not contain any edges in any of the copies of $\mathcal{M}$. Now let Ω_i be the nodes in Ω that are in the ith copy of $\mathcal{N}$. It is not difficult to check that $\Omega_1, ..., \Omega_\ell$ satisfy the conditions of Theorem 7.1. Now let $\delta(A, B) = \{(u, v) \in E : u \in \mathcal{A}, v \in \mathcal{B}\}$. We notice that if $v \in \Omega_j^c \setminus \mathcal{S}$ for some j, for each $u \in \mathcal{I}(v)$ we must either have $u \in \Omega_j^c$ or $(u, v) \in \mathcal{M} \cap \delta(\Omega_j, \Omega_j^c)$ (or else v would be in Ω_j). Hence,

$$H(Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_{\Omega_{j-1}^c}) \leq H(Y_{\Omega_j^c \cap \Omega_{j-1}} | X_{\Omega_j^c}) = H(X_{u,v} : (u, v) \in \mathcal{M} \cap \delta(\Omega_j, \Omega_j^c \cap \Omega_{j-1}))$$

Finally, since the sets $\Omega_1^c, \Omega_2^c \cap \Omega_1, \ldots, \Omega_\ell^c \cap \Omega_{\ell-1}$, are pairwise disjoint, Theorem 7.1 implies that

$$\sum_{i=1}^{K} R_i \leq \sum_{j=1}^{\ell} H(X_{u,v} : (u,v) \in \mathcal{M} \cap \delta(\Omega_j, \Omega_j^c \cap \Omega_{j-1})) \leq |\mathcal{M}| \log |\mathbb{F}|.$$

■

It is easy to check that this bound is equivalent to the GNS bound as stated in [38]. Moreover, the conditions in Corollary 7.4 provide a new interpretation to the bound, illustrated in Fig. 7.5.

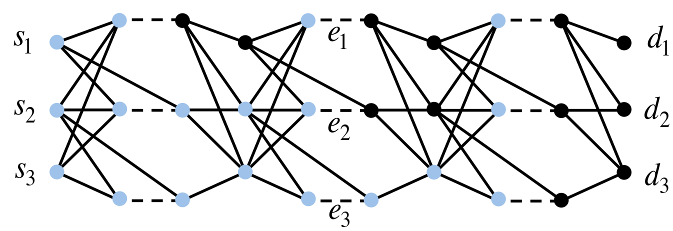


Figure 7.5: Illustration of the GNS bound for a 3-unicast network $\mathcal{N}$. By removing edges e_1, e_2 and e_3 (dashed) from all three copies of $\mathcal{N}$, we disconnect all sources and destinations. The nodes that can be reached from the sources after removing e_1, e_2 and e_3 (in blue) form Ω_1, Ω_2 and Ω_3.

7.2.5 Bounds for Linear Network Coding

Since the bound in Theorem 7.1 holds for general deterministic networks, if one restricts the kinds of relaying operations that can be used (say, to linear), these operations can be absorbed into the network. In this section, we illustrate one such example, where Theorem 7.1 can be used to obtain a bound that is tighter than the GNS bound. Consider the wireline network in Fig. 7.6, first introduced in [39]. With the purpose of finding an upper bound on $2R_1 + R_2$, we consider

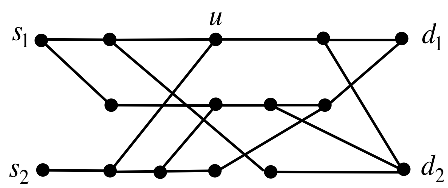


Figure 7.6: Two-unicast network where the GNS bound is not tight [39]

applying the concept of network concatenation but this time in a different fashion. We will concatenate the network in Fig. 7.6 sideways, as shown in Fig. 7.7. It is not difficult to see that if (R_1, R_2) is achieved in the network in Fig. 7.6, then we can achieve rate (R_1, R_2, R_1) in this new network. Moreover, the fact that Theorem 7.1 can be applied to general deterministic networks allows us to consider a mixed network model where the nodes in the junction must "broadcast" the same signal into both copies of the network. Moreover, we can remove the dashed edges, since d_2 should be able to decode its message just using the signals from the first copy of the network. Now applying Theorem 7.1 with Ω_1, Ω_2

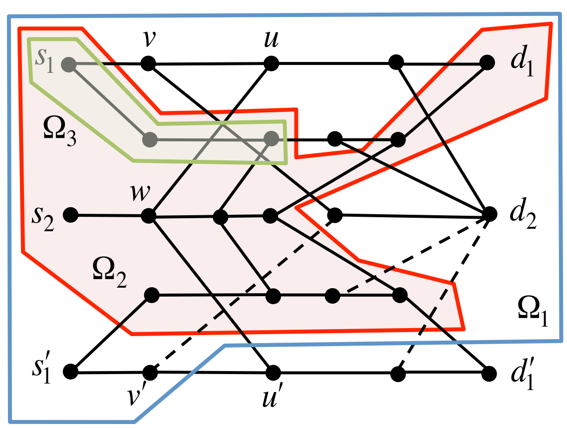


Figure 7.7: Cut choices to obtain a bound on $2R_1 + R_2$

and Ω_3 as shown in Fig. 7.7, we have

$$2R_1 + R_2 \leq 2 + [2 + H(Y_u | X_v, Y_{u'}, X_{v'})] + 0.$$

Finally we notice that, if we restrict ourselves to linear network coding, we can absorb the operation performed at u into the network, and have Y_u be the result of this operation. In this case, since $X_{w,u} = X_{w,u'}$, we will have $H(Y_u|X_v, Y_{u'}, X_{v'}) = 0$, which implies $2R_1 + R_2 \leq 4$. By noticing that $R_2 \leq 1$, this implies $R_1 + R_2 \leq 2.5$, which is achievable by linear network coding, as shown in [39].

7.3 Discussion and Extensions

In this chapter, we described a generalization of the classical cut-set bound for deterministic multi-flow networks. Besides having the potential for applications other than the ones presented in the last section, this work can be extended in several ways and raises several questions and directions for future research.

A simple improvement of Theorem 7.1 can be obtained by not requiring Ω_1 to include all the sources, which results in a bound for a subset of the rates. Notice that all such bounds should hold for the *same* joint distribution on the transmit signals $p(x_V)$. Moreover, by following the proof of the theorem, we notice that the joint distribution $\tilde{X}_V$ that we construct through the time-sharing variable Q satisfies the requirement that, if the sources do not receive any signal, or feedback (i.e., $\mathcal{Y}_{s_i} = \emptyset$), then $\tilde{X}_{s_i}$ is independent of $\tilde{X}_{s_j}$ for $i \neq j$. Hence, this restriction can be added to the statement of the result. This way, we obtain the following strengthened version of Theorem 7.1.

Theorem 7.4 *Consider a K-unicast deterministic network $\mathcal{N}$ with node set V. If a rate tuple $(R_1, ..., R_K)$ is achievable on $\mathcal{N}$, then there exists a joint distribution $p(x_V)$ on the* *transmit signals of the nodes in V, such that*

$$\sum_{i \in A} R_i \leq \sum_{j=1}^{\ell} I(X_{\Omega_j}; Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_{\Omega_{j-1}^c}), \tag{7.20}$$

for all choices of ℓ node subsets $\Omega_1, ..., \Omega_\ell$ such that $\Omega_1 \supseteq \Omega_2 \supseteq ... \supseteq \Omega_\ell$, and $d_i \in \Omega_j \Leftrightarrow s_i \in \Omega_{j+1}$ for $j = 1, ..., \ell - 1$, $i = 1, ..., K$ and any $\ell \geq 1$, where $A = \{i : s_i \in \Omega_1, d_i \notin \Omega_\ell\}$. Moreover, if the sources of $\mathcal{N}$ encode their messages solely based on their messages (i.e., they have no received signal), then we can require that $p(x_V) = p(x_{V-\mathcal{S}}|x_{\mathcal{S}}) \prod_{s_i \in \mathcal{S}} p(x_{s_i})$.

7.3.1 Asymmetric Bounds

One natural question raised by the result in Theorem 7.4 concerns asymmetric bounds, i.e., bounds of the form $\sum_{i=1}^{K} \alpha_i R_i \leq \beta$, with not all $\alpha_i \in \{0, 1\}$. Clearly, bounds of this form can be obtained by adding the bound in (7.20) for different subsets of the rates. However, we know that in several multi-user network scenarios, there are asymmetric bounds on the rates that are not implied by symmetric ones. This is the case, for example, of the class of deterministic two-user interference channels considered by El Gamal and Costa in [22].

One idea that may allow us to generalize Theorem 7.4 further so that non-trivial asymmetric bounds are also implied is an extension of the notion of network concatenation. Similar to what is done in Section 7.2.5, instead of concatenating different copies of the network one after the other, we consider "concatenating" them side by side. In Fig. 7.8(a), we have a deterministic two-user interference channel from [22], for which we assume that we have a sequence of coding schemes $\{C_n\}$ which allows rates R_1 and R_2 to be achieved. Then, in Fig. 7.8(b), we consider a new three-user channel, obtained essentially by taking two copies of the two-user interference channel and connecting them side by side. Notice that we remove the effect of source s_1' on d_2 in order to keep func-

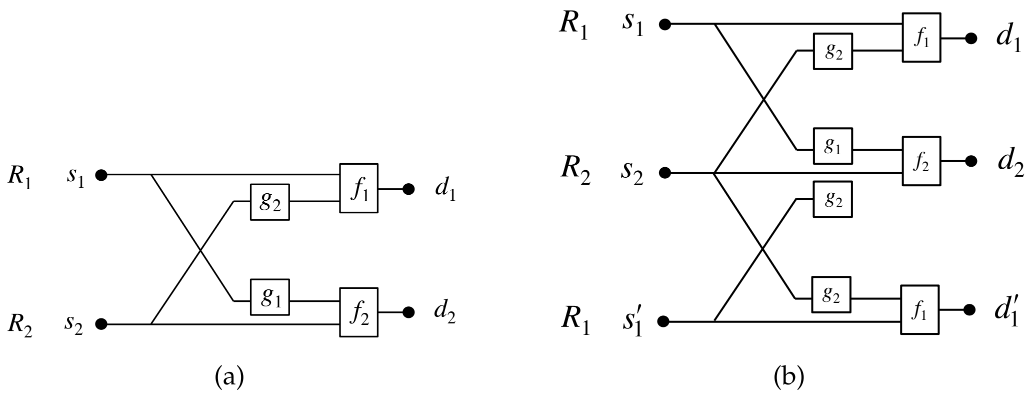


Figure 7.8: (a) The El Gamal-Costa 2-user IC; (b) 3-user IC obtained by combining two copies of the 2-user IC

tion f_2 the same. It is easy to see that by utilizing the encoding function of s_1 and the decoding function of d_1 from coding scheme C_n on source s_1' and destination d_1', we obtain a new sequence of coding schemes capable of achieving rate tuple (R_1, R_2, R_1) on the network in Fig. 7.8(b). Therefore, any bound on the sum rate of this network is also a bound on $2R_1 + R_2$ for the original network. Thus, we

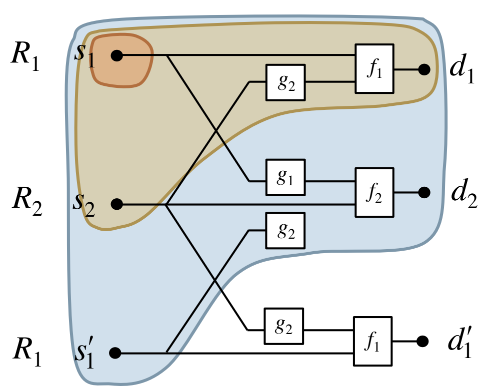


Figure 7.9: Choice of Ω_1, Ω_2 and Ω_3 for the "sideways-concatenated" network

now consider applying Theorem 7.4 on the network in Fig. 7.8(b) with $\ell = 3$ and the three nested cuts shown in Fig. 7.3.1. As a result, we conclude that

$$2R_1 + R_2 \leq H(Y_{d'_1}) + H(Y_{d_2}|Y_{d'_1}, X_{s'_1}) + H(Y_{d_1}|Y_{d'_1}, X_{s'_1}, Y_{d_2}, X_{s_2})$$
$$\leq H(Y_{d'_1}) + H(Y_{d_2}|V_2) + H(Y_{d_1}|V_2, V_1),$$

for some transmit signal distribution $p(x_{s_1,s_2,s'_1}) = p(x_{s_1})p(x_{s_2})p(x_{s'_1})$, where $V_1 = g_1(X_{s_1})$ and $V_2 = g_2(X_{s_2})$, which satisfy $V_1 = h_2(X_{s_2}, Y_{d_2})$ and $V_2 = h_1(X_{s_1}, Y_{d_1})$, according to the assumptions in [22]. Furthermore, based on our previous argument, we know that $2R_1 + R_2$ is achievable in the network in Fig. 7.8(b) with a coding scheme that induces the same distribution on X_{s_1} and $X_{s'_1}$. This would also guarantee that Y_{d_1} and $Y_{d'_1}$ follow the same distribution and therefore, we have

$$2R_1 + R_2 \leq H(Y_{d_1}) + H(Y_{d_2}|V_2) + H(Y_{d_1}|V_2, V_1), \tag{7.21}$$

which recovers the $2R_1 + R_2$ bound from [22].

But how can this kind of argument be generalized to other networks? Similar to what we did to obtain Theorem 7.1 from the intuition of network concatenation, one idea is to view the cuts taken on the concatenated network in Fig. 7.3.1 as cuts on the original network, as shown in Fig. 7.3.1. Notice that if Theo-

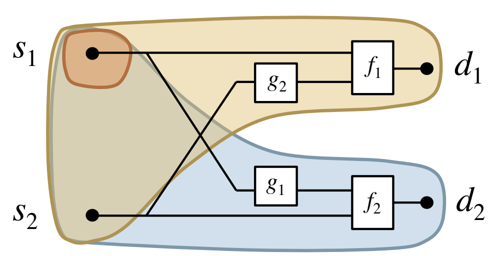


Figure 7.10: Choice of Ω_1, Ω_2 and Ω_3 on the original 2-user IC based on the choices for the "sideways-concatenated" network in Fig. 7.3.1

rem 7.4 allowed us to choose Ω_1, Ω_2 and Ω_3 that do not satisfy the decreasing-sets property, the resulting right-hand side of the bound in (7.20) would equal the right-hand side of the bound in (7.21). Hence, intuitively we would like to say that Theorem 7.4 can be generalized to allow arbitrary choices of $\Omega_1, ..., \Omega_\ell$, each of which results in a bound for a different linear combination of the rates. But what should be the coefficient α_i of each rate R_i? One idea that seems to capture the last example and many others would be to let α_i be the number of cuts Ω_j that contain s_i but not d_i. We conjecture that this is in fact correct, and one can obtain the following result (which can be verified to have Theorem 7.4 as a special case).

Conjecture 7.1 *Consider a K-unicast deterministic network $\mathcal{N}$ with node set V. If a rate tuple $(R_1, ..., R_K)$ is achievable on $\mathcal{N}$, then there exists a joint distribution $p(x_V)$ on the transmit signals of the nodes in V, such that*

$$\sum_{i=1}^{K} \alpha_i R_i \leq \sum_{j=1}^{\ell} I(X_{\Omega_j}; Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_{\Omega_{j-1}^c}), \tag{7.22}$$

for all choices of ℓ node subsets $\Omega_1, ..., \Omega_\ell$ and any $\ell \geq 1$, where $\alpha_i = \left|\{j : s_i \in \Omega_j, d_i \notin \Omega_j\}\right|$. Moreover, if the sources of $\mathcal{N}$ encode their messages solely based on their messages (i.e., they have no received signal), then we can require that $p(x_V) = p(x_{V-\mathcal{S}}|x_\mathcal{S}) \prod_{s_i \in \mathcal{S}} p(x_{s_i})$.

7.3.2 Non-Deterministic Networks

Another immediate question raised by Theorems 7.1 and 7.4 is related to the requirement of deterministic networks. Intuitively speaking, there does not seem to be a fundamental reason why the upper bound in (7.2) would hold only for deterministic networks. In fact, by following the steps in the proof of Theorem 7.1, we notice that the assumption of a deterministic network is only utilized in the "single-letterization" of the mutual information terms. Thus, one could state a multi-letter bound similar to that in (7.2) that holds for non-deterministic networks as well. Moreover, as shown in Section 7.2.2, by bounding the capacity region of AWGN two-hop networks with the capacity region of a deterministic counterpart, one can in fact use Theorem 7.1 to obtain bounds which are degrees-of-freedom tight in many two-hop networks.

We conjecture that the bound in Theorem 7.4 in fact holds for non-determinisic networks as well.

Conjecture 7.2 *Consider a K-unicast network $\mathcal{N}$ with node set V. If a rate tuple $(R_1, ..., R_K)$ is achievable on $\mathcal{N}$, then there exists a joint distribution $p(x_V)$ on the transmit signals of the nodes in V, such that*

$$\sum_{i \in A} R_i \leq \sum_{j=1}^{\ell} I(X_{\Omega_j}; Y_{\Omega_j^c} | X_{\Omega_j^c}, Y_{\Omega_{j-1}^c}), \tag{7.23}$$

for all choices of ℓ node subsets $\Omega_1, ..., \Omega_\ell$ such that $\Omega_1 \supseteq \Omega_2 \supseteq ... \supseteq \Omega_\ell$, and $d_i \in \Omega_j \Leftrightarrow s_i \in \Omega_{j+1}$ for $j = 1, ..., \ell - 1$, $i = 1, ..., K$ and any $\ell \geq 1$, where $A = \{i : s_i \in \Omega_1, d_i \notin \Omega_\ell\}$. Moreover, if the sources of $\mathcal{N}$ encode their messages solely based on their messages (i.e., they have no received signal), then we can require that $p(x_V) = p(x_{V-\mathcal{S}} | x_{\mathcal{S}}) \prod_{s_i \in \mathcal{S}} p(x_{s_i})$.

We point out that, in the non-deterministic setting, one can in general improve capacity bounds for multi-user networks by noticing that the ability of a destination d_i to decode its message depends only on the conditional distribution $p(y_{d_i}|x_V)$ (provided that there is no feedback). Thus, if Conjecture 7.2 holds, one can also improve it by replacing each Y_{d_i} with a $\tilde{Y}_{d_i}$ such that $p(y_{d_i}|x_V)$ and $p(\tilde{y}_{d_i}|x_V)$ are the same conditional distribution. In fact, the resulting conjecture would recover the bound in [40, Theorem 1], which unifies the bounds in [49] and [17].

CHAPTER 8
CONCLUSION

Multi-hop multi-flow wireless networks are becoming increasingly relevant with the advent of several new technologies and applications in wireless communications. However, traditional tools in network information theory are unable to characterize their fundamental capabilities and limitations. In this dissertation, we sought to develop new tools in order to study different aspects of the fundamentals of communication in these network scenarios.

In the first part of the dissertation, we developed new communication schemes and outer bounds that allowed us to establish the high-SNR performance limits in two-unicast layered wireless networks and two-hop K-unicast wireless networks, via degrees-of-freedom characterizations. The new introduced schemes combine relaying strategies such as amplify-and-forward and decode-and-forward with interference management techniques such as interference neutralization and interference alignment, and show that relays can be "game-changers" by reducing the effective interference experienced by the destination nodes and significantly increasing the achievable rates.

In the second part, we studied the robustness of the Gaussian model for multi-hop multi-flow wireless networks. We proved that such a model is in general a worst-case assumption, in the sense that it minimizes the capacity region. This generalizes a classical result in information theory for point-to-point channels and provides theoretical basis for the widespread adoption of Gaussian models in the field. Furthermore, our proof of this result is constructive and suggests an interesting engineering potential: even if the precise statistics of the noises in a network are unknown (which is often the case), it is still possible to design coding schemes which can achieve the Gaussian capacity region.

In the third and final part, we tackled the difficult challenge of obtaining novel capacity outer bounds for multi-hop multi-flow networks. The standard outer-bounding technique in information theory, the cut-set bound, has long been known to be loose, particularly in the case of multi-flow networks, and finding outer bounds that go beyond cut-set is important. We first utilized the worst-case noise result from part II in order to establish the truncated deterministic network model as an outer-bounding model for Gaussian networks. This allows one to look for capacity outer bounds for Gaussian networks by instead focusing on a deterministic counterpart. Then we introduced a new generalization of the cut-set bound, which holds for deterministic multi-hop multi-flow networks. Besides unifying several previously known bounds, by combining this new bound with the result that establishes the truncated deterministic network model as an outer-bounding model, we were able to characterize the degrees of freedom of several two-hop multi-flow wireless networks.

Due to the increasing practical relevance of multi-hop multi-flow wireless networks, in the next few years they should receive more and more attention from the research community. As we move forward, we expect approaches such as analyzing specific SNR regimes, seeking to indentify flow-like structures in multi-flow scenarios, establishing a relationship between different network models, and studying networks' deterministic counterparts, to play a key role in sheding some light on the principles of communication in these new scenarios. In this dissertation, we explored these approaches and developed new techniques that advanced the current understanding of multi-hop multi-flow networks and yielded foundations and tools for future research.

APPENDIX A

SUPPLEMENT FOR PART I

A.1 Proof of Lemma 3.1

Lemma 3.1. *Consider the vector* $\vec{p}(x_1, ..., x_m) = [p_1(x_1, ..., x_m), ..., p_\ell(x_1, ..., x_m)]^\dagger$*, where each* $p_i(x_1, ..., x_m)$ *is a distinct monomial on the variables* $x_1, ..., x_m$*. The determinant of the* $\ell \times \ell$ *matrix*

$$[\vec{p}(x_{1,1}, ..., x_{1,m}), \vec{p}(x_{2,1}, ..., x_{2,m}), ..., \vec{p}(x_{\ell,1}, ..., x_{\ell,m})]$$

is a non-identically zero polynomial on the variables $x_{1,1}, ..., x_{1,m}, ..., x_{\ell,1}, ..., x_{\ell,m}$*.*

Proof: Obviously, the determinant of $[\vec{p}(x_{1,1}, ..., x_{1,m}), \vec{p}(x_{2,1}, ..., x_{2,m}), ..., \vec{p}(x_{\ell,1}, ..., x_{\ell,m})]$ is a polynomial on the variables $x_{1,1}, ..., x_{1,m}, ..., x_{\ell,1}, ..., x_{\ell,m}$. To show that it is non-identically zero, we just need to show that, for some choice of $x_{1,1}, ..., x_{1,m}, ..., x_{\ell,1}, ..., x_{\ell,m}$, the determinant is nonzero. We do this by showing inductively that we can first choose values for $x_{1,1}, ..., x_{1,m}$, then $x_{2,1}, ..., x_{2,m}$ and so on, so that, when we choose $x_{j,1}, ..., x_{j,m}$, the column $\vec{p}(x_{j,1}, ..., x_{j,m})$ is linearly independent from $\vec{p}(x_{1,1}, ..., x_{1,m}), ..., \vec{p}(x_{j-1,1}, ..., x_{j-1,m})$. The base case is trivial. Now take any $j \in \{2, ..., \ell\}$, and suppose $x_{1,1}, ..., x_{1,m}, ..., x_{j-1,1}, ..., x_{j-1,m}$ have been chosen such that the linear space spanned by $\vec{p}(x_{1,1}, ..., x_{1,m}), ..., \vec{p}(x_{j-1,1}, ..., x_{j-1,m})$, $\mathscr{L}$, has dimension $j-1$. Since $j-1 < \ell$, there must be constants $\alpha_1, ..., \alpha_\ell$ (not all zero) such that, for any $(y_1, .., y_\ell) \in \mathscr{L}$, $\sum_{i=1}^{\ell} \alpha_i y_i = 0$. But since each $p_i(x_1, ..., x_m)$ is a distinct monomial on the variables $x_1, ..., x_m$, $\sum_{i=1}^{\ell} \alpha_i p_i(x_1, ..., x_m)$ is not identically zero. Thus, we can choose $x_{j,1}, ..., x_{j,m}$, such that $\sum_{i=1}^{\ell} \alpha_i p_i(x_{j,1}, ..., x_{j,m}) \neq 0$, which implies that $\vec{p}(x_{j,1}, ..., x_{j,m}) \notin \mathscr{L}$, completing the proof. ■

A.2 AND for Constant Channels

A.2.1 Scheme Description

In this subsection, we describe in detail the operations of AND when the $K \times K \times K$ wireless network has constant channels. As in Section 3.2.2, we describe the scheme by first considering the encoding at the sources, followed by the relaying operations and the decoding operations. Then, in the next subsection, we present a performance analysis of the scheme, where we formally prove that it achieves arbitrarily close to K degrees of freedom for almost all values of the channel gains.

Encoding at the sources:

Each source s_i starts by breaking its message W_i into L submessages. Each of the submessages will be encoded in a separate data stream, using a single codebook with codewords of length n, obtained by uniform i.i.d. sampling of the set

$$\mathcal{U} = \mathbb{Z} \cap \left[-P^{\frac{1-\epsilon}{2(d+\epsilon)}}, P^{\frac{1-\epsilon}{2(d+\epsilon)}}\right], \tag{A.1}$$

for a small $\epsilon > 0$, and $d = (N+1)^{K^2}$. The rate of this code, i.e., the number of codewords, will be determined later. Notice that d can be thought of as a parameter which sets the number of degrees of freedom given to each stream to be $(1-\epsilon)/(d+\epsilon) \approx 1/d$. The set of transmit directions $\mathcal{T}_N$ is defined as in (3.6), the only difference being that we drop the time index t, since the channels are constant. We again let $c_{i,\vec{s}}[m]$, for $0 \le m \le n-1$, represent the $(m+1)$th symbol of the codeword associated to the submessage of stream $\vec{s} \in \Delta_N$. At time

$t \in \{1, ..., n\}$, source s_i will thus transmit

$$X_{s_i}[t] = \gamma \sum_{\vec{s} \in \Delta_N} T_{\vec{s}}\, c_{i,\vec{s}}[t]$$

where $\gamma = \beta P^{\frac{d-1+2\epsilon}{2(d+\epsilon)}}$, for a contant β to be determined. Since the maximum power of a transmit signal from s_i can be loosely upper bounded by

$$\beta^2 P^{\frac{d-1+2\epsilon}{d+\epsilon}} \left(\sum_{\vec{s} \in \Delta_N} |T_{\vec{s}}| \right)^2 P^{\frac{1-\epsilon}{d+\epsilon}} = \beta^2 \left(\sum_{\vec{s} \in \Delta_N} |T_{\vec{s}}| \right)^2 P,$$

for any value of γ and N, we can choose the constant β such that the maximum transmit power at the sources is no more than P.

Relaying operations:

Similar to (3.8) and (3.9), in the case of constant channels the received signals can be equivalently written as

$$Y_{u_j}[t] = \gamma \sum_{\vec{s} \in \Delta_N} T_{\vec{s}} \left(\sum_{i=1}^{K} h_{s_i,u_j} c_{i,\vec{s}}[t] \right) + Z_{u_j}[t] \text{ or} \tag{A.2}$$

$$Y_{u_j}[t] = \gamma \sum_{\vec{s} \in \Delta_{N+1}} T_{\vec{s}}\, a_{j,\vec{s}}[t] + Z_{u_j}[t]. \tag{A.3}$$

Since in this case the code symbols $c_{i,\vec{s}}$ (and consequently each $a_{j,\vec{s}}$) are all integers, it makes sense to consider the (noiseless) received constellation at each relay, given by

$$\mathcal{V} = \left\{ \gamma \sum_{\vec{s} \in \Delta_{N+1}} T_{\vec{s}}\, a_{\vec{s}} : a_{\vec{s}} \in \mathbb{Z} \cap \left[-K\gamma P^{\frac{1-\epsilon}{2(d+\epsilon)}}, K\gamma P^{\frac{1-\epsilon}{2(d+\epsilon)}} \right], \forall\, \vec{s} \in \Delta_{N+1} \right\}. \tag{A.4}$$

Each relay u_j will map its received signal $Y_{u_j}[t]$ to the nearest point in $\mathcal{V}$. This point can then be used to obtain the integers $a_{j,\vec{s}}$, for $\vec{s} \in \Delta_{N+1}$, due to the following claim (which is later proven in the next subsection).

Claim A.1 *There exists a one-to-one map between points $v \in \mathcal{V}$ and tuples of integers $(a_{\vec{s}} : \vec{s} \in \Delta_{N+1})$ with entries in $\mathbb{Z} \cap \left[-KP^{\frac{1-\epsilon}{2(d+\epsilon)}}, KP^{\frac{1-\epsilon}{2(d+\epsilon)}} \right]$ such that $v = \gamma \sum_{\vec{s} \in \Delta_{N+1}} T_{\vec{s}}\, a_{\vec{s}}$.*

After decoding $a_{j,\vec{s}}$, for $\vec{s} \in \Delta_{N+1}$, using this one-to-one map, relay u_j will re-encode all these integers using new transmit directions, similar to those described in Section 3.2.2. More precisely, the transmit signal of relay u_j at time $t+1$ will be given by

$$X_{u_j}[t+1] = \gamma' \sum_{\vec{s} \in \Delta_{N+1}} \tilde{T}_{\vec{s}}\, a_{j,\vec{s}}[t], \tag{A.5}$$

where $\gamma' = \beta' P^{\frac{d-1+2\epsilon}{2(d+\epsilon)}}$, and β' is chosen so that the output power constraint is satisfied (similar to β). The new transmit directions $\tilde{T}_{\vec{s}}$ are defined exactly as before, according to (3.14), (C.6) and (3.16). Moreover, the natural equivalent of Claim 7.1 still holds in this case, and the transmit signals can be equivalently written as

$$X_{u_j}[t+1] = \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}} \left(\sum_{i=1}^{K} b_{ij}\, c_{i,\vec{s}}[t] \right). \tag{A.6}$$

Decoding at the destinations:

In order to compute the received signals at the destinations, similar to (C.8), we first express the transmit signals at time t in vector form, as

$$\begin{bmatrix} X_{u_1}[t+1] \\ \vdots \\ X_{u_K}[t+1] \end{bmatrix} = \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}} \begin{bmatrix} b_{11} & \dots & b_{K1} \\ \vdots & \ddots & \vdots \\ b_{1K} & \dots & b_{KK} \end{bmatrix} \begin{bmatrix} c_{1,\vec{s}}[t] \\ \vdots \\ c_{K,\vec{s}}[t] \end{bmatrix}. \tag{A.7}$$

Notice that the main difference between (A.7) and (C.8) is the absence of the noise term, since a decoding operation is performed at the relays in the case of constant channel gains. Then, similar to (3.20), we can obtain

$$\begin{bmatrix} Y_{d_1}[t+1] \\ \vdots \\ Y_{d_K}[t+1] \end{bmatrix} = \begin{bmatrix} h_{u_1,d_1} & \dots & h_{u_K,d_1} \\ \vdots & \ddots & \vdots \\ h_{u_1,d_K} & \dots & h_{u_K,d_K} \end{bmatrix} \begin{bmatrix} X_{u_1}[t+1] \\ \vdots \\ X_{u_K}[t+1] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t+1] \\ \vdots \\ Z_{d_K}[t+1] \end{bmatrix}$$

$$= \begin{bmatrix} b_{11} & \dots & b_{K1} \\ \vdots & \ddots & \vdots \\ b_{1K} & \dots & b_{KK} \end{bmatrix}^{-1} \begin{bmatrix} X_{u_1}[t+1] \\ \vdots \\ X_{u_K}[t+1] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t+1] \\ \vdots \\ Z_{d_K}[t+1] \end{bmatrix}$$

$$= \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}}[t] \begin{bmatrix} c_{1,\vec{s}}[t] \\ \vdots \\ c_{K,\vec{s}}[t] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t+1] \\ \vdots \\ Z_{d_K}[t+1] \end{bmatrix}. \tag{A.8}$$

Thus, the received signal at destination d_j at time $t+1$ is simply given by

$$Y_{d_j}[t+1] = \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}}\, c_{j,\vec{s}}[t] + Z_{d_j}[t+1]. \tag{A.9}$$

The points in the (noiseless) received constellation at each destination, given by

$$\tilde{\mathcal{V}} = \left\{ \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}}\, c_{\vec{s}} : c_{\vec{s}} \in \mathcal{U}, \forall\ \vec{s} \in \Delta_N \right\}, \tag{A.10}$$

can also be uniquely mapped into tuples of integers due to the following claim, also proved in the next subsection.

Claim A.2 *There exists a one-to-one map between points $v \in \tilde{\mathcal{V}}$ and tuples of integers $(c_{\vec{s}} : \vec{s} \in \Delta_N)$ with entries in $\mathcal{U}$ such that $v = \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}}\, c_{\vec{s}}$.*

At each time $t = 2, ..., n$, destination d_i will first map its received signal to the nearest point in $\tilde{\mathcal{V}}$ and then use the one-to-one map between points in $\tilde{\mathcal{V}}$ and tuples $(c_{\vec{s}} : \vec{s} \in \Delta_N)$ with entries in $\mathcal{U}$ to obtain the L integers $c_{i,\vec{s}}$ encoded by source s_i at time $t-1$. At time $n-1$, destination d_i has decoded $L = |\Delta_N|$ data streams of n integers each (in fact, $n-1$ integers, since the integers encoded by the source at time $t = n-1$ do not arrive at the destination within the length-n block), and it applies an individual typicality-based decoder to each of these streams to decode the source message W_i.

A.2.2 Performance Analysis

Next we show that AND for constant channels can in fact achieve K degrees of freedom. In order to do that, we first need to bound the error probability of the hard-decoding operations at the relays and destinations. In the process of doing that, we prove Claims A.1 and A.2.

Error probability of relaying operations:

To bound the error probability of the relaying operations, we need to find a lower bound on the minimum distance between two points in the received constellation $\mathcal{V}$, described in (A.4). Since the directions $T_{\vec{s}}$, for $\vec{s} \in \Delta_{N+1}$, are all distinct monomials of the channel gains of the first hop, they can be viewed as analytic functions of h_{s_i,u_j}, for $1 \leq i, j \leq K$, that are linearly independent over the reals. Moreover, the distance between any two points in $\mathcal{V}$ has the form

$$\gamma \sum_{\vec{s} \in \Delta_{N+1}} T_{\vec{s}}\, a_{\vec{s}},$$

where each $a_{\vec{s}}$ can take values in $\mathbb{Z} \cap \left[-2KP^{\frac{1-\epsilon}{2(d+\epsilon)}}, 2KP^{\frac{1-\epsilon}{2(d+\epsilon)}}\right]$. Thus, we can apply Theorem 5 in [45] (see also its subsequent remarks and inequality (8) in particular) to conclude that, for almost all values of the channel gains, there exists a constant κ, independent of P, such that the minimum distance of $\mathcal{V}$ satisfies

$$d_{\min} > \gamma \frac{\kappa}{\left(2KP^{\frac{1-\epsilon}{2(d+\epsilon)}}\right)^{|\Delta_{N+1}|-1+\epsilon}}.$$

Since $d = |\Delta_{N+1}| = (N+1)^{K^2}$, we have

$$d_{\min} > \frac{\kappa\beta P^{\frac{d-1+2\epsilon}{2(d+\epsilon)}}}{(2K)^{d-1+\epsilon} P^{\frac{(1-\epsilon)(d-1+\epsilon)}{2(d+\epsilon)}}} = \frac{\kappa\beta}{(2K)^{d-1+\epsilon}} P^{\epsilon/2}. \tag{A.11}$$

The fact that the minimum distance between any two points in $\mathcal{V}$ is strictly positive implies that there exists a one-to-one map between points $v \in \mathcal{V}$ and tuples of integers $(a_{\vec{s}} : \vec{s} \in \Delta_{N+1})$ with entries in $\mathbb{Z} \cap \left[-KP^{\frac{1-\epsilon}{2(d+\epsilon)}}, KP^{\frac{1-\epsilon}{2(d+\epsilon)}}\right]$, thus proving Claim A.1. Therefore, after mapping its received signal to the nearest point in $\mathcal{V}$, relay u_j can in fact decode each $a_{j,\vec{s}}$, $\vec{s} \in \Delta_{N+1}$, using this one-to-one map. This procedure will correctly decode each $a_{j,\vec{s}}$, provided that $|Z_{u_j}[t]| < d_{\min}/2$, implying that the probability of error for relay u_j is at most

$$\begin{aligned}\Pr(|Z_{u_j}[t]| \geq d_{\min}/2) &= 2\,Q\left(\frac{d_{\min}}{2\sigma}\right)\\ &\leq \exp\left(-\frac{d_{\min}^2}{8\sigma^2}\right)\\ &= \exp(-\delta P^{\epsilon}), \end{aligned} \tag{A.12}$$

where δ is a positive contant that is independent of P.

Error probability of symbol decoding at the destinations:

Similar to what we did for the received signals at the relays, we would like to lower bound the minimum distance between two points in the destinations' (noiseless) received constellation $\tilde{\mathcal{V}}$, given in (A.10). The following lemma, whose proof we present in Appendix A.3, allows us to use Theorem 5 from [45] as we did before.

Lemma A.1 *The received directions at the destinations, $\tilde{T}_{\vec{s}}$, for $\vec{s} \in \Delta_N$, are analytic functions of h_{u_i,d_j}, $1 \leq i, j \leq K$, that are linearly independent over the reals.*

Theorem 5 from [45] now implies that, for almost all values of the channel gains, the minimum distance $\tilde{d}_{\min}$ between any two points in $\tilde{\mathcal{V}}$ can be lower-bounded as

$$\tilde{d}_{\min} > \gamma' \frac{\tilde{\kappa}}{\left(2P^{\frac{1-\epsilon}{2(d+\epsilon)}}\right)^{|\Delta_N|-1+\epsilon}}$$

for some constant $\tilde{\kappa}$ (which is independent of P). Since $d = |\Delta_{N+1}| > |\Delta_N|$, for $P > 1$, we have

$$\tilde{d}_{\min} > \frac{\tilde{\kappa}\beta' P^{\frac{d-1+2\epsilon}{2(d+\epsilon)}}}{2^{|\Delta_N|-1+\epsilon} P^{\frac{(1-\epsilon)(d-1+\epsilon)}{2(d+\epsilon)}}} = \frac{\tilde{\kappa}\beta'}{2^{|\Delta_N|-1+\epsilon}} P^{\epsilon/2}. \tag{A.13}$$

The fact that the minimum distance between any two points in $\tilde{\mathcal{V}}$ is strictly positive implies that there exists a one-to-one map between points $v \in \tilde{\mathcal{V}}$ and tuples of integers $(c_{\vec{s}} : \vec{s} \in \Delta_N)$ with entries in $\mathbb{Z} \cap \left[-P^{\frac{1-\epsilon}{2(d+\epsilon)}}, P^{\frac{1-\epsilon}{2(d+\epsilon)}}\right]$, thus proving Claim A.2. After mapping its received signal to the nearest point in $\tilde{\mathcal{V}}$, destination d_j can in fact decode each $c_{j,\vec{s}}$, $\vec{s} \in \Delta_N$, using this one-to-one map. As in (A.12), the probability that d_i incorrectly decodes these integers (provided that no relay made an error in the previous step) is at most

$$\Pr(|Z_{d_j}[t]| \geq \tilde{d}_{\min}/2) = \exp(-\tilde{\delta} P^{\epsilon}), \tag{A.14}$$

for some constant $\tilde{\delta} > 0$.

Achievable rates:

To determine the rate of our original codebook, we first notice that each data stream between s_i and d_i effectively creates a discrete memoryless channel with input and output alphabets $\mathcal{U}$ and an error probability which can be upper bounded as

$$\begin{aligned} P_e &\leq 1 - (1 - \exp(-\delta P^{\epsilon}))^K \left(1 - \exp(-\tilde{\delta} P^{\epsilon})\right) \\ &\leq 1 - (1 - \exp(-\delta' P^{\epsilon}))^{K+1} \end{aligned}$$

$$\leq (K+1)\exp(-\delta' P^{\epsilon}), \tag{A.15}$$

where $\delta' = \min(\delta, \tilde{\delta})$. This allows us to lower bound the mutual information between the input U and the output $\hat{U}$ of this channel, for a uniform distribution over the input alphabet. Using Fano's inequality, we have

$$\begin{aligned} I(U;\hat{U}) &\geq H(U) - H(U|\hat{U}) \\ &\geq \log|\mathcal{U}| - (1 + P_e \log|\mathcal{U}|) \\ &= (1-P_e)\log|\mathcal{U}| - 1 \\ &\geq (1-(K+1)\exp(-\delta' P^{\epsilon}))\left(\frac{1-\epsilon}{d+\epsilon}\frac{\log P}{2} + 1\right) - 1, \end{aligned}$$

and we can achieve rate

$$R = (1-(K+1)\exp(-\delta' P^{\epsilon}))\left(\frac{1-\epsilon}{d+\epsilon}\frac{\log P}{2} + 1\right) - 1$$

over each data stream, by having our original codebook have 2^{nR} codewords. This means that each data stream can achieve

$$\lim_{P\to\infty} \frac{R}{\frac{1}{2}\log P} = \frac{1-\epsilon}{d+\epsilon} = \frac{1-\epsilon}{(N+1)^{K^2}+\epsilon}$$

degrees of freedom. Since each source transmits $L = |\Delta_N| = N^{K^2}$ data streams, each source-destination pair achieves a total of

$$\frac{(1-\epsilon)N^{K^2}}{(N+1)^{K^2}+\epsilon} \geq \frac{(1-\epsilon)N^{K^2}}{(1+\epsilon)(N+1)^{K^2}} = \frac{1-\epsilon}{1+\epsilon}\left(\frac{N}{N+1}\right)^{K^2}$$

degrees of freedom, for any large N and any small $\epsilon > 0$, implying that each source-destination pair can achieve arbitrarily close to one degree of freedom. We conclude that the aligned network diagonalization scheme can achieve arbitrarily close to K degrees of freedom for almost all values of the channel gains, which proves Theorem 3.2.

A.3 Proof of Lemma A.1

Lemma A.1. *The received directions at the destinations, $\tilde{T}_{\vec{s}}$, for $\vec{s} \in \Delta_N$, are analytic functions of h_{u_i,d_j}, $1 \leq i, j \leq K$, that are linearly independent over the reals.*

Proof: To prove that each $\tilde{T}_{\vec{s}}$, for $\vec{s} \in \Delta_N$, is an analytic function of h_{u_i,d_j}, $1 \leq i, j \leq K$, we notice that if we let

$$H = \begin{bmatrix} h_{u_1,d_1} & \dots & h_{u_K,d_1} \\ \vdots & \ddots & \vdots \\ h_{u_1,d_K} & \dots & h_{u_K,d_K} \end{bmatrix},$$

then, for $1 \leq i, j \leq K$ we can write $b_{ij} = \frac{C_{ji}}{\det H}$, where C_{ji} is the cofactor of the (j, i) entry of H. This means that each b_{ij} is a ratio of two polynomials with h_{u_i,d_j}, $1 \leq i, j \leq K$, as variables. Since each $\tilde{T}_{\vec{s}}$ is a distinct monomial of the b_{ij}s, each $\tilde{T}_{\vec{s}}$ is an analytic function of h_{u_i,d_j}, $1 \leq i, j \leq K$.

Next, suppose by contradiction that $\tilde{T}_{\vec{s}}$, for $\vec{s} \in \Delta_N$, are not linearly independent over the reals. Then there must be real numbers $\alpha_{\vec{s}}$, for $\vec{s} \in \Delta_N$, not all zero, such that

$$\sum_{\vec{s} \in \Delta_N} \alpha_{\vec{s}} \tilde{T}_{\vec{s}} = 0$$

for all values of h_{u_i,d_j}, for $1 \leq i, j \leq K$. However, since the $\tilde{T}_{\vec{s}}$, for $\vec{s} \in \Delta_N$ are distinct monomials of the b_{ij}s, we have that, for almost all values of the b_{ij}s, $\sum_{\vec{s} \in \Delta_N} \alpha_{\vec{s}} \tilde{T}_{\vec{s}} \neq 0$. Since for almost all values of the b_{ij}s, the matrix

$$B = \begin{bmatrix} b_{11} & \dots & b_{K1} \\ \vdots & \ddots & \vdots \\ b_{1K} & \dots & b_{KK} \end{bmatrix}$$

is invertible, we can find $b_{11}, b_{12}, ..., b_{KK}$ for which B is invertible and $\sum_{\vec{s}\in\Delta_N} \alpha_{\vec{s}}\tilde{T}_{\vec{s}} \neq 0$ (with the $\tilde{T}_{\vec{s}}$s seen as functions of the b_{ij}s). But this means that if we choose the values of h_{u_i,d_j}, for $1 \leq i, j \leq K$, by setting $H = B^{-1}$, we will have $\sum_{\vec{s}\in\Delta_N} \alpha_{\vec{s}}\tilde{T}_{\vec{s}} \neq 0$ (with the $\tilde{T}_{\vec{s}}$s seen as functions of the h_{u_i,d_j}s), which is a contradiction. ■

APPENDIX B

SUPPLEMENT FOR PART II

B.1 Auxiliary Results for the Proof of Lemma 4.2

Lemma B.1 *In the proof of Lemma 4.2, for each* $\mathbf{y} \in \mathcal{Y}$*,* $Q(\mathbf{y})$ *is a convex set.*

Proof: Consider two noise realizations $\mathbf{z}, \mathbf{z}' \in Q(\mathbf{y})$ and fix some $\alpha \in [0, 1]$. We will show that if we replace one of the components of $\mathbf{z}$ with the corresponding component of $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}'$, the resulting noise realization $\mathbf{z}''$ is still in $Q(\mathbf{y})$. Then, by using the same argument with $\mathbf{z}''$ instead of $\mathbf{z}$, another component of $\mathbf{z}''$ is replaced with a component $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}'$, and by repeating this argument, it follows that $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}'$ is itself in $Q(\mathbf{y})$. So let us focus on the component corresponding to node v at time ℓ. Let $y_v[\ell]^*$ be the noiseless version of the received signal at v at time ℓ with its complete binary expansion. Since $\mathbf{z}$ and $\mathbf{z}'$ result in the same $\mathbf{y}$, we have that

$$y_v[\ell] = \lfloor y_v[\ell]^* + z_v[\ell] \rfloor_\rho = \lfloor y_v[\ell]^* + z'_v[\ell] \rfloor_\rho .$$

Now, if we assume wlog that $z_v[\ell] \leq z'_v[\ell]$, we have

$$\begin{aligned} \lfloor y_v[\ell]^* + z_v[\ell] \rfloor_\rho &\leq \lfloor y_v[\ell]^* + \alpha z_v[\ell] + (1 - \alpha) z'_v[\ell] \rfloor_\rho \\ &\leq \lfloor y_v[\ell]^* + z'_v[\ell] \rfloor_\rho . \end{aligned}$$

It follows that $y_v[\ell] = \lfloor y_v[\ell]^* + \alpha z_v[\ell] + (1 - \alpha) z'_v[\ell] \rfloor_\rho$, and by replacing $z_v[\ell]$ with $\alpha z_v[\ell] + (1 - \alpha) z'_v[\ell]$, we obtain a noise realization $\mathbf{z}''$ that is still in $Q(\mathbf{y})$, and the lemma follows. ■

Lemma B.2 *Let* λ *denote the Lebesgue measure. Then, for any convex set* S*,* $\lambda(\partial S) = 0$*.*

Proof: Consider any point $p \in \partial S$. Clearly, $p \notin S^\circ$, and by the Supporting Hyperplane Theorem [7], there exists a hyperplane that passes through p and contains S in one of its closed half-spaces. Let H be such a closed half-space. Since H is closed, it is clear that $\partial S \subset H$. Then, for any closed ball $B_\epsilon(p)$ centered at p, it is clear that

$$\frac{\lambda(B_\epsilon(p) \cap \partial S)}{\lambda(B_\epsilon(p))} \leq \frac{\lambda(B_\epsilon(p) \cap H)}{\lambda(B_\epsilon(p))} = 1/2.$$

By Lebesgue's Density Theorem, the set

$$P = \left\{ p \in \partial S : \liminf_{\epsilon \to 0} \frac{\lambda(B_\epsilon(p) \cap \partial S)}{\lambda(B_\epsilon(p))} < 1 \right\}$$

should have Lebesgue measure zero. But since $P = \partial S$, we conclude that $\lambda(\partial S) = 0$. ■

B.2 Proof of Lemma 4.3

Lemma 4.3. *Suppose Y is a random variable with density f. Let $\tilde{Y}_m = \lfloor Y \rfloor_m + U_m$, where U_m is uniformly distributed in $\left(-2^{-m-1}, 2^{-m-1}\right)$ and independent from Y. Then each $\tilde{Y}_m$ has a density f_m, and f_m converges pointwise almost everywhere to f.*

Proof: Since the density of $U\left(-2^{-m-1}, 2^{-m-1}\right)$ is $g(x) = 2^m \mathbb{1}\{x \in \left(-2^{-m-1}, 2^{-m-1}\right)\}$, $\tilde{Y}_m$ will have a density f_m that can be written, for almost all y, as

$$\begin{aligned}
f_m(y) &= E\left[g\left(y - \lfloor Y \rfloor_m\right)\right] \\
&= 2^m E\left[\mathbb{1}\{y - \lfloor Y \rfloor_m \in \left(-2^{-m-1}, 2^{-m-1}\right)\}\right] \\
&= 2^m \Pr\left[y - \lfloor Y \rfloor_m \in \left(-2^{-m-1}, 2^{-m-1}\right)\right] \\
&= 2^m \Pr\left[\lfloor Y \rfloor_m \in \left(y - 2^{-m-1}, y + 2^{-m-1}\right)\right] \\
&= 2^m \Pr\left[\lfloor 2^m Y \rfloor \in (y2^m - 1/2, y2^m + 1/2)\right]
\end{aligned}$$

$$= 2^m \Pr\left[2^m Y \in (\lceil y2^m - 1/2\rceil, \lceil y2^m + 1/2\rceil)\right]$$

$$= 2^m \Pr\left[Y \in (2^{-m}\lceil y2^m - 1/2\rceil, 2^{-m}\lceil y2^m + 1/2\rceil)\right]$$

$$= 2^m \int_{a_m}^{b_m} f(x)dx, \tag{B.1}$$

where $a_m = 2^{-m}\lceil y2^m - 1/2\rceil$ and $b_m = 2^{-m}\lceil y2^m + 1/2\rceil$. Notice that we can write $b_m = a_m + 2^{-m}$. Moreover, we have that

$$y - 2^{-(m+1)} \leq a_m < y + 2^{-(m+1)}, \tag{B.2}$$

from which we have $a_m \to y$ as $m \to \infty$. If we let $F(y)$ be the cdf of Y, then (B.1) can be written as

$$\frac{F(b_m) - F(a_m)}{2^{-m}} = \frac{F(a_m + 2^{-m}) - F(a_m)}{2^{-m}} \triangleq q_m. \tag{B.3}$$

Our goal is to show that q_m converges to $f(y)$ as $m \to \infty$ for almost all y. Since by assumption Y has an absolutely continuous distribution, $F(y)$ is differentiable almost everywhere, so it suffices to show that q_m converges to $f(y)$ as $m \to \infty$ wherever $F(y)$ is differentiable and the derivative is $f(y)$. Thus, we focus on a y where $F'(y) = f(y)$. Suppose by contradiction that q_m does not converge to $f(y)$. Then there must be an $\epsilon > 0$ and a subsequence $\{q_{m_i}\}_{i=1}^{\infty}$, such that one of the following

$$q_{m_i} > f(y) + \epsilon \tag{B.4}$$

$$q_{m_i} < f(y) - \epsilon \tag{B.5}$$

holds for all $i \geq 1$. Suppose wlog that we have a subsequence $\{q_{m_i}\}_{i=1}^{\infty}$ for which (B.4) holds for all $i \geq 1$. We will now pick a further subsequence of $\{q_{m_i}\}_{i=1}^{\infty}$ in the following way. First, we choose $K \in \mathbb{Z}_+$ large enough so that $f(y)/K < \epsilon$, and we define K subsets of $\{1, 2, ...\}$ as

$$S_j = \left\{ i \geq 1 : y - 2^{-(m_i+1)} + \frac{j-1}{K}2^{-m_i} \leq a_{m_i} < y - 2^{-(m_i+1)} + \frac{j}{K}2^{-m_i} \right\},$$

for $j = 1, 2, ..., K$. From (B.2), the sets $S_1, ..., S_K$ partition $\{1, 2, ...\}$, and we must be able to find some S_j that is infinite. Suppose $|S_t| = \infty$. Then we have a subsequence $\{q_{m_i}\}_{i \in S_t}$, which we re-index as $\{q_{\ell_i}\}_{i=1}^{\infty}$. For each of the elements in this subsequence we have

$$
\begin{aligned}
q_{\ell_i} &= \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(a_{\ell_i})}{2^{-\ell_i}} \\
&= \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{2^{-\ell_i}} + \frac{F(y) - F(a_{\ell_i})}{2^{-\ell_i}} \\
&= \frac{a_{\ell_i} + 2^{-\ell_i} - y}{2^{-\ell_i}} \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{a_{\ell_i} + 2^{-\ell_i} - y} + \frac{y - a_{\ell_i}}{2^{-\ell_i}} \frac{F(y) - F(a_{\ell_i})}{y - a_{\ell_i}} \\
&\overset{(i)}{\leq} \frac{2^{-\ell_i}(1 + t/K - 1/2)}{2^{-\ell_i}} \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{a_{\ell_i} + 2^{-\ell_i} - y} + \frac{2^{-\ell_i}(1/2 - (t-1)/K)}{2^{-\ell_i}} \frac{F(y) - F(a_{\ell_i})}{y - a_{\ell_i}} \\
&= (t/K + 1/2) \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{a_{\ell_i} + 2^{-\ell_i} - y} + (1/2 - (t-1)/K) \frac{F(y) - F(a_{\ell_i})}{y - a_{\ell_i}}, \qquad \text{(B.6)}
\end{aligned}
$$

where (i) follows since $F(y)$ is non-decreasing and $\ell_i \in S_t$. Now, notice that the right-hand side in (B.6) has a limit, and, by taking the lim sup, we obtain

$$
\begin{aligned}
\limsup_{i \to \infty} q_{\ell_i} &\leq (t/K + 1/2) f(y) + (1/2 - (t-1)/K) f(y) \\
&= \left(1 + \frac{1}{K}\right) f(y) < f(y) + \epsilon.
\end{aligned}
$$

But this is a contradiction because all q_{m_i} satisfied $q_{m_i} > f(y) + \epsilon$, and $\{q_{\ell_i}\}_{i=1}^{\infty} \subset \{q_{m_i}\}_{i=1}^{\infty}$. We conclude that we must have

$$
\lim_{m \to \infty} q_m = f(y),
$$

which implies that $f_m(y) \to f(y)$ as $m \to \infty$. ■

B.3 Proof of Lemma 4.4

Lemma 4.4. *Suppose $\{(Z_1[i], ..., Z_k[i])\}_{i=0}^{nb-1}$ is an i.i.d. sequence of length-k zero-mean random vectors with covariance matrix $\mathbf{K}$, and let $\mathbf{Q}$ be the $b \times b$ matrix defined in (4.16)* *and*

$$\begin{bmatrix} \tilde{Z}_1^{(0)}[t] & \cdots & \tilde{Z}_k^{(0)}[t] \\ \vdots & \ddots & \vdots \\ \tilde{Z}_1^{(b-1)}[t] & \cdots & \tilde{Z}_k^{(b-1)}[t] \end{bmatrix} = \mathbf{Q} \begin{bmatrix} Z_1[tb] & \cdots & Z_k[tb] \\ \vdots & \ddots & \vdots \\ Z_1[tb+b-1] & \cdots & Z_k[tb+b-1] \end{bmatrix}$$

for $t = 0, 1, ..., n-1$*. Then, for any sequence* ℓ_b *such that, for* $b = 1, 2, ...$, $\ell_b \in \{0, 1, ..., b-1\}$*, and any* $t \in \{0, 1, ..., n-1\}$*,*

$$\left(\tilde{Z}_1^{(\ell_b)}[t], ..., \tilde{Z}_k^{(\ell_b)}[t]\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{K}), \text{ as } b \to \infty.$$

Proof: Clearly, it suffices to show that $\left(\tilde{Z}_1^{(\ell_b)}[0], ..., \tilde{Z}_k^{(\ell_b)}[0]\right)$ converges in distribution to a jointly Gaussian random vector with covariance matrix $\mathbf{K}$, as $b \to \infty$. In order to use the Cramér-Wold Theorem [9], we fix an arbitrary vector $(t_1, ..., t_k) \in \mathbb{R}^k$ and we notice that

$$\begin{aligned} \sum_{m=1}^{k} t_m \tilde{Z}_m^{(\ell_b)}[0] &= \sum_{m=1}^{k} t_m \sum_{j=0}^{b-1} Z_m[j]\; Q(\ell_b, j) \\ &= \sum_{j=0}^{b-1} \left(\sum_{m=1}^{k} t_m Z_m[j] \right) Q(\ell_b, j). \end{aligned} \tag{B.7}$$

To characterize the convergence in distribution of (B.7), we will need Lindeberg's Central Limit Theorem (Theorem 4.4). To apply Theorem 4.4, we will let, for $j = 0, ..., b-1$,

$$Y_{b,j+1} = \sqrt{b} \left(\sum_{m=1}^{k} t_m Z_m[j] \right) Q(\ell_b, j).$$

Then, if we let $\mathbf{K}_{u,v}$ be the entry in the uth row and vth column of $\mathbf{K}$, we have

$$\begin{aligned} s_b^2 = \sum_{j=1}^{b} E\left[Y_{b,j}^2\right] &= b \sum_{j=1}^{b} Q^2(\ell_b, j-1) E \left(\sum_{m=1}^{k} t_m Z_m[j-1] \right)^2 \\ &= b \sum_{1 \le u,v \le k} t_u t_v \mathbf{K}_{u,v} \sum_{j=1}^{b} Q^2(\ell_b, j-1) \end{aligned}$$

$$= b \sum_{1 \le u,v \le k} t_u t_v \mathbf{K}_{u,v},$$

regardless of the value of ℓ_b. In order to verify Lindeberg's condition, we define $\sigma^2 = \sum_{1 \le u,v \le k} t_u t_v \mathbf{K}_{u,v}$ and we let $U_{b,j} = Y_{b,j}^2 \mathbb{1}\left\{|Y_{b,j}| \ge \varepsilon s_b\right\} = Y_{b,j}^2 \mathbb{1}\left\{|Y_{b,j}| \ge \varepsilon\sigma\sqrt{b}\right\}$. Consider any sequence j_b, for $b = 1, 2, ...,$ such that $j_b \in \{1, ..., b\}$, and any $\delta > 0$. Then we have that

$$\Pr\left(U_{b,j_b} < \delta\right) \ge \Pr\left(|Y_{b,j_b}| < \varepsilon\sigma\sqrt{b}\right) \ge \Pr\left(\left|\sum_{m=1}^{k} t_m Z_m[j_b - 1]\right| \sqrt{2} < \varepsilon\sigma\sqrt{b}\right)$$
$$= \Pr\left(\left|\sum_{m=1}^{k} t_m Z_m[0]\right| < \varepsilon\sigma\sqrt{b/2}\right) \to 1,$$

as $b \to \infty$, which means that $U_{b,j_b} \xrightarrow{p} 0$ (i.e., U_{b,j_b} converges in probability to 0) as $b \to \infty$. Moreover, we have that,

$$|U_{b,j_b}| \le Y_{b,j_b}^2 \le 2\left(\sum_{m=1}^{k} t_m Z_m[j_b - 1]\right)^2$$

for $b = 1, 2, ...,$ and

$$E\left[2\left(\sum_{m=1}^{k} t_m Z_m[0]\right)^2\right] = 2\sigma^2 < \infty.$$

Thus by the Dominated Convergence Theorem (see Appendix B.4), we have that $E[U_{b,j_b}] \to 0$ as $b \to \infty$. We conclude that

$$\frac{1}{s_b^2} \sum_{i=1}^{b} E\left(Y_{b,j}^2 \mathbb{1}\{|Y_i| \ge \varepsilon s_b\}\right) = \frac{1}{\sigma^2 b} \sum_{j=1}^{b} E\left[U_{b,j}\right]$$
$$\le \frac{1}{\sigma^2} \max_{1 \le j \le b} E\left[U_{b,j}\right] \to 0,$$

as $b \to \infty$, and Lindeberg's condition (4.6) is satisfied for any $\varepsilon > 0$. Hence, from Theorem 4.4, we have that

$$\frac{\sum_{i=1}^{b} Y_{b,j}}{\sigma\sqrt{b}} \xrightarrow{d} \mathcal{N}(0, 1),$$

which implies, from (B.7), that

$$\sum_{m=1}^{k} t_m \tilde{Z}_m^{(\ell_b)}[0] = \sum_{j=0}^{b-1} \left(\sum_{m=1}^{k} t_m Z_m[j] \right) Q(\ell_b, j)$$
$$= \frac{\sum_{j=1}^{b} Y_{b,j}}{\sqrt{b}} \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Finally, since for a jointly Gaussian vector $(Z_1^G, ..., Z_k^G)$ with mean zero and covariance matrix $\mathbf{K}$, we have $\sum_{m=1}^{k} t_m Z_m^G \sim \mathcal{N}(0, \sigma^2)$, we conclude, from the Cramér-Wold Theorem that $\left(\tilde{Z}_1^{(\ell_b)}[0], ..., \tilde{Z}_k^{(\ell_b)}[0]\right)$ converges in distribution to a jointly Gaussian random vector with zero mean and covariance matrix $\mathbf{K}$, as $b \to \infty$. ■

B.4 Dominated Convergence Theorem

We require the following variation of the Dominated Convergence Theorem.

Theorem B.1 *Suppose we have a sequence of random vectors $\mathbf{Z}_n \in \mathbb{R}^a$ converging weakly to $\mathbf{Z}$, and two almost-everywhere continuous functions $f, g : \mathbb{R}^a \to \mathbb{R}$ such that $0 \leq f \leq g$. Then, if $E[g(\mathbf{Z}_n)] = E[g(\mathbf{Z})] = c < \infty$ for all n, we have $\lim_{n\to\infty} E[f(\mathbf{Z}_n)] = E[f(\mathbf{Z})]$.*

Proof: If we let $X_n = f(\mathbf{Z}_n)$, $Y_n = g(\mathbf{Z}_n)$, $X = f(\mathbf{Z})$ and $Y = g(\mathbf{Z})$, from the almost everywhere continuity of the functions, we have $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$. From Theorem 25.11 in [9], we have that

$$E[X] \leq \liminf_{n\to\infty} E[X_n].$$

Note that, $Y_n - X_n = g(\mathbf{Z_n}) - f(\mathbf{Z_n})$ is an almost everywhere continuous function of $\mathbf{Z_n}$, hence the sequence of random variables $Y_n - X_n$, converges weakly to $Y - X$.

Therefore, since $Y_n - X_n \geq 0$, a second application of Theorem 25.11 yields

$$c - E[X] = E[Y - X] \leq \liminf_{n\to\infty} E[Y_n - X_n]$$
$$= \liminf_{n\to\infty} c - E[X_n] = c - \limsup_{n\to\infty} E[X_n].$$

Combining both inequalities, we obtain

$$\limsup_{n\to\infty} E[X_n] \leq E[X] \leq \liminf_{n\to\infty} E[X_n],$$

which implies that $\lim_{n\to\infty} E[X_n] = E[X]$. ■

B.5 Proof of Lemma 5.1

Lemma 5.1. *Suppose $(X_1[t], ..., X_k[t])$ has an arbitrary joint distribution with covariance matrix $\mathbf{K}$ and a coding scheme C with block length n achieves distortion vector $(D_1, ..., D_k)$. Then, for any $\epsilon > 0$, one can build another coding scheme $\tilde{C}$ of block length n with decoding functions $\tilde{g}_{d_m}$ such that*

$$\left\|\tilde{g}_{d_j}(y_1, ..., y_n)\right\|_\infty \leq M,$$

for any $(y_1, ..., y_n) \in \mathbb{R}^n$, $j = 1, ..., k$ and a fixed $M > 0$, which achieves distortion vector $(D_1 + \epsilon, ..., D_k + \epsilon)$.

Proof: From a coding scheme C with blocklength n achieving distortion vector $(D_1, ..., D_k)$, we will create a sequence of coding schemes $C^{(m)}$, $m = 1, 2, ...$, obtained by clipping the output of the decoding functions g_{d_j}, $j = 1, ..., k$. More precisely, coding scheme $C^{(m)}$ has the same encoding and relaying functions as

C, and decoding functions $g_{d_j}^{(m)}$ whose ith component is defined as

$$g_{d_j}^{(m)}(y_1, ..., y_n)[i] = \begin{cases} m, & \text{if } g_{d_j}(y_1, ..., y_n)[i] > m \\ -m, & \text{if } g_{d_j}(y_1, ..., y_n)[i] < -m \\ g_{d_j}(y_1, ..., y_n)[i], & \text{otherwise} \end{cases}$$

for $j = 1, ..., k$, and $i = 0, ..., n-1$. Now, consider a fixed $j \in \{1, ..., k\}$, and define, for $i = 0, ..., n-1$, the event B_i as

$$B_i = \left\{X_j[i] > m, g_{d_j}\left(Y_{d_j}^n\right)[i] > m\right\} \cup \left\{X_j[i] < -m, g_{d_j}\left(Y_{d_j}^n\right)[i] < -m\right\}.$$

It is easy to verify that the complementary event is given by

$$\begin{aligned} B_i^c = \left\{\left|X_j[i]\right| \leq m\right\} \cup \left\{\left|g_{d_j}\left(Y_{d_j}^n\right)[i]\right| \leq m\right\} \cup \left\{X_j[i] > m, g_{d_j}\left(Y_{d_j}^n\right)[i] < -m\right\} \\ \cup \left\{X_j[i] < -m, g_{d_j}\left(Y_{d_j}^n\right)[i] > m\right\}. \end{aligned}$$

For each of the four sub-events in B_i^c, it is clear that

$$\left|X_j[i] - g_{d_j}\left(Y_{d_j}^n\right)[i]\right| \geq \left|X_j[i] - g_{d_j}^{(m)}\left(Y_{d_j}^n\right)[i]\right|.$$

Thus, we can upper bound the expected distortion of the output of decoder j of $C^{(m)}$ as

$$\begin{aligned} E\left[\left\|X_j^n - g_{d_j}^{(m)}\left(Y_{d_j}^n\right)\right\|^2\right] &= \sum_{i=0}^{n-1} E\left[\left(X_j[i] - g_{d_j}^{(m)}\left(Y_{d_j}^n\right)[i]\right)^2\right] \\ &= \sum_{i=0}^{n-1} \left\{E\left[\left(X_j[i] - g_{d_j}^{(m)}\left(Y_{d_j}^n\right)[i]\right)^2 \mathbb{1}_{B_i^c}\right]\right. \\ &\qquad \left. + E\left[\left(X_j[i] - g_{d_j}^{(m)}\left(Y_{d_j}^n\right)[i]\right)^2 \mathbb{1}_{B_i}\right]\right\} \\ &\leq \sum_{i=0}^{n-1} \left\{E\left[\left(X_j[i] - g_{d_j}\left(Y_{d_j}^n\right)[i]\right)^2\right] + E\left[\left(X_j[i] - g_{d_j}^{(m)}\left(Y_{d_j}^n\right)[i]\right)^2 \mathbb{1}_{B_i}\right]\right\} \\ &= E\left[\left\|X_j^n - g_{d_j}\left(Y_{d_j}^n\right)\right\|^2\right] + \sum_{i=0}^{n-1} E\left[\left(X_j[i] - g_{d_j}^{(m)}\left(Y_{d_j}^n\right)[i]\right)^2 \mathbb{1}_{B_i}\right] \\ &\leq E\left[\left\|X_j^n - g_{d_j}\left(Y_{d_j}^n\right)\right\|^2\right] + \sum_{i=0}^{n-1} E\left[\left(X_j[i]\right)^2 \mathbb{1}_{B_i}\right] \end{aligned}$$

$$= nD_j + nE\left[\left(X_j[0]\right)^2 \mathbb{1}_{B_0}\right].$$

Since $\left|X_j[0]^2\mathbb{1}_{B_0}\right| \leq X_j[0]^2$, $E\left[X_j[0]^2\right] < \infty$, and $X_j[0]^2\mathbb{1}_{B_0} \xrightarrow{p} 0$ as $m \to \infty$, by the Dominated Convergence Theorem (see Appendix B.4),

$$\lim_{m\to\infty} E\left[\left(X_j[0]\right)^2 \mathbb{1}_{B_0}\right] = 0.$$

Therefore, for any $\epsilon > 0$, we can pick $m = M$ large enough so that

$$\tfrac{1}{n}E\left[\left\|X_j^n - g_{d_j}^{(M)}\left(Y_{d_j}^n\right)\right\|^2\right] \leq D_j + \epsilon \qquad \text{and} \qquad \left\|g_{d_j}^{(M)}(y_1, ..., y_n)\right\|_\infty \leq M,$$

for all $j = 1, ..., K$, and we may let $\tilde{C} = C^{(M)}$. ■

B.6 Proof of Lemma 5.2

Lemma 5.2. *Suppose the distortion tuple $(D_1, \cdots, D_k)$ is achievable over the (k, N)-memoryless network. Then for any $\epsilon > 0$, there exists a coding scheme with finite encoding precision that achieves distortion tuple $(D_1 + \epsilon, \cdots, D_k + \epsilon)$.*

Proof: Achievability of the distortion tuple $(D_1, \cdots, D_k)$ implies the existence of a coding scheme C with block length n, such that,

$$\frac{1}{n}E\Big[\parallel \mathbf{X}_m - \hat{\mathbf{X}}_m \parallel^2 \Big] \leq D_m, \; \forall\, m = [1:k]. \tag{B.8}$$

Using Lemma 5.1, without loss of generality we will suppose that,

$$\left\|g_{d_j}(y_1, ..., y_n)\right\|_\infty \leq M,$$

for each destination $d_j \in \mathcal{D}$, for a fixed $M > 0$. Note that, using Lemma 5.4, the memoryless channel $f_{Y_1,\cdots,Y_N|U_1,\cdots,U_N}$ can be equivalently represented as a deterministic channel $Y_i = h_i(U_1, \cdots, U_N, \mathbf{Z})$, $\forall i = [1 : N]$ where $\mathbf{Z}$ is a random vector, independent of the channel inputs, $(U_1, \cdots, U_N)$. Thus for a fixed block length n, given the description of our encoding procedure, we can write, for some functions F_i depending on h_i, $\mathbf{Y}_i = F_i(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k, \mathbf{Z})$, $\forall\ i \in [1 : N]$, as the evolution of the system depends only on the sources and the random vector $\mathbf{Z}$. Thus, noting that the reconstruction for the mth source is $\hat{\mathbf{X}}_m = g_{d_m}(\mathbf{Y}_m)$, the above equation on distortion constraints can be equivalently written as,

$$\frac{1}{n}E\Big[\parallel \mathbf{X}_m - g_{d_m}(F_m(\mathbf{X}_1, \cdots, \mathbf{X}_k, \mathbf{Z})) \parallel^2 \Big] \leq D_m,\ \forall\ m = [1 : k], \tag{B.9}$$

To prove this lemma we have to show that, given an $\epsilon > 0$, we can construct a scheme C_ρ for some $\rho = [\rho_1, \cdots, \rho_k] \in \mathcal{N}^k$, where the encoding function at each source $s_m \in \mathcal{S}$ satisfies

$$\tilde{f}_{s_m,t}(x_m^n, y^{t-1}) = \tilde{f}_{s_m,t}(\lfloor x_m^n \rfloor_{\rho_m}, y^{t-1}),\ \forall\ m \in [1 : k]$$

for any $x_m^n \in \mathbb{R}^n$, any $y^{t-1} \in \mathbb{R}^{t-1}$, and any time t, such that,

$$\frac{1}{n}E\Big[\parallel \mathbf{X}_m - g_{d_m}(F_m(\lfloor \mathbf{X}_1 \rfloor_{\rho_1}, \cdots, \lfloor \mathbf{X}_k \rfloor_{\rho_k}, \mathbf{Z})) \parallel^2 \Big] \leq D_m + \epsilon,\ \forall\ m = [1 : k], \tag{B.10}$$

To prove this, we consider the following randomized encoding scheme C_ρ. Note the disclaimer that, in our definition of schemes, the encoding, relaying and decoding operations were defined to be deterministic, but for the time being we will allow for randomization and later show that it can be dispensed with. The scheme C_ρ, operated in blocks of length n, uses the same relaying encoding and destination encoding and decoding functions, the only change being in the source encoding. At the source node s_m the source is encoded as, $U_{s_m,t} = f_{s_m,t}(\tilde{\mathbf{X}}_m, Y^{t-1})$, $\forall\ t \in [1 : N]$, where $\tilde{\mathbf{X}}_m = \{\tilde{\mathbf{X}}_m[t]\}_{t=0}^{n-1}$, such that $\tilde{\mathbf{X}}_m[t] = \lfloor \mathbf{X}_m[t] \rfloor_{\rho_m} + V_{\rho_m}$, where V_{ρ_m} is a random variable independent of the sources in the network, uniformly distributed in $(-2^{-\rho_m - 1}, 2^{-\rho_m - 1})$. Consider

$$\frac{1}{n}E\Big[\parallel \mathbf{X}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel^2 \Big]$$

$$\leq \underbrace{\frac{1}{n}E\Big[\parallel \mathbf{X}_m - \tilde{\mathbf{X}}_m \parallel^2\Big]}_{(I)} + \underbrace{\frac{1}{n}E\Big[\parallel \tilde{\mathbf{X}}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel^2\Big]}_{(II)}$$
$$+ \underbrace{\frac{1}{n}E\Big[\parallel \mathbf{X}_m - \tilde{\mathbf{X}}_m \parallel\parallel \tilde{\mathbf{X}}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel\Big]}_{(III)}.$$

Note that

$$\begin{aligned} |\mathbf{X}_m[t] - \tilde{\mathbf{X}}_m[t]| &= |-V_{\rho_m} + \mathbf{X}_m[t] - \lfloor \mathbf{X}_m[t] \rfloor_{\rho_m}| \\ &= |-V_{\rho_m} + 2^{-\rho_m}(2^{\rho_m}\mathbf{X}_m[t] - \lfloor 2^{\rho_m}\mathbf{X}_m[t] \rfloor)| \\ &\leq |V_{\rho_m}| + 2^{-\rho_m}|2^{\rho_m}\mathbf{X}_m[t] - \lfloor 2^{\rho_m}\mathbf{X}_m[t] \rfloor| \\ &\leq 2^{-\rho_m - 1} + 2^{-\rho_m} \leq 2^{-\rho_m + 1}, \end{aligned} \tag{B.11}$$

which implies $\parallel \mathbf{X}_m - \tilde{\mathbf{X}}_m \parallel \leq \sqrt{n}2^{1-\rho_m}$. This further implies that the term (I) of (B.15) is bounded as

$$\frac{1}{n}E\Big[\parallel \mathbf{X}_m - \tilde{\mathbf{X}}_m \parallel^2\Big] \leq 2^{2-2\rho_m}, \tag{B.12}$$

implying that, in the limit, term (I) vanishes. Define the (measurable) functions $\mathbf{H}_m(\cdots) : \underbrace{\mathbf{R}^n \times \cdots \times \mathbf{R}^n}_{k+1 \text{ times}} \to \mathbf{R}, \forall\, m \in [1:k]$ as

$$\mathbf{H}_m(\mathbf{y}_1, \cdots, \mathbf{y}_k, \mathbf{z}) = \parallel \mathbf{y}_m - g_{d_m}(F_m(\mathbf{y}_1, \cdots, \mathbf{y}_k, \mathbf{z})) \parallel. \tag{B.13}$$

Since $\mathbf{Z}$ is independent of the sources, using Lemma B.3 in Appendix B.9, we have the following convergence of the joint densities,

$$\lim_{\rho_m \to \infty} f(\mathbf{X}_1, \cdots, \tilde{\mathbf{X}}_m, \cdots, \mathbf{X}_k, \mathbf{Z}) = f(\mathbf{X}_1, \cdots, \mathbf{X}_m, \cdots, \mathbf{X}_k, \mathbf{Z}),\ \forall\, m \in [1:k]. \tag{B.14}$$

Using the above result we have that term (II) in (B.15) satisfies

$$\lim_{\rho_1 \to \infty} \cdots \lim_{\rho_k \to \infty} \frac{1}{n}E\Big[\parallel \tilde{\mathbf{X}}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel^2\Big]$$
$$= \lim_{\rho_1 \to \infty} \cdots \lim_{\rho_k \to \infty} \frac{1}{n}E\Big[H(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, Z)\Big]$$

$$
\begin{aligned}
&\overset{(a)}{=} \lim_{\rho_1\to\infty} \cdots \lim_{\rho_{k-1}\to\infty} \frac{1}{n} E\Big[H(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_{k-1}, \mathbf{X}_k, Z)\Big] \\
&\overset{(b)}{=} \frac{1}{n} E\Big[H(\mathbf{X}_1, \cdots, \mathbf{X}_k, Z)\Big] \\
&\leq \frac{1}{n} E\Big[\parallel \mathbf{X}_m - g_{d_m}(F_m(\mathbf{X}_1, \cdots, \mathbf{X}_k, \mathbf{Z})) \parallel^2 \Big] \leq D_m
\end{aligned} \tag{B.15}
$$

where (a) follows from the fact that pointwise convergence of the density implies convergence in distribution of a (measurable) function of the random variable and this implies convergence in expectation via the Dominated Convergence Theorem (see Appendix B.4), as we have from the fact that $g_m(\cdot)$ is bounded (say by M),

$$
\begin{aligned}
\frac{1}{n} E \left\| \mathbf{X}_m - g_{d_m}\left(F_m\left(\mathbf{X}_1 ..., \mathbf{X}_k, \vec{Z}\right)\right)\right\|^2 &\leq \frac{2}{n} E\left(\|\mathbf{X}_m\|^2 + \left\| g_{d_m}\left(F_m\left(\mathbf{X}_1, ..., \mathbf{X}_k, \vec{\tilde{Z}}\right)\right)\right\|^2 \right) \\
&\leq 2E\left(\|\mathbf{X}_m\|^2 + M^2\right) = 2\mathbf{K}_{m,m} + 2M^2 < \infty,
\end{aligned} \tag{B.16}
$$

and (b) follows from similarly repeating (a) by taking one limit at a time.

Now bounding the cross term (III) in (B.15),

$$
\begin{aligned}
&\frac{1}{n} E\Big[\parallel \mathbf{X}_m - \tilde{\mathbf{X}}_m \parallel \parallel \tilde{\mathbf{X}}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel \Big] \\
&\qquad \leq \frac{1}{\sqrt{n}} 2^{1-\rho_m} E\Big[\parallel \tilde{\mathbf{X}}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel \Big] \\
&\qquad \leq \frac{1}{\sqrt{n}} 2^{1-\rho_m} \sqrt{E\Big[\parallel \tilde{\mathbf{X}}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel^2 \Big]}
\end{aligned}
$$

and using the bound on the term (II), implies that in limit this term is bounded as $2^{1-\rho_m}\sqrt{D_m}$ which vanishes. Hence we have proved that

$$
\begin{aligned}
&\lim_{\rho_1\to\infty} \cdots \lim_{\rho_k\to\infty} \frac{1}{n} E\Big[\parallel \mathbf{X}_m - g_{d_m}(F_m(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_k, \mathbf{Z})) \parallel^2 \Big] \\
&\qquad \leq \frac{1}{n} E\Big[\parallel \mathbf{X}_m - g_{d_m}(F_m(\mathbf{X}_1, \cdots, \mathbf{X}_k, \mathbf{Z})) \parallel^2 \Big] \\
&\qquad \leq D_m.
\end{aligned}
$$

Thus for any $\epsilon > 0$, we can choose $\rho \in \mathcal{N}^k$, with components large enough so C_ρ achieves the distortion tuple, $(D_1 + \epsilon, \cdots, D_k + \epsilon)$. What is left is to show we can dispense away with random encoders. This is argued in a standard manner by choosing the best randomizations $\mathbf{V}_i$'s at respective encoders, as done in [56]. ■

B.7 Proof of Lemma 5.3

Lemma 5.3. *If, for some $\rho \in \mathcal{N}$, $f : \mathbb{R}^a \to \mathbb{R}^b$ satisfies*

$$f(\mathbf{x}) = \mathbf{f}(\lfloor \mathbf{x} \rfloor_\rho)$$

for any $\mathbf{x} \in \mathbb{R}^a$, f is locally constant (and thus continuous) almost everywhere.

Proof: Denote the set $\mathscr{S}(\rho) = \{x \in \mathbb{R}^a : 2^\rho x \in \mathbb{Z}^a\}$, where $\mathbb{Z}$ is the set of integers. Note that the function in the theorem can take values $f(y)$ where $y \in \mathscr{S}(\rho)$. Now for each $y \in \mathscr{S}(\rho)$, define the set $S(y) = \{x \in \mathbb{R}^a : x \neq y, \lfloor x \rfloor_\rho = y\}$, which are disjoint for different values of $y \in \mathscr{S}(\rho)$ and cover the whole space $\mathbb{R}^a$. Since f takes a constant value in each of the sets $S(\cdot)$, the only regions of discontinuity are the boundaries of these regions. But these boundaries are disjoint bounded rectangles each of which has Lebesgue measure zero, implying the total region of discontinuity has zero measure. Thus f is locally constant almost-everywhere (and hence continuous). ■

B.8 Proof of Lemma 5.4

Lemma 5.4. *For any two random vectors $\mathbf{Y}$ and $\mathbf{U}$, there exist a (deterministic, measurable) function F and a random vector $\mathbf{Z}$, independent of $\mathbf{U}$, for which the pair $(F(\mathbf{U}, \mathbf{Z}), \mathbf{U})$ has the same distribution as $(\mathbf{Y}, \mathbf{U})$.*

Proof: We prove the lemma by induction on the size t of the random vector $\mathbf{Y}$. If $\mathbf{Y}$ is a scalar, i.e., $t = 1$, let $g_{\mathbf{u}}(\mathbf{y}) = G_{\mathbf{Y}|\mathbf{U}}(\mathbf{y}|\mathbf{u})$, where $G_{\mathbf{Y}|\mathbf{U}}$ is the conditional distribution function of $\mathbf{Y}$ given $\mathbf{U}$. Then we let Z be a uniform random variable on $[0, 1]$ (independent of $\mathbf{U}$), and we let $F(\mathbf{u}, z) = g_{\mathbf{u}}^{-1}(z)$ (where $^{-1}$ represents the generalized inverse). It is then clear that $F(\mathbf{u}, Z)$ is distributed as $\mathbf{Y}$ conditioned on $\mathbf{U} = \mathbf{u}$ for any $\mathbf{u}$, which implies that $(F(\mathbf{U}, Z), \mathbf{U})$ is distributed as $(\mathbf{Y}, \mathbf{U})$.

Now suppose the lemma is true when the size of $\mathbf{Y}$ is t. Consider a random vector $\mathbf{Y} = (\mathbf{Y}', \tilde{\mathbf{Y}})$, where $\mathbf{Y}'$ has size t and $\tilde{\mathbf{Y}}$ is a scalar. Then there exists a random vector $\mathbf{Z}'$ and a function F' such that $(F'(\mathbf{U}, \mathbf{Y}'), \mathbf{U})$ is distributed as $(\mathbf{Y}', \mathbf{U})$. Now let $g_{\mathbf{u},\mathbf{y}'}(\mathbf{y}) = G_{\tilde{\mathbf{Y}}|\mathbf{U},\mathbf{Y}'}(\mathbf{y}|\mathbf{u}, \mathbf{y}')$ be the conditional distribution function of $\tilde{\mathbf{Y}}$ given $\mathbf{U}$ and $\mathbf{Y}'$. Then we let $\mathbf{Z} = (\mathbf{Z}', Z'')$, where Z'' is a uniform random variable on $[0, 1]$ (independent of $\mathbf{U}$ and $\mathbf{Z}'$), and we let $F(\mathbf{u}, (\mathbf{z}', z'')) = g^{-1}_{\mathbf{u},F'(\mathbf{u},\mathbf{z}')}(z'')$. Then $(F(\mathbf{U}, \mathbf{Z}), \mathbf{U})$ is distributed as $(\tilde{\mathbf{Y}}, \mathbf{U})$, and $(F'(\mathbf{U}, \mathbf{Z}'), F(\mathbf{U}, \mathbf{Z}), \mathbf{U})$ is distributed as $(\mathbf{Y}, \mathbf{U}) = (\mathbf{Y}', \tilde{\mathbf{Y}}, \mathbf{U})$. ■

B.9 Density Lemma

Lemma B.3 *Suppose* $\mathbf{Y} = (Y_1, \ldots, Y_i, \ldots, Y_k)$ *is a random vector with density* $f_{Y_1,\ldots,Y_i,\ldots,Y_k}$. *Consider some* $\rho \in \mathcal{N}$. *For some* $i \in [1 : k]$, *let* $\tilde{Y}_i^{(\rho)} = \lfloor Y_i \rfloor_\rho + U_\rho$, *where* U_ρ *is uniformly distributed in* $(-2^{-\rho-1}, 2^{-\rho-1})$ *and is independent of* $\mathbf{Y}$. *Then*

$$\lim_{\rho\to\infty} f_{Y_1,\ldots,\tilde{Y}_i^{(\rho)},\ldots,Y_k}(y_1, \ldots, y_i, \ldots, y_k) = f_{Y_1,\ldots,Y_i,\ldots,Y_k}(y_1, \ldots, y_i, \ldots, y_k), \ \forall\, i \in [1 : k], \tag{B.17}$$

for almost every $(y_1, \ldots, y_i, \ldots, y_k) \in \mathbb{R}^k$.

Proof: For the sake of simplicity, we will consider the case $k = 2$ and $i = 2$ (i.e, Y_1 is quantized to $\tilde{Y}_1$). The proof for $k > 2$ follows via a straightforward generalization. The proof follows similar lines of thought as Lemma 3 in [56], we state here the required steps for completeness. The density $f_{\tilde{Y}_1,Y_2}(y_1,y_2)$ can be written for almost all tuples (y_1,y_2) as,

$$\begin{aligned}
f_{\tilde{Y}_1,Y_2}(y_1,y_2) &= 2^{\rho} E[\mathbf{1}_{\{y_1-\lfloor Y_1\rfloor_\rho \in(-2^{-\rho-1},2^{-\rho-1})\}}|Y_2=y_2]\ f_{Y_2}(y_2) \\
&= 2^{\rho}\Pr[y_1-\lfloor Y_1\rfloor_\rho \in(-2^{-\rho-1},2^{-\rho-1})|Y_2=y_2]\ f_{Y_2}(y_2) \\
&= 2^{\rho}\Pr[\lfloor Y_1\rfloor_\rho \in(y_1-2^{-\rho-1},y_1+2^{-\rho-1})|Y_2=y_2]\ f_{Y_2}(y_2) \\
&= 2^{\rho}\Pr[\lfloor 2^{\rho}Y_1\rfloor \in(y_12^{\rho}-\frac{1}{2},y_12^{\rho}+\frac{1}{2})|Y_2=y_2]\ f_{Y_2}(y_2) \\
&= 2^{\rho}\Pr[2^{\rho}Y_1 \in(\lceil y_12^{\rho}-\frac{1}{2}\rceil,\lceil y_12^{\rho}+\frac{1}{2}\rceil)|Y_2=y_2]\ f_{Y_2}(y_2) \\
&= 2^{\rho}\Pr[Y_1 \in(2^{-\rho}\lceil y_12^{\rho}-\frac{1}{2}\rceil,2^{-\rho}\lceil y_12^{\rho}+\frac{1}{2}\rceil)|Y_2=y_2]\ f_{Y_2}(y_2) \\
&= 2^{\rho}\int_{a_\rho}^{b_\rho} f_{Y_1,Y_2}(x_1,y_2)dx_1, \qquad \text{(B.18)}
\end{aligned}$$

where $a_\rho = 2^{-\rho}\lceil y_12^{\rho}-\frac{1}{2}\rceil$ and $b_\rho = 2^{-\rho}\lceil y_12^{\rho}+\frac{1}{2}\rceil$, such that $b_\rho = a_\rho + 2^{-\rho}$ which implies, $a_\rho \to y_1$. What is left to prove is that

$$\lim_{\rho\to\infty} 2^{\rho}\int_{a_\rho}^{b_\rho} f_{Y_1,Y_2}(x_1,y_2)dx_1 = f_{Y_1,Y_2}(y_1,y_2)$$

for almost all tuples (y_1,y_2). But this follows using the proof of Lemma 3 in [56], replacing the integrand function appropriately. ■

APPENDIX C

SUPPLEMENT FOR PART III

C.1 Proof of Claim 7.1

Claim 7.1. *Let $C_{\mathcal{N}}$ and $C_{\mathcal{N}^\ell}$ be the capacity regions of a K-unicast memoryless network $\mathcal{N}$ and of the concatenation of ℓ copies of $\mathcal{N}$. Then $C_{\mathcal{N}} \subseteq C_{\mathcal{N}^\ell}$.*

Proof: We prove the case $\ell = 2$. The general case follows similarly. We show that if $(R_1, ..., R_K)$ is achievable in $\mathcal{N}$, then $(R_1(1-\delta), ..., R_K(1-\delta))$ is achievable in $\mathcal{N}^2$ for any $\delta > 0$. Consider a coding scheme C_n for $\mathcal{N}$ with rate tuple $(R_1, ..., R_K)$ and error probability $P_{\text{error}}(C_n) = \epsilon_n$. For an arbitrary $\delta > 0$, we construct a new coding scheme with rate tuple $(R_1(1-\delta), ..., R_K(1-\delta))$ and block length nL for the concatenated network $\mathcal{N}^2$, where we let $L = \lfloor \epsilon_n^{-1/2} \rfloor$, as follows. Each source s_i will view its message $W_i \in \{1, ..., 2^{nLR_i(1-\delta)}\}$ as $L(1-\delta)$ messages $W_i^{(1)}, ..., W_i^{(L(1-\delta))}$ in $\{1, ..., 2^{nR}\}$. In the jth block of length n, the sources and relays in the first copy of $\mathcal{N}$ behave as if they were simply using coding scheme C_n with messages $W_1^{(j)}, ..., W_K^{(j)}$, and the nodes in $\mathcal{U} = \{u_1, ..., u_K\}$ behave as destinations, outputting $\hat{W}_1^{(j)}, ..., \hat{W}_K^{(j)}$ at the end of the block. In the $(j+1)$th block, the nodes in $\mathcal{U}$ operate as sources for the second copy of $\mathcal{N}$, re-encoding the decoded messages from the previous block $\hat{W}_1^{(j)}, ..., \hat{W}_K^{(j)}$, and all the remaining nodes in the second copy of $\mathcal{N}$ simply operate according to coding scheme C_n. Provided that ϵ_n is small enough, $L^{-1} < \delta$, and at the end of the $[L(1-\delta)+1]$th block, each destination d_i obtains an estimate for all $L(1-\delta)$ messages from s_i. By the union bound, the error probability of this code over $\mathcal{N}^2$ is at most $2L(1-\delta)\epsilon_n \leq 2\epsilon_n^{1/2}$, which tends to zero as $\epsilon_n \to 0$. ■

C.2 Auxiliary Results for Section 7.2

Lemma C.1 *If* $\mathbf{x}$ *is a d-dimension random vector with entries in a finite field* $\mathbb{F}$, *then*

$$H(A\mathbf{x}|B\mathbf{x}) \leq \left(\operatorname{rank}\begin{bmatrix} A \\ B \end{bmatrix} - \operatorname{rank} B\right) \log|\mathbb{F}|.$$

Proof: Let B' be a $(\operatorname{rank} B) \times d$ matrix made up of $\operatorname{rank} B$ linearly independent rows of B. Clearly, $H(B\mathbf{x}) = H(B'\mathbf{x})$. Let A' be a matrix obtained by removing rows of A until $\operatorname{rank}\begin{bmatrix} A' \\ B' \end{bmatrix}$ is full rank. We then have

$$\begin{aligned}
H(A\mathbf{x}|B\mathbf{x}) &\leq H(A\mathbf{x}|B'\mathbf{x}) \\
&= H(A\mathbf{x}, B'\mathbf{x}) - H(B'\mathbf{x}) \\
&= H\left(\begin{bmatrix} A \\ B' \end{bmatrix}\mathbf{x}\right) - H(B'\mathbf{x}) \\
&= H\left(\begin{bmatrix} A' \\ B' \end{bmatrix}\mathbf{x}\right) - H(B'x) \\
&= H(A'\mathbf{x}, B'\mathbf{x}) - H(B'x) \\
&\leq H(A'\mathbf{x}) \\
&\leq \operatorname{rank}(A') \log|\mathbb{F}|.
\end{aligned}$$

Moreover, we have that

$$\operatorname{rank}\begin{bmatrix} A \\ B \end{bmatrix} - \operatorname{rank} B = \operatorname{rank}\begin{bmatrix} A' \\ B' \end{bmatrix} - \operatorname{rank} B' = \operatorname{rank} A',$$

concluding the proof. ■

Lemma C.2 *For a vector* $\mathbf{y}$*, let* $\lfloor \mathbf{y} \rfloor$ *be obtained by applying the floor function to each component of* $\mathbf{y}$*. If* $\mathbf{x}$ *is a d-dimension zero-mean continuous random vector with* $E[x_i^2] \leq P$*, then*

$$H\left(\lfloor A\mathbf{x} \rfloor \,|\, \lfloor B\mathbf{x} \rfloor\right) \leq \left(\operatorname{rank}\begin{bmatrix} A \\ B \end{bmatrix} - \operatorname{rank}B\right) \tfrac{1}{2}\log P + c,$$

where $c = o(\log P)$*.*

Proof: Following the proof of Lemma C.1, we let B' be a $(\operatorname{rank}B) \times d$ matrix made up of $\operatorname{rank}B$ linearly independent rows of B and A' be a matrix obtained by removing rows of A until $\operatorname{rank}\begin{bmatrix} A' \\ B' \end{bmatrix}$ is full rank. Furthermore, we let $\tilde{A}$ be the matrix containing the t rows removed from A to obtain A'. Notice that there exists a matrix M such that $\tilde{A} = M\begin{bmatrix} A' \\ B' \end{bmatrix}$. We then have

$$\begin{aligned} H(\lfloor A\mathbf{x} \rfloor \,|\, \lfloor B\mathbf{x} \rfloor) &\leq H(\lfloor A\mathbf{x} \rfloor \,|\, \lfloor B'\mathbf{x} \rfloor) \\ &= H\left(\left\lfloor \begin{bmatrix} A \\ B' \end{bmatrix} \mathbf{x} \right\rfloor\right) - H(\lfloor B'\mathbf{x} \rfloor) \\ &= I\left(\mathbf{x}; \left\lfloor \begin{bmatrix} A \\ B' \end{bmatrix} \mathbf{x} \right\rfloor\right) - I(\mathbf{x}; \lfloor B'\mathbf{x} \rfloor). \end{aligned} \tag{C.1}$$

Now if we let $\mathbf{z}_{A'}$, $\mathbf{z}_{B'}$ and $\mathbf{z}_{\tilde{A}}$ be independent random vectors of dimensions $\operatorname{rank}A'$, $\operatorname{rank}B'$ and t respectively with i.i.d. $\mathcal{N}(0,1)$ entries. Then, from Lemma 7.2 in [6], we can upper-bound (C.1) by

$$I\left(\mathbf{x}; \begin{bmatrix} \tilde{A} \\ A' \\ B' \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{z}_{\tilde{A}} \\ \mathbf{z}_{A'} \\ \mathbf{z}_{B'} \end{bmatrix}\right) - I(\mathbf{x}; B'\mathbf{x} + \mathbf{z}_{B'}) + c_1$$

$$
\begin{aligned}
&= I\left(\mathbf{x}; A'\mathbf{x} + \mathbf{z}_{A'} | B'\mathbf{x} + \mathbf{z}_{B'}\right) + I\left(\mathbf{x}; \tilde{A}\mathbf{x} + \mathbf{z}_{\tilde{A}} \left| \begin{bmatrix} A' \\ B' \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{z}_{A'} \\ \mathbf{z}_{B'} \end{bmatrix} \right.\right) + c_1 \\
&\overset{(ii)}{\leq} I\left(\mathbf{x}; A'\mathbf{x} + \mathbf{z}_{A'}\right) + I\left(\mathbf{x}; \mathbf{z}_{\tilde{A}} - M \begin{bmatrix} \mathbf{z}_{A'} \\ \mathbf{z}_{B'} \end{bmatrix} \left| \begin{bmatrix} A' \\ B' \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{z}_{A'} \\ \mathbf{z}_{B'} \end{bmatrix} \right.\right) + c_1 \\
&\leq I\left(\mathbf{x}; A'\mathbf{x} + \mathbf{z}_{A'}\right) + h\left(\mathbf{z}_{\tilde{A}} - M \begin{bmatrix} \mathbf{z}_{A'} \\ \mathbf{z}_{B'} \end{bmatrix}\right) - h\left(\mathbf{z}_{\tilde{A}} \left| \begin{bmatrix} A' \\ B' \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{z}_{A'} \\ \mathbf{z}_{B'} \end{bmatrix}, \mathbf{x} \right.\right) + c_1 \\
&= I\left(\mathbf{x}; A'\mathbf{x} + \mathbf{z}_{A'}\right) + h\left(\mathbf{z}_{\tilde{A}} - M \begin{bmatrix} \mathbf{z}_{A'} \\ \mathbf{z}_{B'} \end{bmatrix}\right) - h\left(\mathbf{z}_{\tilde{A}}\right) + c_1 \\
&\leq I\left(\mathbf{x}; A'\mathbf{x} + \mathbf{z}_{A'}\right) + c_1 + c_2,
\end{aligned}
$$

where (i) follows from $A'\mathbf{x} + \mathbf{z}_{A'} \leftrightarrow \mathbf{x} \leftrightarrow \mathbf{B}'\mathbf{x} + \mathbf{z}_{\mathbf{B}'}$ and c_1 and c_2 are scalars independent of P. Since a MIMO channel with transfer matrix A' has $\operatorname{rank} A'$ degrees of freedom, we have that

$$I\left(\mathbf{x}; A'\mathbf{x} + \mathbf{z}_{A'}\right) \leq (\operatorname{rank} A') \tfrac{1}{2} \log P + o(\log P).$$

Moreover, from the proof of Lemma C.1, we know that $\operatorname{rank} A' = \operatorname{rank} \begin{bmatrix} A \\ B \end{bmatrix} - \operatorname{rank} B$, which concludes the proof. ■

Lemma C.3 *Let A be an $n \times n$ invertible matrix. If A' is an $(n-1) \times (n-1)$ submatrix obtained by removing the ith row and jth column of A for some i and j, then* $\operatorname{rank} A' \geq n - 2$.

Proof: Suppose by contradiction that $\operatorname{rank} A' < n - 2$. Consider the cofactor expansion of the determinant of A along the ith row. For each element (i, k), for $k \neq j$, the (i, k)th cofactor of A corresponds to the determinant of a matrix A'', obtained by replacing one of the columns of A' with the jth column of A without the ith entry. Since rank$A' < n-2$, rank$A'' \leq n-1$ and $\det A'' = 0$. Moreover, the (i, j)th cofactor of A is simply $\det A' = 0$. But this implies that $\det A = 0$, which is a contradiction. ■

C.3 Proof of Lemma 7.2

Lemma 7.2. *If a $K \times K \times K$ AWGN network is diagonalizable (in the sense of Definition 7.1), then for almost all values of the channel gains, $D_\Sigma = K$.*

Proof: The achievability scheme used to achieve K sum degrees of freedom is nearly identical to the Aligned Network Diagonalization scheme from Section 3.2.2 in the case of constant channel gains (see also Section 3.2.3). We will point out the main differences and refer the reader to Sections 3.2.2 and 3.2.3 and [57] for the technical details.

Each source s_i starts by breaking its message W_{s_i} into L submessages. Each of the submessages will be encoded in a separate data stream, using a single codebook with codewords of length n and only integer symbols. Now, let E_1 and E_2 be the edges from the first and second hops respectively. Then we define $\Delta_N = \{0, ..., N-1\}^{|E_1|}$ and

$$T_{\vec{m}} = \prod_{(s_i,u_j)\in E_1} F(s_i, u_j)^{m(s_i,u_j)}, \tag{C.2}$$

for some $\vec{m} = (m(e) : e \in E_1) \in \mathbb{N}^{|E_1|}$, and the set of transmit directions for the first hop will be given by

$$\mathcal{T}_N = \{T_{\vec{m}} : \vec{m} \in \Delta_N\}, \tag{C.3}$$

for some arbitrary N. Notice that the number of transmit directions (which is also the number of data streams) is $L = |\mathcal{T}_N| = |\Delta_N| = N^{|E_1|}$. We will let $c_{i,\vec{m}}[1]$, $c_{i,\vec{m}}[2], ..., c_{i,\vec{m}}[n]$ be the n symbols of the codeword associated to the submessage to be sent by source s_i over the transmit direction indexed by $\vec{m}$. At time $t \in \{1, ..., n\}$, source s_i will thus transmit

$$X_{s_i}[t] = \gamma \sum_{\vec{m}\in\Delta_N} T_{\vec{m}}\; c_{i,\vec{m}}[t]$$

where γ is chosen to satisfy the power constraint.

The received signal at relay u_j can be written as

$$\begin{aligned} Y_{u_j}[t] &= \gamma \sum_{\vec{m}\in\Delta_N} T_{\vec{m}} \left(\sum_{i=1}^{K} F_{s_i,u_j} c_{i,\vec{m}}[t] \right) + Z_{u_j}[t] \\ &= \gamma \sum_{\vec{m}\in\Delta_{N+1}} T_{\vec{m}}\; a_{j,\vec{m}}[t] + Z_{u_j}[t], \end{aligned} \tag{C.4}$$

where $a_{j,\vec{m}}[t] = \sum_{i=1}^{K} c_{i,\vec{m}_{ij}}[t]$ and we define $m_{ij}(s_k, u_\ell) = m(s_k, u_\ell)$ if $(s_k, u_\ell) \neq (s_i, u_j)$, $m_{ij}(s_i, u_j) = m(s_i, u_j) - 1$ and $c_{i,\vec{m}}[t] = 0$ if any component of $\vec{m}$ is -1 or N. As explained in [57], for almost all values of the channel gains, relay u_j can decode each integer $a_{j,\vec{m}}$ with high probability. These integers will be re-encoded by u_j using new transmit directions. To describe the new set of transmit directions, we first define

$$B(\mathcal{S}, \mathcal{U}) = \begin{bmatrix} B(s_1, u_1) & ... & B(s_K, u_1) \\ \vdots & \ddots & \vdots \\ B(s_1, u_K) & ... & B(s_K, u_K) \end{bmatrix} = F(\mathcal{U}, \mathcal{D})^{-1}. \tag{C.5}$$

Since we are considering a diagonalizable $K \times K \times K$ network according to Definition 7.1, for almost all values of the channel gains, $B(s_i, u_j) \neq 0$ if and only if $(s_i, u_j) \in E_1$. Thus, we may let

$$\tilde{T}_{\vec{m}} = \prod_{(s_i,u_j)\in E_1} B(s_i, u_j)^{m(s_i,u_j)}, \tag{C.6}$$

and, similar to (C.3), we can define the set of transmit directions for the relays to be

$$\tilde{\mathcal{T}}_{N+1} = \left\{ \tilde{T}_{\vec{m}} : \vec{m} \in \Delta_{N+1} \right\}.$$

Relay u_j will re-encode the $a_{j,\vec{m}}$ s by essentially replacing each received direction $T_{\vec{m}}$ in (C.4) with the direction $\tilde{T}_{\vec{m}}$. We highlight that this is only possible under the assumption of a diagonalizable $K \times K \times K$ network. The transmit signal of relay u_j at time $t+1$ will be given by

$$\begin{aligned} X_{u_j}[t+1] &= \gamma' \sum_{\vec{m} \in \Delta_{N+1}} \tilde{T}_{\vec{m}}\, a_{j,\vec{m}}[t] \\ &= \gamma' \sum_{\vec{m} \in \Delta_N} \tilde{T}_{\vec{m}} \left(\sum_{i=1}^{K} B(s_i, u_j)\, c_{i,\vec{m}}[t] \right) \end{aligned} \tag{C.7}$$

where γ' is chosen so that the output power constraint is satisfied.

In order to compute the received signals at the destinations, we first notice that, from (C.7), the vector of the K relay transmit signals at time $t+1$ can be written as

$$\gamma' \sum_{\vec{m} \in \Delta_N} \tilde{T}_{\vec{m}} \begin{bmatrix} B(s_1, u_1) & \dots & B(s_K, u_1) \\ \vdots & \ddots & \vdots \\ B(s_1, u_K) & \dots & B(s_K, u_K) \end{bmatrix} \begin{bmatrix} c_{1,\vec{m}}[t] \\ \vdots \\ c_{K,\vec{m}}[t] \end{bmatrix}. \tag{C.8}$$

Since the $\tilde{T}_{\vec{s}}$ s are just scalars, we can write the vector of the K received signals at the destinations as

$$\begin{aligned} \begin{bmatrix} Y_{d_1}[t+1] \\ \vdots \\ Y_{d_K}[t+1] \end{bmatrix} &= F(\mathcal{U}, \mathcal{D}) \begin{bmatrix} X_{u_1}[t+1] \\ \vdots \\ X_{u_K}[t+1] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t+1] \\ \vdots \\ Z_{d_K}[t+1] \end{bmatrix} \\ &= B(\mathcal{S}, \mathcal{U})^{-1} \begin{bmatrix} X_{u_1}[t+1] \\ \vdots \\ X_{u_K}[t+1] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t+1] \\ \vdots \\ Z_{d_K}[t+1] \end{bmatrix} \end{aligned}$$

$$= \gamma' \sum_{\vec{s} \in \Delta_N} \tilde{T}_{\vec{s}} \begin{bmatrix} c_{1,\vec{m}}[t] \\ \vdots \\ c_{K,\vec{m}}[t] \end{bmatrix} + \begin{bmatrix} Z_{d_1}[t+1] \\ \vdots \\ Z_{d_K}[t+1] \end{bmatrix}.$$

Thus, the received signal at destination d_j at time $t+1$ is simply given by

$$Y_{d_j}[t+1] = \gamma' \sum_{\vec{m} \in \Delta_N} \tilde{T}_{\vec{m}} \, c_{j,\vec{m}}[t] + Z_{d_j}[t+1], \tag{C.9}$$

and we see that all the interference has been cancelled, and destination d_j receives only the data streams originated at source s_j. Following the arguments in [57], it can be shown that such a scheme can indeed achieve K DoF. ■

BIBLIOGRAPHY

[1] M. J. Abdoli and A. S. Avestimehr. On degrees of freedom scaling in layered interference networks with delayed csi. *Proceedings of International Symposium on Information Theory*, 2013.

[2] M. J. Abdoli and A. S. Avestimehr. Layered interference networks with delayed csi: Dof scaling with distributed transmitters. *IEEE Transactions on Information Theory*, 60(3):1822–1839, March 2014.

[3] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, July 2000.

[4] V. S. Annapureddy and V. V. Veeravalli. Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region. *IEEE Transactions on Information Theory*, 55(7):3032–3050, July 2009.

[5] H. Asnani, I. Shomorony, T. Weissman, and A. S. Avestimehr. Network compression: Worst-case analysis. *Submitted to IEEE Transactions on Information Theory*, 2013.

[6] A. S. Avestimehr, S. Diggavi, and D. Tse. Wireless network information flow: a deterministic approach. *IEEE Transactions on Information Theory*, 57(4):1872–1905, April 2011.

[7] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[8] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[9] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 3rd edition, 1995.

[10] G. Bresler, D. Cartwright, and D. Tse. Geometry of the 3-user Gaussian MIMO interference channel. *Proc. Allerton Conf. on Communication, Control, and Computing, Monticello, IL*, September 2011.

[11] G. Bresler, A. Parekh, and D. Tse. The approximate capacity of the many-to-one and one-to-many Gaussian interference channels. *IEEE Transactions on Info. Theory*, 56(9):4566–4592, September 26-28 2010.

[12] Guy Bresler and David Tse. The two-user gaussian interference channel: a deterministic view. *European Transactions on Telecommunications*, 19(4):333–354, June 2008.

[13] V. R. Cadambe and S. A. Jafar. Interference alignment and degrees of freedom for the K-user interference channel. *IEEE Transactions on Information Theory*, 54(8):3425–3441, August 2008.

[14] V. R. Cadambe and S. A. Jafar. Interference alignment and degrees of freedom of wireless x networks. *IEEE Trans. on Information Theory*, 55(9):3893–3908, Sept. 2009.

[15] K. Cai, K. B. Letaief, P. Fam, and R. Feng. On the solvability of 2-pair unicast networks — a cut-based characterization. *ArXiv:1007.0465*, 2010.

[16] A. B. Carleial. Interference channels. *IEEE Transactions on Information Theory*, 24:60–70, January 1978.

[17] A. B. Carleial. Outer bounds on the capacity of interference channels. *IEEE Trans. on Info. Theory*, 29:602–606, July 1983.

[18] S. N. Diggavi and T. M. Cover. The worst additive noise under a covariance constraint. *IEEE Transactions on Info. Theory*, 47(7):3072–3081, November 2001.

[19] R. Etkin and D. Tse. Degrees of freedom in underspread mimo fading channels. *IEEE Transactions on Info. Theory*, 52:1576–1608, April 2006.

[20] R. Etkin, D. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Transactions on Information Theory*, 54(12):5534–5562, December 2008.

[21] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.

[22] A. El Gamal and M. Costa. The capacity region of a class of deterministic interference channels. *IEEE Transactions on Information Theory*, 28(2):343–346, March 1982.

[23] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2012.

[24] T. Gou, S.A. Jafar, S.-W. Jeon, and S.-Y. Chung. Aligned interference neutralization and the degrees of freedom of the $2 \times 2 \times 2$ interference channel. *IEEE Transactions on Information Theory*, 58(7):4381–4395, July 2012.

[25] T. Gou, C. Wang, and S. A. Jafar. Degrees of freedom of a class of non-layered two unicast wireless networks. *Proc. of Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA*, Nov. 2011.

[26] T. Gou, C. Wang, and S. A. Jafar. Interference neutralization and the degrees of freedom of the 2x2x2 interference channel with interfering relays. *Proceedings of Allerton Conference*, September 2011.

[27] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Trans. on Info. Theory*, 27:49–60, January 1981.

[28] N.J.A. Harvey, R. Kleinberg, and A.R. Lehman. On the capacity of information networks. *IEEE Transactions on Information Theory*, 52(6):2445–2464, June 2006.

[29] R.W. Heath, Jr. and A.J. Paulraj. Switching between diversity and multiplexing in mimo systems. *IEEE Trans. on Communications*, 53(6):962– 968, June 2005.

[30] S.-N. Hong and G. Caire. Two-unicast two-hop interference network: Finite-field model. *Information Theory Workshop*, 2013.

[31] S.-N. Hong and G. Caire. On interference networks over finite fields. *IEEE Transactions on Information Theory*, 60(8):4902–4921, August 2014.

[32] I. Issa, S. L. Fong, and A. S. Avestimehr. Two-hop interference channels: Impact of linear schemes. *Submitted to IEEE Transactions on Information Theory*, 2013.

[33] I. Issa, S. L. Fong, and A. S. Avestimehr. Two-hop interference channels: Impact of linear time-varying schemes. *Proceedings IEEE International Symposium on Information Theory*, 2013.

[34] S. A. Jafar and M. Fakhereddin. Degrees of freedom for the mimo interference channel. *IEEE Transactions on Info. Theory*, 53(7):2637–2642, July 2007.

[35] S.A. Jafar and S. Shamai (Shitz). Degrees of freedom region for the mimo x channel. *IEEE Transactions on Info. Theory*, 54(1):151–170, January 2008.

[36] S.-W. Jeon, S.-Y. Chung, and S. A. Jafar. Degrees of freedom region of a class of multi-source gaussian relay networks. *IEEE Transactions on Information Theory*, 57(5):3032–3044, May 2011.

[37] S.-W. Jeon, C.-Y. Wang, and M. Gastpar. Approximate ergodic capacity of a class of fading two-user two-hop networks. *IEEE Transactions on Information Theory*, 60(2):866, February 2014.

[38] S. Kamath and D. N. C. Tse. On the generalized network sharing bound and edge-cut bounds for network coding. *Proceedings of the International Symposium on Network Coding*, 2013.

[39] S. Kamath, D. N. C. Tse, and V. Anantharam. Generalized network sharing outer bound and the two-unicast problem. *Proceedings of the International Symposium on Network Coding*, 2011.

[40] G. Kramer. Outer bounds on the capacity of Gaussian interference channels. *IEEE Transactions on Information Theory*, 50(3):581–586, March 2004.

[41] A. Lapidoth. Nearest neighbor decoding for additive non-Gaussian noise channels. *IEEE Transactions on Information Theory*, 42(5):1520–1529, September 1996.

[42] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung. Noisy network coding. *IEEE Transactions on Information Theory*, 57(5):3132–3152, May 2011.

[43] M. A. Maddah-Ali, S. A. Motahari, and Amir K. Khandani. Communication over mimo x channels: Interference alignment, decomposition, and performance analysis. *IEEE Transactions on Information Theory*, 54(8):3457–3470, August 2008.

[44] A. S. Motahari and A. K. Khandani. Capacity bounds for the Gaussian interference channel. *IEEE Transactions on Information Theory*, 55(2):620–643, February 2009.

[45] A. S. Motahari, S. Oveis-Gharan, M. A. Maddah-Ali, and A. K. Khandani. Real interference alignment: Exploiting the potential of single antenna systems. *IEEE Transactions on Information Theory*, 60(8):4799–4810, August 2014.

[46] B. Nazer, M. Gastpar, S. A. Jafar, and S. Vishwanath. Ergodic interference alignment. *IEEE Transactions on Information Theory*, 58(10):6355–6371, September 2012.

[47] A. Ozgur and S. Diggavi. Approximately achieving Gaussian relay network capacity with lattice-based QMF codes. *IEEE Transactions on Information Theory*, 59(12):8275–8294, December 2013.

[48] B. Rankov and A. Wittneben. Spectral efficient protocols for half-duplex fading relay channels. *IEEE Journal on Selected Areas in Communications*, 25(2):379–389, Feb. 2007.

[49] H. Sato. Two-user communication channels. *IEEE Trans. on Info. Theory*, 23:295–304, May 1977.

[50] H. Sato. The capacity of the Gaussian interference channel under strong interference. *IEEE Transactions on Information Theory*, 27:786–788, November 1981.

[51] S. Shamai and S. Verdu. Worst-case power-constrained noise for binary-input channels. *IEEE Transactions on Information Theory*, 38(5):1494–1511, September 1992.

[52] X. Shang, G. Kramer, and B. Chen. A new outer bound and the noisy-interference sum–rate capacity for Gaussian interference channels. *IEEE Transactions on Information Theory*, 55(2):689–699, February 2009.

[53] S. Shenvi and B. K. Dey. A simple necessary and sufficient condition for the double unicast problem. *in Proceedings of ICC*, 2010.

[54] I. Shomorony and A. S. Avestimehr. On the role of deterministic models in $K \times K \times K$ wireless networks. *Proc. Information Theory Workshop (ITW)*, September 2012.

[55] I. Shomorony and A. S. Avestimehr. Two-unicast wireless networks: Characterizing the degrees of freedom. *IEEE Transactions on Information Theory*, 59(1):353–383, January 2013.

[56] I. Shomorony and A. S. Avestimehr. Worst-case additive noise in wireless networks. *IEEE Transactions on Information Theory*, 59(6):3833–3847, June 2013.

[57] I. Shomorony and A. S. Avestimehr. Degrees-of-freedom of two-hop wireless networks: "everyone gets the entire cake". *IEEE Transactions on Information Theory*, 60(5):2417–2431, May 2014.

[58] I. Shomorony, A. S. Avestimehr, H. Asnani, and T. Weissman. Worst-case source for distributed compression with quadratic distortion. *In Proc. of Information Theory Workshop (ITW)*, 2012.

[59] O. Simeone, O. Somekh, Y. Bar-Ness, H. V. Poor, and S. Shamai (Shitz). Capacity of linear two-hop mesh networks with rate splitting, decode-and-forward relaying and cooperation. *Proceedings of the Allerton Conference*, 2007.

[60] S. Sridharan, A. Jafarian, S. Vishwanath, and S. A. Jafar. Capacity of symmetric k-user gaussian very strong interference channels. *IEEE Global Telecommunications Conference*, 2008.

[61] C. Suh and D. Tse. Feedback capacity of the Gaussian interference channel to within two bits. *IEEE Transactions on Information Theory*, 57(5):2667–2685, May 2011.

[62] S. Thakhor, A. Grant, and T. Chan. Network coding capacity: A functional dependence bound. *Proc. of IEEE International Symposium on Information Theory*, 2009.

[63] C. S. Vaze and M. K. Varanasi. The degrees of freedom of the 2 x 2 x 2 interference network with delayed csit and with limited shannon feedback. *Proceedings of the Allerton Conference*, September 2011.

[64] A.B. Wagner, S. Tavildar, and P. Viswanath. Rate region of the quadratic Gaussian two-encoder source-coding problem. *IEEE Transactions on Information Theory*, 54(5):1938–1961, May 2008.

[65] C. Wang, T. Gou, and S. A. Jafar. Multiple unicast capacity of 2-source 2-sink networks. *IEEE Global Telecommunications Conference*, 2011.

[66] I.-H. Wang. On two unicast wireless networks with destination-to-source feedback on two unicast wireless networks with destination-to-source feedback. *Proceedings of IEEE International Symposium on Information Theory*, July 2012.

[67] I.-H. Wang and S. Diggavi. On degrees of freedom of layered two unicast networks with delayed csit. *Proceedings of IEEE International Symposium on Information Theory*, July 2012.

[68] I.-H. Wang, S. Kamath, and D. N. C. Tse. Two unicast information flows over linear deterministic networks. *Proc. of IEEE International Symposium on Information Theory*, 2011.

[69] J. Xie and S. Ulukus. Sum secure degrees of freedom of two-unicast layered wireless networks. *IEEE Journal on Selected Areas in Communications*, 31(9):1931–1943, September 2013.

[70] X. Yan, J. Yang, and Z. Zhang. An outer bound for multisource multisink network coding with minimum cost consideration. *IEEE Transactions on Information Theory*, 52(6):2373–2385, June 2006.

[71] M. Yuksel and E. Erkip. Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective. *IEEE Transactions on Information Theory*, 53(10):3371–3393, October 2007.

[72] L. Zheng and D.N.C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Trans. on Information Theory*, 49(5):1073–1096, May 2003.