Opportunistic Scheduling, Cooperative Relaying and Multicast in Wireless Networks

by

Chandrashekhar Thejaswi Pataguppe Suryanarayan Bhat

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved February 2011 by the
Graduate Supervisory Committee:

Junshan Zhang, Chair
Douglas Cochran
Yu Hui
Thomas Taylor
Tolga Duman

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

This dissertation builds a clear understanding of the role of information in wireless networks, and devises adaptive strategies to optimize the overall performance. The meaning of information ranges from channel/network states to the structure of the signal itself. Under the common thread of characterizing the role of information, this dissertation investigates opportunistic scheduling, relaying and multicast in wireless networks.

To assess the role of channel state information, the problem of opportunistic distributed opportunistic scheduling (DOS) with incomplete information is considered for ad-hoc networks in which many links contend for the same channel using random access. The objective is to maximize the system throughput. In practice, link state information is noisy, and may result in throughput degradation. Therefore, refining the state information by additional probing can improve the throughput, but at the cost of further probing. Capitalizing on optimal stopping theory, the optimal scheduling policy is shown to be threshold-based and is characterized by either one or two thresholds, depending on network settings.

To understand the benefits of side information in cooperative relaying scenarios, a basic model is explored for two-hop transmissions of two information flows which interfere with each other. While the first hop is a classical interference channel, the second hop can be treated as an interference channel with transmitter side information. Various cooperative relaying strategies are developed to enhance the achievable rate. In another context, a simple sensor network is considered, where a sensor node acts as a relay, and aids fusion center in detecting an event. Two relaying schemes are considered: analog relaying and digital relaying. Sufficient conditions are provided for the optimality of analog relaying over digital relaying in this network.

To illustrate the role of information about the signal structure in joint source-channel coding, multicast of compressible signals over lossy channels is studied. The focus is on the network outage from the perspective of signal distortion across all receivers. Based on extreme value theory, the network outage is characterized in terms of key parameters. A new method using subblock network coding is devised, which prioritizes resource allocation based on the signal information structure.

To *Anna* (my late father) and *Pachanna* (my late brother)

who instilled an unquenchable thirst for knowledge in me.

ACKNOWLEDGEMENTS

I am greatly indebted to my advisor, Prof. Junshan Zhang, for his support (financial, professional and moral) and his close guidance. I have always admired his passion and enthusiasm towards research and formulating interesting problems. His unquenchable thirst and excitement towards solving new problems and deriving useful insights has always inspired me. Through several insightful discussions, he has taught me a lot about what it takes to be a good researcher and a good writer. Moreover, he is a very kind person and has been always accessible to discuss technical as well as personal problems with him. He has been a great philosopher, friend and guide to me during my stay as a PhD student. I am very fortunate to have him as my advisor.

Another individual who deeply influenced me is Prof. Cochran, with whom I had the opportunity to work on a part of my research. I thank him for giving me an opportunity to work with him and teaching me how to ask simple but fundamental questions while doing research. I truly appreciate his patience and am very grateful to him for all support and time he gave to me despite his busy schedule.

I thank Prof. Duman for teaching me the fundamentals of wireless communications, and space-time coding theory. Several technical discussions I had with him were really valuable and helped me in strengthening my fundamentals, and in building a strong foundation for my research. I would also like to thank Prof. Hui and Prof. Taylor for serving my thesis committee and for their valuable suggestions to shape my research.

Many thanks to my current and former colleagues: Dong, Weiyan, Qinghai, Shanshan, Lei, Miao, Dajun, Sugumar, Tuan, Xiaowen, Brian and Eric for the pleasant and inspiring discussions. I also thank my collaborators: Simon and Amir, for the insightful discussions I had with them.

I am grateful to my wife Smitha for her patience, understanding, support and for putting up with me during my PhD. I am grateful to my mother, sisters, brothers, brothers-in-law and sisters-in-law for their constant encouragement and support throughout my life. I also thank my father-in-law and mother-in-law for the excitement and interest they showed in the progress of my research. It has been a great pleasure to interact, rag and mentor my nephews Manja, Pallu and my sister-in-law Priyanka. It has been equally blissful to play with my cute nieces Manasa, Spoorthi, Shreya and Raksha. I am really fortunate to have such a wonderful family.

Finally, I would also like thank all my teachers, friends and relatives without whose support and encouragement, I could not have come this far.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Recent years have witnessed a tremendous growth in the demand for ubiquitous information access. Indeed, our society hinges heavily on reliable and efficient operations of large-scale networks, e.g., the Internet and wireless ad-hoc/sensor networks, for collecting, processing, analyzing and managing information in adverse noisy environments. The unique characteristics of wireless links, together with the bursty nature of traffic flows, create both new challenges and opportunities in our attempts to implement this vision. Different from the wireline counterpart, the design of wireless networks faces a number of unique challenges, particularly, co-channel interference and multipath fading.

1) *Co-channel interference:* When more than one information flow occupies the same channel, the shared nature of wireless medium results in co-channel interference. The receiver, apart form the message of its own source, also receives transmissions from other unintended sources. This may lead to collisions and transmission outages. Wireless systems deal with interference by incorporating time/frequency orthogonality, or by handling at the medium access control (MAC)-layer level, typically, via scheduling or random access protocols.

2) *Time varying channel conditions over fading channels:* Fading is the time variation of the wireless channel, and can often be characterized with two types of effects: large-scale path loss and shadowing effects that cause the signal to attenuate with distance; and multipath scattering effects that result in delayed copies of the signal adding up constructively or destructively at the receiver. Fading effects are often mitigated at the physical layer using coding/modulation and diversity techniques.

Besides the above impediments, wireless networks often operate under hostile conditions that include bursty traffic and changing network topology. Meeting the quality of services (QoS) requirements of the end users can be extremely difficult in such hostile operating conditions.

Further, for practical reasons, wireless systems are often constrained in terms of resources such as bandwidth, power, time etc. The impetus for the optimal design and operation of any wireless system lies in the efficient management of these limited resources. Effective management of limited resources, in turn, hinges heavily on the availability of information about the system states in the presence of complicated dynamics in the network. For instance, in multimedia systems, availability of *a priori* information about the parameters, such as the distribution and "compressibility" of the signals, helps in efficient data compression, thereby saving storage and transmission bandwidth. In multiuser wireless networks, the availability of information about the dynamics of the channel state, traffic flow, topology etc., at each epoch, can be leveraged using the notion of opportunistic scheduling to significantly enhance the QoS provisioning. Also, in networks where interference is the main impediment, access to information about the interference and its structure enables efficient design of transmission/reception techniques toward optimal interference management.

Thus motivated, one primary goal of this work is to understand the "value of information" in enhancing the system performance. As noted above, the meaning of information ranges from state information about the channel/network dynamics to the structure of the source signal itself. By developing a clear understanding of various forms of information availability, we demonstrate that indeed one can gain substantially, in terms of overall performance, when compared to the methods which do not exploit the availability of information.

Under the common theme of characterizing the value of information, this thesis is broadly organized into four chapters. In Chapter 2, distributed opportunistic scheduling (DOS) is studied with the objective of quantifying the trade-off between throughput gain with refined channel state information and the corresponding overhead. Based on optimal stopping theory, a framework for PHY-aware scheduling is developed for exploiting rich diversities at the physical-layer and medium access control (MAC) layer. In Chapter 3, a two-hop interference network is studied, in which the second hop is viewed as an interference channel with side information at the transmitters. Information theoretic techniques are applied to study the value of this side information in enhancing the achievable rate. In Chapter 4, the problem of multicasting compressible signals is considered, where one principal objective is to exploit the information available on the signal structure in developing joint source/channel coding strategy to improve the QoS.

In Chapter 5, a sensory relay network model is considered, and optimality conditions for analog vs digital relaying are characterized. In what follows, we briefly summarize our contributions along with the motivations and techniques involved.

1.1 Distributed Opportunistic Scheduling

As noted above, two key challenges to the wireless communications are interference and fading. In the design of wireless ad-hoc networks, the traditional approach has been to separate packet losses caused by fading from those caused by collision, neglecting the structure pertaining to the dynamics of these parameters in the networks. That is, the PHY layer addresses the issues of fading, and the MAC layer addresses the issue of contention. However, as shown in [1, 2], fading can often adversely affect the MAC layer in many realistic scenarios. The coupling between the temporal dynamics of fading and MAC calls for a unified PHY/MAC design for wireless ad-hoc networks in order to achieve optimal throughput and latency. Indeed, lurking beneath this dynamic nature of the wireless medium and traffic are the joint PHY/MAC diversities (including multiuser, time and spatial diversities), which are available for exploitation in a wide range of wireless scenarios. Nevertheless, one has to carefully understand these dynamics and build channel-aware scheduling approaches for efficient handling of the information flow. It is therefore of critical importance to develop a rigorous understanding of state-aware scheduling that can resolve contention and mitigate interference efficiently while exploiting diversities.

Notably, there has recently been a surge of interest in channel-aware scheduling and channel-aware access control. Channel-aware opportunistic scheduling was first developed for downlink transmissions in multiuser wireless networks ([3],[4],[5],[6]). While these studies assumed centralized scheduling, distributed opportunistic scheduling was initiated by Zheng *et al.* [7], where, using optimal stopping theory, authors devise scheduling strategies for ad-hoc networks under the assumption that the nodes have perfect channel information. Generalization of [7] to the noisy estimation case is carried out in [8]. It has been observed in [9] that the errors in channel estimation results in outages, and they propose backoff schemes to avoid transmission outages. However, backoff may lead to severe throughput degradation, especially in the low SNR regime, due to a more conservative rate reduction. Therefore, a plausible solution is to mitigate rate estimation errors by performing further channel probing. Clearly, the improved rate estimation obtained with second-level probing enables the desired link to make more accurate decisions. However, the advantages of second-level probing come at the price of additional overhead. There is, therefore, a tradeoff between the throughput gain from better channel conditions and the cost for further probing.

With this insight, in Chapter 2, we investigate DOS with two-level channel probing by optimizing the tradeoff between the throughput gain from more accurate rate estimation and the resulting additional probing overhead. Based on the recent advances in OST, namely OST with two-level incomplete information [10] and statistical versions of "prophet inequalities" [11], we show that the optimal scheduling policy is threshold-based and is characterized by either one or two thresholds, depending on network settings. Necessary and sufficient conditions for both cases are rigorously established. In particular, our analysis reveals that performing second-level channel probing is optimal when the first-level estimated channel condition falls in between the two thresholds. Numerical results are provided to illustrate the effectiveness of the proposed DOS with two-level channel probing. We also extend our study to the case with limited feedback, where the feedback from the receiver to its transmitter takes the form of $(0, 1, e)$.

1.2 Two-Hop Interference Flows

Relay channels and interference channels have been basic building blocks of wireless networks. Particularly, interference in networks is modeled by a basic two user interference channel (IC), which consists of two transmitter-receiver pairs involved in simultaneous communication, and their transmissions interfere with each other. As noted earlier, interference is a central phenomenon in wireless communications. Most state-of-the-art wireless systems deal with interference in one of two ways: orthogonalize the communication links in time or frequency, so that they do not interfere with each other; or, allow the communication links to share the same degrees of freedom, but treat each other's interference as adding to the noise floor. It is clear that both approaches can be sub-optimal. The first approach entails an a priori loss of degrees of freedom in both links, no matter how weak the potential interference is. The second approach treats interference as pure noise while it actually carries information and has structure that can potentially be exploited in mitigating its effect.

Unfortunately, the problem of characterizing the capacity region of a general IC has been open. The only case in which the capacity is known is in the strong interference case, where each receiver has a better reception of the other user's signal than the intended receiver [12]. The best known strategy for the general case is due to Han and Kobayashi (HK) [13]. This strategy is a natural one and involves splitting the transmitted information of both users into two parts: private information to be decoded only at its own receiver and common information that can be decoded at both receivers. By decoding the common information, part of the interference can be canceled off, while the remaining private information from the other user is treated as noise. Recently, Etkin and Tse [14], have proposed a specific HK type scheme where it is shown that the proposed scheme achieves to within a single bit of the capacity region of an IC.

In a relay channel, as introduced by van der Meulen [15], a relay node assists the source in communicating data to a destination. Relay networks are instrumental in harnessing the dynamic nature of the wireless medium to yield a form of diversity called cooperative diversity [16].

An interesting model under consideration can be viewed as a combination of relay channels and interference channels. Specifically, this model involves two sources which communicate simultaneously with two destinations, and are aided by two relay nodes. While the first hop is akin to the traditional interference channel, the second hop has an interesting feature. Specifically, as a by-product of the Han-Kobayashi [13] transmission scheme applied to the first hop, each of the relays (in the second hop) has access to some of the data that is intended to the other destination, in addition to its own data. Thus, the second hop represents an interference channel with side information at the transmitters. This feature opens the door to cooperation between the relays. A clear understanding towards exploiting side information at the relays in the second hop enables one to devise relaying strategies to manage interference and efficient information flow, thereby enhancing the achievable rate region of the network.

In Chapter 3, we explore some cooperative strategies among relays that systematically exploit the availability of side information to maximize the achievable rates. Specifically, we observe that the availability of side information at each relay opens the door for cooperation which can take the form of distributed multiple-input and multiple-output (MIMO) broadcast, thus greatly enhancing its effectiveness at high SNR. However, since each relay has only *partial* side information of the data beyond its own, full cooperation is not possible. Thus, a key feature of this network is that it has elements of both the broadcast channel and the interference channel.

In light of this, we study strategies based on the nontrivial marriage of MIMO-BC techniques for cooperative relaying, rate-splitting and superposition coding, to enable interference cancelation at the receivers, and the Gelfand-Pinsker (GP) coding [17] at each transmitter to reduce the interference to its own receivers. Specifically, we propose two types of layered schemes that combine MIMO broadcast and HK coding. The first one is layered coding with binning which mainly hinges on a novel interplay between binning (DPC) and HK coding. The second one is a layered coding with superposition strategy that involves superposition coding over different tiers. Numerical results are provided that indicate our approaches provide substantial benefits at high SNR.

1.3 Multicast of Compressible Signals

Reliable delivery of compressible information to many destinations is an important scenario in modern day wireless systems. This transmission scenario, for instance, is useful for multimedia streaming over a WiFi or WiMAX network with multiple receivers. It is also applicable to sensor networks, where each sensor node wishes to communicate its sensed observations to multiple coordinating agents. Needless to say, this problem is quite challenging due to the lossy nature of wireless channels and the heterogeneity in the amount of information received across different receivers. Another challenge is the "*bottleneck*" effect with the shared transmission medium among many receivers; i.e., the overall performance is limited by the receiver that has the worst channel condition. Therefore, ensuring that the QoS for the bottleneck user is on par with the others clearly makes the multicast transmission more challenging. In light of this, it is of paramount importance to come up with transmission schemes that can mitigate channel erasures so that data loss is minimized. Notably, network coding (NC), a recent breakthrough in this avenue by Ahlswede *et al.* [18], offers a promising platform for multicast transmissions.

When we consider source signals for transmission, many of these signals are known to be sparse or compressible. The conventional method of data compression involves sampling the signal at the Nyquist's rate, storing the samples, and compressing them in an appropriate domain prior to the transmission, which would incur heavy sampling and storing burden at the sender. On the other hand, recent developments of compressive sensing theory [19, 20, 21] have provided methods not only for lower-rate signal acquisition, but also for accurate signal reconstruction.

When it comes to transmission of compressible data over wireless links, the traditional wisdom has been to adopt separation principle. That is, while the structure of the source signal is considerably exploited in sampling and compression, it is seldom exploited in minimizing the "information loss" incurred due to transmission over the hostile channel. This calls for joint design of compression and transmission strategies to enhance the flow of information.

With this motivation, in Chapter 4, we study multicasting compressively sampled signals from a source to many receivers, over lossy wireless channels. Our focus is on the network outage from the perspective of signal distortion across all receivers, for both cases where the transmitter may or may not be capable of reconstructing the compressively sampled signals. Capitalizing on extreme value theory, we characterize the network outage in terms of key system parameters, including the erasure probability, the number of receivers and the sparse structure of the signal. We show that when the transmitter can reconstruct the compressively sensed signal, the strategy of using network coding to multicast the reconstructed signal coefficients can reduce the network outage significantly. We observe, however, that the traditional network coding could result in suboptimal performance for power-law decay signals. Thus motivated, we devise a new method, namely subblock network coding, that exploits the knowledge about the signal structure. Essentially, subblock coding involves fragmenting the data into subblocks, and allocating time slots to different subblocks, based on their priorities. We formulate the corresponding optimal allocation as an integer programming problem. Since integer programming is often intractable, we develop a heuristic algorithm that prioritizes the time slot allocation by exploiting the inherent priority structure of power-law decay signals. Numerical results show that the proposed schemes outperform the traditional methods with significant margins.

1.4 Analog versus Digital Relaying in a Sensor Network

In Chapter 5, we consider a simple model of wireless sensor network for hypothesis testing, where a sensor node acts as a relay and aids a fusion center to perform hypothesis testing on an event. The sensor performs noisy observations on the underlying event, performs a local processing and then relays the processed data to the fusion center. We compare two relay schemes: "estimate-and-forward" (analog relaying) and "detect-and-forward" (digital relaying), in terms of the ultimate detection performance they support at the fusion center. With this performance criterion, the relative merit of the two schemes is shown to depend on the observation SNR at the sensor and the SNR of the communication link connecting the sensor and the fusion center. A sufficient condition, in terms of these SNRs, for the superiority of digital relaying over analog relaying is derived for this simple network model. Although the network model used here is highly simplified, it is hoped that this work will contribute to a foundation for analysis of more realistic scenarios, leading to advances in sensor placement strategies and inference algorithms in sensor networks. This study can also impact the development of cooperative relaying strategies in wireless relay networks.

Chapter 2

DISTRIBUTED OPPORTUNISTIC SCHEDULING WITH INCOMPLETE STATE INFORMATION

2.1 Introduction

Channel-aware scheduling has recently emerged as a promising technique to harness the rich diversities inherent in wireless networks. In channel-aware scheduling, joint physical layer (PHY)/medium access control (MAC) optimization is utilized to improve network throughput by scheduling links with good channel conditions for data transmissions [3, 22, 6]. While most existing studies focus on centralized scheduling (see, e.g., [4, 23, 24, 22, 6]), some initial steps have been taken in [25] to develop distributed opportunistic scheduling (DOS) to reap multiuser diversity and time diversity in wireless ad-hoc networks.

The DOS framework considers an ad-hoc network in which many links contend for the same channel using random access, e.g., carrier-sense multiple-access (CSMA). However, random access protocols provide no guarantee that a successful channel contention is necessarily attained by a link with good channel conditions. From a holistic perspective, a successful link with poor channel conditions should forgo its data transmission and let all links re-contend for the channel. This is because after further channel probing, it is more likely for a link with better channel conditions to take the channel, yielding possibly higher throughput. In this way, multiuser diversity across links and time diversity across time can be exploited in a joint manner. However, each channel probing incurs a cost of contention time. The desired tradeoff between the throughput gain from better channel conditions and the cost for further probing reduces to judiciously choosing an optimal rule for stopping channel probing for throughput maximization. Using optimal stopping theory (OST), it is shown in [25] that the optimal scheduling scheme turns out to be a pure threshold policy: The successful link proceeds to transmit data only if its supportable rate is higher than the pre-designed threshold; otherwise, it skips the transmission opportunity and lets all other links re-contend. In general, threshold-based scheduling uses local information only, and hence is amenable to easy distributed implementation in practical systems.

The initial study of DOS [25] hinges upon a key assumption that the channel state information (CSI) is perfectly available at the receiver. In practice, the link rates are estimated from noisy observations. It is shown in [9] that the signal-to-noise ratio (SNR) estimated by the minimum mean squared error (MMSE) method is larger than the "actual SNR" due to the estimation error noise. Thus, the transmission rate must be backed off from the estimated rate in order to avoid transmission outages. Our initial steps in [8] show that the optimal scheduling policy under noisy channel estimation still has a threshold structure.

Despite their robust performance under noisy channel estimation, the linear backoff schemes proposed in [8] are reactive in nature and back off the rate by a factor proportional to the channel estimation errors, which may lead to severe throughput degradation, especially in the low SNR regime (where a more conservative rate backoff is needed). Recently, wideband communications (e.g., ultra-wideband) has attracted significant attention [26], owing to its low-power operation and the ability to co-exist with other legacy networks, etc. The great potential of wideband communications gives an impetus to address the problem of throughput degradation due to estimation errors, in the low-SNR (wideband) regime. More specifically, to circumvent this drawback, a plausible solution is to mitigate the rate estimation errors by performing further channel probing. In the sequel, we refer to the initial rate estimation performed *during* the channel contention as "*first-level probing*", whereas the subsequent probing performed *after* the successful contention is referred to as "*second-level probing*". Clearly, the improved rate estimation obtained with second-level probing enables the desired link to make more accurate decisions. However, the advantages of second-level probing come at the price of additional delay. This gives rise to two important questions: 1) Is it worthwhile for the link with successful contention to perform further channel probing to refine the rate estimate, at the cost of additional probing? 2) While there is always a gain in the transmission rate due to the refinement, how much can one bargain with the additional probing overhead?

We shall answer these questions by considering distributed opportunistic scheduling with two-level channel probing. Based on two recent advances in optimal stopping theory, namely optimal stopping with two-level incomplete information [10] and statistical versions of "prophet inequalities" [11], we provide a rigorous characterization of the scheduling strategy that optimizes the tradeoff between the throughput gain achieved by second-level channel probing and the resulting additional delay. It is shown that the optimal scheduling strategy is threshold-based and is characterized by either one or two thresholds, depending on the system parameters. By establishing the corresponding necessary and sufficient conditions for these two cases, we show that the second-level probing can significantly improve the system throughput when the estimated rate via first-level probing falls in between the two thresholds. In such scenarios, the cost of additional delay can be well justified by the throughput enhancement using the second-level channel probing. We elaborate further on this in Section 2.3. Finally, through numerical results, we illustrate the effectiveness of the proposed scheduling scheme.

Before proceeding further, the main contributions distinguishing this work from other existing works should be emphasized. OST under two levels of incomplete information is addressed with the objective of *maximizing the net-return* in [10]; in contrast, we study OST with two levels of probing as applied to DOS with the objective of *maximizing the rate of return* (i.e., the throughput). We study distributed opportunistic scheduling for ad-hoc communications under noisy conditions where the rate estimate is available only after a successful channel contention; and this is clearly different from [9], which considers centralized scheduling assuming that the rate estimates of all links are available a priori at the base station. Despite the fact that both this work and [8] study distributed opportunistic scheduling with imperfect information, this work concentrates on proactively improving throughput by enhancing rate estimation, whereas [8] proposes to passively reduce data rate to avoid transmission outages. Another related work [27] uses optimal stopping theory to investigate the intrinsic trade-off between energy and delay in distributed data aggregation and forwarding in sensor networks.

The rest of the chapter is organized as follows. In Section 2.2, we provide a brief introduction to the optimal stopping theory, discuss the system model, and provide background on DOS with only first-level probing in noisy environments. In Section 2.3, we present second-level channel probing and characterize the optimal DOS with two-level probing. We also present numerical results to illustrate the gain due to two-level probing. In Section 2.4, we extend our study to the case where there is limited feedback from the receiver to its transmitter. Finally, Section 2.5 contains our conclusions.

Notation: $|\cdot|$ denotes the amplitude of the enclosed complex-valued quantity. $\mathcal{R}^+$ denotes the set of non-negative real numbers. We use $[x]^+$ for $\max[x, 0]$, and $\mathbb{E}[\cdot]$ for expectation.

2.2 Background and System Model

Preliminaries on optimal stopping theory

As noted above, in an ad-hoc communication network with many links, when a link discovers that its channel condition is "relatively poor" after a successful channel contention, it can either transmit or skip this opportunity so that, in the next round, some link with a better condition would have the chance to transmit. This is intimately related to the optimal stopping problem in sequential analysis [28]. Simply put, optimal stopping theory is concerned with the problem of choosing a strategy for deciding when to take a given action based on the past events in order to maximize an average return, where return is the net gain (the difference between the reward and the cost). The corresponding strategy is called an optimal stopping rule.

More specifically, let $Z_1, Z_2, ...$ denote a sequence of random variables, and let $Y_0, Y_1(Z_1), Y_2(Z_1, Z_2), \ldots, Y_\infty(Z_1, Z_2, \ldots)$ a sequence of real-valued reward functions. The reward is $Y_n(Z_1, ..., Z_n)$ if the strategy chooses to stop at time n. The theory of optimal stopping is concerned with determining the stopping time N to maximize the expected reward $\mathbb{E}[Y_N]$; and in general, a *stopping rule (or a stopping time)* (cf. [28]) is defined to be a random variable N such that $\{N = n\} \in \mathcal{F}_n$, where $\mathcal{F}_n$ is the σ-algebra generated by $\{Z_1, \ldots, Z_n\}$. This is equivalent to saying that the decision to transmit at a slot n depends only on the sequence $\{Z_1, \ldots, Z_n\}$. A good introduction to optimal stopping theory can be found in [28],[29], or [30].

Consider a single-hop ad-hoc network in which L links contend for the channel using random access. A collision model is assumed for random access, in which a channel contention of a link is said to be successful if no other links transmit at the same time. Let p_ℓ be the probability that link ℓ contends for the channel, $\ell = 1, \ldots, L$. Then the overall successful contention probability, p_s, is given by $p_s = \sum_{\ell=1}^{L} \left(p_\ell \prod_{i \neq \ell} (1 - p_i) \right)$ (cf. [31]). For ease of exposition, we assume that the contention probabilities, $\{p_\ell\}$, remain fixed (see [32] for a study with adaptive contention probabilities). We define the random duration of achieving one successful channel contention as one round of channel probing. Clearly, the number of slots in each probing round, K, is a geometric random variable, i.e., $K \sim G(p_s)$. Denoting the slot duration by τ, the corresponding random duration for one probing round thus becomes $K\tau$, with its expected value being τ/p_s.

In a nutshell, each round of channel probing consists of two phases, namely, channel contention and channel estimation. We assume that a link can estimate its link conditions (hence the transmission rate) after a successful contention[1].

Let $s(n)$ denote the successful link in the n-th round of channel probing, and R_n denote the corresponding transmission rate. Due to the time-varying nature of wireless channels, R_n is random. Following the standard assumption on block fading channels in wireless communications [33], we assume that the channel remains constant for a duration of T. When an estimate of the transmission rate is available, the successful link may decide to transmit over a duration of T, if the rate is high enough, or may skip it[2] and allow all links to re-contend, in the hope that another link with a better channel will take the channel later.

To get a more concrete sense of joint channel probing and distributed scheduling, we depict, in Fig. 2.1, an example with N rounds of channel probing and one single data transmission. Specifically, suppose after the first round of channel probing with a duration of K_1 slots, the rate, R_1, of link $s(1)$ is very small (indicating a poor channel

[1]The successful link can carry out its rate estimation via a training phase during the request-to-send/clear-to-send (RTS/CTS) handshake, which follows a successful contention. This procedure is fairly standard in the literature, and is not dealt with here.

[2]This decision can be broadcast to all users in the one-hop neighborhood (e.g., NCTS).

condition); and as a result, $s(1)$ gives up this transmission opportunity and lets all links re-contend. Then, after the second successful contention with a duration of K_2 slots, link $s(2)$ also gives up the transmission because its rate, R_2, is also small. This continues for N rounds until link $s(N)$ transmits because its transmission rate, R_N, is good. Clearly, there exists a tradeoff between the throughput gain from better channel conditions and the cost for further probing.

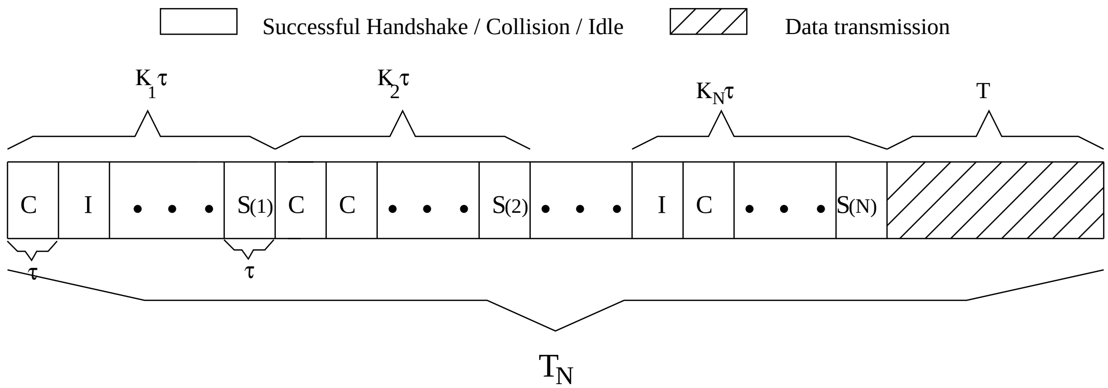

Figure 2.1: A sample realization of channel probing and data transmission.

In [25], it is shown that the process of joint channel probing and distributed scheduling can be treated as a team optimization problem in which all links collaborate to *maximize rate of return* (the average throughput). Specifically, as illustrated in Fig. 2.1, after one round of channel probing, a stopping rule N decides whether the successful link carries out data transmission, or simply skips this opportunity to let all links re-contend. Let $T_n = \sum_{j=1}^{n} K_j\tau + T$ be the total system time, defined as the sum of the contention time and the transmission duration, where K_j is the number of slots in the jth probing round. It turns out that the optimal DOS strategy achieving the maximum throughput hinges on the optimal stopping rule N^*, which yields the maximal rate of return θ^*. That is,

$$\theta^* \triangleq \sup_{N \in Q} \frac{\mathbb{E}[R_N T]}{\mathbb{E}[T_N]}, \tag{2.1}$$

and

$$N^* \triangleq \arg\max_{N \in Q} \frac{\mathbb{E}[R_N T]}{\mathbb{E}[T_N]}, \tag{2.2}$$

where

$$Q \triangleq \{N : N \geq 1, \mathbb{E}[T_N] < \infty\}. \tag{2.3}$$

It is clear that R_n plays a critical role in distributed opportunistic scheduling. In practice, rate estimates are seldom perfect. It is shown in [9] that the rate corresponding to the estimated SNR tends to be greater than the actual rate, and subsequently the transmission rate must be backed off from the estimated rate to avoid outages. Then, a natural question to ask is whether it is worthwhile for the link with successful contention to perform further channel probing to refine the channel estimate, at the cost of additional probing overhead, and how much can one bargain?

Intuitively speaking, when the transmission rate is small, it makes sense to give up the transmission, since the gain due to rate refinement would be marginal due to the poor link conditions. On the other hand, when the rate is large enough, it may not be advantageous to perform additional probing as the improvement is meager. It is natural to expect that there exists a "gray area" between these extremes where significant gains are possible by refining the rate estimate with additional probing. In what follows, we seek a clear understanding of the above fundamental issues.

To this end, we present the PHY-layer signal model first. The received signal corresponding to $s(n)$ can be written as[3]

$$Y_{s(n)}(n) = \sqrt{\rho} h_{s(n)}(n) X_{s(n)}(n) + \xi_{s(n)}(n), \tag{2.4}$$

where ρ is the *normalized* receiver SNR, $h_{s(n)}(n)$ is the channel gain for link $s(n)$, $X_{s(n)}(n)$ is the transmitted signal with $\mathbb{E}[|X_{s(n)}(n)|^2] = 1$, and $\xi_{s(n)}(n)$ is additive white Gaussian noise (AWGN) with unit variance. Further, $h_{s(n)}(n)$ and $h_{s(m)}(m)$ are the channel coefficients corresponding to the link with successful contention in the n-th round of probing and that in the mth round of probing. In this work, we consider a homogeneous network in which all links are subject to independent Rayleigh fading, with identical channel statistics. *With this observation, we assume that $h_{s(n)}(n)$ and $h_{s(m)}(m)$ are independent for $n \neq m$.* This is a practically valid assumption because the likelihood of one link (say link m) achieving two consecutive successful channel probings, $p_m^2 \prod_{i \neq m}(1 - p_i)^2$, is fairly small, especially when the number of links in the network is large. Furthermore, even if the same link successfully obtains two consecutive

[3]We note that the results reported here can be extended to frequency-selective fading channels by replacing scalar fading parameters with vectors.

channel contentions, the channel conditions corresponding to the two consecutive successful channel probings are independent since the channel probing duration in between is designed to be comparable to the channel coherence time. As shown in [25], when $p_m = \frac{1}{L}, L = 10$ and $\pi = 0.9$, the probability that the correlation across two adjacent successful contentions is no greater than 0.1 is 0.903. Summarizing, it is quite reasonable to impose the assumption on the channel independence between two successful channel contentions.

Without loss of generality, to simplify our exposition, we make the following simplifications: We focus on the n-th probing round and omit the temporal index n, whenever possible. We use Y_n, X_n, ξ_n and h_n to denote $Y_{s(n)}(n)$, $X_{s(n)}(n)$, $\xi_{s(n)}(n)$ and $h_{s(n)}(n)$, respectively, in the sequel. For convenience, the parameter T is normalized to unity; i.e., $T = 1$.

When perfect CSI is available to the link, as assumed in [25], the instantaneous supportable data rate is given by the Shannon channel capacity:

$$R_n = W \log(1 + \rho |h_n|^2), \tag{2.5}$$

where W is the bandwidth. Observe that $\{R_n,\ n = 1, \ldots\}$ are independent and identically distributed (i.i.d.) due to the assumption that h_n are independent and homogeneous.

To facilitate our analysis, we concentrate our following investigation in the low SNR (wideband) regime, assuming $\rho \to 0$ and $W = \Theta(\frac{1}{\rho})$. It is well known that a decrease in SNR estimation error can only increase the rate of communication. For cases with wideband signaling (e.g., in the low SNR regime), where an increase in the SNR results in a linear increase in the throughput, obtaining more accurate estimates of the SNR can yield substantial benefits.

DOS with one-level probing

In this section, we briefly examine DOS with one-level channel probing in the low SNR regime [8]. Let M be the training length, and $\tau_t = MT_s$, where T_s is the symbol duration. We assume that $\tau_t = \Theta(1)$ as $\rho \to 0$. We further assume that the rate estimation is performed via minimum mean square error (MMSE) estimates of the channel coefficient

h_n. It follows that, $\hat{h}_n^{(1)}$, the MMSE estimate of h_n, is given by [34]:

$$\hat{h}_n^{(1)} = \frac{\sqrt{\rho}}{\rho M + 1} \sum_{m=1}^{M} Y_m, \tag{2.6}$$

Accordingly, we can express h_n in terms of $\hat{h}_n^{(1)}$ and the estimation error $\tilde{h}_n^{(1)}$ as follows:

$$h_n = \hat{h}_n^{(1)} + \tilde{h}_n^{(1)}, \tag{2.7}$$

where

$$\hat{h}_n^{(1)} \sim \mathcal{CN}\left(0, \frac{\rho M}{\rho M + 1}\right) \tag{2.8}$$

and

$$\tilde{h}_n^{(1)} \sim \mathcal{CN}\left(0, \frac{1}{\rho M + 1}\right). \tag{2.9}$$

Based on the orthogonality principle, $\hat{h}_n^{(1)}$ and $\tilde{h}_n^{(1)}$ are uncorrelated.

Without perfect CSI, the link employs the *estimated* SNR $\{\rho|\hat{h}_n^{(1)}|^2, n = 1, \ldots\}$ as the basis for distributed scheduling. However, since the channel estimation error, $\tilde{h}_n^{(1)}$, behaves as additive Gaussian noise, the *actual* instantaneous SNR of the link is given by [9, Eq.(3)]:

$$\lambda_n^{(1)} = \frac{\rho|\hat{h}_n^{(1)}|^2}{1 + \rho|\tilde{h}_n^{(1)}|^2}, \tag{2.10}$$

where the effect due to channel estimation errors is subsumed in the noise term. [4]

Inspection of (2.10) reveals that $\lambda_n^{(1)}$ is always *smaller* than the estimated SNR $\{\rho|\hat{h}_n^{(1)}|^2\}$, in the presence of channel estimation errors. As a result, an outage occurs if the link transmits at a data rate specified by $\{\rho|\hat{h}_n^{(1)}|^2\}$. To circumvent this problem, a linear backoff scheme is proposed in [8] to reduce the data rate. More specifically, the estimated SNR is linearly backed off to $\sigma_M \rho|\hat{h}_n^{(1)}|^2$, where σ_M is the backoff factor with $0 < \sigma_M < 1$. Under imperfect information, the transmission rate in the low-SNR wideband region simplifies to

$$R_n^{(1)} \approx \rho W \sigma_M |\hat{h}_n^{(1)}|^2. \tag{2.11}$$

The steps leading to the above equation are discussed in Appendix A.1. [5]

[4] For example, the maximum likelihood decoding method yields that $\hat{X} = \sqrt{\rho}\hat{h}^* Y = \rho|\hat{h}|^2 X + \rho\hat{h}^*\tilde{h}X + \sqrt{\rho}\hat{h}^*\xi$, where $\rho\hat{h}^*\tilde{h}X$, is effectively additive noise.

[5] For further discussion regarding the design of the backoff factor, σ_M, we refer the reader to [8].

For convenience, let θ be the *cost per unit system time*, where the system time encompasses the contention time, the probing time and the transmission time. It follows that a successful channel contention incurs an average cost of $\theta\tau/p_s$, whereas the data transmission for a duration of T entails a cost θT. It takes a total duration of $\sum_{j=1}^{n} K_j\tau$ to reach the n-th round of probing. After the n-th round of probing and computing its rate $R_n^{(1)}$, the successful link has the following options:

1. Transmit at rate $R_n^{(1)}$ for a time duration of $T = 1$ (the corresponding reward is $R_n^{(1)} - \theta$); or

2. Defer transmission and let all nodes re-contend (the corresponding reward is the expected return).

Note that the cost of probing, $\theta\tau/p_s$, is common to both options. Clearly, the basis for distributed opportunistic scheduling with one-level probing is the observation sequence $\{R_n^{(1)}\}_n$. Using Proposition 3.1 of [25], we can show that the optimal DOS policy with noisy channel estimation still has a threshold structure, given by

$$N^* = \min_n \{n \geq 1 : \quad R_n^{(1)} \geq \hat{\theta}\},$$

where the optimal threshold $\hat{\theta}$ is given as the solution to the following Bellman optimality equation:

$$\mathbb{E}\left[R_n^{(1)} - \theta\right]^+ = \frac{\theta\tau}{p_s}. \tag{2.12}$$

Furthermore, $\hat{\theta}$ is the corresponding throughput.

The above result reveals that the optimal stopping rule, N, is a pure threshold policy, and the stopping decision can be made based on the current rate only. Accordingly, the optimal channel probing and scheduling strategy takes the following simple form: If the successful link discovers that its current rate $R_n^{(1)}$ is higher than the threshold $\hat{\theta}$, it transmits the data with rate $R_n^{(1)}$; otherwise, it skips this transmission opportunity (e.g., by skipping the CTS), and then the links re-contend.

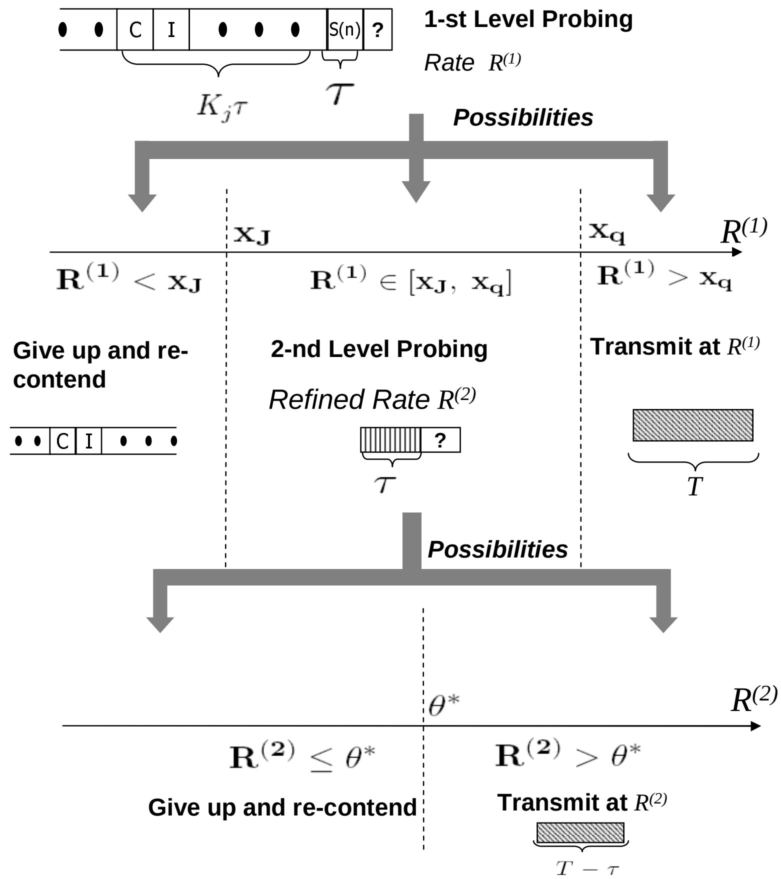

Figure 2.2: A sketch of DOS with two-level probing.

2.3 DOS with Two-Level Probing

In this section, we characterize the optimal DOS with two-level probing, i.e., the links may choose to refine their rate estimates before making a decision on whether to transmit or not. We illustrate, in Fig. 2.2, the underlying rationale behind DOS with two-level probing. In the following, we detail the procedure with second-level probing, and then cast DOS with two-level probing as a problem of maximal rate of return, using optimal stopping theory with incomplete information. We then characterize the corresponding structure and provide a complete description of the optimal strategy.

Second-level probing

To improve the estimation accuracy, the receiver of the successful link can request its transmitter to send another pilot packet, at the cost of a part of the data transmission duration allotted to it. More specifically, in addition to the pilot symbols sent during the first-level probing, the receiver obtains a refined MMSE estimate of h_n by exploiting the newly transmitted pilot symbols of length τ_t during second-level probing (of duration τ). Then, the link uses the remaining $1-\tau$ of the time for the data transmission. We let $\hat{h}_n^{(2)}$ denote this refined estimate of h_n, obtained via two-level probing. We can show that

$$\hat{h}_n^{(2)} = \frac{\sqrt{\rho}}{2\rho M + 1} \sum_{i=1}^{2M} Y_i, \tag{2.13}$$

Furthermore, the estimate $\hat{h}_n^{(2)}$, and the corresponding estimation error, $\tilde{h}_n^{(2)}$, are uncorrelated, where

$$\hat{h}_n^{(2)} \sim \mathcal{CN}\left(0, \frac{\rho 2M}{\rho 2M + 1}\right) \tag{2.14}$$

and

$$\tilde{h}_n^{(2)} \sim \mathcal{CN}\left(0, \frac{1}{\rho 2M + 1}\right). \tag{2.15}$$

Finally, the resulting data rate is computed as

$$R_n^{(2)} \approx \rho W \sigma_{2M} |\hat{h}_n^{(2)}|^2, \tag{2.16}$$

where σ_{2M} is the corresponding linear rate backoff factor.

Next, we establish the relationship between the estimates due to first-level and second-level probings. Simply put, we are interested in obtaining an estimate of h_n from $(\hat{h}_n^{(1)}, \sum_{n=M+1}^{2M} Y_n)$. Applying the Gram-Schmidt orthogonalization procedure, we can transform $(\hat{h}_n^{(1)}, \sum_{n=M+1}^{2M} Y_n)$ into orthogonal components. Then, we project h_n onto these components to represent $\hat{h}_n^{(2)}$ as (see [35, Ch.4, p.130] for details):

$$\hat{h}_n^{(2)} = \hat{h}_n^{(1)} + h_e, \tag{2.17}$$

where $h_e \sim \mathcal{CN}(0, \sigma_e^2)$, with $\sigma_e^2 = \frac{M\rho}{(M\rho+1)(2M\rho+1)}$. Note that $\hat{h}_n^{(1)}$ and h_e are orthogonal. By orthogonality, we have

$$\mathbb{E}[|\hat{h}_n^{(2)}|^2] = \mathbb{E}[|\hat{h}_n^{(1)}|^2] + \sigma_e^2. \tag{2.18}$$

Thus, it follows that the expected rate of the second-level probing conditioned on the rate due to first-level probing, obeys the following relationship:

$$\mathbb{E}[R_n^{(2)} | R_n^{(1)}] = c_r R_n^{(1)} + R_e,$$

where $R_e = \sigma_{2M} W \rho \sigma_e^2$, and $c_r = \frac{\sigma_{2M}}{\sigma_M}$. We note that R_e can be interpreted as the expected relative rate gain due to the second level probing.

Scheduling options and rewards

In what follows, we devise DOS with two levels of probing using optimal stopping theory. Drawing on the ideas from [28], we show that optimizing the network throughput via DOS can be cast as a *maximal rate of return* problem.

Consider the example in Fig. 2.1. It takes a total duration of $\sum_{j=1}^{n} K_j \tau$ to reach the n-th round of probing. After the n-th round of probing, the successful link has the following three options after computing its rate $R_n^{(1)}$:

1. Transmit at rate $R_n^{(1)}$ for a time duration of $T = 1$;

2. Defer transmission and let all nodes re-contend; or

3. Perform second-level probing to obtain the new rate $R_n^{(2)}$, and then decide whether to transmit at $R_n^{(2)}$ for a time duration of $1 - \tau$, or to defer and re-contend.

Clearly, the basis for distributed opportunistic scheduling with two-level probing is the observation sequence $\{R_n^{(1)}, R_n^{(2)}\}_n$, with the option of skipping $R_n^{(2)}$. *We emphasize that the transmission duration after second-level probing reduces to $1 - \tau$, in contrast to the duration of one after first-level probing.*

Let $\phi_n : \mathcal{R}^+ \rightarrow \{0, 1, 2\}$ and $\psi_n : \mathcal{R}^+ \rightarrow \{0, 1\}$ be the decision sequences after $R_n^{(1)} = x$ is observed. In particular, $\phi_n(x) = 1$ refers to transmitting at the current rate, $\phi_n(x) = 0$ means giving up the transmission and re-contending, while $\phi_n(x) = 2$ indicates engaging in the second-level probing. Furthermore, when $\phi_n(x) = 2$, the final decision hinges on $R_n^{(2)} = y$: if $\psi_n(y) = 1$, the link transmits at the refined rate, whereas if $\psi_n(y) = 0$, the link gives up the transmission and lets all nodes re-contend.

Let N be a stopping rule such that $\{N = n\} \in \mathcal{F}_n$, where $\mathcal{F}_n$ is the σ-algebra generated by $\{R_j^{(1)}, R_j^{(2)}\}_{j \le n}$. Stopping rule N is given by

$$N = \inf\{n \ge 1 | \phi_n = 1, \text{ or } \phi_n = 2 \text{ and } \psi_n = 1\}.$$

Let T_n be the total time, given by

$$T_n = \sum_{j=1}^{n} K_j \tau + 1,$$

which is the sum of total contention time (and time due to second-level probing, when performed) and the data transmission duration, $T_{d,n} = 1 - \mathbf{I}(\phi_n = 2)\mathbf{I}(\psi_n = 1)\tau$ with $\mathbf{I}(\cdot)$ being the indicator function.

Let θ be the cost per unit system time. Then successful contention, with first-level probing, incurs an average cost of $\theta\tau/p_s$. Second-level probing incurs a further cost of $\theta\tau$, whereas the data transmission for a duration of T_d entails a cost θT_d.

The expected net reward (*expected return*) is given by

$$r = \mathbb{E}\left[R_N T_{d,N} - \theta T_N\right],$$

where R_n is the transmission rate after the n-th probing round and is given by

$$R_n = \mathbf{I}(\phi_n = 1) \cdot R_n^{(1)} + \mathbf{I}(\phi_n = 2)\mathbf{I}(\psi_n = 1) \cdot R_n^{(2)}.$$

The corresponding *rate of return* is $\mathbb{E}[R_N T_{d,N}]/\mathbb{E}[T_N]$. The maximal expected return is given by

$$r_0 = \sup_{N \in Q} \mathbb{E}\left[R_N T_{d,N} - \theta T_N\right].$$

Note that the expected return, r, depends on the decision functions ϕ, ψ, and the cost θ. The principal objective is to maximize the rate of return (i.e., the throughput) of the DOS with two-level probing, defined as

$$\theta^* = \sup_{N \in Q} \frac{\mathbb{E}\left[R_N T_{d,N}\right]}{\mathbb{E}\left[T_N\right]}.$$

Summarizing, we are interested in seeking a stopping rule $N \in Q$ that obtains θ^*. The following lemma relates the optimal throughput θ^* to the expected optimal return r_0, and guarantees the existence of such an optimal stopping rule.

Lemma 2.1. *For DOS with two-level probing, the optimal stopping rule N^* exists. Furthermore, θ^* is attained at N^*, and θ^* satisfies*

$$r_0 = \sup_{N \in Q} \mathbb{E}\left[R_N T_{d,N} - \theta^* T_N\right] = 0.$$

Proof. See Appendix A.2. □

Next, we derive the optimality equation for DOS with two-level probing.

We begin by considering the option of second-level probing and introducing its associated reward function. Suppose after observing $R_n^{(1)} = x$, the link performs a second-level probing to obtain $R_n^{(2)}$, and then uses an optimal strategy thereafter. Then, with $R_n^{(2)} = y$, it may choose to transmit at rate y, for a duration of $1-\tau$; or it could defer and re-contend. Note that the reward associated with the transmission is $(y-\theta)(1-\tau)$, and the reward associated with forgoing the transmission is the expected return, r. Therefore, the link engages in a transmission if $(y-\theta)(1-\tau) > r$, and defers its transmission if $(y-\theta)(1-\tau) \leq r$. The expected net reward corresponding to the second-level probing is thus given by

$$J_\theta(x, r) \triangleq rG\left(\frac{r}{1-\tau} + \theta | x\right) + (1-\tau)\int_{\frac{r}{1-\tau}+\theta}^{\infty} (y-\theta)G(dy|x) - \theta\tau, \tag{2.19}$$

where $G(y|x)$ is the conditional cumulative distribution function (cdf) of $R_n^{(2)}$ given $R_n^{(1)} = x$. Note that $G(y|x)$ is a non-central χ^2 distribution with two degrees of freedom. Furthermore, both $R_n^{(1)}$ and $R_n^{(2)}$ are exponentially distributed. We use F to denote the cdf of $R_n^{(1)}$. Finally, it can be shown that $\lim_{x\to\infty} G(y|x) = 0$ and $\mathbb{E}\left[y|x\right] = c_r x + R_e$.

Upon observing $R_n^{(1)}$ after the n-th probing round, the link $s(n)$ can obtain one of the following three rewards:

1. $R_n^{(1)} - \theta$: the reward obtained by transmitting at a rate $R_n^{(1)}$;

2. r_0: the reward obtained by forgoing the current opportunity and re-contending (the maximal expected return); or

3. $J_\theta(R_n^{(1)}, r_0)$: the reward obtained by resorting to refining the rate via second-level probing.

The optimal strategy for the link is to choose the option that yields the maximum of the above rewards. Therefore, the optimality equation of DOS with two-level probing can be represented by the following Bellman optimality equation:

$$\mathbb{E}\left[\max\left\{R^{(1)} - \theta, r_0, J_\theta(R^{(1)}, r_0)\right\}\right] - \frac{\theta\tau}{p_s} = r_0, \tag{2.20}$$

where $R^{(1)}$ has same distribution as $R_n^{(1)}$. Note that, in the discussions above, we have factored out the cost for obtaining the first successful channel probing, i.e. $\theta\tau/p_s$, since it is common to all three returns. From Lemma 2.1, when the throughput, as a function of θ, reaches its maximum, we have that $r_0 = 0$ at $\theta = \theta^*$. Thus, (2.20) can be rewritten as

$$\mathbb{E}\left[\max\left\{R^{(1)} - \theta^*, J_{\theta^*}(R^{(1)}, 0)\right\}\right]^+ = \frac{\theta^*\tau}{p_s}. \tag{2.21}$$

Inspection of (2.21) indicates that the second-level probing is optimal when $J_{\theta^*}(x, 0) > 0$ and $J_{\theta^*}(x, 0) > x - \theta^*$ for some x.

It is worth noting that

$$\theta^* > \theta_L \triangleq \frac{\mathbb{E}[R^{(1)}]}{\frac{\tau}{p_s} + 1}. \tag{2.22}$$

Note that θ_L corresponds to the throughput of *PHY-oblivious scheduling*, which is a single-level probing scheme with zero threshold. This can be achieved by the degenerate stopping rule, which stops at the very first time.

Structure of optimal scheduling strategy

We next proceed to study the structure of the optimal scheduling strategy. Essentially, the optimal strategy takes a threshold form. Depending on the specific network setting, the optimal strategy may admit one of the two intuitively reasonable types, which we will call Strategy A and Strategy B. Generally speaking, under Strategy A, it is always optimal to demand additional information when the estimated rate lies between two thresholds. This is the case when the gain due to second-level probing is comparable with the additional overhead. In contrast, under Strategy B, there is never a need to appeal to second-level probing. This case occurs for example, when the improvement due to the refinement is dominated by the probing overhead. An extreme example of this case is when perfect CSI is available to the transmitter.

Before we state the main result on the optimal strategy, we define $q(x) \triangleq J_{\theta^*}(x, 0) - x + \theta^*$. Intuitively speaking, $q(x)$ represents the expected gain achieved by second-level probing compared to directly transmitting at the current rate. Thus, if $q(x) > 0$, performing second-level probing is a better option than directly proceeding to data transmission. We need the following lemmas before characterizing the structure of the optimal scheduling strategy.

Lemma 2.2. *$J_{\theta^*}(x,0)$ and $q(x)$ are characterized by the following properties:*

i. $J_{\theta^*}(x,0)$ is monotonically increasing in x with $\lim_{x\to\infty} J_{\theta^*}(x,0) = \infty$ and $\lim_{x\to 0} J_{\theta^*}(x,0) < 0$ when $\frac{R_e}{\theta^*} e^{-\frac{\theta^*}{R_e}} < \frac{\tau}{1-\tau}$.

ii. For $c_r < \frac{1}{1-\tau}$, $q(x)$ is monotonically decreasing in x with $\lim_{x\to 0} q(x) > 0$ and $\lim_{x\to\infty} q(x) = -\infty$.

Proof. See Appendix A.3. □

Remarks: Observe that the above conditions are stated in terms of the design variables (e.g., τ and c_r). It is clear that $R_e \leq \theta^*$, since R_e is the relative gain due to rate refinement and cannot be greater than the optimal throughput θ^*. Thus, in the extreme case, where $R_e = \theta^*$, we have the pessimistic bound $\frac{R_e}{\theta^*} e^{-\frac{\theta^*}{R_e}} < e^{-1}$, based on which it suffices to have $\tau > 1/(1+\exp(1))$ to guarantee that Condition i) holds. We, however, caution that $\tau > 1/(1+\exp(1))$ is only a sufficient condition. Also, it is easy to satisfy the condition in ii) by choosing $c_r \leq 1/(1-\tau) - \delta$, where $\delta > 0$.

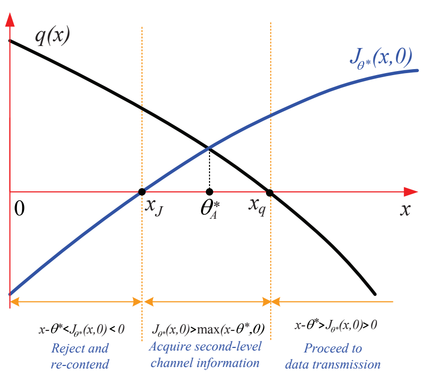

Figure 2.3: A structural sketch for Strategy A.

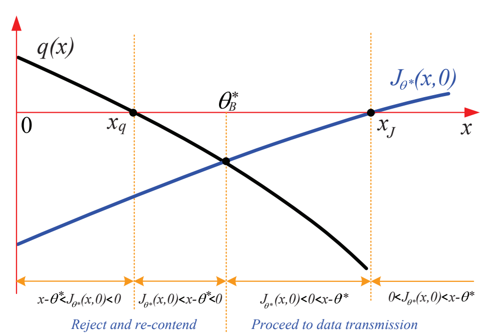

Figure 2.4: A structural sketch for Strategy B.

Lemma 2.3. *There exists at most one solution, in terms of* $\{x_J, x_q, \theta^*\}$*, to the following system of equations:*

$$\begin{cases} \int_{\theta^*}^{\infty}(1 - G(u|x_J))du = \frac{\theta^*\tau}{1-\tau}, \\ (c_r(1-\tau) - 1)x_q + (1-\tau)\left(R_e + \int_0^{\theta^*} G(u|x_q)du\right) = 0, \\ \int_{x_J}^{x_q} J_{\theta^*}(u,0)\, dF(u) + \int_{x_q}^{\infty}(u - \theta^*)\, dF(u) = \frac{\theta^*\tau}{p_s}. \end{cases} \tag{2.23}$$

Recall that x_J and x_q are the solutions to $J_{\theta^*}(x,0) = 0$ and $q(x) = 0$, respectively. From Lemma 2.2, it is easy to see that there is at most one pair $\{x_J, x_q\}$ satisfying (2.23). Similarly, since $J_{\theta^*}(x,0)$ and $q(x)$ intercept at $x = \theta^*$, there exists at most one θ^* due to the monotonic nature of $J_{\theta^*}(x,0)$ and $q(x)$.

For convenience, let $\{x_J, x_q, \theta_A^*\}$ denote the solution to (2.23) with $x_J \le x_q$, and θ_B^* be the solution to (2.12). Using the above lemmas, we obtain the following result on the structure of optimal scheduling strategy.

Theorem 2.1. *The optimal strategy for DOS with two-level probing, takes one of the two forms:*

[Strategy A] *It is optimal for the successful link*

i. to transmit immediately after the first-level probing if $R_n^{(1)} > x_q$*; or*

ii. to give up the transmission and let all links re-contend if $R_n^{(1)} < x_J$*; or*

iii. to engage in second-level probing if $R_n^{(1)} \in [x_J, x_q]$*; upon computing the new rate* $R_n^{(2)}$*, to transmit at rate* $R_n^{(2)}$ *if* $R_n^{(2)} > \theta_A^*$*, or to give up the transmission otherwise.*

Furthermore, the throughput under Strategy A is θ_A^**.*

[Strategy B] *There is never a need to perform second-level probing. That is, it is optimal for the successful link to transmit at the current rate* $R_n^{(1)}$ *if* $R_n^{(1)} > \theta_B^*$*, or to defer its transmission and re-contend otherwise. Furthermore, the throughput under Strategy B is* θ_B^**.*

Proof. See Appendix A.4. □

Optimality conditions

In previous sections, we have studied DOS with two-level probing within the OST framework, and characterized the structure of optimal scheduling strategies. Our findings reveal that optimal scheduling may take either of two forms: Strategy A or Strategy B. The next key step is to determine the conditions under which it is optimal to use Strategy A or Strategy B. We show that this can be easily determined by performing a threshold test on the function $J_{\theta^*}(\cdot,\cdot)$. We have the following theorem.

Theorem 2.2. *Strategy A is optimal if* $J_{\theta_A^*}(\theta_A^*, 0) \geq 0$*; else, Strategy B is optimal.*

Proof. See Appendix A.5. □

Numerical results

In this section, we provide a numerical example to illustrate the effectiveness of the proposed DOS with two-level probing under noisy estimation. Specifically, we compare the performance of the proposed *DOS with two-level probing*, with that of *DOS with one-level probing* and *PHY-oblivious scheduling*. The baseline for comparison is the PHY-oblivious scheduling that does not make use of any link-state information. We focus on the *relative gain* over PHY-oblivious scheduling, which is a function of ρM, and is defined as

$$\Gamma(\rho M) = \frac{\theta - \theta_L}{\theta_L}.$$

We set $p_s = \exp(-1)$, $M = 300$ and $W = 3000$, so that $\tau_t = 0.1$ and $\tau = 0.2$. Fig. 2.5 depicts the performance comparison. It is clear that the relative gain achieved by DOS with two-level probing substantially outperforms that obtained by DOS with one-level probing. Observe that the performance gain is significant in the low SNR regime (i.e., smaller values of α). As α increases, the relative gain of DOS with two-level probing approaches that of DOS with one-level probing, and our intuition is that, for higher values of α, the cost of overhead offsets that of the rate gain due to additional probing. Accordingly, the "gray area" between two thresholds (x_h and x_q) collapses, and the optimal strategy degenerates to Strategy B, which is essentially DOS with first-level probing.

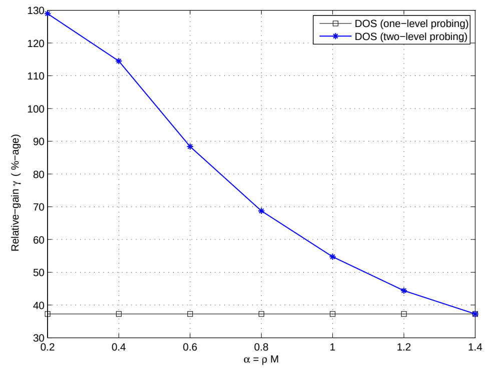

Figure 2.5: Relative gain Γ as a function of $\alpha = \rho M$.

2.4 DOS with Two-Level Probing: A Case with Limited Feedback

In the above studies, it is assumed that for the link with successful contention, its transmitter knows the rate estimate for data transmissions. In some practical scenarios, there is only limited feedback from the receiver to the transmitter. With this motivation, we extend the study of DOS with two-level probing to the case where the feedback from the receiver to its transmitter takes the form $(0, 1, e)$. More specifically, the decisions from the receiver to the transmitter are conveyed by using "NACK/ACK/ERASURE" signaling, where "NACK" is represented by "0" corresponding to the decision of *defer and re-contend*, "ACK" by "1" corresponding to the decision of *transmit*, and "ERASURE" by "e" indicating that the rate estimate falls in the gray area.

One-level probing

We first consider DOS with one-level probing, with one-bit feedback from the receiver to its transmitter. The basic idea is as follows. A constant transmission rate, denoted as R_1, is pre-determined and known to the transmitter, and the data transmission takes place only when the one-bit feedback is "1". A central problem here is to design the transmission strategy for maximal throughput. Let γ be the price function per unit time. Then, given its current rate estimate $R_n^{(1)}$, the successful link in the the $n-$th probing has two options:

- "1"— transmit at rate R_1, and the corresponding reward is $R_1\mathbf{I}(R_n^{(1)} > R_1) - \gamma$; or

- "0"— defer and re-contend, with the expected reward of r_0.

Clearly, there is an average cost of $\gamma\tau/p_s$ for every successful contention.

Let $\hat{\gamma} \triangleq \sup_{N \in Q} \mathbb{E}[R_N T]/\mathbb{E}[T_N]$ be the optimal throughput. Then, based on Lemma 2.1, the optimality equation is given by

$$\mathbb{E}\left[R_1 \mathbf{I}(R^{(1)} > R_1) - \hat{\gamma}\right]^+ = \frac{\hat{\gamma}\tau}{p_s}. \tag{2.24}$$

As a result, we can show that the optimal policy in this case still has a threshold structure with R_1 being the threshold. Furthermore, noting that $R_n^{(1)} \sim \exp(\mathbb{E}[R^{(1)}])$, we conclude that the average throughput is given as

$$\hat{\gamma} = \frac{R_1 e^{-\frac{R_1}{\mathbb{E}[R^{(1)}]}}}{\frac{\tau}{p_s} + e^{-\frac{R_1}{\mathbb{E}[R^{(1)}]}}}.$$

Observe that $\hat{\gamma}$ is a function of R_1. For a given stopping rule, R_1 can be chosen to maximize the throughput, i.e., the optimal transmission rate $\hat{R}_1$ and the corresponding throughput obey

$$\hat{R}_1 = \arg\max_{R_1} \hat{\gamma} \text{ and } \hat{\gamma}_{max} = \hat{\gamma}(\hat{R}_1).$$

It can be shown that $\hat{R}_1$ is the solution to

$$\left(\frac{R_1}{\mathbb{E}[R^{(1)}]} - 1\right) e^{\frac{R_1}{\mathbb{E}[R^{(1)}]}} = \frac{p_s}{\tau}. \tag{2.25}$$

It follows that the optimal throughput is given by

$$\hat{\gamma}_{max} = \hat{R}_1 - \mathbb{E}[R^{(1)}] = \frac{p_s}{\tau} \mathbb{E}[R^{(1)}] e^{\frac{-\hat{R}_1}{\mathbb{E}[R^{(1)}]}}. \tag{2.26}$$

Two-level probing

Next, we study DOS with two-level probing, with the feedback taking the form of $(0, 1, e)$. Along the same line as in the studies in Section 2.3, the receiver of the successful link, depending on its rate estimate $R_n^{(1)}$, presents three options to its transmitter:

- "1"— transmit at the rate R_1; or

- "0"— defer and re-contend; or

- "e"— perform a second-level probing to obtain $R_n^{(2)}$, and then decide:

– "1"— to transmit at rate R_1; or

– "0"— to defer and re-contend.

Define $\gamma^* = \sup_{N \in Q} \mathbb{E}[R_N T_{d,N}]/\mathbb{E}[T_N]$, which represents the optimal throughput for the given R_1. By Theorem 2.1, this corresponds to $r_0 = 0$. Since, γ^* is the function of the rate R_1, we further maximize the throughput over all choices of R_1, by defining $\gamma^*_{max} = \max_{R_1} \gamma^*$.

We can write the expected net reward function corresponding to the second-level probing as

$$V_{\gamma^*}(x, R_1) = (1-\tau)(R_1 - \gamma^*) \int_{R_1}^{\infty} G(dy|x) - \gamma^* \tau,$$

which can be further simplified as

$$V_{\gamma^*}(x, R_1) \quad = \quad (1-\tau)(R_1 - \gamma^*)\left(1 - G\left(R_1|x\right)\right) - \gamma^* \tau.$$

The optimality equation in this case is given by

$$\mathbb{E}\left[\max\left\{R_1 I\left(R^{(1)} \geq R_1\right) - \gamma^*, V_{\gamma^*}(R^{(1)}, R_1)\right\}\right]^+ = \gamma^* \frac{\tau}{p_s}. \qquad (2.27)$$

The following lemma gives useful bounds on the optimal throughput.

Lemma 2.4. For a given transmission rate R_1, the optimal throughput satisfies the inequalities

$$\gamma_L \leq \gamma^* \leq \gamma_U,$$

where

$$\gamma_L \triangleq \frac{(1-\tau)R_1}{(1-\tau) + \tau\left(1 + \frac{1}{p_s}\right) e^{\frac{R_1}{\mathbb{E}[R^{(2)}]}}}; \;\; \gamma_U \triangleq \frac{R_1}{1 + \frac{\tau}{p_s}}.$$

Remarks:

a) The lower bound γ_L is obtained by using a strategy where the successful link always performs a second level probing, and then decides to transmit for a duration of $1-\tau$ or to re-contend.

b) The upper bound γ_U is achieved by a genie aided scheme, where the successful link contends only when its channel is good and there is no transmission outage.

Next we turn our attention to structural properties of the optimal strategy. For convenience, define the relative gain function as

$$q_{\gamma^*}(x, R_1) \triangleq V_{\gamma^*}(x, R_1) - R_1 \mathbf{I}(x \geq R_1) + \gamma^*.$$

Lemma 2.5. *$V_{\gamma^*}(x,R_1)$ and $q_{\gamma^*}(x,R_1)$ are characterized by the following properties:*

i. $V_{\gamma^*}(x,R_1)$ is monotonically increasing in x. Furthermore, $\lim_{x\to\infty} V_{\gamma^*}(x,R_1) = c_1 > 0$, if $\tau \le 1-p_s$, and $\lim_{x\to 0} V_{\gamma^*}(x,R_1) < 0$, when $\tau \ge \frac{1}{2}(\ln(1+\frac{1}{p_s})-1)$.

ii. $q_{\gamma^*}(x,R_1) \ge 0$ for $x < R_1$; and $q_{\gamma^*}(x,R_1) < 0$ for $x \ge R_1$.

Proof. See Appendix A.6. □

The above lemma serves as the basis to determine the optimal DOS scheduling under the feedback of $(0,1,e)$. Specifically, from the properties of $V_{\gamma^*}(\cdot,R_1)$, there exists an x_v such that

$$V_{\gamma^*}(x,R_1) \ge 0,\ \forall x \ge x_v, \tag{2.28}$$

which, in turn, gives a threshold below which the option of "defer and re-contend" is optimal. From the properties of $q_{\gamma^*}(\cdot,R_1)$, it is also clear that for all $x \ge R_1$, it is optimal to transmit immediately without a second-level probing. Therefore, the interval $[x_v,R_1]$ defines the gray area where one could benefit from performing a second-level probing.

We note that the throughput, denoted by γ^*, is the parameter to be optimized over the thresholds x_v and R_1. Combining (2.27) and (2.28), we establish

$$\gamma^* = \frac{(1-\tau)R_1\int_{x_v}^{R_1}(1-G(R_1|u))dF(u) + R_1 e^{-\frac{R_1}{\mathbb{E}[R^{(1)}]}}}{(1-\tau)\left(\int_{x_v}^{R_1}(1-G(R_1|u))dF(u) + e^{-\frac{R_1}{\mathbb{E}[R^{(1)}]}}\right) + \tau(\frac{1}{p_s} + e^{-\frac{x_v}{\mathbb{E}[R^{(1)}]}})} \tag{2.29}$$

$$\gamma^*\tau = (1-\tau)(R_1-\gamma^*)\left(1-G\left(R_1|x_v\right)\right). \tag{2.30}$$

which relate three key parameters, namely the lower threshold x_v, the transmission rate R_1, and the throughput γ^*. It can be seen from (2.29) and (2.30) that $x_v = f_1(R_1,\gamma^*)$ and $\gamma^* = f_2(x_v,R_1)$, indicating that $\gamma^* = g(R_1)$. Then, R_1 can be chosen to maximize $g(R_1)$; i.e.,

$$R_1^* = \arg\max_{R_1} g(R_1);\ \text{and}\ \gamma^*_{max} = g(R_1^*).$$

Accordingly, the optimal x_v^* is given by

$$x_v^* = f_1(R_1^*,\gamma^*_{max}).$$

Let $\{x_v^*, R_1^*, \gamma_{max}^*\}$ be the set of parameters obtained as outlined above. Also, let $\hat{R}_1$ be the solution to (2.25). The optimal strategy in the case with limited feedback is given by the following result.

Theorem 2.3. *The optimal strategy for DOS with two-level probing, with $(0, 1, e)$ feedback, takes one of the two forms:*

[Strategy A] *It is optimal for the receiver of the successful link to*

i. feed back "1" if $R_n^{(1)} \geq R_1^*$, indicating to transmit at rate R_1^* immediately after the first-level probing; or

ii. feed back "0" if $R_n^{(1)} < x_v^*$, indicating to give up the transmission and let all links re-contend; or

iii. feed back "e" if $R_n^{(1)} \in [x_v^*, R_1^*)$, indicating to engage in second-level probing; and upon computing the new rate $R_n^{(2)}$,

a) feed back "1" if $R_n^{(2)} \geq R_1^*$, indicating to transmit at rate R_1^*; or

b) feed back "0" if $R_n^{(2)} < R_1^*$, indicating to give up the transmission and re-contend.

Furthermore, the throughput under Strategy A is γ_{max}^*.

[Strategy B] *There is never a need to perform second-level probing. That is, it is optimal for receiver of the successful link to*

i. feed back "1" if $R_n^{(1)} \geq \hat{R}_1$, indicating to transmit at rate $\hat{R}_1$; or

ii. feed back "0" if $R_n^{(1)} < \hat{R}_1$, indicating to give up the transmission and re-contend.

Furthermore, the throughput under Strategy B is $\hat{\gamma}$.

Proof. The proof follows the same line of that for Theorem 2.1. □

2.5 Conclusion

We have considered distributed opportunistic scheduling for single-hop ad-hoc networks in which many links contend for the same channel using random access. Specifically, we have investigated DOS with two-level channel probing by optimizing the tradeoff between the throughput gain from more accurate rate estimation and the corresponding probing overhead. Capitalizing on optimal stopping theory with two-level incomplete information, we have shown that the optimal scheduling policy is threshold-based and is characterized by either one or two thresholds, depending on system settings. We have also identified optimality conditions. In particular, our analysis reveals that DOS with second-level channel probing is optimal when the first-level estimated rate falls in between the two thresholds. By a numerical example, we have illustrated the effectiveness of the proposed DOS with two-level channel probing. Finally, we considered the extension of DOS with two-level probing to the case where there is limited feedback, of the form $(0, 1, e)$, from the receiver to its transmitter.

In this work, we have considered DOS with two-level probing, where it is assumed that the refinement of the rate estimate is carried out once, via second-level probing of duration τ. However, we can further extend this to $L-$level probing, where for $k = 1, \ldots, L-1$, a successfully contending transmitter has the options 1) to transmit, or 2) to defer and re-contend, or 3) to resort to $(k+1)-$st level training at the cost of additional overhead. In this situation, it is of interest to devise well-structured, yet simple policies.

We note that the proposed distributed scheduling with two-level probing provides a new framework to study joint PHY/MAC optimization in practical networks where noisy probing is often the case and imperfect information is inevitable. We believe that this line of study provides some initial steps towards opening a new avenue for exploring the intrinsic tradeoffs between probing (sensing) and scheduling to enhance spectrum utilization; and this is potentially useful for enhancing MAC protocols for wireless mesh networks and cognitive radio networks. Notably, a very recent work [36] has applied our methods [25] to study optimal selection of channel sensing order in cognitive radio networks.

Chapter 3

TWO-HOP INTERFERENCE FLOWS: A CASE OF INTERFERENCE CHANNELS WITH PARTIAL SIDE INFORMATION

3.1 Introduction

Relay channels and interference channels model two fundamental forms of networked communications. In a relay channel, as introduced by van der Meulen [15], a relay node assists the source in communicating data to a destination. In an interference channel, as introduced by Ahlswede [37], two pairs of nodes, each consisting of a source and a destination, wish to communicate simultaneously. A defining property of this channel is that each of the two destinations experiences interference, resulting from the signal transmitted to the other destination.

In this chapter, we consider a model building on these two basic models In our model, two sources wish to communicate simultaneously with two destinations, and are aided by two relay nodes. We confine our attention to a two-hop scenario, i.e., communication is performed in two consecutive "hops" occurring in two time intervals. During the first hop, the two sources communicate data to their respective relays, and during the second hop, the relays forward the data to the destinations. For ease of exposition, we focus on the model where the relays are half-duplex (i.e., the relays are not able to transmit data during the first hop), and the destinations are not able to hear the signal transmitted by the source. For example, this models a situation where the destinations are far away.

In designing communication strategies for this scenario, it is useful to begin with communication strategies for the interference channel. During the first hop in the model under consideration, the two sources simultaneously communicate their messages, each to its corresponding relay. That is, the relays function as virtual destinations, and each relay experiences interference resulting from the signal communicated to the other relay. The communication problem at hand is thus equivalent to transmission over an interference channel. The best known transmission strategy for this channel is due to Han and Kobayashi [13], known as Han-Kobayashi (HK) coding. During the second hop, each relay functions as a virtual source, and communicates its corresponding message to the final destination. While HK coding can be applied to this hop as well, the availability of side information at the relays leaves room to achieve better performance. We elaborate further on this in the following. The HK coding achieves its remarkable performance by requiring each destination (of the interference channel) to decode, beyond its own message, also a part of the message intended for the other destination. That is, each source splits its message into two sub-messages: a *common* sub-message and a *private* sub-message. Each destination decodes both of the sub-messages transmitted by its corresponding source, and also the common sub-message communicated to the other destination. In the context of the interference channel scenario considered by Han and Kobayashi, the destination has no use for the additional common sub-message it decodes, and is expected to discard it once decoding is complete. However, in the context of the two-hop scenario considered here, this additional knowledge, obtained by the relays in their roles as destinations in the first hop, can be further exploited to enable cooperation between them, in their roles as sources during the second hop. Thus, the requirement that each node decode part of the message intended for the other, provides a benefit in both hops.

How can cooperation between the relays be achieved? In the approach by Simeone *et al.* [38], each relay transmits superimposed signals, corresponding to its own two sub-messages (the private one and the common one), as well as the alternate relay's common sub-message[1]. In this chapter, we explore a different set of new approaches based on the following observations.

Consider one extreme case where, during the first hop, each of the common sub-messages constitutes the entire message transmitted by the corresponding source, and the private sub-messages degenerate to null. Accordingly, after the first hop communications, the two relays have full knowledge of *all* of the data intended for *both* destinations. They may, therefore, function as a virtual antenna array, and the communication problem (during the second hop) coincides with multi-antenna (MIMO) broadcast[2]. Co-

[1] More precisely, their model considers many sets of source-relay-destinations, and each relay is aware of *two* common sub-messages, in addition to their own.

[2] Strictly speaking, each relay is subject to an individual power-constraint, rather than a total power

operation of this form, between nodes, is typically known as distributed MIMO (see, e.g., [39, 40]). In particular, we can apply broadcast coding techniques based on the Gelfand-Pinsker (GP) binning [17] or Marton's coding technique [41] which yields the best known achievable rate region for general broadcast channels. Indeed, broadcast transmissions with dirty-paper coding (DPC) [42] (an application of Morton's coding to the Gaussian case) have recently been shown to achieve the capacity region of Gaussian multi-antenna broadcast channels [43]. Additional multi-antenna broadcast strategies are possible as well. The above case, where the common sub-messages constitute all of the data transmitted during the first hop, is one extreme. At the other extreme, the private messages constitute all of the data during the first hop, and the common messages are reduced to null. In this case, cooperation between the relays during the second hop is not possible, and the best known communication strategy is again HK coding.

A natural question to ask is what signaling strategy would work well for the continuum between the two extremes discussed above? A key feature of this network is that it has elements of both the broadcast channel and the interference channel. In light of this, we study strategies based on the nontrivial marriage of MIMO-BC techniques for cooperative relaying, rate-splitting and superposition coding, to enable interference cancelation at the receivers, and the Gelfand-Pinsker (GP) coding [17] at each transmitter to reduce the interference to its own receivers. One key contribution of our work lies in addressing the above challenge by developing strategies that are carefully crafted with a flavor of layered coding. Specifically, we propose two types of layered schemes that combine MIMO broadcast and HK coding. The first one is the *layered coding with binning* that mainly hinges on a novel interplay between binning (DPC) and HK coding. The second one is the *layered coding with superposition* strategy that involves superposition coding over different tiers.

More specifically, our relaying schemes involve each relay splitting its message into three parts:

1. Common message: This part essentially encompasses the "side-information" available

constraint as typically assumed in multi-antenna communication problems. However, in this chapter, we focus on symmetric interference channels, and use a simple time-sharing argument to circumvent this problem.

at both relays. This is used towards efficient transmission cooperation between the relays. Depending on the cooperative techniques used, each receiver may be required to decode either the full common message or only the relevant part.

2. Public message: This part is known only to the intended relay, used for canceling a part of interference at the receivers. This message needs to be decoded at both the receivers.

3. Sub-private message: This part is also known only to the intended relay, but it is required to be decoded solely at the intended receiver.

We use the term "private message" to refer to public and sub-private messages collectively. The details of these schemes will be presented in Sections 3.3 and 3.3.

In related work, some cooperative strategies for interference channels with/without relays have been devised. Work [38] focuses on linear relaying strategies (particularly, beamforming), to maximize the end-to-end rate in a symmetric two-hop network, with the optimal power allocation being performed jointly by the sources and the relays. In [39], Oyman *et al.*apply distributed MIMO techniques for dense MIMO interference networks with multiple MIMO relay nodes. They quantify the tradeoff between the power and the bandwidth efficiency, under the restriction to linear processing at the relays. Mohajer *et al.* [44] study two-hop interference networks (a.k.a. the Gaussian relay network), and establish an approximate characterization of the rate region for some specific cases, known as ZS and ZZ networks. Cao *et al.* [45] propose upper bounds on the end-to-end capacity of a symmetric two-hop interference network. In contrast to the above works, based on our initial steps on two-hop interference flows [46], the focus of this study is mainly on the second hop communication, where we explore non-linear cooperative strategies at the relays to enhance the achievable rate region.

Maric *et al.* [47] consider the case of an interference channel where the transmitters are interested in sending common information to both destinations along with their own independent messages and presented the capacity region under the assumption of strong interference. The authors [47] also examine the case where the available knowledge is asymmetric, in the sense that one transmitter has complete knowledge of the other's message, while the other transmitter knows only its own message. In contrast, in our model, the two receivers are both interested in distinct portions of the "common information" and our focus is on the case with moderate interference (as defined in [14]) in the second hop. In the work by Prabhakaran *et al.* [48], the authors consider a two-user Gaussian interference channel under full-duplex scenario, where each of the sources, apart from transmitting its own message, is also able to hear the transmissions from the other source. Under such scenario, the authors characterize the sum capacity up to a constant number of bits, and inner and outer bounds are derived. Wang *et al.* [49] study the two-user Gaussian interference channel with conferencing transmitters characterize the capacity region to within 6.5 bits/s/Hz. Interference channels with conferencing encoders have also been considered in [50, 51].

The rest of the chapter is organized as follows. In Section 3.2, we describe the system model and give the relevant background on the two-hop communications. In Section 3.3, we consider the transmissions from relay in the second hop, and present coding strategies based on layered binning and superposition coding. In Section 3.4, we illustrate our coding strategies on a symmetric Gaussian interference network. Numerical results are provided in Section 3.5, and we conclude this study in Section 3.6.

Notation: We use x^N to denote the sequence, $[x[1], \ldots, x[n]]$, $[\cdot]^\dagger$ to denote the transpose operator, $\mathbb{E}[\cdot]$ to denote the expectation operator and $C(x)$ to denote $\frac{1}{2}\log(1+x)$. $T_\epsilon^N(X)$ and $T_\epsilon^N(X|Y)$, respectively, refer to the set of N-length sequences that are ϵ-typical with respect to the distribution P_X and conditional distribution $P_{X|Y}$ [52, 53].

3.2 System Model and Background

As illustrated in Fig. 3.1, the communications in both hops are slotted and synchronized. The mode of communication is half-duplex, i.e., in the first slot, sources (S_1, S_2) transmit and relays (L_1, L_2) listen; and in the second slot, relays transmit and the mobile stations (D_1, D_2) listen. The source to relay $(S \to L)$ (relay to destination $(L \to D)$) transmission of one information flow interferes with the $S \to L$ $(L \to D)$ transmission of the other information flow. Specifically, we assume that each slot spans for a period of N channel uses, where N is chosen large enough to achieve reliable transmission. During the first slot, sources transmit and relays listen.

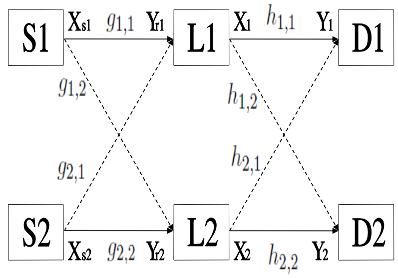

Figure 3.1: A sketch of two-hop interference flows.

For ease of exposition, we focus on the Gaussian case in this study. It is worth pointing out that our achievable schemes presented in Section 3.3 are more general, in the sense that they can be applicable to general discrete memoryless channels as well. It follows that the received signal at the relay L_1, for $n = 1, \dots, N$, is given as

$$Y_{r,1}[n] = g_{1,1}X_{s,1}[n] + g_{1,2}X_{s,2}[n] + Z_{r,1}[n], \tag{3.1}$$

where $Z_{r,1}[n] \sim \mathcal{N}(0, \sigma_r^2)$ is the AWGN process at the relay L_1. Further, $g_{1,1}$ and $g_{1,2}$, respectively, are the channel gains on the direct and interfering links for user 1. Transmitters are subject to an average power constraint, i.e., $\mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}|X_{s,i}|^2] \leq \Gamma, \quad i = 1, 2$. In the second slot, the received symbol at the destination, for $n = N+1, \dots, 2N$, is given by

$$Y_1[n] = h_{1,1}X_1[n] + h_{1,2}X_2[n] + Z_1[n], \tag{3.2}$$

where $Z_1[n] \sim \mathcal{N}(0, 1)$ is the AWGN process at the receiver D_1. Further, $h_{1,1}$ and $h_{1,2}$, respectively, are the channel gains on the direct and interfering links for user 1. A similar set-up is employed for the second flow. Relays are subject to an average power constraint, i.e., $\mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}|X_i|^2] \leq P, \quad i = 1, 2$. We assume that the transmitters as well as the receivers in both hops have the complete state information about their direct links and interfering links.

A primary goal is to characterize the achievable rate regions for the second-hop communications. In the following sections, we briefly describe the transmission schemes and coding strategies adopted by $S \to L \to D$ links in the network for the cooperative relaying and for the interference minimization.

Transmission from sources: the first-hop communications

In the first hop, the two parallel $S \to L$ channels represent the classical interference channel introduced by [37]. The capacity achieving strategy for a general interference channel is still unknown, and the best known strategy is due to Han and Kobayashi [13]. The HK coding involves splitting the transmitted information of both users into two parts: private information to be decoded only at own receiver and common information that can be decoded at both receivers. Therefore, at the end of first slot, each relay has its own message estimates and the estimate of the common message intended for the other user. Source S_i, $i = 1, 2$, draws its message $w_i \in \{1, \ldots, 2^{NR'_i}\}$ and splits it into $(\mathrm{w}_{p,i}, \mathrm{w}_{c,i})$, where $\mathrm{w}_{p,i} \in \{1, \ldots, 2^{NR'_{p,i}}\}$, $\mathrm{w}_{c,i} \in \{1, \ldots, 2^{NR'_{c,i}}\}$ and $R'_i = R'_{c,i} + R'_{p,i}$. Source S_i generates two codebooks: $\mathcal{C}_{p,i}$, with $|\mathcal{C}_{p,i}| = 2^{NR'_{p,i}}$, to carry the private message to the intended relay, and $\mathcal{C}_{c,i}$, with $|\mathcal{C}_{c,i}| = 2^{NR'_{c,i}}$, to carry the common message to be decoded by both relays. For a given block length N, source S_i chooses codewords, $C^N_{p,i}$ and $C^N_{c,i}$, from $\mathcal{C}_{p,i}$ and $\mathcal{C}_{c,i}$ respectively, and transmits their superposition: $X^N_{s,i} = C^N_{p,i} + C^N_{c,i}$. Each receiver decodes its own message along with the common message of the other user. Thus, after decoding the transmitted signals, L_1 has $(\mathrm{w}_{p,1}, \mathrm{w}_{c,1}, \mathrm{w}_{c,2})$ and L_2 has $(\mathrm{w}_{p,2}, \mathrm{w}_{c,2}, \mathrm{w}_{c,1})$. For details on the mechanism of decoding and the analysis, we refer to [13].

Observe that, at one extreme, when the interference in the first hop is very low, the rate of common messages diminishes. Thus, it follows that $\mathrm{w}_i = \mathrm{w}_{p,i}$. On the other hand, when the first hop has strong (or very strong) interference, each relay can completely decode both the intended and interfering messages [12, 54]. Therefore, we have $\mathrm{w}_i = \mathrm{w}_{c,i}$ in this case.

The relays process the data during the first slot before transmitting in the second hop. Relay L_1 has $(\mathrm{w}_1, \mathrm{w}_{c,2})$, while relay L_2 has $(\mathrm{w}_2, \mathrm{w}_{c,1})$. Observe that each relay has a piece of common information intended for the other receiver. Therefore, the second hop transmission can be viewed as an interference channel where each transmitter has partial information intended for the other receiver. An important goal of our study here is to devise relaying schemes for the second hop that effectively harness the side information available. We assume that, relay L_i, depending on its forward channel conditions, splits

its decoded message into three parts $(\mathrm{v}_{p,i}, \mathrm{v}_{c,1}, \mathrm{v}_{c,2})$, where $\mathrm{v}_{p,i}$ is the private message, known only to L_i, and $(\mathrm{v}_{c,1}, \mathrm{v}_{c,2})$ is the set of common messages, known to both the relays. Note that L_i is interested in conveying $\mathrm{v}_i = (\mathrm{v}_{p,i}, \mathrm{v}_{c,i})$ to D_i. Next, we present some definitions and background for the second hop.

Transmission from relays with partial side information: the second hop communications
The model for second hop consists of two transmitters (L_1 and L_2), and two receivers, (D_1 and D_2), and a discrete memoryless channel with input alphabets $\mathcal{X}_1, \mathcal{X}_2$ and output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$ and a conditional distribution $p(y_1, y_2|x_1, x_2)$, where $(x_1, x_2) \in (\mathcal{X}_1, \mathcal{X}_2)$ and $(y_1, y_2) \in (\mathcal{Y}_1, \mathcal{Y}_2)$.

Relay L_i, $i = 1, 2$, wishes to send the message, $\mathrm{v}_i \in \mathcal{V}_i = \{1, \ldots, 2^{NR_i}\}$, to the receiver D_i, $i = 1, 2$, over N channel uses. Let $\mathcal{V}_i = \mathcal{V}_{p,i} \times \mathcal{V}_{c,i}$, with $\mathcal{V}_{p,i} = \{1, \ldots, 2^{NR_{p,i}}\}$ and $\mathcal{V}_{c,i} = \{1, \ldots, 2^{NR_{c,i}}\}$. Message v_i is split such that $\mathrm{v}_i = (\mathrm{v}_{p,i}, \mathrm{v}_{c,i})$ with $\mathrm{v}_{p,i} \in \mathcal{V}_{p,i}$, $\mathrm{v}_{c,i} \in \mathcal{V}_{c,i}$. Observe that L_1 knows $\mathrm{v}_{c,2}$ while L_2 knows $\mathrm{v}_{c,1}$. The channel is memoryless; that is

$$p(y_1^N, y_2^N|x_1^N, x_2^N) = \prod_{n=1}^{N} p(y_1[n], y_2[n]|x_1[n], x_2[n]).$$

A $(2^{NR_1}, 2^{NR_2}, N, P_e^{(N)})$ code for the channel has two encoding functions

$$f_1 : \mathcal{V}_1 \times \mathcal{V}_{c,2} \to \mathcal{X}_1^N,\ f_2 : \mathcal{V}_2 \times \mathcal{V}_{c,1} \to \mathcal{X}_2^N$$

and two decoding functions $g_i : \mathcal{Y}_i^N \to \mathcal{V}_i,\quad i = 1, 2$, and an error probability $P_e^{(N)} = \max\{P_{e,1}^{(N)}, P_{e,2}^{(N)}\}$, where, for $i = 1, 2$, we have

$$P_{e,i}^{(N)} = \frac{1}{2^{N(R_1+R_2)}} \sum_{\mathrm{v}_1, \mathrm{v}_2} P(g_i(Y_i^N) \neq \mathrm{v}_i|\mathrm{v}_1, \mathrm{v}_2),\ i = 1, 2.$$

A rate pair (R_1, R_2) is achievable if, for any $\epsilon > 0$, there is an $(2^{NR_1}, 2^{NR_2}, N, P_e^{(N)})$ code, for N sufficiently large, such that $P_e^{(N)} < \epsilon$.

Our focus is on the moderate interference regime. That is, for all $(X_1, X_2) \sim p(x_1)p(x_2)$, we assume that

$$I(X_1; Y_1|X_2) > I(X_1; Y_2|X_2) \text{ and } I(X_2; Y_2|X_1) > I(X_2; Y_1|X_1).$$

3.3 Layered Coding for Second-Hop Communications: IC with Partial Side-Information

In this section, we study some strategies based on iterative binning and superposition coding. Before proceeding further with achieveability strategies, we briefly comment on the encoding procedure employed by the relays for the second-hop transmission. Throughout, we assume that the relay $L_i,\ i = 1, 2$ splits its own private message $\mathrm{v}_{p,i} \in \mathcal{V}_{p,i}$ further into two parts as $\mathrm{v}_{p,i} = (\mathrm{v}_{pp,i}, \mathrm{v}_{pc,i})$ such that $\mathrm{v}_{pp,i} \in \mathcal{V}_{pp,i} = \{1, \ldots, 2^{NR_{pp,i}}\}$, $\mathrm{v}_{pc,i} \in \mathcal{V}_{pc,i} = \{1, \ldots, 2^{NR_{pc,i}}\}$, with $\mathcal{V}_{p,i} = \mathcal{V}_{pp,i} \times \mathcal{V}_{pc,i}$ and $R_{p,i} = R_{pp,i} + R_{pc,i}$. We refer to $\mathrm{v}_{pp,i}$ as the sub-private message (which is decoded only by the intended receiver) and $\mathrm{v}_{pc,i}$ is the public message (which is decoded by both receivers). Thus each relay encodes the message tuple $(\mathrm{v}_{pp,i}, \mathrm{v}_{pc,i}, \mathrm{v}_{c,1}, \mathrm{v}_{c,2})$ and transmits the corresponding codewords.

Layered binning for interference cancelation

We start with the application of binning strategies to mitigate interference by exploiting knowledge available at the transmitters. One key observation is that, when the focus is just on the transmission of common information, the transmitters can form a virtual two-antenna transmitter and use the Gelfand-Pinsker (GP) binning/Marton's coding (DPC for Gaussian channels) which is known to be the best strategy in such cases. On the other hand, when all the information available at the relays is private, it is beneficial to perform HK type coding at the transmitters. Clearly, the idea is to exploit the capabilities inherent in both HK coding and binning techniques. In light of these observations, we propose "layered binning" as a novel combination of GP binning and HK coding.

Simply put, layered binning consists of a combination of binning and superposition coding over different tiers, with the following steps:

i. The common message codebooks are binned against each other in line with Morton's coding.

ii. Public messages are encoded such that they are decoded by both the receivers, to enable interference cancelation (in line with HK coding).

iii. Observe that the interfering part of common messages is still interference at each receiver. However, this is the "side-information" known at each relay. Each relay bins its own sub-private message codebook against the codebook of the interfering common message.

A superposition of sub-private message over public message, along with the common message is transmitted. Observe that clubbing the transmission of sub-private message with the common messages is related to the problem of broadcast channel with random parameters [55, 56]. We have the following result.

Theorem 3.1. *The rate pair* $(R_1, R_2) = (R_{p,1}+R_{c,1}, R_{p,2}+R_{c,2})$ *is achievable if, for some joint distribution such that*

$$p(x_{pc1})p(x_{pc2})p(\mathbf{u}_{c1}, \mathbf{u}_{c2})p(x_1, u_{pp,1}|\mathbf{u}_{c,1}, \mathbf{u}_{c,2}, x_{pc,1})p(x_2, u_{pp,2}|\mathbf{u}_{c,1}, \mathbf{u}_{c,2}, x_{pc,2}),$$

the rate tuple $(R_{p,1}, R_{c,1}, R_{p,2}, R_{c,2})$ *satisfies the following three sets of constraints, denoted* $\mathcal{B}_1$*,* $\mathcal{B}_2$ *and* $\mathcal{B}_3$*:*

Constraint set $\mathcal{B}_1$*: (constraints at the transmitters)*

$$r_{c1} + r_{c2} \geq I(\mathbf{U}_{c1}; \mathbf{U}_{c2}) \quad (3.3)$$

$$r_{p1} \geq I(U_{pp1}; \mathbf{U}_{c2}|\mathbf{U}_{c1}, X_{pc1}) \quad (3.4)$$

$$r_{p2} \geq I(U_{pp2}; \mathbf{U}_{c1}|\mathbf{U}_{c2}, X_{pc1}) \quad (3.5)$$

$$R_{c,1} \leq R'_{c,1} \quad (3.6)$$

$$R_{c,2} \leq R'_{c,2}; \quad (3.7)$$

Constraint set $\mathcal{B}_2$***:***

$$\begin{aligned}
R_{p1} &\leq I(Y_1; X_{pc1}, U_{pp,1}|\mathbf{U}_{c1}, X_{pc2}) - r_{p1}\\
R_{p1} + R_{c1} &\leq I(Y_1; X_{pc1}, \mathbf{U}_{c1}, U_{pp,1}|X_{pc2}) - r_{c1} - r_{p1}\\
R_{p1} + R_{p2} &\leq \min\{\\
&\quad I(Y_1; X_{pc2}, U_{pp,1}|X_{pc1}, \mathbf{U}_{c1}) - r_{p1} + I(Y_2; X_{pc1}, U_{pp,2}|X_{pc2}, \mathbf{U}_{c2}) - r_{p2},\\
&\quad I(Y_1; U_{pp,1}|X_{pc1}, \mathbf{U}_{c1}, X_{pc2}) - r_{p1} + I(Y_2; X_{pc1}, X_{pc2}, U_{pp,2}|\mathbf{U}_{c2}) - r_{p2},\\
&\quad I(Y_1; X_{pc1}, X_{pc2}, U_{pp,1}|\mathbf{U}_{c1}) - r_{p1} + I(Y_2; U_{pp,2}|X_{pc2}, \mathbf{U}_{c2}, X_{pc1}) - r_{p2}\\
&\quad \}
\end{aligned}$$

$$
\begin{aligned}
2R_{p1}+R_{p2} &\leq I(Y_1;X_{pc1},X_{pc2},U_{pp,1}|\mathbf{U}_{c1})-r_{p1}+I(Y_1;U_{pp,1}|X_{pc1},\mathbf{U}_{c1},X_{pc2})-r_{p1}\\
&\quad +I(Y_2;X_{pc1},U_{pp,2}|X_{pc2},\mathbf{U}_{c2})-r_{p2}\\
R_{p1}+R_{p2}+R_{c1} &\leq \min\{\\
&\quad I(Y_2;X_{pc1},X_{pc2},U_{pp,2}|\mathbf{U}_{c2})-r_{p2}+I(Y_1;\mathbf{U}_{c1},U_{pp,1}|X_{pc1},X_{pc2})-r_{c1}-r_{p1},\\
&\quad I(Y_2;U_{pp,2}|X_{pc2},\mathbf{U}_{c2},X_{pc1})-r_{p2}+I(Y_1;X_{pc1},X_{pc2},\mathbf{U}_{c1},U_{pp,1})-r_{c1}-r_{p1},\\
&\quad I(Y_2;X_{pc1},U_{pp,2}|X_{pc2},\mathbf{U}_{c2})-r_{p2}+I(Y_1;X_{pc2},\mathbf{U}_{c1},U_{pp,1}|X_{pc1})-r_{c1}-r_{p1}\\
&\quad \}\\
R_{p1}+2R_{p2}+R_{c1} &\leq I(Y_1;X_{pc2},\mathbf{U}_{c1},U_{pp,1}|X_{pc1})-r_{c1}-r_{p1}+I(Y_2;X_{pc1},X_{pc2},U_{pp,2}|\mathbf{U}_{c2})-r_{p2}\\
&\quad +I(Y_2;U_{pp,2}|X_{pc2},\mathbf{U}_{c2},X_{pc1})-r_{p2}\\
2R_{p1}+R_{p2}+R_{c1} &\leq \min\{\\
&\quad I(Y_1;X_{pc1},X_{pc2},\mathbf{U}_{c1},U_{pp,1})-r_{c1}-r_{p1}+I(Y_1;U_{pp,1}|X_{pc1},\mathbf{U}_{c1},X_{pc2})-r_{p1}\\
&\quad +I(Y_2;X_{pc1},U_{pp,2}|X_{pc2},\mathbf{U}_{c2})-r_{p2},\\
&\quad I(Y_1;X_{pc1},X_{pc2},U_{pp,1}|\mathbf{U}_{c1})-r_{p1}+I(Y_1;\mathbf{U}_{c1},U_{pp,1}|X_{pc1},X_{pc2})-r_{c1}-r_{p1}\\
&\quad +I(Y_2;X_{pc1},U_{pp,2}|X_{pc2},\mathbf{U}_{c2})-r_{p2}\\
&\quad \}\\
2R_{p1}+R_{p2}+2R_{c1} &\leq I(Y_2;X_{pc1},U_{pp,2}|X_{pc2},\mathbf{U}_{c2})-r_{p2}+I(Y_1;X_{pc1},X_{pc2},\mathbf{U}_{c1},U_{pp,1})-r_{c1}-r_{p1}\\
&\quad +I(Y_1;\mathbf{U}_{c1},U_{pp,1}|X_{pc1},X_{pc2})-r_{c1}-r_{p1}\\
R_{p1}+R_{p2}+R_{c1}+R_{c2} &\leq \min\{\\
&\quad I(Y_1;X_{pc1},X_{pc2},\mathbf{U}_{c1},U_{pp,1})-r_{c1}-r_{p1}+I(Y_2;\mathbf{U}_{c2},U_{pp,2}|X_{pc1},X_{pc2})-r_{c2}-r_{p2},\\
&\quad I(Y_2;X_{pc1},X_{pc2},\mathbf{U}_{c2},U_{pp,2})-r_{c2}-r_{p2}+I(Y_1;\mathbf{U}_{c1},U_{pp,1}|X_{pc1},X_{pc2})-r_{c1}-r_{p1},\\
&\quad I(Y_1;X_{pc2},\mathbf{U}_{c1},U_{pp,1}|X_{pc1})-r_{c1}-r_{p1}+I(Y_2;X_{pc1},\mathbf{U}_{c2},U_{pp,2}|X_{pc2})-r_{c2}-r_{p2}\\
&\quad \}\\
2R_{p1}+R_{p2}+R_{c1}+R_{c2} &\leq \min\{\\
&\quad I(Y_1;X_{pc1},X_{pc2},\mathbf{U}_{c1},U_{pp,1})-r_{c1}-r_{p1}+I(Y_1;U_{pp,1}|X_{pc1},\mathbf{U}_{c1},X_{pc2})-r_{p1}\\
&\quad +I(Y_2;X_{pc1},\mathbf{U}_{c2},U_{pp,2}|X_{pc2})-r_{c2}-r_{p2},\\
&\quad I(Y_1;X_{pc1},X_{pc2},U_{pp,1}|\mathbf{U}_{c1})-r_{p1}+I(Y_1;\mathbf{U}_{c1},U_{pp,1}|X_{pc1},X_{pc2})-r_{c1}-r_{p1}\\
&\quad +I(Y_2;X_{pc1},\mathbf{U}_{c2},U_{pp,2}|X_{pc2})-r_{c2}-r_{p2}\\
&\quad \}\\
2R_{p1}+R_{p2}+2R_{c1}+R_{c2} &\leq I(Y_2;X_{pc1},\mathbf{U}_{c2},U_{pp,2}|X_{pc2})-r_{c2}-r_{p2}\\
&\quad +I(Y_1;X_{pc1},X_{pc2},\mathbf{U}_{c1},U_{pp,1})-r_{c1}-r_{p1}\\
&\quad +I(Y_1;\mathbf{U}_{c1},U_{pp,1}|X_{pc1},X_{pc2})-r_{c1}-r_{p1};
\end{aligned}
$$

Constraint set $\mathcal{B}_3$*: Same as constraint set* $\mathcal{B}_3$*, with the indices "1" and "2" interchanged.*

Proof. Refer to Appendix B.1. □

Remarks: Observe that the above rate region yields HK and MIMO-BC regions as the special cases.

1. HK Region: $R_{c,i} = R_{p,i} = r_{c,i} = r_{p,i} = 0, R_{p,i} = R_i \; i = 1, 2, \; \mathbf{U}_{c1} = \mathbf{U}_{c2} = \phi$.

$$\begin{aligned}
R_1 &\leq I(Y_1; X_{pc1}, U_{pp,1}|X_{pc2}) \\
R_2 &\leq I(Y_2; X_{pc2}, U_{pp,2}|X_{pc1}) \\
R_1 + R_2 &\leq \min\{I(Y_1; X_{pc2}, U_{pp,1}|X_{pc1}) + I(Y_2; X_{pc1}, U_{pp,2}|X_{pc2}), \\
&\quad I(Y_1; U_{pp,1}|X_{pc1}, X_{pc2}) + I(Y_2; X_{pc1}, X_{pc2}, U_{pp,2}), \\
&\quad I(Y_1; X_{pc1}, X_{pc2}, U_{pp,1}) + I(Y_2; U_{pp,2}|X_{pc2}, X_{pc1})\} \\
2R_1 + R_2 &\leq I(Y_1; X_{pc1}, X_{pc2}, U_{pp,1}) + I(Y_1; U_{pp,1}|X_{pc1}, X_{pc2}) + I(Y_2; X_{pc1}, U_{pp,2}|X_{pc2}) \\
R_1 + 2R_2 &\leq I(Y_2; X_{pc1}, X_{pc2}, U_{pp,2}) + I(Y_2; U_{pp,2}|X_{pc2}, X_{pc1}) + I(Y_1; X_{pc2}, U_{pp,1}|X_{pc1}),
\end{aligned}$$

2. MIMO-BC: $R_{p1} = R_{p2} = r_{p1} = r_{p2} = 0, \; R_i = R_{c,i}; X_{pc1} = X_{pc2} = X_{pp1} = X_{pp2} = \phi$.

$$\begin{aligned}
R_1 &\leq I(Y_1; \mathbf{U}_{c1}) - r_{c1} \\
R_2 &\leq I(Y_2; \mathbf{U}_{c2}) - r_{c2} \\
R_1 + R_2 &\leq I(Y_1; \mathbf{U}_{c1}) + I(Y_2; \mathbf{U}_{c2}) - r_{c1} - r_{c2}
\end{aligned}$$

which, from (3.3), is equivalent to the Marton's region for the vector BC, given by

$$\begin{aligned}
R_1 &\leq I(Y_1; \mathbf{U}_{c1}) \\
R_2 &\leq I(Y_2; \mathbf{U}_{c2}) \\
R_1 + R_2 &\leq I(Y_1; \mathbf{U}_{c1}) + I(Y_2; \mathbf{U}_{c2}) - I(\mathbf{U}_{c1}; \mathbf{U}_{c2}).
\end{aligned}$$

Superposition coding for interference cancelation

The layered binning strategy discussed above has high complexity in general. Observe that, when public messages are decoded, the interference due to the common message still remains. This may not be beneficial always. On the other hand, we also observe that simple coding strategies based on superposition coding have much lower complexity, and hence are much easier to implement. With this insight, in the following sections, we propose coding strategies based on superposition coding.

Strategy I:

In this method, the common messages are encoded such that they are decoded by both the receivers. The receivers, after decoding the common messages, cancel out the corresponding interference from the received signal so that other messages can be decoded free of interference due to common messages. The coding strategy takes the following steps:

i. Each common message is encoded first using vector codewords.

ii. Each transmitter encodes its public message using superposition coding treating the common message set as the cloud center.

iii. Each transmitter generates sub-private codeword which superposed over the combination of its own public message and common messages.

We have the following theorem.

Theorem 3.2. *The rate pair* $(R_1, R_2) = (R_{p,1}+R_{c,1}, R_{p,2}+R_{c,2})$ *is achievable if, for some joint distribution such that*

$$p(\mathbf{x}_{c1})p(\mathbf{x}_{c2})p(x_{pc1}|\mathbf{x}_{c2},\mathbf{x}_{c1})p(x_{pc2}|\mathbf{x}_{c2},\mathbf{x}_{c1})p(x_1|x_{pc1},\mathbf{x}_{c2},\mathbf{x}_{c1})p(x_1|x_{pc1},\mathbf{x}_{c2},\mathbf{x}_{c1}).$$

the rate tuple $(R_{p,1}, R_{c,1}, R_{p,2}, R_{c,2})$ *satisfies the following two sets of constraints, denoted* $\mathcal{C}_1$ *and* $\mathcal{C}_2$*:*

Constraint set $\mathcal{C}_1$***:*** *(constraints at the encoders)*

$$\begin{aligned} R_{c,1} &\leq R'_{c,1} \\ R_{c,2} &\leq R'_{c,2}; \end{aligned}$$

Constraint set $\mathcal{C}_2$ ***:***

$$\begin{aligned} R_{p1} &\leq I(Y_1; X_1|X_{pc2}\mathbf{X}_{c,1},\mathbf{X}_{c,1}) \\ 2R_{p1}+R_{p2} &\leq I(Y_1; X_1, X_{pc2}|\mathbf{X}_{c,1},\mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc2}, X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \\ &\quad +I(Y_2; X_2, X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \\ 2R_{p1}+R_{p2}+R_{c1} &\leq I(Y_1; X_1, X_{pc2}|\mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc2}, X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \\ &\quad +I(Y_2; X_2, X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \\ 2R_{p1}+R_{p2}+R_{c2} &\leq I(Y_1; X_1, X_{pc2}|\mathbf{X}_{c,1}) + I(Y_1; X_1|X_{pc2}, X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \\ &\quad +I(Y_2; X_2, X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \\ 2R_{p1}+R_{p2}+R_{c1}+R_{c2} &\leq I(Y_1; X_1, X_{pc2}) + I(Y_1; X_1|X_{pc2}, X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \\ &\quad +I(Y_2; X_2, X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2}) \end{aligned}$$

$$
\begin{aligned}
R_{p1}+R_{p2} &\leq \min\{I(Y_1;X_1|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,1},\mathbf{X}_{c,2}),\\
&\quad I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2}),\\
&\quad I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_1;X_1,X_{pc2}|\mathbf{X}_{c,1},\mathbf{X}_{c,2})\}\\
R_{p1}+R_{p2}+R_{c1} &\leq \min\{I(Y_1;X_1|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,2}),\\
&\quad I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_1;X_1,X_{pc2}|\mathbf{X}_{c,2})\}\\
R_{p1}+R_{p2}+R_{c2} &\leq \min\{I(Y_1;X_1|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,1}),\\
&\quad I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_1;X_1,X_{pc2}|\mathbf{X}_{c,1})\}\\
R_{p1}+R_{p2}+R_{c1}+R_{c2} &\leq \min\{I(Y_1;X_1|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}),\\
&\quad I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_1;X_1,X_{pc2})\}\\
2R_{p1}+2R_{p2}+R_{c1} &\leq I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,2})\\
&\quad +I(Y_1;X_1|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
2R_{p1}+2R_{p2}+R_{c2} &\leq I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,1})\\
&\quad +I(Y_1;X_1|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
2R_{p1}+2R_{p2}+R_{c1}+R_{c2} &\leq I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1})\\
&\quad +I(Y_1;X_1|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|X_{pc2},\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
2R_{p2}+R_{p1}+R_{c1}+R_{c2} &\leq I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1})\\
&\quad +I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
2R_{p2}+R_{p1}+R_{c2} &\leq I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,1})\\
&\quad +I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
2R_{p2}+R_{p1}+R_{c1} &\leq I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,2})\\
&\quad +I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
2R_{p2}+R_{p1} &\leq I(Y_1;X_1,X_{pc2}|X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})+I(Y_2;X_2,X_{pc1}|\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
&\quad +I(Y_2;X_2|X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})\\
R_{p2} &\leq I(Y_2;X_2|X_{pc1}\mathbf{X}_{c,1},\mathbf{X}_{c,1}).
\end{aligned}
$$

Proof. Refer to Appendix B.2. □

Strategy II:

In Strategy I discussed above, each receiver needs to decode all the common messages. This might restrict the rate of common information when the interference is moderately low. This problem can be circumvented by relaxing the condition and just requiring that each receiver only decodes its own common message. A typical instance of this method is the zero-forcing beamforming for Gaussian networks, where transmitters presubtract the known interference due to the interfering part. The coding strategy takes the following steps:

i. Each common message is encoded first using vector codewords.

ii. Each transmitter encodes its public message independent of common messages.

iii. Each transmitter generates sub-private codeword which superimposed over the combination of its own public message and the common message.

We have the following theorem.

Theorem 3.3. *The rate pair* $(R_1, R_2) = (R_{p,1} + R_{c,1}, R_{p,2} + R_{c,2})$ *is achievable if, for some joint distribution such that*

$$p(\mathbf{x}_{c,1})p(\mathbf{x}_{c,2})p(x_{pc,1})p(x_{pc,2})p(x_1|\mathbf{u}_{c,1}, x_{pc,1})p(x_2|\mathbf{u}_{c,2}, x_{pc,2}),$$

the rate tuple $(R_{p,1}, R_{c,1}, R_{p,2}, R_{c,2})$ *satisfies the following three sets of constraints, denoted* $\mathcal{D}_1$, $\mathcal{D}_2$ *and* $\mathcal{D}_3$*:*

Constraint set $\mathcal{D}_1$***:*** *constraints at the transmitters*

$$R_{c,1} \leq R'_{c,1} \tag{3.8}$$

$$R_{c,2} \leq R'_{c,2}; \tag{3.9}$$

Constraint set $\mathcal{D}_2$

$$\begin{aligned}
R_{p,1} &\leq I(Y_1; X_1|X_{pc,2}, \mathbf{X}_{c,1}) \\
R_{p,1} + R_{c,1} &\leq I(Y_1; X_1|X_{pc,2}) \\
R_{p,1} + R_{p,2} &\leq \min\{I(Y_2; X_2, X_{pc1}|\mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc,1}, X_{pc,2}, \mathbf{X}_{c,1}), \\
&\quad I(Y_1; X_1, X_{pc,2}|X_{pc,1}, \mathbf{X}_{c,1}) + I(Y_2; X_2, X_{pc,1}|X_{pc,2}, \mathbf{X}_{c,2})\} \\
2R_{p,1} + R_{p,2} &\leq I(Y_1; X_1, X_{pc2}|\mathbf{X}_{c,1}) + I(Y_2; X_2, X_{pc,1}|X_{pc,2}, \mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc,1}, X_{pc,2}, \mathbf{X}_{c,1}) \\
R_{p,1} + R_{p,2} + R_{c,1} &\leq \min\{I(Y_2; X_2, X_{pc,1}|X_{pc,2}, \mathbf{X}_{c,2}) + I(Y_1; X_1, X_{pc,2}|X_{pc,1}), \\
&\quad I(Y_2; X_2, X_{pc1}|\mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc,1}, X_{pc,2}), \\
&\quad I(Y_1; X_1, X_{pc,2}) + I(Y_2; X_2|X_{pc,2}, X_{pc,1}, \mathbf{X}_{c,2})\} \\
2R_{p,1} + R_{p,2} + R_{c,1} &\leq \min\{I(Y_1; X_1, X_{pc,2}) + I(Y_2; X_2, X_{pc,1}|X_{pc,2}, \mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc,1}, X_{pc,2}, \mathbf{X}_{c,1}), \\
&\quad I(Y_1; X_1, X_{pc2}|\mathbf{X}_{c,1}) + I(Y_2; X_2, X_{pc,1}|X_{pc,2}, \mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc,1}, X_{pc,2})\} \\
R_{p,1} + 2R_{p,2} + R_{c,1} &\leq I(Y_2; X_2, X_{pc1}|\mathbf{X}_{c,2}) + I(Y_1; X_1, X_{pc,2}|X_{pc,1}) + I(Y_2; X_2|X_{pc,2}, X_{pc,1}, \mathbf{X}_{c,2}) \\
2R_{p,1} + R_{p,2} + 2R_{c,1} &\leq I(Y_1; X_1, X_{pc,2}) + I(Y_2; X_2, X_{pc,1}|X_{pc,2}, \mathbf{X}_{c,2}) + I(Y_1; X_1|X_{pc,1}, X_{pc,2})
\end{aligned}$$

$$
\begin{aligned}
R_{p,1}+R_{p,2}+R_{c,1}+R_{c,2} &\leq \min\{I(Y_1;X_1,X_{pc,2}|X_{pc,1})+I(Y_2;X_2,X_{pc,1}|X_{pc,2}),\\
&\quad I(Y_2;X_2,X_{pc,1})+I(Y_1;X_1|X_{pc,1},X_{pc,2})\}\\
2R_{p,1}+R_{p,2}+R_{c,1}+R_{c,2} &\leq \min\{I(Y_1;X_1,X_{pc,2})+I(Y_2;X_2,X_{pc,1}|X_{pc,2})+I(Y_1;X_1|X_{pc,1},X_{pc,2},\mathbf{X}_{c,1}),\\
&\quad I(Y_1;X_1,X_{pc2}|\mathbf{X}_{c,1})+I(Y_2;X_2,X_{pc,1}|X_{pc,2})+I(Y_1;X_1|X_{pc,1},X_{pc,2})\}\\
2R_{p,1}+R_{p,2}+2R_{c,1}+R_{c,2} &\leq I(Y_1;X_1,X_{pc,2})+I(Y_2;X_2,X_{pc,1}|X_{pc,2})+I(Y_1;X_1|X_{pc,1},X_{pc,2});
\end{aligned}
$$

Constraint set $\mathcal{D}_3$***:*** *Same as* $\mathcal{D}_2$ *with the indices "1" and "2" interchanged.*

Proof. The proof is similar to the proof of Theorem 3.2, and is omitted. □

3.4 The Gaussian Case with Symmetric Channel Gains

In this section, we illustrate the coding strategies discussed above for the case of a second-hop modeled as a symmetric Gaussian interference channel; i.e., $h_{1,1}=h_{2,2}=h$ and $h_{1,2}=h_{2,1}=g$. We note that determining the complete rate region, optimal coding schemes and input distributions for the interference channel is a complex task, and does not yield much insights. Therefore, we restrict our attention to characterizing the *symmetric achievable rate*, defined as $R=\max\min\{R_1,R_2\}$, where the rate pair (R_1,R_2) belongs to the best known achievable rate region for the interference channel of interest. Further, we limit our discussions to the specific signaling schemes, encoding/decoding order, as will be discussed below.

Let the received signal at each receiver is given by

$$
\begin{aligned}
Y_1 &= hX_1+gX_2+Z_1\\
Y_2 &= hX_2+gX_1+Z_2,
\end{aligned}
$$

where X_1 and X_2 are the transmitted signals from S_1 and S_2 respectively, and $Z_i\sim\mathcal{N}(0,1), i=1,2$, is the additive Gaussian noise at receiver D_i. Each transmitter has a power constraint $\mathbb{E}[|X_i[n]|^2]\leq P$. We assume that each user has access to the same amount of side-information, i.e., $R'_{c,1}=R'_{c,2}=R'_c$.

Before proceeding further with the coding strategies, we remark on the notation used in this section. For convenience, we use $\mathbf{X}_a$ to denote any vector of the form $[X_{a,1},X_{a,2}]$. Further, we let $\mathbf{h}_1=[h,g]$, $\mathbf{h}_2=[g,h]$ and $\|\mathbf{h}\|^2=h^2+g^2$. Observe that $\mathbf{h}_i$ is the vector channel formed by the transmitter (S_1,S_2) at the receiver D_i .

Let $(\mathrm{v}_1, \mathrm{v}_{c,2}) = (\mathrm{v}_{pc,1}, \mathrm{v}_{pp,1}, \mathrm{v}_{c,1}, \mathrm{v}_{c,2})$ and $(\mathrm{v}_2, \mathrm{v}_{c,1}) = (\mathrm{v}_{pc,1}, \mathrm{v}_{pp,1}, \mathrm{v}_{c,1}, \mathrm{v}_{c,2})$ be the message set to be transmitted from S_1 and S_2 respectively. Encoder 1 employs $X^N_{pc,1}$ to encode $\mathrm{v}_{pc,1}$, $U^N_{pp,1}$ to encode $\mathrm{v}_{pp,1}$, and $U^N_{c,1,1}$ and $U^N_{c,2,1}$ to encode $\mathrm{v}_{c,1}$ and $\mathrm{v}_{c,2}$ respectively. Similarly, encoder 2 employs $X^N_{pc,2}$ to encode $\mathrm{v}_{pc,2}$, $U^N_{pp,2}$ to encode $\mathrm{v}_{pp,2}$, and $U^N_{c,1,2}$ and $U^N_{c,2,2}$ to encode $\mathrm{v}_{c,1}$ and $\mathrm{v}_{c,2}$ respectively. Note that $\mathbf{U}^N_{c,1}$ and $\mathbf{U}^N_{c,2}$ are generated cooperatively by both encoders. Each encoder splits its power P into three parts, P_c, P_{pp} and P_{pc}. All codewords are drawn from a zero mean Gaussian ensemble. Both encoders compute $\mathbf{X}_{c,1} = f_{c1}(\mathbf{U}_{c,1}, \mathbf{U}_{c,2})$, $\mathbf{X}_{c,2} = f_{c2}(\mathbf{U}_{c,1}, \mathbf{U}_{c,2})$ and encoder S_i computes $X_{pp,i} = f_{p,i}(U_{pp,i})$ as the transmit signals. Then S_i transmits

$$X_i = X_{c,1,i} + X_{c,2,i} + X_{pc,i} + X_{pp,i}$$

The received signal at D_i can be written as

$$Y_i = \mathbf{h}_i^\dagger \left(\mathbf{X}_{c,1} + \mathbf{X}_{c,2} + \mathbf{X}_{pc} + \mathbf{X}_{pp}\right) + Z_1.$$

We let $\zeta_i = \mathbf{h}_i^\dagger \left(\mathbf{X}_{pc} + \mathbf{X}_{pp}\right) + Z_1$ where ζ_i is the equivalent additive noise at D_i in decoding the common messages, treating the interference due to private messages as additive noise. Observe that $\zeta_i \sim \mathcal{N}(0, \sigma^2_\zeta)$, where $\sigma^2_\zeta = 1 + \|\mathbf{h}\|^2 P_p$.

Binning for interference cancelation: layered dirty paper coding (DPC)

In this section we demonstrate the application of GP binning techniques which, when specialized to the Gaussian case, boil down to dirty paper coding (DPC) [42]. Essentially, DPC involves pre-cancelation of interference at the transmitter without incurring any power cost. From the perspective of transmitting common information only, the interference channel can be modeled as a Gaussian *vector broadcast channel*, except that it has the interference due to private messages. Gaussian MIMO broadcast channels have been well studied and their capacity regions have been characterized in [43, 57] by exploiting DPC. Therefore, it is clear that DPC's performance in the second hop is superior when entire information at relays is common.

Encoding:

1. *Encoding the public messages:* Encoder S_i chooses a codeword $X^N_{pc,i}$, $X_{pc,i} \sim \mathcal{N}(0, P_{pc})$ to relay its own public message. Public messages are decoded first at the

receivers, and their effects are subtracted from the received signals before decoding common and sub-private messages.

2. *Encoding the common messages:* Relays collaborate to encode the common messages sequentially. Suppose at any time the common message pertaining to one user is encoded, DPC is performed on the common message of the other user treating the interference due to the first user's common message as the known information. Consider the general case, where relays L_1 and L_2 allocate $P_{pp,1}$ and $P_{pp,2}$ for the transmission of their private messages. Then, it is clear that the sum power budget is $2P - P_{p,1} - P_{p,2}$ for transmitting common messages, where $P_{p,i} = P_{pp,i} + P_{pc,i}$. While the second-hop is analogous to a vector BC, we should point out that, relay L_1 has to use only $P - P_{p,1}$ and relay L_2 has to use only $P - P_{p,2}$ to transmit the common information. We will develop a time sharing mechanism to overcome this difficulty. Depending on the encoding order of the common information, DPC is performed using following strategies.

• <u>Encoding strategy π_1</u> For private messages, set $P_{pp,1} = p_1^*$ and $P_{pp,2} = p_2^*$. Relays first encode $\mathrm{v}_{c,2}$ using $\mathbf{U}_{c,2}^N(\mathrm{v}_{c,2})$. Then, set $\mathbf{X}_{c,2}^N = \mathbf{U}_{c,2}^N$. Encode $\mathrm{v}_{c,1}$ using $\mathbf{U}_{c,1}^N(\mathrm{v}_{c,1})$, and apply DPC on $\mathbf{U}_{c,1}^N(\mathrm{v}_{c,1})$ treating $\mathbf{X}_{c,2}^N$ as the known interference while treating the effect due to sub-private messages, $\zeta_1^N = \mathbf{h}_1^\dagger \mathbf{X}_{pp}^N + Z_1^N$, as the noise. That is, choose the vector $\mathbf{X}_{c,1} \sim \mathcal{N}(0, \Sigma_{c,1})$, independent of $\mathbf{X}_{c,2}$, such that

$$\mathbf{U}_{c,1} = \mathbf{X}_{c,1} + \mathbf{B}_c \mathbf{X}_{c,2},$$

where $\mathbf{B}_c$ is chosen as per the DPC requirement [42]; i.e., $\mathbf{B}_c$ corresponds to the optimal non-causal MMSE estimate of $\mathbf{X}_{c,1}$ given $\mathbf{h}_1^\dagger \mathbf{X}_{c,1} + \zeta_1$. Encoders jointly compute their signals for relaying the common message set as $\mathbf{X}_c = \mathbf{X}_{c,1} + \mathbf{X}_{c,2}$. Relays choose covariance matrices to maximize the sum-rate; i.e, $\Sigma_1 = \Sigma_1^{\pi_1}$, $\Sigma_2 = \Sigma_2^{\pi_1}$; where $\{\Sigma_1^{\pi_1}, \Sigma_2^{\pi_1}\}$ is the solution to (3.10).

$$R_{\pi_1}(p_1^*, p_2^*) = \max_{\Sigma_1, \Sigma_2} \left(C\left(\frac{\mathbf{h}_1^\dagger \Sigma_1 \mathbf{h}_1}{1 + h^2 p_1^* + g^2 p_2^*} \right) + C\left(\frac{\mathbf{h}_2^\dagger \Sigma_2 \mathbf{h}_2}{1 + \mathbf{h}_2^\dagger \Sigma_1 \mathbf{h}_2 + h^2 p_2^* + g^2 p_1^*} \right) \right) \tag{3.10}$$

$$\text{subject to} \quad \mathsf{tr}\,(\Sigma_1 + \Sigma_2) \le 2P - P_{pc,1} - P_{pc,2} - p_1^* - p_2^*.$$

• <u>Encoding strategy π_2</u> For private messages, set $P_{pp,1} = p_2^*$ and $P_{pp,2} = p_1^*$. Relays encode $\mathrm{v}_{c,1}$ in $\mathbf{X}_{c,1}^N$, then generate $\mathbf{X}_{c,2}^N$ by encoding $\mathrm{v}_{c,2}$ and then performing DPC treating $\mathbf{h}_2^\dagger \mathbf{X}_{c,1}^N$

as the known side information. As before, the optimal sum-rate, $R_{\pi_2}(p_1^*, p_2^*)$, is obtained by solving (3.10) swapping indexes 1 and 2, with the corresponding covariance matrices being $\{\Sigma_1, \Sigma_2\} = \{\Sigma_1^{\pi_2}, \Sigma_2^{\pi_2}\}$.

• Time sharing Observe that the power allocation for common messages at each relay, corresponding to the strategy π_j, is determined by $P_{c,1}^{\pi_j} \triangleq \left[\Sigma_1^{\pi_j} + \Sigma_2^{\pi_j}\right]_{1,1}$; $P_{c,2}^{\pi_j} \triangleq \left[\Sigma_1^{\pi_j} + \Sigma_2^{\pi_j}\right]_{2,2}$. To meet the power constraint at each relay, relays perform time sharing by using π_1 for the first half of the slot and π_2 for the second half. Note that, due to symmetry in channel conditions, relay L_i obeys the average power constraint per slot: $\frac{1}{2}(P_{c,i}^{\pi_1} + P_{c,1}^{\pi_i} + P_{pc,1} + P_{pc,2} + p_1^* + p_2^*) = P$.

3. *Encoding the sub-private messages:* Observe that, due to DPC, each user is able to decode its own common message but not the interfering part. Therefore, the effect of the common message due the other user is still interference. Fortunately, this is known side-information to the encoders. Therefore, each encoder performs another round of DPC on its own sub-private messages, treating the effect of interfering part of the common codeword as the known side-information. To this end, D_i chooses the DPC signal $X_{pp,i} \sim \mathcal{N}(0, P_{pp})$, independent of $\mathbf{U}_{c,j},\ j \neq i$, such that,

$$U_{pp,i} = X_{pp,i} + \mathbf{b}_c^\dagger \mathbf{U}_{c,j},$$

where $\mathbf{b}_c$ corresponds to the MMSE estimator of $X_{pp,i}$ given $\mathbf{h}_1^\dagger X_{pp,i} + Z_i$. It is interesting to note that $\mathbf{b}_c$ does not depend on the variance of the interference. Therefore, due to DPC, each individual receiver can decode its private message as though there were no interference due to common messages.

Decoding: Receivers decode messages in the following order. Public messages are decoded first, treating $\mathbf{h}_i^\dagger(\mathbf{X}_{c,1}^N + \mathbf{X}_{c,2}^N + \mathbf{X}_{pp}^N)$, the interference from common messages and private messages, as noise. Once the public messages are decoded, each receiver subtracts their effect, $\mathbf{h}_i^\dagger \mathbf{X}_{pc}$, from the received signal, Y_i. In the second step, each receiver decodes the part of common message intended for it. In decoding the common message, each receiver treats $\mathbf{h}_i^\dagger \mathbf{X}_{pp}^N$, the interference due to sub-private messages, as noise. Finally, receivers decode their sub-private messages individually, treating the interference due to the sub-private message of other user as noise. Note that the interference from public messages disappears due to interference cancelation at the receiver, and the interference from common messages is canceled out due to the DPC performed on sub-private messages.

Analysis: Next, we evaluate the achievable rates due to this scheme. Let a power split (P_{pc}, P_c, P_{pp}) at relays be given such that $P_{pc}+P_c+P_{pp} = P$. As mentioned earlier, public messages are decoded first, treating the effect of common and sub-private messages as noise. Therefore, the equivalent noise at receiver D_i in decoding the public message is given by

$$\Psi_i = \mathbf{h}_i^{\dagger} \left(\mathbf{X}_{c,1} + \mathbf{X}_{c,2} + \mathbf{X}_{pp}\right) + Z_{d,i}.$$

We note that $\Psi \sim \mathcal{N}(0, \sigma_\Psi^2)$, where

$$\begin{aligned} \sigma_\Psi^2 &= E[|\Psi|^2] \\ &= \mathbf{h}_i^{\dagger} \left(\Sigma_1 + \Sigma_2 + P_{pp} \cdot \mathbf{I}\right) \mathbf{h}_i^{\dagger} + 1 \\ &= \|\mathbf{h}\|^2 \left(P - P_{pc}\right) + \frac{4P_c g^2 h^2}{\|\mathbf{h}\|^2 + \frac{P_c h_\Delta^2}{\sigma_p^2}} + 1, \end{aligned}$$

where $\sigma_p^2 = \|\mathbf{h}\|^2 P_{pp} + 1$.

Next, consider the transmission of common messages. For the general case, a closed-form solution for the optimal power allocation is still unknown. However, under the symmetric power allocation for the sub-private message set $P_{pp,1} = P_{pp,2} = P_{pp}$, it can be shown that the optimal sum-rate is given by (see [58] for details)

$$R_c^{\pi_1} = C\left((h^2 - g^2)^2 \frac{P_c^2}{\sigma_p^4} + \frac{2P_c}{\sigma_p^2} \|\mathbf{h}\|^2\right), \tag{3.11}$$

and the corresponding DPC matrices are given by (refer to Appendix B.4, for more details)

$$\begin{aligned} \Sigma_{c,1}^{\pi_1} &= \frac{P_c}{\left(\sigma_p^2 1 + \frac{P_c}{\sigma_p^2}\|\mathbf{h}\|^2\right)\left(\|\mathbf{h}\|^2 + \frac{P_c}{\sigma_p^2} h_\Delta^2\right)} \begin{bmatrix} h^2\left(1 + \frac{P_c}{\sigma_p^2} h_\Delta\right)^2 & gh\left(1 - \frac{P_c}{\sigma_p^2}^2 h_\Delta^2\right) \\ gh\left(1 - P_c'^2 h_\Delta^2\right) & g^2\left(1 - P_c' h_\Delta\right)^2 \end{bmatrix}, \\ \Sigma_{c,2}^{\pi_1} &= \frac{P_c\left(1 + 2\frac{P_c}{\sigma_p^2}\|\mathbf{h}\|^2 + \frac{P_c^2}{\sigma_p^4}\right)}{\left(1 + \frac{P_c}{\sigma_p^2}\|\mathbf{h}\|^2\right)\left(\|\mathbf{h}\|^2 + \frac{P_c}{\sigma_p^2} h_\Delta^2\right)} \begin{bmatrix} g^2 & hg \\ hg & h^2 \end{bmatrix}, \end{aligned} \tag{3.12}$$

where $h_\Delta^2 = h^2 - g^2$.

Next, consider sub-private messages. Observe that DPC is performed on the private message treating the interference due to common information as the known side-information. Further, each receiver subtracts out the interference due to public messages. Therefore, each receiver can decode its own sub-private message as though there were no interference due to common messages and public messages. Therefore, each relay can transmit its private information at a rate

$$R_{pp} = C\left(\frac{h^2 P_{pp}}{1 + g^2 P_{pp}}\right). \tag{3.13}$$

Therefore, when the proposed layered coding with DPC is used, the total achievable rate per link is given by

$$R_{\mathsf{LAY-DPC}} = \max_{\left\{\substack{P_{pc}, P_c, P_{pp}: \\ P_c + P_p = P}\right\}} R_{pc} + R_c + R_{pp};\ \text{s. t. } R_c \leq R_c',$$

where R_c is the common information rate per link, R_{pc} is the public information rate per link, and R_{pp} the sub-private information rate per link, given by

$$R_c = \frac{1}{2} C\left((h^2 - g^2)^2 \frac{P_c^2}{\sigma_p^4} + \frac{2P_c}{\sigma_p^2} \|\mathbf{h}\|^2)\right),$$

$$R_{pc} = \min\left\{C\left(\frac{g^2 P_{pc}}{\sigma_\Psi^2}\right), \frac{1}{2} C\left(\frac{\|\mathbf{h}\|^2 P_{pc}}{\sigma_\Psi^2}\right)\right\}, \text{ and } R_{pp} = C\left(\frac{h^2 P_{pp}}{1 + g^2 P_{pp}}\right).$$

We remark that our choice of covariance matrices, in performing DPC for the common messages, is greedy, in the sense that it maximizes the rates for transmission of the common message without taking into account their effect on the public messages. We observe that due to the above mentioned decoding order, apart from private messages, common messages also appear as interference for public messages. Since common messages are aired using vector broadcast techniques, the signal power of the common messages at the receivers is enhanced significantly. Therefore, public messages may incur significant amounts of interference from common messages. Thus when R_c is large, the contribution due to public messages vanishes, and consequently, it is beneficial to perform DPC without resorting to splitting the private message toward interference cancelation. Indeed, when R_c and hence P_c increases, a small amount of power $P_1 = P - P_c$ is left to rely private and public messages. Concordant to the results obtained for weak interference channels [59, 60, 61], when P_1 satisfies $P_1 \leq \frac{(h-2g)}{2g^3}$, splitting of the private message is no longer necessary. Thus, to decoding one user's private message while treating other user's private message as noise becomes optimal. When R_c is very small, a relatively smaller amount of power is spent on transmitting common messages. Therefore, it is beneficial to perform message splitting of the private message to aid in interference cancelation.

Superposition coding: layered beamforming

In the previous section, we proposed a layered coding strategy that incorporates DPC to relay common messages. However, we note that DPC has high complexity. The non-linearity inherent in DPC inhibits us from avoiding interference due to the common messages when decoding public messages. This may not be beneficial always. On the other hand, linear schemes like beamforming are much easier to implement, and reduce the burden of common messages before decoding the private messages. In light of the above observations, we propose coding schemes based on a marriage of HK coding and BF schemes.

Strategy I: As noted in the previous section, this scheme requires each receiver to decode both the common messages. Since information rate in such cases is determined by the weaker link, this scheme yields benefits when the gain on the interfering link is higher.

Encoding: For transmitting common messages, transmitters apply coherent beamforming by letting each encoder choose its symbols such that $U_{c,1,1} = U_{c,1,2} = U_{c,1}$ and $U_{c,2,1} = U_{c,2,2} = U_{c,2}$, with $U_{c,i} \sim \mathcal{N}(0,1),\ i = 1,2$. That is, $\Sigma_{u,1} = \Sigma_{u,2} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then, set

$$X_{c,1,1} = \sqrt{P_{cm}}U_{c,1};\quad X_{c,2,1} = \sqrt{P_{cu}}U_{c,2}$$

$$X_{c,1,2} = \sqrt{P_{cu}}U_{c,1};\quad X_{c,2,2} = \sqrt{P_{cm}}U_{c,2}$$

where P_{cm} and P_{cu} are chosen such that $P_c = P_{cm}+P_{cu}$. Observe that P_{cm} is the portion of the power used by each transmitter to transmit the intended part of its common message, while P_{cu} is the power used to transmit the "interfering part" as the side-information.

For transmitting public and sub-private messages, each encoder uses codewords $X_{pc,i}^N$ and $X_{pp,i}^N$ respectively. To transmit its message set, transmitter i computes the signal X_i as

$$X_i = X_{c,1,i} + X_{c,2,i} + X_{pc,i} + X_{pp,i}, \quad i = 1, 2.$$

Observe that $\mathbb{E}[|X_i|^2] = P$.

Decoding: Receivers decode messages in the following order. First, each receiver decodes common information. To decode the common information, the effect of private messages is treated as noise. Once the common messages are decoded, each receiver subtracts the effect of the common messages from its received signal in order to decode its private messages Thus, there is no interference from the common messages on the private messages. Observe that each splitting of private message into sub-private and public sub-messages is according to HK coding.

Analysis: We represent the received signal as

$$Y_i = \mathbf{h}_i^\dagger \left(\mathbf{X}_{c,1} + \mathbf{X}_{c,2} + \mathbf{X}_{pc} + \mathbf{X}_{pp}\right) + Z_i.$$

Considering the transmission of only common messages, we equivalently model the channel as

$$Y_1 = \mathbf{h}_1^\dagger \mathbf{a} U_{c,1} + \mathbf{h}_2^\dagger \mathbf{a} U_{c,2} + \zeta_1, \tag{3.14}$$

where, for convenience, we have defined $\mathbf{a} = [\sqrt{P_{cm}}, \sqrt{P_{cu}}]$. Furthermore, ζ_1 is the equivalent noise at D_1 in decoding the common messages, with $\zeta_{d,1}[n] \sim \mathcal{N}(0, \sigma_\zeta^2)$, where $\sigma_\zeta^2 = \|\mathbf{h}\|^2 P_p + 1$.

Similarly, for the second receiver D_2,

$$Y_2 = \mathbf{h}_1^\dagger \mathbf{a} U_{c,2} + \mathbf{h}_2^\dagger \mathbf{a} U_{c,1} + \zeta_2, \tag{3.15}$$

where $\zeta_2 \sim \mathcal{N}(0, \sigma_\zeta^2)$.

Consider the transmission of common information. We observe that (3.14) represents a virtual MAC formed by $(U_{c,1}, U_{c,2}, Y_1)$, and (3.15) represents another virtual MAC formed by $(U_{c,1}, U_{c,2}, Y_2)$. The achievable common information rate region is the intersection of the regions formed by these two MACs. The following proposition gives the optimal power allocation vector for the common information transmission.

Proposition 3.1. *The common information rate per link, under coherent beamforming, is maximized by equal power allocation at the relays, given by $P_{cm} = P_{cu} = P_c/2$, and the corresponding common information rate per link is given by*

$$R_c = \frac{1}{2}C\left(\frac{(g+h)^2 P_c}{\|\mathbf{h}\|^2 P_p + 1.}\right). \tag{3.16}$$

The proof can be found in Appendix B.3.

Next, consider the private messages. Since sub-common messages and sub-private messages are decoded by canceling the interference due to the common messages, the encoding and decoding of the public and sub-private messages mimic the version of HK coding considered in [14]. The information rate for the private message for each user is given by

$$R_p = \max_{\left\{\substack{P_{pc}, P_{pp}: \\ P_{pc}+P_{pp}=P_p}\right\}} R_{pp} + R_{pc}, \tag{3.17}$$

with

$$R_{pp} = C\left(\frac{h^2 P_{pp}}{1+g^2 P_{pp}}\right) \text{ and } R_{pc} = \min\left\{C\left(\frac{g^2 P_{pc}}{1+\|\mathbf{h}\|^2 P_{pp}}\right), \frac{1}{2}C\left(\frac{\|\mathbf{h}\|^2 P_{pc}}{1+\|\mathbf{h}\|^2 P_{pp}}\right)\right\}.$$

Thus, the rate that can be achieved per user in the second hop due to CBF is

$$R_{\mathsf{CBF}} = \max_{\left\{\substack{P_c, P_p: \\ P_c+P_p=P}\right\}} R_p + R_c; \text{ s. t. } R_c \leq R'_c.$$

Strategy II:

In the coherent beamforming case, each receiver needs to decode all the common messages. This might restrict the rate of common information when the interference is moderately low. Under these circumstances, it can be beneficial to relax this requirement which is in spirit with the strategy II discussed in Section 3.3. For the Gaussian case, the design of an optimal beamforming vector for the common message codewords is crucial. In our illustration, for simplicity, we consider zero forcing beamforming (ZFBF), which involves transmitters pre-subtracting the known interference due to the interfering part of the common message. Essentially, ZFBF consists of inverting the channel matrix by the relays to create orthogonal channels between the relays and the receivers.

Encoding: For transmitting common messages, each encoder chooses its symbols such that $U_{c,1,1} = U_{c,1,2} = U_{c,1}$ and $U_{c,2,1} = U_{c,2,2} = U_{c,2}$, with $U_{c,i} \sim \mathcal{N}(0,1), i = 1,2$. That is, $\Sigma_{u,1} = \Sigma_{u,2} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then, set

$$X_{c,1,1} = \frac{h\sqrt{P_c}}{\|\mathbf{h}\|}U_{c,1}; \quad X_{c,2,1} = -\frac{g\sqrt{P_c}}{\|\mathbf{h}\|}U_{c,2}$$

$$X_{c,1,2} = -\frac{g\sqrt{P_c}}{\|\mathbf{h}\|}U_{c,1}; \quad X_{c,2,2} = \frac{h\sqrt{P_c}}{\|\mathbf{h}\|}U_{c,2}$$

For transmitting public and sub-private messages, each encoder uses codewords $X_{pc,i}^N$ and $X_{pp,i}^N$ respectively. To transmit its message set, transmitter i computes the signal X_i as

$$X_i = X_{c,1,i} + X_{c,2,i} + X_{pc,i} + X_{pp,i}, \quad i = 1,2.$$

Observe that $\mathbb{E}[|X_i|^2] = P$.

Decoding: Receivers decode messages in the following order. First, each receiver decodes the common information. Observe that each user is able to decode only his own common message and the interfering part of the common messages vanishes due to zero-forcing. To decode the common information, the effect of private messages is treated as noise. Once the common messages are decoded, each receiver subtracts the effect of the common messages from its received signal in order to decode its private message. Thus, there is no interference from the common messages on the private messages. Observe that each splitting of private message into sub-private and public sub-messages is similar to HK coding.

Analysis: With the decoding order fixed as mentioned above, the rate of common information per user is given by

$$R_c = C\left(\frac{\left(h^2 - g^2\right)^2 P_c}{\|\mathbf{h}\|^2 \left(1 + \|\mathbf{h}\|^2 P_p\right)}\right). \tag{3.18}$$

The information rate for the private message per user is given by

$$R_p = \max_{\left\{\substack{P_{pc}, P_{pp}: \\ P_{pc} + P_{pp} = P_p}\right\}} R_{pp} + R_{pc}, \tag{3.19}$$

with

$$R_{pp} = C\left(\frac{h^2 P_{pp}}{1 + g^2 P_{pp}}\right) \text{ and } R_{pc} = \min\left\{C\left(\frac{g^2 P_{pc}}{1 + \|\mathbf{h}\|^2 P_{pp}}\right), \frac{1}{2}C\left(\frac{\|\mathbf{h}\|^2 P_{pc}}{1 + \|\mathbf{h}\|^2 P_{pp}}\right)\right\}.$$

Thus, the rate that can be achieved per user in the second hop due to ZFBF is

$$R_{\mathsf{ZFBF}} = \max_{\left\{\substack{P_c, P_p: \\ P_c + P_p = P}\right\}} R_p + R_c; \text{ s. t. } R_c \leq R_c'.$$

3.5 Numerical Examples

We perform numerical evaluation of the schemes discussed above. The metric of interest is the total achievable symmetric rate per user. For ease of evaluation, we consider a symmetric interference $L \to D$ link with a link SNR of 20dB, where we define link $\mathsf{SNR} = h^2 P$. We are interested in studying the effect of common information availability, R_c at the relays, on the achievable information rate per user I_{LD} over the $L \to D$ link. Define the parameter $\alpha = g^2/h^2$. Figs. 3.2 and 3.3 show I_{LD} versus R_c plots for different schemes. For comparison purposes, we also consider the naive strategies without layering: HK-scheme (does not exploit the common information), DPC and beamforming (do not split private message for interference cancelation). The optimal power allocation policies and the corresponding achievable rates of different schemes were determined numerically by exhaustive search.

We have the following observations:

- Layered coding with DPC outperforms all other schemes for all ranges of R_c examined and for relatively small values of α. This is because, for relatively smaller values of α's, the interference of common messages on public messages is low, en-

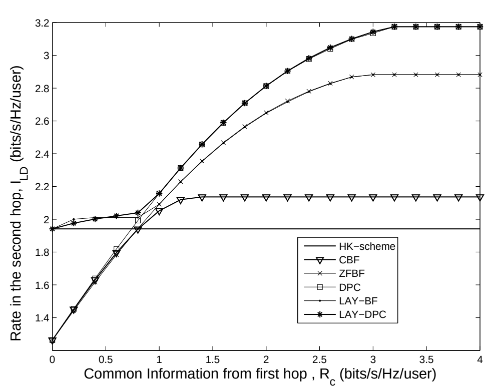

Figure 3.2: I_{LD} vs. R_c curves ($\alpha = 0.2$).

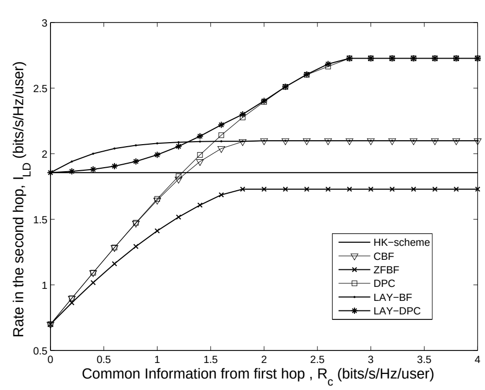

Figure 3.3: I_{LD} vs. R_c curves ($\alpha = 0.6$).

abling efficient message splitting for smaller values of R_c. When R_c is large, the layered coding scheme boils down to DPC which is optimal when R_c is large.

- Layered coding with BF is superior to its DPC counterpart for larger values of α in the regime when R_c is small. While both DPC and BF enhance the SNR of the common messages, for the layered DPC scheme, the interference due to common messages may drastically affect the rate of the public messages to an extent that it may become beneficial for relays to allot more power to transmitting the private message than to transmitting the common messages. However, for the layered BF scheme, successive cancelation of interference from common messages at the receivers proves to be useful.

- HK based scheme outperforms DPC and BF schemes when the amount of common information available is relatively low. This indicates that in such cases, it is beneficial to allocate more power for unknown interference cancelation than relaying the common information.

- The performance of layered coding schemes boils down to that of HK coding when there is no side information available at the relays; i.e., $R_c = 0$. This is due to the fact that HK coding is the best known scheme for ICs where the transmitters have no side information.

- The beamforming method is not superior in any regime. For higher values of α, performance of CBF improves whereas the performance of ZFBF becomes worse. When the values of α is low, ZFBF performs better when compared to the CBF. This is due to the fact that the power spent on the information bits in ZFBF decreases with the interference and the gain due to coherent combining in CBF increases with the strength of interfering link.

3.6 Conclusions

In this chapter, we considered a basic model for two-hop transmissions of two information flows that interfere with each other. The main focus has been on the second hop transmission, where one intrinsic feature is that each relay has access to a part of the information intended for the other destination. A key observation is that when each relay has only partial side-information, neither distributed MIMO broadcast nor HK coding would be optimal, and a naive combination of these two schemes would not work either. Further, since each receiver receives a superposition of different sub-messages that are relayed, the decoding order of the signals at the receiver plays a critical role. Therefore, efficient relaying strategies entail a careful marriage of distributed MIMO broadcast and HK coding.

We addressed the above challenge by investigating different relaying schemes and developing novel layered strategies built on distributed MIMO broadcast and HK coding. Specifically, we proposed two types of layered schemes: layered coding with binning and layered superposition coding. We illustrated the applications of proposed strategies for a symmetric Gaussian case. We presented numerical examples to demonstrate that our schemes can yield a substantial rate gain.

Chapter 4

WHEN COMPRESSIVE SAMPLING MEETS MULTICAST: OUTAGE ANALYSIS AND SUBBLOCK NETWORK CODING

4.1 Introduction

Multicasting is an emerging technology in wireless networks in supporting a plethora of applications, including text, teleconferencing, and multimedia streaming, to a group of users. Notably, many of these signals are known to be sparse or compressible, where the sparsity (or compressibility) refers to the property that the "*essential information rate*" of a signal is significantly smaller than what is suggested by its bandwidth (in case of continuous signals), or that a signal essentially depends on a number of degrees of freedom which is much smaller than its length (in case of discrete-time signals). Images, for instance, usually admit a compact representation in the wavelet domain, and audio signals can be often compressed in the frequency domain. As a result, in data communication systems, data transmission is often preceded by its compression, to ensure efficient utilization of resources by minimizing redundancy.

As shown in Fig. 4.1(a), the conventional method of data compression involves sampling the signal at the Nyquist rate, storing the samples, and compressing them in an appropriate domain prior to the transmission. One "drawback" of the traditional approach when applied to sparse signals is that the size of measurements is of the same order as of the size of the signal itself, independent of its sparsity/compressibility. This would incur heavy sampling and storage burden at the sender. On the other hand, recent developments in compressive sensing theory [19, 20, 21] have provided methods not only for lower-rate signal acquisition, but also for accurate signal reconstruction, as shown in Fig. 4.1(b). One basic idea behind compressive sensing theory is that it exploits the *compressibility* of the signals by only sensing its essential information. We note that a scheme involving compressive sensing was considered in [62], for point-to-point transmission.

In this study, a primary objective is to understand the reliable delivery of compressible information to many destinations over wireless channels. This transmission scenario, for instance, is useful for multimedia streaming over a WiFi or WiMAX network with multiple receivers. It is also applicable to sensor networks, where each sensor node wishes to communicate its sensed observations to multiple coordinating agents. Needless to say, this problem is quite challenging due to the lossy nature of wireless channels and the heterogeneity in the amount of information received across different receivers. Another challenge is the "*bottleneck*" effect with the shared transmission medium among many receivers, i.e., the overall performance is limited by the receiver that has the worst channel condition. Therefore, ensuring that the QoS for the bottleneck user is on par with the others clearly makes the multicast transmission more challenging.

We note that traditionally, in order to mitigate data loss over wireless channels, retransmission schemes are widely used, e.g., Automatic Repeat reQuest (ARQ) [63]. However, retransmissions would not work well in some practical situations, particularly, in multicast scenarios with many destinations (receivers) [64] or when the applications are delay sensitive. Alternatively, forward error correction (FEC) can be used to reduce the amount of feedback [65], but a significant challenge in using FEC is to determine the right amount of redundancy because too much FEC redundancy would unnecessarily slow down the data dissemination process, whereas insufficient redundancy leaves many receivers unable to decode the information. Furthermore, the inherent heterogeneity of information conveyed to multiple receivers significantly complicates the task of redundancy estimation. Notably, network coding (NC), a recent breakthrough by Ahlswede *et al.* [18], offers a promising platform for multicast transmissions. With this insight, we will study joint compressive sensing and network coding for multicasting compressive signals, as illustrated in Fig. 4.1(c). To the best of our knowledge, there has been little effort on exploring compressive sensing over a multicast scenario; and this work is among the first few to examine the interplay between compressive sensing and network coding.

Specifically, we consider a single-hop wireless network in which the sender multicasts a compressively sensed signal to L receivers, over wireless channels. We note that many natural and man-made signals can be modeled as power-law decay signals, a class of signals whose coefficients decay according to the power law, when sorted in the decreasing order. (For example, video signals admit power-law decay in the wavelet domain.) Without loss of generality, we assume that the compressible signal belongs to the class of power-law decay signals. We are particularly interested in the distortion of the reconstructed signal, in terms of the mean squared error (MSE), at each individual receiver. For a given target MSE, we shall characterize the outage probability, where the outage is defined as the event in which the MSE in signal reconstruction exceeds the target MSE. Due to multiple receivers, the network outage is dominated by the "*bottleneck*" receiver.

Roughly speaking, depending on the computing capability of senders and receivers, there are two possible approaches for transmitting the compressible data: (A) transmit raw measurements or (B) transmit the compressed coefficients. Approach (A) of transmitting raw samples is more applicable to situations where senders are less sophisticated than the receivers, e.g., in a sensor network where a sensor multicasts its sensed observation to multiple coordinating agents. A typical sensor node, being a tiny and battery-operated device, could not afford sophisticated on-board data processing. Further, since several sensor nodes are deployed in the network and many of them are frequently refreshed/replaced, it might become highly costly for each sensor node to be cognizant of the underlying signal structure. In this case, a plausible approach is to transmit the compressed samples, as illustrated in Fig. 4.1(b). It is clear that, in this scheme, the onus of faithful reconstruction of the compressive signal lies with the receiver, and the sender just needs to sample the source signal using an appropriate kernel prior to the transmission. A significant advantage is that a random sampling kernel (or sampling matrix) at the sender often suffices and no knowledge about the source signal's structure is needed. In contrast, in Approach (B), the sender samples the signal, stores it, compresses it in an appropriate domain, and transmits the compressed data to the receiver. In this case, the sender needs to be cognizant of the signal structure and be sufficiently sophisticated to carry out the data compression, e.g., as seen in multimedia streaming in wireless networks and television broadcasting. Obviously, Approach (B) would outperform Approach (A). Nevertheless, Approach (A) has the advantage that it entails no strict requirements on the computational power at the sender.

In our study, we consider both approaches outlined above. First, we analyze the outage performance corresponding to different transmission strategies, and quantify the behavior of the outage probability as a function of key parameters, including the signal structure, channel erasure, and the number of receivers. As expected, when the sender has the capability to sample and compress the signal before transmitting the coefficients, the required number of transmissions can be greatly reduced to meet the outage requirement.

Next, we explore subblock network coding, a new network coding method tailored towards transmitting power-law decay signals. This is motivated by the observation that the traditional network coding could result in undesirable performance for power-law decay signals. Then, we formulate the subblock network coding as an integer programming problem, and develop a heuristic algorithm that exploits the inherent priority structure of the power-law decay signal. We show that by using the proposed subblock network coding scheme, one can gain substantially in terms of reconstruction performance when compared to the traditional network coding.

In related work, recent developments in compressive sensing theory have spurred a wide range of interest in different applications of compressive signals [19, 20, 66]. In particular, there have been some recent advances dealing with the applications of compressive sensing in communications, and the transmission of compressible signals over wireless channels. The transmission of compressible data over wireless channel has been considered in [67, 68] which focus on the estimation of sensor data under a joint source-channel coding framework and analyze the power-latency relationship with the number of nodes. We note that, these works do not consider multicasting and the effect of channel erasures. Along a different avenue, network coding has received much attention since the pioneering work of Ahlswede *et al.* [18]. In particular, in the context of reliable multicast of data over wired networks, Ho *et al.* [69] have proposed a random linear network coding used at an intermediate node to generate the coded packets. By doing this, the authors have shown that the network capacity can be achieved asymptotically. Recently, network coding has also been applied successfully to increase throughput in wireless networks [70, 71]. In these works, by appropriately combining transmission packets, the sender can reduce the number of required transmissions. Also, Nguyen *et*

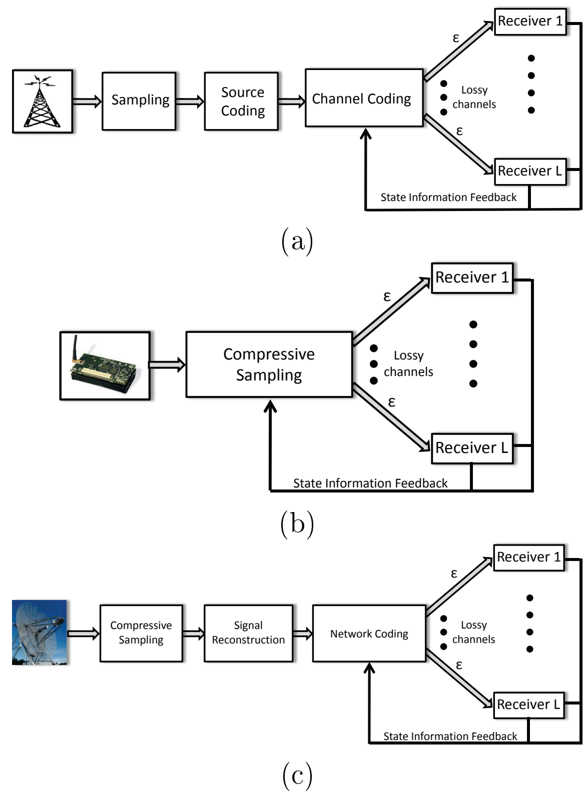

Figure 4.1: Multicast transmission of compressible data (a) conventional method; (b) transmission of compressed measurements (c) proposed joint compressive sensing and network coding.

al. [72] have proposed an XOR-based network coding for single-hop wireless broadcast networks to reduce the number of retransmissions. In contrast to these works, we investigate subblock network coding for transmitting power-law decay signal in multicast networks. The work related most to ours is that of Stankovic *et al.* [73]. However, this work studied only unequal error protection algorithms for progressive image transmission in unicast transmissions.

The organization of this chapter is as follows. We provide in Section 4.2 the system model and some background. In Section 4.3, we investigate in detail a variety of transmission strategies and analyze the corresponding outage performance. In Section 4.4, we explore subblock network coding for power-law decay signals. Numerical results and discussions are presented in Section 4.5. Finally, we draw conclusions in Section 4.6.

4.2 System Model and Background

We consider a multicast network where a sender wishes to transmit compressible signals to L receivers over lossy wireless channels. Let $\mathbf{x} \in \mathbb{R}^N$ be the source signal of interest. For example, $\mathbf{x}$ can be an audio/video signal in the context of multimedia and television networks, or it can be a sensed physical stimulus at a sensor node in the context of sensor networks. The link between the sender and each receiver is lossy; i.e., each transmitted packet to a receiver R_j is subject to an erasure with probability ϵ. We assume that the system is time-slotted and each slot length corresponds to one packet transmission.

Each receiver is required to meet a certain target MSE. We say that an *outage* occurs, at receiver R_j, if the corresponding MSE_j exceeds the target MSE. The outage probability determines the QoS for each receiver. For a given target MSE, say mse, the outage probability is given by

$$P_{\text{out}}^{j} = \mathbb{P}(\mathsf{MSE}_j > \mathsf{mse}), \quad j = 1, \dots, L.$$

For convenience, we say there is a network outage when at least one of the receivers suffers an outage, i.e., $\{\exists j \in \{1, \dots, L\} : \ \mathsf{MSE}_j > \mathsf{mse}\}$. Accordingly, the network outage probability can be given by

$$P_{\text{out}}^{*} = \mathbb{P}(\max_j \mathsf{MSE}_j > \mathsf{mse}).$$

Clearly, the outage performance is governed by the receiver with the worst channel conditions, i.e., the "bottleneck" receiver.

Next, we briefly remark on the structure of the source signal under consideration. Suppose that $\mathbf{x} \in \mathbb{R}^N$, in some $N-$dimensional orthonormal basis $\mathbf{\Phi} = \{\phi\}_{n=1}^{N}, \ \phi_n \in \mathbb{R}^N$, can be represented as

$$\mathbf{x} = \sum_{n=1}^{N} \theta_n \phi_n, \quad \theta_n = \langle \mathbf{x}, \phi_n \rangle.$$

The vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_N]^T$, is the coefficient vector of $\mathbf{x}$ in the basis $\mathbf{\Phi}$. We assume that $\mathbf{x}$ belongs to the class of power-law decay signals, in $\mathbf{\Phi}$, in the sense that its coefficients in this basis $\mathbf{\Phi}$ decay according to a power-law, which is formally defined as follows [19].

Let $\mathbf{x} \in \mathbb{R}^N$, and, for some orthonormal basis $\mathbf{\Phi}$, assume that the coefficients $\{\theta\}_{n=1}^N : \ \theta_n = \langle \mathbf{x}, \phi_n \rangle$ are re-indexed in a descending order such that

$$|\theta_{(1)}| \geq |\theta_{(2)}| \geq \ldots |\theta_{(n)}|.$$

Then, we have the following definition.

Definition 2.1. A signal $\mathbf{x}$ is a power-law decay signal in the orthonormal basis $\mathbf{\Phi}$, if the vector $\boldsymbol{\theta}$ obeys power-law decay, that is,

$$|\theta_{(n)}| \leq R n^{-\frac{1}{p}}, \ \ R > 0, \ \text{for some fixed } p \in (0, 1]. \tag{4.1}$$

The corresponding error due to the best $K-$term approximation, $\mathbf{x}^{(K)} \triangleq \sum_{k=1}^{K} \theta_{(k)} \phi_{(k)}$, is given by

$$\mathsf{MSE}(K) = \left\| \mathbf{x} - \mathbf{x}^{(K)} \right\|^2 = \sum_{n=K+1}^{N} |\theta_{(n)}|^2 \leq C_r K^{-r}, \ \ r = \frac{2}{p} - 1,$$

where $C_r = C_p R^2$ and C_p depends only on p.

Power-law decay signals admit compressibility in the sense that the MSE in their best $K-$term approximation can always be upper bounded by $C_r K^{-r}$.

We now restate an important result of compressive sensing regarding the reconstruction of power-law decay signal $\mathbf{x}$. In assessing the practical significance of power-law decay signals, a natural question arises; i.e., can we reconstruct the power-law decay signal, $\mathbf{x}$, with arbitrarily small reconstruction error with only a few number of measurements of $\mathbf{x}$? It turns out that the answer is positive. An important result by Candes *et al.* [21] asserts that, with only $\mathcal{O}(K \log N)$ random measurements (i.e., measurement matrices are random matrices with i.i.d. entries) from a of $\mathbf{x}$, one can reconstruct $\mathbf{x}$ with $\mathsf{MSE} \leq C_r K^{-r}$. For the sake of completeness, we present the following result [21].

Theorem 4.1. *(Optimal recovery of power-law decay signals from random measurements) Suppose that* $\mathbf{x} \in \mathbb{R}^N$ *is a power-law decay signal in the basis* $\mathbf{\Phi}$ *obeying (4.1) for some fixed* $0 < p < 1$, *and let* $a > 0$ *be a sufficiently small number. Assume that we are given* $M = \mathcal{O}(K \log N)$ *random measurements of* $\mathbf{x}$: $\mathbf{y} = \mathbf{\Psi}\mathbf{x}$, *where* $\mathbf{\Psi}$ *is the* $M \times N$ *measurement matrix. Then with probability* 1, *the minimizer* $\boldsymbol{\theta}^*$ *to the optimization* *problem*

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^N} |\boldsymbol{\theta}|_{\ell_1}$$
$$s.\ t. \qquad \mathbf{y} = \boldsymbol{\Psi\Phi\theta}$$

is unique. Furthermore, with probability at least $1 - \mathcal{O}(N^{-\frac{\rho}{a}})$, *we have*

$$\|\mathbf{x} - \mathbf{x}^*\|^2 \leq C_{p,a} R^2 K^{-r},\ r = 2/p - 1.$$

Here, $C_{p,a}$ *is a fixed constant depending on* p *and* a*, but not on anything else. The implicit constant in* $\mathcal{O}(N^{-\rho/a})$ *is allowed to depend on* a*.*

The above theorem asserts that the power-law decay signals can be "sampled accurately" using the techniques of compressive sensing.

Without loss of generality, we assume that the coefficients of $\mathbf{x}$ follow the order: $|\theta_1| \geq |\theta_2| \geq \ldots \geq |\theta_N|$. Further, we assume that, prior to transmissions, the data are quantized and encapsulated into packets. Since this study aims at obtaining a fundamental understanding of the interplay between compressive sensing and network coding, we assume that data quantization errors are negligible, which could be accounted for further by using the same method in [74].

4.3 Transmission Strategies: Outage Analysis

In this section, we investigate in detail a variety of strategies for multicasting compressible data to many receivers, depending on whether the transmitter is capable of reconstructing the compressively sensed signal.

Transmission of compressed measurements (TCM)

In this method, the sender performs the compressed measurements of the source signal using an $M \times N$ ($M << N$) sensing matrix $\boldsymbol{\Psi}$. Each measurement is quantized with a fine precision and transmitted in the form of a packet. This scenario is of interest where the sender could not afford the sophistication to reconstruct the source signal from compressed measurements via ℓ_1 optimization, and the receiver takes on more processing burden. Further, in this case, the sender does not even need to know the compressible basis of the source signal. This is typical in the case of sensor networks, where the tiny sensor node has less computational power and is oblivious of the structure of the source signal. The system diagram is illustrated in Fig. 4.1(b). In this case, the sender transmits the raw, compressed measurements that are obtained via compressive sensing, as outlined in [21].

Let K be the desired level of sparsity, and $M = c_1 K \log N$ be the corresponding threshold on the number of measurements for reconstruction of the signal with the desired accuracy. Let $\mathbf{y}$ be the M-length measurement vector obtained by sensing $\mathbf{x}$ using $M \times N$ measurement matrix $\mathbf{\Psi}$, i.e., $\mathbf{y} = \mathbf{\Psi}\mathbf{x}$, where $\mathbf{\Psi}$ is chosen such that the solution to the ℓ_1 optimization problem (4.2) will enable a satisfactory reconstruction of the signal $\mathbf{x}$ in the sense of Theorem 4.1. It has been shown in [21] that this is possible as long as the measurement matrix, $\mathbf{\Psi}$, obeys a *Uniform Uncertainty Principle* (UUP) and *Exact Reconstruction Principle* (ERP). Matrix $\mathbf{\Psi}$ should be chosen so as to capture sufficient information with a high probability. More importantly, it is easy to find such a projection matrix $\mathbf{\Psi}$, since easily constructible ensembles like binary, Gaussian and Fourier ensembles obey both UUP and ERP with high probability. Thus, in our case, we assume that $\mathbf{\Psi}$ is generated by drawing its elements, independently and identically, from a zero mean Gaussian distribution, i.e.,$[\mathbf{\Psi}]_{i,j} \sim \mathcal{N}(0, \frac{1}{N})$.

Let $M > M_{cs}$ be the number of measurements transmitted. Let X_m^j be a random variable representing the event of loss of the mth packet at the jth receiver. X_m^j is a binary random variable given by $X_m^j \sim Bin(\epsilon)$. Let $N_j = \sum_{m=1}^{M} X_m^j$, i.e., the number of samples lost due to erasure at receiver R_j. Then, N_j is a binomial random variable with parameters (M, ϵ). Observe that the outage is the event when a receiver loses more than $M - M_{cs}$ samples, i.e.,

$$P_{\text{out}}^j = \mathbb{P}\left(N_j > M - M_{cs}\right).$$

Next, we evaluate the outage probability of the "bottleneck" receiver:

$$P_{\text{out}}^* = \mathbb{P}\left(\max_j N_j > M - M_{cs}\right).$$

Since M is large, it is reasonable to approximate N_j as a Gaussian random variable.

Define $N_j^* = \frac{N_j - \mathbb{E}[N_j]}{\sqrt{Var(N_j)}}$. It follows that, $N_j^* \sim \mathcal{N}(0,1)$. Then, we have

$$\begin{aligned} P_{\text{out}}^* &= \mathbb{P}\left(\max_j N_j \geq M - M_{cs}\right) \\ &= 1 - \mathbb{P}\left(\max_j N_j^* \leq M_1\right), \end{aligned} \tag{4.2}$$

where $M_1 = \dfrac{M(1-\epsilon) - M_{cs}}{\sqrt{\epsilon(1-\epsilon)M}}$.

Next, we need the following result from extreme value theory [75]: Let $N_1^*, \ldots N_L^*$ be i.i.d. with $N_1^* \sim \mathcal{N}(0,1)$, a_L and b_L be two sequences of numbers such that $a_L \to \infty$, $b_L \to 0$, as $L \to \infty$. In particular,

$$a_L = \sqrt{2\log L} - \frac{1}{2}(2\log L)^{-\frac{1}{2}}\left(\log(4\pi) + \log\log L\right),$$

$$b_L = (2\log L)^{-\frac{1}{2}}.$$

Then,

$$\frac{\max_j N_j^* - a_L}{b_L} \overset{\mathcal{D}}{\Longrightarrow} \zeta,$$

where $\overset{\mathcal{D}}{\Longrightarrow}$ denotes convergence in distribution, and $\zeta(x) = \exp(-e^{-x})$ is the Gumbel CDF. Thus, we conclude that, as $L \to \infty$,

$$\mathbb{P}\left(\max_j N_j^* \leq M_1\right) \Longrightarrow \zeta\left(\frac{M_1 - a_L}{b_L}\right).$$

Based on (4.2), we obtain the outage probability as

$$P_{\text{out}}^* = 1 - \zeta\left(\frac{M_1 - a_L}{b_L}\right).$$

For convenience, define $M_2 \triangleq \dfrac{M_1 - a_L}{b_L}$. Equivalently,

$$M_2 = M_1\sqrt{2\log L} - 2\log L - \frac{1}{2}\log(4\pi\log L).$$

Then, we can represent the outage probability as

$$P_{\text{out}}^* = 1 - \exp(-e^{-M_2}). \tag{4.3}$$

After some algebra, it follows that

$$P_{\text{out}}^* \approx 1 - \exp\left(-e^{-\left(\frac{1-\epsilon-\frac{M_{cs}}{M}}{\sqrt{\epsilon(1-\epsilon)}}\sqrt{2M\log L} - 2\log L\right)}\right). \tag{4.4}$$

Observe that, in many practical scenarios, M is larger than L. We can then approximate (4.4) as

$$\boxed{P_{\text{out}}^{*} \approx \exp\left(-\left[\frac{1-\epsilon-\frac{M_{cs}}{M}}{\sqrt{\epsilon(1-\epsilon)}}\sqrt{2M\log L} - 2\log L\right]\right).} \tag{4.5}$$

Observe that, for a fixed value of L, P_{out}^{*} decays exponentially with the square root of the number of samples M.

Transmission of coefficients (TC)

In this transmission strategy, the sender performs compressed sensing of $\mathbf{x}$ via measurement matrix $\mathbf{\Psi}$, and then estimates the K largest elements of the coefficient vector $\boldsymbol{\theta}$. That is to say, the sender is assumed to be capable of bearing the computational burden of signal reconstruction via compressed measurements. Furthermore, each coefficient is quantized and encapsulated in a packet. There are a variety of ways in which these coefficients can be protected against the channel erasure. To address this issue, we consider several coding methods for the transmission of coefficients.

Round robin transmission of coefficients (TC-RR):

In this "naive" scheme, the sender transmits Kt packets with K coefficients being transmitted in a round-robin fashion with t repetitions. We assume that $\mathcal{O}(K\log N)$ measurements are made at the sender to obtain a K-term approximation for $\mathbf{x}$. Without loss of generality, we assume that the coefficients $\theta_1,\ldots,\theta_K$ are transmitted t times each.

For ease of exposition, first consider the case when $t=1$. The number of packets received by the receiver R_j is given by $N_j = \sum_{k=1}^{K} X_k^j$. Further, the kth packet received is denoted by $\hat{\theta}_k^j = X_k^j\theta_k$. The MSE due to this scheme can then be upper bounded as follows:

$$\begin{aligned}
\mathsf{MSE}_j &= \sum_{k=1}^{N} |\theta_k - \hat{\theta}_k^j|^2 \\
&\leq \sum_{k=1}^{K} |\theta_k - X_k^j\theta_k|^2 + C_r K^{-r}.
\end{aligned}$$

Observe that

$$|\theta_k - X_k^j\theta_k|^2 = \begin{cases} 0 & \text{w. p. } 1-\epsilon \\ |\theta_n|^2 & \text{w. p. } \epsilon. \end{cases}$$

Next, we evaluate the outage performance. We are particularly interested in the "bottleneck" receiver. Therefore,

$$\begin{aligned} P^*_{\text{out}} &= \mathbb{P}\left(\max_j \mathsf{MSE}_j > \mathsf{mse}\right) \\ &= 1 - \left(\mathbb{P}\left(\mathsf{MSE}_j \leq \mathsf{mse}\right)\right)^L. \end{aligned}$$

Consider the desired threshold $\mathsf{mse} = C_r K^{-r}$. This yields

$$\begin{aligned} P^*_{\text{out}} &= 1 - \left(\mathbb{P}\left(\sum_{k=1}^{K} |\theta_k - X_k^1 \theta_k|^2 = 0\right)\right)^L \\ &= 1 - (1-\epsilon)^{KL}. \end{aligned}$$

It is clear that when each coefficient is transmitted repeatedly for $t > 1$ slots, the equivalent packet erasure probability is ϵ^t (this is a naive way of introducing redundancy). We obtain the network outage probability as

$$P^*_{\text{out}} = 1 - \left(1 - \epsilon^t\right)^{KL}. \tag{4.6}$$

Since ϵ^t is very small, using binomial expansion, (4.6) can be approximated as

$$\boxed{P^*_{\text{out}} \approx KL\epsilon^t.} \tag{4.7}$$

Observe that, for a fixed value of L and t, P^*_{out} decays geometrically with t.

Transmission of coefficients by random network coding (TC-RNC)

In this scheme, the sender wishes to transmit K coefficients to L receivers. Again, we assume that each coefficient is encapsulated into one packet. The sender uses random network coding [76] for encoding the packets. In particular, each transmitted packet C is generated by a linear combination of the K original packets, $C = \sum_{i=1}^{K} \alpha_i c_i$, where α_i denote coefficients drawn randomly from a large finite field. We assume the field size is large enough that the generated coded packets are independent. In order to recover the original packets, a receiver needs to receive at least K coded packets.

Assume that Kt packets, with $t > 1$, are transmitted by performing network coding on K coefficients. Similar to Section 4.3, we let N_j be the number of samples lost due to erasure at receiver R_j; i.e., $N_j = \sum_{m=1}^{Kt} X_m^j$. Consequently, N_j is a binomial random variable with parameters (Kt, ϵ). Since Kt is large, we have the approximation $N_j \sim \mathcal{N}\left(Kt\epsilon, \epsilon(1-\epsilon)Kt\right)$. Again, appealing to extreme value theory, we have the network outage probability as

$$P_{\text{out}}^{*} \approx 1 - \exp\left(-e^{-\left(\frac{1-\epsilon-\frac{1}{t}}{\sqrt{\epsilon(1-\epsilon)}}\sqrt{2Kt\log L}-2\log L\right)}\right). \tag{4.8}$$

Again, (4.8) can be approximated as

$$\boxed{P_{\text{out}}^{*} \approx \exp\left(-\left[\frac{1-\epsilon-\frac{1}{t}}{\sqrt{\epsilon(1-\epsilon)}}\sqrt{2Kt\log L}-2\log L\right]\right).} \tag{4.9}$$

We note that, for a fixed value of L and K, P_{out}^{*} decays exponentially with the number of transmitted packets, Kt.

Remarks: Observe that, network coding requires at least K packets to recover any information regarding the data, and it would incur complete information loss otherwise. In contrast, for the round-robin method, even when the number of packets received is less than K, the recovery is still possible, but with a higher MSE. Thus, when the quantity M/K, the ratio of the number of transmitted packets to that required, is small, round-robin can outperform network coding in terms of average MSE. With this observation, we propose "subblock network coding" that amalgams the robustness of network coding with the redundancy present in the round-robin scheme. Simply put, subblock network coding exploits the power-law decay nature of the signal coefficients, partitions the coefficient vector accordingly, and performs random network coding within each subblock. We detail this subblock network coding in the following section.

4.4 Network Coding for Power-law Decay Signals

Traditional network coding for power-law decay signals

It is well known that network coding can help to improve the network bandwidth efficiency in wireless multicast scenarios [77]. By combining the data before transmitting them out, the sender may reduce the required number of transmissions. In this study, we consider power-law decay signals which yield unequal priority data. Observe that, different coefficients contribute differently to the signal distortion, indicating that combining the coefficients blindly without considering their priorities would lead to undesirable performance. In contrast, subblock network coding takes into account the priority of the data and appropriately allocates time slots to each subblock. To get a more concrete sense, we present the following example for illustration.

Example: Consider a transmission scenario in which the sender has two packets a and b that it wants to deliver to a receiver. Assume that packet a is more important than packet b. This happens in multimedia transmissions; e.g., an I frame is more important than a P or B frame. Furthermore, assume that the losses of packets a and b, respectively, contribute 75% and 25% to the total distortion. We assume that the packet erasure probability is ϵ, and the sender has 4 time slots for transmissions. If the sender uses random network coding blindly, it simply combines these two packets to generate coded packets and transmits them in the four time slots. In order to recover the transmitted data, the receiver needs to receive at least 2 coded packets (given that the finite field is large enough so that all the coded packets are independent). In this case, the expected distortion of the received signal at the receiver is given by

$$\mathbb{E}(D_1) = \epsilon^4 + 4\epsilon^3(1-\epsilon). \tag{4.10}$$

Now, if the sender takes into account the priorities of the packets, instead of combining the two packets together, it allocates 3 time slots for transmitting packet a and 1 time slot for transmitting packet b. Therefore, in this case, the expected distortion at the receiver is given by

$$\mathbb{E}(D_2) = \frac{3\epsilon^3 + \epsilon}{4}. \tag{4.11}$$

From (4.10) and (4.11), we observe that $\mathbb{E}(D_1) > \mathbb{E}(D_2)$ when $\epsilon > 1/3$.

As shown above, the traditional network coding method may result in suboptimal performance for power-law decay signal. On the other hand, better performance is attainable by appropriately fragmenting the transmitted data into subblocks and carrying out network coding within each subblock. In the following subsections, we explore subblock network coding for power-law decay signals.

Recall that the first K coefficients of $\mathbf{x}$ have the order $|\theta_1| \geq |\theta_2| \geq \cdots \geq |\theta_K|$. Let T be the total number of time slots available for transmissions. Without loss of generality, we assume that K ordered coefficients are partitioned into d subblocks. Let m_i be the number of coefficients in subblock i; thus, $\sum_{i=1}^{d} m_i = K$. The problem of subblock network coding now boils down to finding an optimal allocation of time slots for subblocks, such that the average MSE of the network is minimized. Let $\pi = [t_1, t_2, \ldots t_d]$ be a time slot allocation policy for d unequal priority subblocks. We then have $\sum_{i=1}^{d} t_i = T$.

Let $\triangle D_i$ be the MSE incurred due to the loss of subblock i. With $M_i = \sum_{j=1}^{i} m_j$, it follows that

$$\triangle D_i = R^2 \sum_{l=M_{i-1}+1}^{M_i} l^{-2/p}, \tag{4.12}$$

where R and p are constants. Furthermore, assume that the sender applies random network coding within each subblock. Let $\mathbb{P}(i)$ be the probability that subblock i is lost at a receiver. We have

$$\mathbb{P}(i) = \sum_{j=0}^{m_i-1} \binom{t_i}{j} \epsilon^{t_i-j}(1-\epsilon)^j. \tag{4.13}$$

For a given allocation policy π, the average MSE is

$$\begin{aligned} \overline{\mathsf{MSE}}(\pi) &= \sum_{i=1}^{d} \triangle D_i \mathbb{P}(i) \\ &= R^2 \sum_{i=1}^{d} \sum_{l=M_{i-1}+1}^{M_i} \sum_{j=0}^{m_i-1} l^{-2/p} \binom{t_i}{j} \epsilon^{t_i-j}(1-\epsilon)^j. \end{aligned}$$

To minimize $\overline{\mathsf{MSE}}(\pi)$ over the set $\mathcal{T}$ of all feasible time slot allocation policies, we have the following optimization problem:

$$\begin{aligned} &\min_{\pi \in \mathcal{T}} \left\{ \overline{\mathsf{MSE}}(\pi) \right\} \\ \text{s.t.} \quad &\sum_{i=1}^{d} m_i = K \\ &\sum_{i=1}^{d} t_i = T \\ &m_i, t_i \geq 0 \quad i = 1, \ldots, d \\ &m_i, t_i \text{ are integers} \quad i = 1, \ldots, d, \end{aligned}$$

This is an integer programming problem with a large search space, and its optimal solution is often intractable. In what follows, we propose a heuristic greedy algorithm.

Two-step heuristic greedy algorithm

In this subsection, we explore an algorithm that greedily exploits the inherent priority (weight) structure embedded in power-law decay signals. Specifically, the algorithm has two steps: coarse allocation and refinement.

Coarse allocation: In this step, each subblock is allocated a number of time slots, proportional to its weight. Roughly speaking, the larger a subblock's weight, the more protection it needs. However, this procedure may end up with overly prioritizing subblocks with higher weights, while leaving the other subblocks unprotected. This calls for an allocation refinement.

Allocation refinement: In this step, the time slot allocation is refined greedily by exploiting an inherent property of the power-law decay signal, summarized as follows. Let $\pi^0 = [t_1^0, t_2^0, \ldots, t_d^0]$ be the time slot allocation that is obtained in the first step. In order to find a refined allocation, we start from the first subblock t_1^0, i.e., the subblock with the highest priority coefficients.

Algorithm 1 : Time Slot Allocation Algorithm (TSA).

Input: d, m_i, θ_i, K, T, ϵ.
Output: $\pi^* = [t_1^*, t_2^*, \ldots, t_d^*]$

STEP 1: (Coarse allocation) The number of time slots allotted to subblock i is proportional to its weight over the total weight of all K coefficients.
$\pi^0 := [t_1^0, t_2^0, \ldots, t_d^0]$
for $i \leftarrow 1$ to $d-1$ **do**
  $M_i := \sum_{j=0}^{i} m_i$
  $t_i^0 \leftarrow \lfloor \frac{\sum_{l=M_{i-1}+1}^{M_i} |\theta_l|}{\sum_{j=1}^{K} |\theta_j|} \times T \rfloor$
end for
$t_d^0 \leftarrow T - \sum_{i=1}^{d-1} t_i^0$
$\mathsf{MSE}^* \leftarrow \mathbb{E}[\mathsf{MSE}(\pi^0)]$
STEP 2: (Allocation refinement) Move one time slot at a time from t_i to t_{i+1} or vice versa until it increases the average MSE or $t_i = m_i$.
for $i \leftarrow 1$ to $d-1$ **do** {For each subblock i}
  if adding one more slot to the current subblock reduces $\mathbb{E}[\mathsf{MSE}]$ **then**
    while $t_{i+1} > m_{i+1}$ **do**
      $\pi^0 \leftarrow [t_1, t_2, \ldots, t_i + 1, t_{i+1} - 1, \ldots, t_d]$ {Move one time slot from the subblock i to subblock $i+1$}
      if $\mathsf{MSE}^* > \mathbb{E}[\mathsf{MSE}(\pi^0)]$ **then**
        $\mathsf{MSE}^* \leftarrow \mathbb{E}[\mathsf{MSE}(\pi^0)]$
        $\pi^* \leftarrow \pi^0$
      else
        $\pi^0 \leftarrow [t_1, t_2, \ldots, t_i + 1, t_{i-1} - 1, \ldots, t_d]$
        break;
      end if
    end while
  else
    while $t_i > m_i$ **do**
      $\pi^0 \leftarrow [t_1, t_2, \ldots, t_i - 1, t_{i+1} + 1, \ldots, t_d]$ {Move one time slot from the subblock i to subblock $i+1$}
      if $\mathsf{MSE}^* > \mathbb{E}[\mathsf{MSE}(\pi^0)]$ **then**
        $\mathsf{MSE}^* \leftarrow \mathbb{E}[\mathsf{MSE}(\pi^0)]$
        $\pi^* \leftarrow \pi^0$
      else
        $\pi^0 \leftarrow [t_1, t_2, \ldots, t_i + 1, t_{i-1} - 1, \ldots, t_d]$
        break;
      end if
    end while
  end if
end for

We first examine whether increasing or decreasing the number of time slots allocated for the first subblock reduces the expected MSE. For example in the case of decreasing, if $t_1^0 > m_1$, then we move one time slot from t_1^0 to t_2^0. Now, the new time slot allocation is $\pi^1 = [t_1^0 - 1, t_2^0 + 1, \ldots, t_d^0]$. If the new time slot allocation π^1 reduces the expected MSE of the network, i.e., $\mathbb{E}[\mathsf{MSE}(\pi^0)] > \mathbb{E}[\mathsf{MSE}(\pi^1)]$, then we record π^1 as the new allocation. However, if that is not the case, the allocation procedure is carried out in the reversed direction by adding one more time slot to the first subblock from the second one. Once the appropriate direction is found, the procedure is repeated in the same manner until $t_1^0 = m_1$, or until the new time slot allocation increases the expected MSE of the network. The same procedure is applied to $t_2^0, t_3^0, \ldots, t_{d-1}^0$. The pseudo-code for this greedy algorithm is provided in Algorithm 1.

Time complexity analysis: We now analyze the time complexity of the greedy algorithm. As indicated in the algorithm, given the number of subblocks d, there are two steps: coarse allocation and refinement. In the coarse allocation, the algorithm takes $\mathcal{O}(d)$ to allocate time slots to all subblocks. In the second step involving allocation refinement, the algorithm loops d times, and in each loop i, it runs at most $T - m_i$ times for moving time slots to the next highest priority subblock. The runtime of this step is $\mathcal{O}(Td)$. Consequently, the running time of the TSA algorithm is $\mathcal{O}(Td)$. Since we assume that the sizes of all subblocks are equal, there are at most $2\sqrt{K}$ possibilities that we can divide K packets into equal size subblocks. Also, we have the number of subblocks $d \leq K/2$. Thus, the running time of the greedy algorithm is upper bounded by $\mathcal{O}(K^{1.5}T)$.

4.5 Numerical Example and Discussions

In this section, we illustrate the performance gain of the subblock network coding via a numerical example. We assume that each coefficient is encapsulated into one packet, i.e., hereinafter, "coefficient" and "packet" are used inter-changeably.

Basic setup

For the sake of comparison, we introduce two more schemes: TC-SysRNC and TC-PRO. Intuitively, the TC-SysRNC scheme is of interest because it would reduce the average MSE compared to the TC-RNC scheme by allowing the receivers to recover partial data. Whereas the TC-PRO scheme is an intuitive time slot allocation method for unequal priority data. The transmission protocols of these two new schemes are described as follows.

- *Transmission of coefficients with systematic random network coding* (TC-SysRNC): In this scheme, the transmission is divided into two phases: the basic phase and the augmentation phase. In the basic phase, all K original packets are transmitted. Next, in the augmentation phase, the sender transmits the coded packets generated as in the TC-RNC scheme. The receivers, which might have lost some original packets, can receive more coded packets to recover their own data by solving the system of linear equations [76]. Note that, because packets are not mixed in the basic phase, a receiver is able to obtain partial of the data, even if it cannot recover all the mixed packets in the augmentation phase.

- *Transmission of coefficients with proportional time slot allocation* (TC-PRO): In this scheme, each of the K coefficients is allocated time slots, proportional to its weight. For example, let $|\theta_i|$ be the weight of coefficient θ_i. Then the number of time slots allocated to θ_i is $t_i = \lfloor T|\theta_i|/\sum_j |\theta_j| \rfloor$. This allocation procedure starts from the largest weight coefficient to the smallest weight coefficient, or stops when there is no available time slot. The coefficients are transmitted repeatedly in their allocated time slots. This transmission scheme coarsely exploits the structure of the power-law decay signal.

Before delving into the details, we should note that the TCM scheme requires more bandwidth (by a factor of $\mathcal{O}(\log N)$) compared to the schemes that transmit coefficients. Clearly, this is an unfair comparison since the two transmission strategies cater to different scenarios, as discussed before. In particular, the TCM scheme is used in scenarios where the sender has limited resources. In contrast, the schemes that transmit coefficients are applied when the sender is more sophisticated. In this section, we evaluate and compare the performances of the schemes that transmit coefficients only.

In our example, we compare the proposed subblock network coding schemes with TC-RR, TC-RNC, TC-SysRNC, and TC-PRO, in terms of outage probability and the average MSE. We consider the multicast of a power-law decay signal of length $N = 1000$

to $L = 500$ receivers. Assume that the number of time slots available for transmissions is $T = 40$. For simplicity, we set $R = 1$, $C_p = 1$, and $p = 0.9$. Let the desired level of sparsity be $K = 30$. That is, we set the target mean squared error as the best K-term approximation, $\mathsf{mse} = C_r K^{-r}$. We assume that the erasure channels between the sender and the receivers are independent, and all have the same erasure probability ϵ. The average MSE is determined as the mean of the average MSEs across all receivers over 1000 trials.

Numerical results

First, we investigate the impact of packet erasure probability on the network outage of different schemes in Fig. 4.2. As expected, the network outage of each scheme increases with ϵ. This is because there are more packet losses due to higher erasure probability. We observe that TC-PRO and TC-RR schemes perform worse. In particular, the TC-PRO scheme, which overprotects the higher priority coefficients while leaving many lower priority coefficients unprotected, results in the worst performance. It is clear that the subblock network coding scheme, TC-SubRNC, outperforms the non-subblock network coding schemes with considerable margin. Intuitively, by using network coding within each subblock, the sender reduces the overwhelming redundancy in the received data, while still ensuring appropriate protection for each subblock.

Next, Fig. 4.3 depicts the normalized average MSE of the different schemes with respect to ϵ. The normalized average MSE is defined as the ratio of the average MSE to the total signal power. First, we note that a scheme with smaller normalized average MSE is not necessarily a scheme with smaller outage. This is because the network outage is dominated by the bottleneck receiver. Next, as seen in Fig. 4.3, the normalized average MSE of the TC-RNC scheme increases significantly with ϵ, especially, when $\epsilon > 0.15$. Our intuition is that in the TC-RNC scheme, the sender treats all coefficients uniformly when combining them to generate coded packets. Consequently, at the receiver, in order to recover the transmitted coefficients, it needs to receive at least K coded packets. This condition, however, would not be satisfied in the high-erasure regime, i.e., $\epsilon > 0.15$. In contrast, the TC-SubRNC scheme divides the transmitted coefficients into smaller subblocks and appropriately allots time slots to each subblock. As a result, the larger weight coefficients would be received successfully with higher probability, leading to only a slight increase in the average MSE.

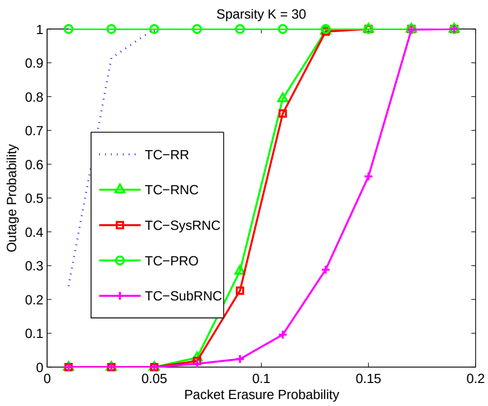

Figure 4.2: P^*_{out} vs. ϵ.

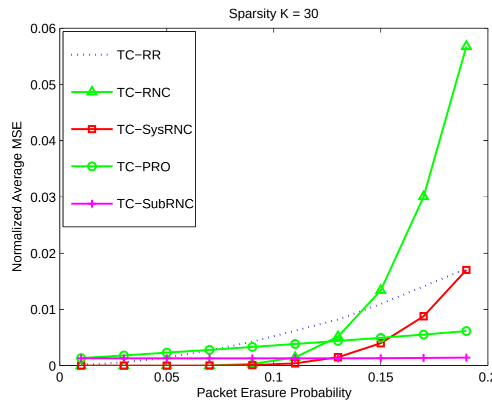

Figure 4.3: Normalized Avg. MSE vs. ϵ.

4.6 Conclusions

In this chapter, we considered transmission scenarios in which the sender multicasts a compressively sampled signal to many receivers over lossy wireless channels. First, using extreme value theory, we quantified the network outage for both cases where the transmitter may or may not be capable of reconstructing the compressively sampled signals. We showed that when the sender can reconstruct the compressively sampled signals, the strategy of using network coding to multicast the reconstructed signal coefficients can reduce the network outage significantly. Next, we showed that the traditional network coding could result in suboptimal performance with the power-law decay signals. With this insight, we devised a new method, namely subblock network coding, which involves fragmenting the data into subblocks, and allocating time slots to different subblocks, based on their priorities. We formulated the corresponding optimal allocation as an integer programming problem, and developed a heuristic algorithm that exploits the inherent priority structure of the power-law signals. We showed that by using subblock network coding one can gain substantially in terms of network outage compared with the other schemes. We are currently exploring the Markov Chain Monte Carlo method to quantify the potential gain of subblock network coding.

To the best of our knowledge, there has been little work on multicasting compressible signals, and this study here presents some initial steps to understand the interplay between compressive sensing and network coding.

Chapter 5

DIGITAL RELAYING VERSUS ANALOG RELAYING: A SUFFICIENT CONDITION FOR OPTIMALITY

5.1 Introduction

Recently, wireless sensor networks (WSNs) have attracted much attention of the research community. Due to their high flexibility, enhanced surveillance coverage, robustness, mobility, and cost effectiveness, WSNs have wide applications and high potential in military surveillance, security, monitoring of traffic, and environment. Usually, a WSN consists of a large number of low-cost and low-power sensors, which are deployed in the environment to collect observations and preprocess the observations. Each sensor node has limited communication capability that allows it to communicate with other sensor nodes via a wireless channel. Often, there is a fusion center that processes data from sensors and forms a global situational assessment.

The ability to detect events of interest is a key capability of the sensor network technology. Detection is the "kick-start" procedure for the operation of any sensor network. Indeed, the physical attributes of a target, like its position and velocity, can be ascertained only after having detected its presence. Furthermore, in some applications such as surveillance, industrial monitoring etc., the detection of an intruder or a hazard is the prime goal.

The decision-making problem, where each sensor sends to the fusion center a summary of its own observations, in the form of a message taking values from a finite alphabet, is termed decentralized detection. Decentralized detection scenarios for sensor networks typically entail geographically dispersed sensors that collect observations about an "event" of interest and transmit information about these individual observations to a fusion center. The fusion center produces an estimate about the event, based on the data it receives from the sensors. If the event is binary with a known prior probability distribution, the problem falls into the Bayesian detection framework, and the probability of error at the fusion center is a typical metric of performance. When no prior distribution on the event is available, Neyman-Pearson detection seeks to minimize the probability of false alarm subject to a constraint on the probability of missed detection [78]. A key challenge in decentralized detection is to come up with decision rules at the sensors, and fusion rules at the fusion center so as to optimize the detection performance at the fusion center.

Wireless sensor nodes are typically tiny devices powered by batteries, and hence are subject to stringent constraints on the resources such as, bandwidth and power. To design an efficient system for detection in sensor networks, it is imperative to understand the interplay between data processing at the sensor nodes, resource allocation, and overall performance in distributed sensor systems. A clear understanding of this interplay can be leveraged to gain insight into the efficient design of sensor networks. Decentralized detection has received much attention in the literature due to its importance in the event driven sensor networks. In-depth treatments of this field can be found in the work of Tsitsiklis [79], Viswanathan *et al.* [80] and Blum *et al.* [81], and the references therein. A recent survey on channel-aware distributed detection, where sensors communicate to the fusion center over noisy (and fading) links, can be found in [82].

Early treatments of distributed detection adopted various simplifying assumptions, particularly about independence of observations at the sensors, and about communication between the sensor nodes and the fusion center, which was often assumed to be perfect. When these assumptions hold, the performance of a decentralized system is optimized by each sensor transmitting its likelihood ratio to the fusion center. The fusion center forms a global likelihood ratio from the product of these and achieves performance equivalent to what would be possible if all the raw sensor data were available at the fusion center. If the communication links between the sensors and the fusion center are imperfect, performance may actually be improved by quantizing the individual sensor likelihood ratios prior to transmission to the fusion center. Some observations along this line have been made by other authors in various contexts. In [83], for example, it is observed that analog sensors perform better than their digital counterpart below a certain threshold SNR in large sensor networks where the metric of performance is the Chernoff exponent. In [84], a distributed detection method based on the method of types is considered. Again in this setting, it is observed that hard-decision fusion

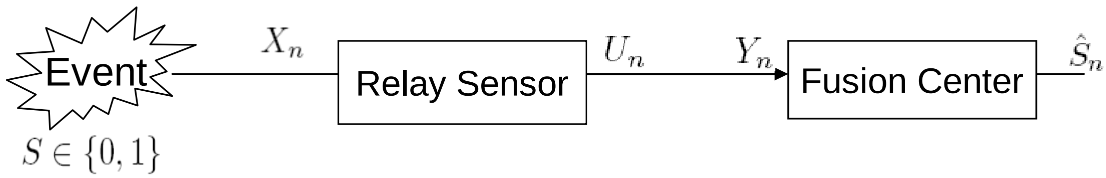

Figure 5.1: Block diagram of a typical sensor relay network.

(digital relaying) can outperform soft-decision fusion (analog relaying) in the single sensor case. In this chapter, comparison of analog and digital relay schemes is undertaken from a classical decision theoretic perspective for a simple network model consisting of a single sensor communicating with the fusion center over an additive Gaussian noise channel. A Bayesian situation is assumed and probability of error at the fusion center is the performance metric. In this setting, a sufficient condition for superiority of digital relaying, given a fixed average transmit power, is derived analytically in terms of the observation SNR at the sensor and the SNR of the communication channel between the sensor node and the fusion center.

Although the network model used here is highly simplified, it is hoped that this work will contribute to a foundation for analysis of more realistic scenarios, leading to advances in sensor placement strategies and inference algorithms in sensor networks as well as the design of wireless relay networks [85] in which a relay assists communication between a source-destination pair.

The rest of this chapter is organized is as follows. Section 5.2 presents the system model and assumptions. Section 5.3 introduces the relaying methods and presents the mathematical framework. Section 5.4 compares the performance of relaying schemes. Section 5.5 discusses simulation results, and some concluding remarks are given in Section 5.6.

5.2 System Model

We consider a single sensory relay network, consisting of a relay-sensor and fusion center, with a power constraint at the relay. The system is as shown in the Fig. 5.1. The prime objective of the network is to make a decision on a binary event at the fusion center with minimum detection error. The job of relay sensor is to assist the fusion center in making a decision on the event. We assume that the underlying phenomenon is an equally-likely binary event, $S \in \{0, 1\}$. In terms of null-hypothesis (H_0) and target-hypothesis (H_1), we write

$$H_0 : S = 0, \quad P(H_0) = 0.5$$
$$H_1 : S = 1, \quad P(H_1) = 0.5.$$

The sensor observes a noisy stochastic process X_n associated with the event. We represent the observation process at the sensor as

$$X_n = hS_n + Z_n, \; n = 1, \ldots, \tag{5.1}$$

where h represents the scaling factor accounting for the uncertainties in the amplitude of the event sensed by the sensor, and Z_n is an i.i.d additive Gaussian process with $Z_n \sim \mathcal{N}(0, \sigma_z^2)$. We assume that the sensor has knowledge of h, and hence, the sensor observations are independent conditioned on the hypothesis (S_n) and h. Upon observing, the sensor processes X_n to obtain a statistic U_n, which is relayed to the fusion center, adhering to an average transmit power constraint of P. The local processing rule for the observations at the sensor node, denoted by γ, is defined as

$$\gamma : \mathcal{X} \to \mathcal{U}, \tag{5.2}$$

where $\mathcal{X}$ is the observation space and $\mathcal{U}$ denotes the output space of the sensor. Due to the transmit power constraint P at the sensor, it is clear that $\mathbb{E}[|U|^2] \leq P$, for all $U \in \mathcal{U}$. Observe that the processing rule at the sensor node, γ, depends on the relaying scheme.

The relay sensor, after computing a statistic U_n from its observations, transmits U_n to the fusion center over a wireless link for decision making. We assume that fusion center has no access to the observations of the event, and has to make a decision solely based on the information sent from the relay. We represent the signal received by the fusion center as

$$Y_n = gU_n + W_n, \; n = 1, \ldots, \tag{5.3}$$

where g represents channel attenuation, and W_n is an i.i.d. additive Gaussian process with $W_n \sim \mathcal{N}(0, \sigma_w^2)$. We assume that the fusion center is cognizant of its channel state g, and hence the observations are conditionally independent given g and the hypothesis. The fusion center has to make a decision $\hat{S}_n$ based on the information relayed by the sensor. The objective is to minimize the detection error under the transmit power constraint at the relay. In what follows, we formulate and analyze the problem in hand under classical binary hypothesis testing framework, and develop a fair idea about the relaying schemes. We assume that the detection is performed on a sample-by-sample basis. In what follows, for presentational convenience, we omit the use of temporal subscript n.

5.3 Relaying Schemes

Upon obtaining the noisy observations of the event, the sensor is posed with the problem of computing a statistic (that is, to design γ) so as to aid fusion center in decision making. The senors can be "digital" or "analog" depending on the processing schemes. A digital sensor may quantize its observations and communicate its decisions in the form of bits, while an analog sensor may compute the sufficient statistics of its observations, which is a real number, and may transmit it as an analog signal. Assessing the performances of these schemes in different regimes of network parameters is key to the design of sensor and relay networks. One wishes to deploy analog or digital sensors depending on which scheme is superior in a given regime.

Digital relaying (detect and forward)

The relay sensor performs binary hypothesis testing on the observations and the decision is conveyed in the form of a bit. Thus, the decision variable U is binary. That is, with the power constraint, $U \in \{-\sqrt{P}, \sqrt{P}\}$. We express the likelihood function at the sensor as

$$\begin{aligned} H_0 : X &\sim f_X(z_i|H_0, h), \\ H_1 : X &\sim f_X(z_i|H_1, h). \end{aligned}$$

where

$$f_X(x|H_0,h) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp -\frac{x^2}{2\sigma_z^2};$$
$$f_X(x|H_1,h) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp -\frac{(x-h)^2}{2\sigma_z^2}.$$

From the fundamentals of detection theory [78], it is clear that the optimal processing rule is given by

$$U = \gamma(X) = \begin{cases} +\sqrt{P} & \text{if} \quad f_X(x|H_1,h) \geq f_X(x|H_0,h) \\ -\sqrt{P} & \text{if} \quad f_X(x|H_1,h) < f_X(x|H_0,h). \end{cases} \tag{5.4}$$

More succinctly, we have the sensor's processing rule as

$$\gamma(X) = \sqrt{P}\left[2\mathbf{I}(X \geq h/2) - 1\right], \tag{5.5}$$

where $\mathbf{I}(\cdot)$ is the indicator function.

Next, consider the fusion rule at the fusion center. The fusion center receives Y, a noisy version of the statistic U communicated by sensor over the noisy link. It is faced with the task of making decision $\hat{S}$, on S, based on the received signal Y. Observe that the distribution of Y, and therefore the fusion rule, depend on the decision rule γ, of the sensor. Therefore, to derive the decision rule at the fusion center, it is essential first to evaluate the detection error probabilities at the sensor. To this end, we define $\mathtt{SNR}_o \triangleq \frac{1}{4}\frac{h^2}{\sigma_z^2}$, as the observation SNR, which is the "signal-to-noise ratio" of the observations at the sensor. Then, for a given $\mathtt{SNR}_o$, we note that the probability of error at the sensor, p, is given by:

$$p = P(X \geq h/2 | S = 0, \mathtt{SNR}_o) = Q\left(\sqrt{\mathtt{SNR}_o}\right), \tag{5.6}$$

where $Q(\cdot)$ is the $Q-$function given by $Q(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{u^2}{2}} du$. We note that the received signal Y obeys:

$$\begin{aligned} H_0: \quad & Y = -\sqrt{P}g + W \ \ w.p. \ 1-p \\ & Y = \sqrt{P}g + W \ \ w.p. \ p \\ H_1: \quad & Y = \sqrt{P}g + W \ \ w.p. \ 1-p \\ & Y = -\sqrt{P}g + W \ \ w.p. \ p. \end{aligned}$$

Particularly, Y is conditionally distributed as a mixture-Gaussian distribution given by

$$
\begin{aligned}
H_0: \qquad & f_Y(y|H_0,g) = \frac{1}{\sqrt{2\pi\sigma_w^2}}\left((1-p)\exp-\frac{(y+g\sqrt{P})^2}{2\sigma_w^2} + p\exp-\frac{(y-g\sqrt{P})^2}{2\sigma_w^2}\right) \\
H_1: \qquad & f_Y(y|H_1,g) = \frac{1}{\sqrt{2\pi\sigma_w^2}}\left((1-p)\exp-\frac{(y-g\sqrt{P})^2}{2\sigma_w^2} + p\exp-\frac{(y+g\sqrt{P})^2}{2\sigma_w^2}\right).
\end{aligned} \tag{5.7}
$$

We point out that, in reality, the optimal detection strategy involves coupling between the decision procedures at the sensor and the fusion center. However, the evaluation of thresholds and the analysis of the optimal scheme is intractable in general. Therefore, in the developments of this section, we adapt the so-called "person-by-person optimization" approach (PBPO) [79], where the decision at the fusion center is optimized assuming fixed decision rules at the sensor.

Therefore, following PBPO approach, the decision process at the fusion center boils down to classical hypothesis testing with

$$\text{Choose } \hat{S} = 1, \text{ if } L(Y) \geq 1,$$

where $L(y)$ is the likelihood ratio, given by

$$L(y) = \frac{p\exp-(\frac{2g\sqrt{P}y}{\sigma_w^2}) + 1 - p}{(1-p)\exp-(\frac{2g\sqrt{P}y}{\sigma_w^2}) + p}. \tag{5.8}$$

After some steps, the decision rule simplifies to

$$\text{Choose } \hat{S} = 1 \text{ if } gY \geq 0.$$

Next, we evaluate the detection error due to this "digital relaying" scheme. We let $\mathtt{SNR}_l \triangleq g^2P/\sigma_w^2$ denote the link SNR, i.e., signal-to-noise ratio of the link between sensor and the fusion center. For convenience, define $\mathtt{SNR} = [\mathtt{SNR}_o, \mathtt{SNR}_l]$. The detection error at the fusion center, when aided by a digital sensor, is given by

$$P_{ED}(\mathtt{SNR}) \quad = \quad p + (1-2p)Q\left(\sqrt{\mathtt{SNR}_l}\right), \tag{5.9}$$

where p is given by the equation (5.6).

Analog relaying (estimate and forward)

This case employs analog sensor to aid fusion center. Relay sensor constructs a sufficient statistic of the phenomenon, scales the message for the given power constraint and transmits it to the sink. Essentially, the sensor node acts as an analog relay amplifier with the output space, $\mathcal{U} = \mathbb{R}$ with $\mathbb{E}[|U|^2] \leq P$ for all $U \in \mathcal{U}$. It is clear that the sufficient statistic T is given by the log likelihood ratio,

$$T(x) = \log \frac{f_X(x|H_1, h)}{f_X(x|H_0, h)} = \frac{h}{2\sigma_z^2} (2x - h). \tag{5.10}$$

It follows that T is a Gaussian RV with variance $\text{Var}(T) = 4\texttt{SNR}_o$, and $\mathbb{E}[T|H_0] = -2\texttt{SNR}_o$ and $\mathbb{E}[T_i|H_1] = 2\texttt{SNR}_o$.

Accounting for the average power constraint, the message sent from the relay is given by

$$U(X) = \sqrt{\frac{P}{4\texttt{SNR}_o(1 + \texttt{SNR}_o)}} T(X) \tag{5.11}$$

so that

$$\begin{aligned} H_0 : U &\sim \mathcal{N}\left(-\sqrt{\frac{P\texttt{SNR}_o}{1 + \texttt{SNR}_o}}, \frac{P}{1 + \texttt{SNR}_o}\right) \\ H_1 : U &\sim \mathcal{N}\left(\sqrt{\frac{P\texttt{SNR}_o}{1 + \texttt{SNR}_o}}, \frac{P}{1 + \texttt{SNR}_o}\right). \end{aligned}$$

The fusion center receives Y, a noisy version of U, and performs decision on the event based on Y, so as to minimize the probability of detection error. The decision rule at the fusion center is the likelihood ratio test, given as:

$$\text{Choose } \hat{S} = 1, \text{ if } \frac{f_Y(y|H_1, g)}{f_Y(y|H_0, g)} \geq 1.$$

We note that

$$\begin{aligned} H_0 : Y &\sim \mathcal{N}\left(-\sqrt{\frac{g^2 P\texttt{SNR}_o}{1 + \texttt{SNR}_o}}, g^2 \frac{P}{1 + \texttt{SNR}_o} + \sigma_w^2\right) \\ H_1 : Y &\sim \mathcal{N}\left(\sqrt{\frac{g^2 P\texttt{SNR}_o}{1 + \texttt{SNR}_o}}, g^2 \frac{P}{1 + \texttt{SNR}_o} + \sigma_w^2\right). \end{aligned}$$

It follows that the optimal rule simplifies to

$$\text{Choose } \hat{S} = 1 \text{ if } gY \geq 0.$$

Consequently, for a given set of SNR parameters, $\mathtt{SNR}$, the detection error at the fusion center when aided by an analog sensor can be shown to be

$$P_{EA}(\mathtt{SNR}) = Q\left(\sqrt{\frac{\mathtt{SNR}_l\mathtt{SNR}_o}{\mathtt{SNR}_l + \mathtt{SNR}_o + 1}}\right). \tag{5.12}$$

5.4 Comparison of Analog and Digital Relaying

In this section, we compare the performance of analog and digital relaying schemes just described, in terms of observation and link SNRs. It is unclear, *a priori*, in which regimes one scheme outperforms the other. We partly answer the above question by providing a sufficient condition for the optimality of digital scheme over the analog scheme. Before presenting our main result, we state the following lemma.

Lemma 5.1. *Let $a, b \geq 0$ such that $a + b \leq \frac{1}{2}$, then*

$$Q\left(\frac{1}{\sqrt{a+b}}\right) \geq Q\left(\frac{1}{\sqrt{a}}\right) + Q\left(\frac{1}{\sqrt{b}}\right).$$

Proof. Let $a \geq b$, so that $\frac{1}{\sqrt{a+b}} \leq \frac{1}{\sqrt{a}} \leq \frac{1}{\sqrt{b}}$. We have

$$\begin{aligned}
Q\left(\frac{1}{\sqrt{a+b}}\right) &- \left[Q\left(\frac{1}{\sqrt{a}}\right) + Q\left(\frac{1}{\sqrt{b}}\right)\right] \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{1}{\sqrt{a+b}}}^{\frac{1}{\sqrt{a}}} e^{-\frac{x^2}{2}}\,dx - \frac{1}{\sqrt{2\pi}}\int_{\frac{1}{\sqrt{b}}}^{\infty} e^{-\frac{y^2}{2}}\,dy \\
&\geq \frac{1}{\sqrt{2\pi}}\int_{\frac{1}{\sqrt{a+b}}}^{\frac{1}{\sqrt{a}}} \sqrt{a}xe^{-\frac{x^2}{2}}\,dx - \frac{1}{\sqrt{2\pi}}\int_{\frac{1}{\sqrt{b}}}^{\infty} \sqrt{b}ye^{-\frac{y^2}{2}}\,dy \\
&= \frac{e^{-\frac{1}{2(a+b)}}}{\sqrt{2\pi}}\left[\sqrt{a} - \left(\sqrt{a}\,e^{-\frac{b}{2a(a+b)}} + \sqrt{b}\,e^{-\frac{a}{2b(a+b)}}\right)\right] \\
&\geq \frac{e^{-\frac{1}{2(a+b)}}}{\sqrt{2\pi}}\left[\sqrt{a} - \left(\sqrt{a}\,e^{-\frac{b}{a}} + \sqrt{b}\,e^{-\frac{a}{b}}\right)\right] \\
&\geq \frac{e^{-\frac{1}{2(a+b)}}}{\sqrt{2\pi}}\sqrt{a}\left[1 - \left(e^{-\frac{b}{a}} + e^{-\frac{a}{b}}\right)\right] \\
&\geq 0,
\end{aligned}$$

where the first inequality follows from the fact that $x \leq \frac{1}{\sqrt{a}}$ for the first integral and $y \geq \frac{1}{\sqrt{b}}$ for the second integral. Second inequality follows from the assumption that $a + b \leq \frac{1}{2}$, third inequality is by $a \geq b$, and the last inequality is due to fact that $\left(e^{-x} + e^{-\frac{1}{x}}\right) \leq 1$ for $x \geq 0$. □

The following result establishes that for the values of SNRs beyond a certain threshold, digital relaying is superior to analog relaying.

Proposition 5.1. ***[Optimality of Digital Relaying]:*** *For all observation and link SNRs,* $\mathtt{SNR} = [\mathtt{SNR}_o, \mathtt{SNR}_l]$ *such that* $1/\mathtt{SNR}_o + 1/\mathtt{SNR}_l \leq 1/2$*, it is optimal to use digital relaying over analog relaying, i.e.,* $P_{ED}(\mathtt{SNR}) \leq P_{EA}(\mathtt{SNR})$*.*

Proof. From (5.9) and (5.12), and using the fact that $Q(\cdot)$ is a decreasing function of its argument, it is clear that

$$P_{ED}(\mathtt{SNR}) \leq Q\left(\sqrt{\mathtt{SNR}_o}\right) + Q\left(\sqrt{\mathtt{SNR}_l}\right) \tag{5.13}$$

and

$$P_{EA}(\mathtt{SNR}) \geq Q\left(\sqrt{\frac{\mathtt{SNR}_l \mathtt{SNR}_o}{\mathtt{SNR}_l + \mathtt{SNR}_o}}\right). \tag{5.14}$$

Setting $a = \frac{1}{\mathtt{SNR}_o}$ and $b = \frac{1}{\mathtt{SNR}_l}$, and applying Lemma 5.1, we obtain

$$P_{EA}(\mathtt{SNR}) - P_{ED}(\mathtt{SNR}) \quad \geq \quad Q\left(\frac{1}{\sqrt{a+b}}\right) - \left[Q\left(\frac{1}{\sqrt{a}}\right) + Q\left(\frac{1}{\sqrt{b}}\right)\right] \geq 0.$$

□

Proposition 5.1 asserts that for the single sensor case, under the regime of moderately high observation and link SNRs, the digital scheme performs better than the analog scheme. This behavior can be explained intuitively as follows. Under digital scheme, since only the decision bits are communicated, sensor just acts as a amplifier-repeater on the event and hence the power spent per symbol is more, which, in turn, manifests the distortions suffered by the signal over the forward link. In case of analog scheme, the sufficient statistic, which is a real number, is communicated and the power spent on the symbol is insufficient to overcome the deleterious effects of the channel. Thus, when the SNR either at the backward link or at the forward link grows large, analog and digital schemes become equivalent.

5.5 Simulation Results

In this section, we evaluate the performance of digital and analog schemes. The performance metric is the decision error as a function of link SNR, for a fixed observation

SNR. Fig. 5.2 shows the performance curves as a function of $\texttt{SNR}_l$, for an observation SNR of 3dB. It can be seen that the digital scheme clearly outperforms analog scheme. As $\texttt{SNR}_l$ increases, the difference in the performance diminishes due to the fact that

$$\lim_{\texttt{SNR}_l \to \infty} P_{ED}(\texttt{SNR}) = \lim_{\texttt{SNR}_l \to \infty} P_{EA}(\texttt{SNR}) = Q(\sqrt{\texttt{SNR}_o}). \tag{5.15}$$

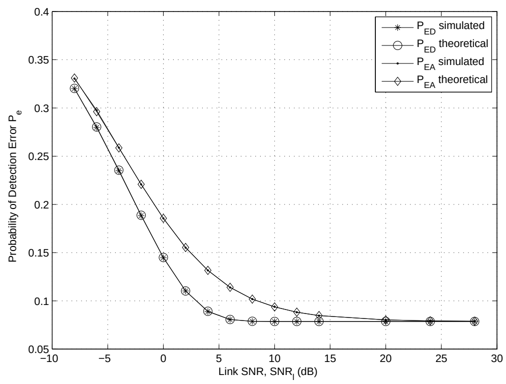

Figure 5.2: P_e vs. $\texttt{SNR}_l$ for $\texttt{SNR}_o = 3$ dB.

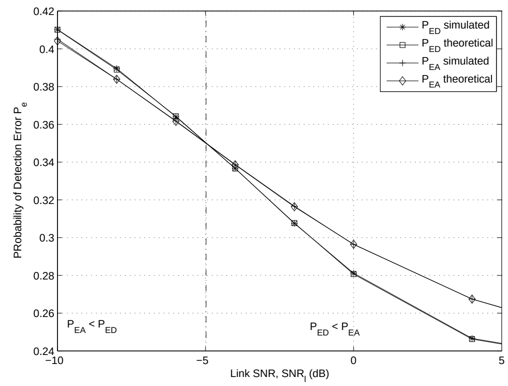

Figure 5.3: P_e vs. $\texttt{SNR}_l$ for $\texttt{SNR}_o = -3$ dB.

5.6 Conclusions

In this work, we considered the problem binary hypothesis testing in a single sensory relay network. In particular, we studied and compared two schemes, namely: detect-and-forward (digital relaying) and estimate-and-forward (analog relaying). We observed that the performance of relaying scheme depends on observation as well as link SNRs. We established a sufficient condition for the optimality digital relaying over analog relaying. This can have important implications on the design of sensor and communication relay networks. As a part of the future work, we wish to address the necessary condition for the optimality of digital sensors. Furthermore, the developments in the paper are carried under a classical hypothesis testing framework. It will be interesting to carry out a similar analysis and comparison with respect to different classes detectors like NP-detector, min-max detector etc.

Chapter 6

CONCLUSIONS AND DISCUSSIONS

In this thesis, we have laid out a framework to build a clear understanding of the value of information in wireless networks. The scope of information in this thesis is broad, ranging from channel/network states to the structure of the signal itself. Under the common thread of characterizing the value of information, we have investigated three important scenarios of wireless networks, namely opportunistic scheduling, cooperative relaying and multicasting in wireless networks. In the case of opportunistic scheduling, the information was in form of channel state information; in the case of cooperative relaying, it was in the form of partial knowledge of the concurrent interfering message; and in the multicasting scenario, information was the knowledge about the physical nature of the signal itself. Exploiting the available information to devise adaptive strategies is a key ingredient in these studies. Through rigorous analysis and numerical examples, we demonstrated that, by suitably exploiting the available information, one can gain substantially in terms of system performance. In what follows, we briefly outline our key contributions with discussions, and possible future directions.

6.1 Distributed Opportunistic Scheduling

Channel-aware scheduling is a promising technique to harness the rich diversities inherent in wireless networks. In channel-aware scheduling, a joint physical layer (PHY)/medium access control (MAC) optimization is utilized to improve network throughput by scheduling links with good channel conditions for data transmissions. However, to reap rich dividends from channel-aware scheduling, one has to carefully understand these dynamics of PHY/MAC diversities. Therefore, in this context, the availability of information at the participating nodes, about the PHY/MAC dynamics, becomes critical. In ad-hoc networks where there is no centralized decision maker, making this information available becomes more challenging. We addressed this challenge by considering a framework for distributed opportunistic scheduling (DOS).

DOS involves a process of joint channel probing and distributed scheduling for ad-hoc (peer-to-peer) communications. The objective is to maximize the system throughput. While related works in DOS hinge on the assumption of perfect channel state information (CSI), in practice, the CSI is often imperfect due to noisy estimation. When the CSI is imperfect, the transmission rate has to be backed off from the estimated rate to avoid transmission outages. This may result in throughput degradation, which is proportional to the estimation error. Clearly, a plausible solution to enhance the throughput is to mitigate the rate estimation errors by performing further channel probing.

With this motivation, our DOS framework is centered around cleverly tackling incomplete channel information, with the option of refining the CSI for the scheduled user. However, the advantages of second-level probing come at the price of additional overhead. Thus, there is a tradeoff between the throughput gain from better channel conditions and the cost for further probing. We showed that this tradeoff reduces to judiciously choosing an optimal stopping rule for channel probing and the transmission rate for throughput maximization. Capitalizing on optimal stopping theory with incomplete information, we showed that the optimal scheduling policy is threshold-based, and is characterized by either one or two thresholds, depending on network settings. We rigorously established necessary and sufficient conditions for both cases. In particular, we observed that performing second-level channel probing is optimal when the first-level estimated channel condition falls in between the two thresholds. Numerical results illustrate the effectiveness of the proposed DOS with two-level channel probing. We also extended our study to the case with limited feedback, where the feedback from the receiver to its transmitter takes the form of $(0, 1, e)$. We note that the proposed distributed scheduling with two-level probing provides a new framework to study joint PHY/MAC optimization in practical networks where noisy probing is often the case and imperfect information is inevitable. We believe that this line of study provides some initial steps towards opening a new avenue for exploring the intrinsic tradeoffs between probing (sensing) and scheduling to enhance spectrum utilization; and this is potentially useful for enhancing MAC protocols for wireless mesh networks and cognitive radio networks.

In this work, we have considered DOS with two-level probing, where it is assumed that the refinement of the rate estimate is carried out once, via second-level probing of duration τ. However, we can further extend this to $L-$level probing, where for $k = 1, \ldots, L-1$, a successfully contending transmitter has the options 1) to transmit, or 2) to defer and re-contend, or 3) to resort to $(k+1)-$st level training at the cost of additional overhead. In this situation, it is of interest to devise well-structured, yet simple policies. Extending the current results on single-hop ad hoc networks to multi-hop models is also of considerable interest for future study.

6.2 Cooperative Relaying

It is well-known that interference is a detrimental phenomenon in wireless communication, limiting the system performance. Interference management is a key issue in the design of wireless systems. Most state-of-the-art wireless systems deal with interference in one of two ways: To orthogonalize the communication links in time or frequency, so that they do not interfere with each other at all; or, to allow the communication links to share the same degrees of freedom, but treat each other's interference as adding to the noise floor. It is clear that both approaches can be sub-optimal. The first approach entails an a priori loss of degrees of freedom in both links, no matter how weak the potential interference is. The second approach treats interference as pure noise while it actually carries information and has structure that can potentially be exploited in mitigating its effect. These treatments are based on the assumption that the each transmitter/receiver pair is isolated from other. In various practical scenarios, however, they are not isolated, and cooperation among transmitters or receivers can be induced. Further, other nodes in the network can serve as relays, and cooperate with the active links in mitigating the interference. A key outcome of this view is that, due to the broadcast nature of the wireless medium, each node in the network obtains partial side information about the interference. The main challenge then is exploit this side information to improve the system performance.

With this motivation, we considered a two-hop interference network in which two sources wish to communicate simultaneously with two destinations, and are aided by two relay nodes. The transmissions interfere with each other in each hop. A key observation is that, while the transmission during the first hop is essentially over a classical interference channel, the transmission in the second hop enjoys an interesting advantage. Specifically, the second hop represents an interference channel with partial side-information at the transmitters. Consequently, this opens the door to cooperation between the relays. In light of this, we proposed various cooperative relaying strategies based on distributed MIMO broadcast, to enhance the achievable rate of the information flows. Indeed, we observed that, by exploiting the side information available at the transmitters, the proposed strategies gain substantially over other strategies considered in the literature.

Next, we discuss some possible future extensions. In this work, we have considered a two-hop model comprising two information flows. It will be interesting to consider more than two flows (i.e., more than two source-destination pairs aided by relays) in a two-hop network. Further, throughout this study, we have assumed that all nodes the network have perfect knowledge of their direct and interfering link conditions. An interesting direction would be to relax the assumption of perfect state information and study its impact on the throughput of the network. Also, it will be interesting to consider the more realistic case where the transmitter nodes have no channel state information. In this case, possible transmission strategies could involve cooperation as well as competition.

6.3 Digital versus Analog Relaying

We considered a basic scenario in sensor networks, where a single sensor node relays a statistic of its observations of an event to the fusion center which performs the final inference on the event. We considered two relaying schemes: analog relaying and digital relaying. With ultimate detection error at the fusion center being the performance metric, we provided sufficient conditions for the optimality of analog relaying over digital relaying in this network. This observation has potential applications in the design and placement of sensor nodes in sensor networks.

There are several directions for extending this work. In this work, we have given only a sufficient condition for the optimality of digital relaying. However, one can strengthen this result by establishing the necessary conditions. Furthermore, the developments in this work are carried under a classical hypothesis testing framework. It will be interesting to carry out a similar analysis and comparison with respect to different classes detectors like NP-detector, min-max detector etc.

6.4 Multicasting

To understand the value of information in the multicast scenario, we considered multicasting compressively sampled signals from a source to many receivers, over lossy wireless channels. We focused on the network outage from the perspective of signal distortion across all receivers, for both cases where the transmitter may or may not be capable of reconstructing the compressively sampled signals. Based on extreme value theory, we characterized the network outage in terms of key system parameters, including the erasure probability, the number of receivers and the sparse structure of the signal. When the transmitter can reconstruct the compressively sensed signal, the strategy of using network coding to multicast the reconstructed signal coefficients can reduce the network outage significantly. However, the traditional network coding could result in suboptimal performance for power-law decay signals. Thus motivated, we devised a new method, using subblock network coding, which involves fragmenting the data into subblocks, and allocating time slots to different subblocks, based on their priorities. The corresponding optimal allocation is formulated as an integer programming problem. Since integer programming is often intractable, we developed a heuristic algorithm that prioritizes the time slot allocation by exploiting the inherent priority structure of power-law decay signals. Through numerical results, we demonstrated that the proposed schemes outperform the traditional methods with significant margins.

As a part of the future work, we are currently exploring the Markov Chain Monte Carlo method to quantify the potential gain of subblock network coding. In this work, we have considered a centralized scheme for compressing/sampling the information. A potentially explorable direction would be to look into distributed ways of exploiting the structure of the signals in sampling and compressing them. This can have significant implications in reducing transmission and processing burdens in sensor networks.

REFERENCES

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris, "Link-level measurements from an 802.11b mesh network," in *In SIGCOMM*, 2004, pp. 121–132.

[2] M. Cao, V. Raghunathan, and P. Kumar, "Cross-layer exploitation of MAC layer diversity in wireless networks," in *Proc. ICNP '06*, USA, 2006, pp. 332–341.

[3] M. Andrews, K. Kumaran, K. Ramannan, A. Stolyar, R. Vijaykumar, and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Labs Tech. Memo.*, April 2000.

[4] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," in *Proc. IEEE INFOCOM'03*, San Francisco, CA, 2003.

[5] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC'95*, vol. 1, USA, 1995, pp. 331–335.

[6] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[7] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad-hoc communications: An optimal stopping approach," in *Proc. Mobihoc 2007*, Montreal, Canada, Sept. 2007.

[8] D. Zheng, M. O. Pun, W. Ge, J. Zhang, and H. V. Poor, "Distributed opportunistic scheduling for ad hoc communications with imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 5450–5460, Dec. 2008.

[9] A. Vakili, M. Sharif, and B. Hassibi, "The effect of channel estimation error on the throughput of broadcast channels," in *Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'06)*, Toulouse, France, May 2006.

[10] W. Stadje, "An optimal stopping problem with two levels of incomplete information," *Mathematical Methods of Operations Research*, vol. 45, no. 1, pp. 119–131, 1997.

[11] D. Assaf, L. Goldstein, and E. Samuel-Cahn, "A statistical version of prophet inequalities," *Annals of Statistics*, vol. 26, pp. 1190–1197, 1996.

[12] M. H. M. Costa and A. A. E. Gamal, "The capacity region of the discrete memoryless interference channe with strong interference (corresp.)," *IEEE Trans. Info. Theory*, vol. 33, no. 5, Sept. 1987.

[13] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Info. Theory*, vol. 27, no. 1, Jan 1981.

[14] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. Info. Theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[15] E. C. van der Meulen, "A survey of multiway channels in information theory: 1961-1976," *IEEE Trans. Info. Theory*, vol. 23, no. 1, Jan. 1977.

[16] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Info. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[17] S. I. Gelf́and and M. S. Pinsker, "Coding for channel with random parameters," *Problems in Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[18] R. Ahlswede, N. Cai, R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Info. Theory*, vol. 46, pp. 1204–1216, July 2000.

[19] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289 – 1306, 2006.

[20] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[21] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Info. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[22] X. Qin and R. Berry, "Exploiting multiuser diversity for medium access control in wireless networks," in *Proc. IEEE INFOCOM'03*, San Francisco, CA, Apr. 2003.

[23] Z. Ji, Y. Yang, J. Zhou, M. Takai, and R. Bagrodia, "Exploiting medium access diversity in rate adaptive wireless LANs," in *Proc. ACM/IEEE MOBICOM'04*, Philadelphia, PA, Sept. 2004.

[24] X. Liu, E. K. Chong, and N. B. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, no. 4, pp. 451–474, Mar. 2003.

[25] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad hoc networks with random access: An optimal stopping approach," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 205–222, Jan. 2009.

[26] B. Wang, J. Zhang, and L. Zheng, "Achievable rates and scaling laws of power-constrained wireless sensory relay networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4084–4104, Sept. 2006.

[27] Z. Ye, A. Abouzeid, and J. Ai, "Optimal stochastic policies for distributed data aggregation in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1494–1507, Oct. 2009.

[28] T. Ferguson, *Optimal Stopping and Applications.* Electronic text, available at http://www.math.ucla.edu/~tom/Stopping/Contents.html, 2006.

[29] Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: Theory of Optimal Stopping.* Houghton Mifflin, 1971.

[30] A. N. Shiryayev, *Optimal Stopping Rules.* New York, NY: Springer-Verlag, 1978.

[31] D. P. Bertsekas and R. Gallager, *Data Networks.* Upper Saddle River, NJ: Prentice-Hall, 1992.

[32] W. Ge, J. Zhang, J. Wieselthier, and S. Shen, "PHY-aware distributed scheduling for ad hoc communications with physical interference model," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2682 – 2693, May 2009.

[33] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," in *Proc. ACM/IEEE MOBICOM'02*, Atlanta, GA, 2002.

[34] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[35] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation.* Englewood Cliffs, NJ: Prentice Hall, 2000.

[36] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297–307, Jan. 2009.

[37] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *The Annals of Probability*, vol. 2, no. 5, 1974.

[38] O. Simeone, O. Somekh, Y. Bar-Ness, H. V. Poor, and S. Shamai, "Capacity of linear two-hop mesh networks with rate splitting, decode-and-forward relaying and

cooperation," in *Proc. 45th Annual Allerton Conference on Communication, Control, and Computing*, 2007.

[39] O. Oyman and A. J. Paulraj, "Leverages of distributed MIMO relaying: A Shannon-theoretic perspective," in *1st IEEE Workshop on Wireless Mesh Networks (WiMesh'05)*, 2005.

[40] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE Vehicular Technology Conf.*, May 2001.

[41] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Info. Theory*, vol. 25, no. 3, pp. 306 – 311, may. 1979.

[42] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. 29, no. 3, May 1983.

[43] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple input multiple output broadcast channel," *IEEE Trans. Info. Theory*, vol. 52, no. 9, Sept. 2006.

[44] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. N. C. Tse, "Approximate capacity of Gaussian interference-relay networks with weak cross links," *CoRR*, vol. Arxiv: 1005.0404, 2010.

[45] Y. Cao and B. Chen, "Capacity bounds for two-hop interference networks," in *47th Annual Allerton Conference on Communication, Control, and Computing, 2009*, Sep. 2009, pp. 272 –279.

[46] P. Thejaswi, A. Bennatan, J. Zhang, R. Calderbank, and D. Cochran, "Rate-achievability strategies for two-hop interference flows," in *2008 46th Annual Aller-*

ton Conference on Communication, Control, and Computing, Sep. 2008, pp. 1432 –1439.

[47] I. Maric, R. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Info. Theory*, vol. 53, no. 10, Oct. 2007.

[48] V. Prabhakaran and P. Viswanath, "Interference channels with source cooperation," *Submitted to IEEE Trans. Info. Theory*, 2010.

[49] I. Wang and D. N. C. Tse, "Interference Mitigation through Limited Transmitter Cooperation," *ArXiv: 1004.5421*, Apr. 2010.

[50] Y. Cao and B. Chen, "An achievable rate region for interference channels with conferencing," in *IEEE International Symposium on Information Theory, 2007. ISIT 2007.*, Jun. 2007, pp. 1251 –1255.

[51] S. Yang and D. Tuninetti, "A new achievable region for interference channel with generalized feedback," in *42nd Annual Conference on Information Sciences and Systems, 2008. CISS 2008.*, Mar. 2008, pp. 803 –808.

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

[53] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.

[54] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Info. Theory*, vol. 27, no. 6, pp. 786–788, Nov. 1981.

[55] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Info. Theory*, vol. 51, no. 8, pp. 2867–2877, Aug. 2005.

[56] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai, "On the capacity of interference channels with one cooperating transmitter," *European Transactions on Telecommunications*, vol. 19, pp. 405–420, Apr. 2008.

[57] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Info. Theory*, vol. 49, no. 7, July 2003.

[58] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Info. Theory*, vol. 49, no. 10, Oct. 2003.

[59] V. S. Annapureddy and V. V. Veeravalli, "Sum capacity of the Gaussian interference channel in the low interference regime," in *Proceedings of ITA Workshop*, San Diego, CA, 2008.

[60] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," in *International Symposium on Information Theory (ISIT'08)*, Toronto, Canada, July 2008.

[61] X. Shang, G. Kramer, and B. Chen, "New outer bounds on the capacity region of Gaussian interference channels," in *International Symposium on Information Theory (ISIT'08)*, Toronto, Canada, July 2008.

[62] Z. Charbiwala, S. Chakraborty, S. Zahedi, Y. Kim, M. B. Srivastava, T. He, and C. Bisdikian, "Compressive oversampling for robust data transmission in sensor networks," in *Proceedings of IEEE INFOCOM'10*, 2010, pp. 1190–1198.

[63] S. Chandran and S. Lin, "Selective-repeat-ARQ schemes for broadcast links," *IEEE Transactions on Communications*, vol. 40, pp. 12–19, 1992.

[64] W. Xiao, S. Agarwal, D. Starobinski, and A. Trachtenberg, "Reliable wireless broadcasting with near-zero feedback," in *Proceedings of IEEE INFOCOM'10*, 2010, pp.

2543–2551.

[65] L. Rizzo and L. Vicisano, “RMDP: An FEC-based reliable multicast protocol for wireless environments,” in *Mobile Computing and Communications Review*, vol. 2, April 1998.

[66] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[67] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, “Compressive wireless sensing,” in *Proceedings of IPSN ’06*, 2006, pp. 134–142.

[68] ——, “Joint source-channel communication for distributed estimation in sensor networks,” *IEEE Trans. Info. Theory*, vol. 53, pp. 3629–3653, October 2007.

[69] T. Ho, M. Mèdard, D. R. Karger, M. Effros, J. Shi, and B. Leong, “A random linear network coding approach to multicast,” *IEEE Trans. Info. Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.

[70] S. Katti, H. Rahul, W. Hu, D. Katabi, and M. M. J. Crowcroft, “XORs in the air: Practical wireless network coding,” in *Proceedings of ACM SIGCOMM*, September 2006.

[71] T. Tran, T. Nguyen, B. Bose, and V. Gopal, “A hybrid network coding technique for single-hop wireless networks,” *IEEE Journal on Selected Areas in Communications (JSAC)*, Jun. 2009.

[72] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, “Wireless broadcast using network coding,” *IEEE Transactions on Vehicular Technology*, vol. 58, February 2009.

[73] V. Stankovic, R. Hamzaoui, Y. Charfi, and Z. Xiong, "Real-time unequal error protection algorithms for progressive image transmission," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1526–1535, 2003.

[74] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "A deterministic approach to wireless relay networks," in *Proceedings of 45th Allerton Conference on Communication, Control, and Computing*, 2007.

[75] D. Aldous, *Probability approximations via the Poisson clumping heuristic.* New York: Springer-Verlag, 1989.

[76] T. Ho, M. Mèdard, J. Shi, M. Effros, and D. R. Karger, "On randomized network coding," in *Proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing*, October 2003.

[77] C. Fragouli, J. L. Boudec, and J. Widmer, "Network coding: An instant primer," in *ACM SIGCOMM Computer Communication Review, Vol. 36, Issue 1*, January 2006.

[78] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* Englewood Cliffs, NJ: Prentice Hall, 1998.

[79] J. N. Tsitsiklis, "Decentralized detection," in *In Advances in Statistical Signal Processing.* JAI Press, 1993, pp. 297–344.

[80] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: I. Fundamentals," in *In Proc. of IEEE*, vol. 85, no. 1, Jan. 1997, pp. 54–63.

[81] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: II. Advanced Topics," in *In Proc. of the IEEE*, vol. 85, no. 1, Jan. 1997, pp. 64–79.

[82] B. Chen, L. Tong, and P. K. Varshney, "Channel aware distributed detection in wireless sensor networks," *IEEE Signal Processing Mag*, vol. 23, pp. 16–25, Jul. 2006.

[83] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks." *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[84] K. Liu and A. M. Sayeed, "Type-Based Decentralized Detection in Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 55, pp. 1899–1910, May 2007.

[85] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Info. Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[86] A. El Gamal and E. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Info. Theory*, vol. 27, no. 1, pp. 120–122, Jan. 1981.

[87] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization.* Athena Scientific, 1997.

APPENDIX A

PROOFS OF RESULTS FROM CHAPTER 2

A.1 Derivation of Rate Equation (2.11)

On setting $\beta^{(1)} \triangleq E\left[|\tilde{h}^{(1)}|^2\right]$, we follow the approach proposed in [9] and normalize $|\hat{h}^{(1)}|^2$ and $|\tilde{h}^{(1)}|^2$ as

$$\hat{\lambda}^{(1)} = \frac{|\hat{h}^{(1)}|^2}{1-\beta^{(1)}} \tag{A.1}$$

and

$$\zeta^{(1)} = \frac{|\tilde{h}^{(1)}|^2}{\beta^{(1)}}, \tag{A.2}$$

where both $\hat{\lambda}^{(1)}$ and $\zeta^{(1)}$ are exponentially distributed with unit variance.

Define the "effective channel SNR" and "normalized error variance" as

$$\rho_{eff}^{(1)} \triangleq (1-\beta^{(1)})\rho \tag{A.3}$$

and

$$\alpha^{(1)} \triangleq \frac{\beta^{(1)}}{1-\beta^{(1)}}, \tag{A.4}$$

respectively. Substituting (A.1),(A.3) and (A.4) in (2.10) results in

$$\lambda^{(1)} = \frac{\rho_{eff}^{(1)}\hat{\lambda}^{(1)}}{1+\alpha^{(1)}\rho_{eff}^{(1)}\zeta^{(1)}}. \tag{A.5}$$

It has been shown in [9] that the conditional probability distribution function (pdf) of $\lambda^{(1)}$ given $\hat{\lambda}^{(1)}$ takes the following form:

$$f\left(\lambda^{(1)}\middle|\,\hat{\lambda}^{(1)}\right) = \frac{\hat{\lambda}^{(1)}}{\alpha^{(1)}\left[\lambda^{(1)}\right]^2}\exp\left\{-\frac{1}{\alpha^{(1)}}\left(\frac{\hat{\lambda}^{(1)}}{\lambda^{(1)}}-\frac{1}{\rho_{eff}^{(1)}}\right)\right\}\mathbf{I}\left(\frac{\hat{\lambda}^{(1)}}{\lambda^{(1)}} \geq \frac{1}{\rho_{eff}^{(1)}}\right). \tag{A.6}$$

The following linear backoff function is employed to prevent channel outage.

$$\lambda_c(\hat{\lambda}^{(1)}) = \sigma_M \rho_{eff}\hat{\lambda}^{(1)}, \tag{A.7}$$

where σ_M is the backoff factor with $0 < \sigma_M < 1$. Let $R_n^{(BK)}$ be the instantaneous rate with backoff, which is given by

$$R_n^{(BK)} = \log\left(1+\lambda_c(\hat{\lambda}_n)\right)\mathbf{I}\left(\lambda_c(\hat{\lambda}_n) \leq \lambda_n\right). \tag{A.8}$$

We note that, due to the estimation errors, the instantaneous rate, $R_n^{(BK)}$ defined in (A.8), is now a random variable, and is not observable at time n. Moreover, since

$\{(\rho|\hat{h}_j|^2, K_j)\}_{j\le n}$ is the only observable sequence, the decision has to be made solely based on $\mathcal{F}'$, the σ-field generated by $\{(\rho|\hat{h}_j|^2, K_j)\}_{j\le n}$. However, it can be shown that the optimal scheduling strategy and the optimal throughput remain the same if the random "reward" $R_n^{(BK)}$ is replaced with its conditional expectation, denoted as $R_n^{(1)}$ [28, Page 1.3] [11]. As a result, the scheduling can now be based on $R_n^{(1)}$ instead of $R_n^{(BK)}$, where $R_n^{(1)} \triangleq E\left[R_n^{(BK)}|\mathcal{F}'\right]$. Using (A.6) and (A.7), the conditional expectation $R_n^{(1)}$ can be computed as

$$\begin{aligned} R_n^{(1)} &= E\left[R_n^{(BK)}|\mathcal{F}'\right] \\ &= E\left[\log\left(1+\lambda_c(\hat{\lambda}^{(1)})\right)\mathbf{I}\left(\lambda_c(\hat{\lambda}^{(1)}) \le \lambda^{(1)}\right)|\mathcal{F}'\right], \\ &= \left[1-\exp\left\{\frac{-\left(\frac{1}{\sigma_M}-1\right)}{\alpha^{(1)}\rho_{eff}^{(1)}}\right\}\right] W \log\left(1+\sigma_M\rho|\hat{h}^{(1)}|^2\right). \end{aligned} \tag{A.9}$$

For the low SNR wideband regime where $\rho \to 0$ and $W = \Theta(\frac{1}{\rho})$, $R_n^{(1)}$ can be well approximated by

$$R_n^{(1)} \approx \rho W \sigma_M |\hat{h}^{(1)}|^2. \tag{A.10}$$

A.2 Proof of Lemma 2.1

For a given θ, let $N(\theta)$ be a stopping rule such that

$$N(\theta) = \arg\sup_{N\in Q} E\left[R_N T_{d,N} - \theta T_N\right].$$

Let $Z_n \triangleq R_n T_{d,n} - \theta T_n$. Then, it follows from Theorem 1 in [28, Chapter 3] that $N(\theta)$ exists if the following conditions are satisfied:

$$\text{(A1)}\ E[\sup_n Z_n] < \infty,\ \text{and (A2)}\ \limsup_{n\to\infty} Z_n = -\infty,\ \text{a.s.}$$

Since, it is clear that $\limsup_{n\to\infty} Z_n = -\infty$, we can easily verify (A2).

For some $0 < \mu < 1/p_s$, we introduce

$$Z_n' = \max\{R_n^{(1)}, R_n^{(2)}\}T - n\theta\tau\left(\frac{1}{p_s} - \mu\right)$$

and

$$Z_n'' = \sum_{j=1}^{n} \theta\tau\left(\frac{1}{p_s} - K_j - \mu\right).$$

Then, we note that

$$E\left[\sup_n Z_n\right] \leq E\left[\sup_n Z'_n\right] + E\left[\sup_n Z''_n\right].$$

Appealing to Theorem 1 and Theorem 2 of [28, Chapter 4], we conclude that $E[Z'_n] < \infty$ and $E[Z''_n] < \infty$, respectively. Therefore (A1) holds.

The second part of the lemma follows directly from Theorem 1 in [28, Ch.6].

A.3 Proof of Lemma 2.2

a) Using integration by parts, we rewrite $J_{\theta^*}(x,r)$ as

$$J_{\theta^*}(x,0) = (1-\tau)\int_{\theta^*}^{\infty}(1-G(u|x))du - \theta^*\tau. \tag{A.11}$$

Since $G(y|x)$ decreases monotonically with x, $J_{\theta^*}(x,0)$ is also monotonically increasing in x. Note that $\lim\limits_{x\to\infty}(1-G(u|x)) = 1$. Then, by Lebesgue's convergence theorem, we have $\lim_{x\to\infty} J_{\theta^*}(x,0) = \infty$. Let $z = \sqrt{\sigma_{2M}W\rho}h_e$, where $z \sim \mathcal{CN}(0, R_e)$, with $R_e = \sigma_{2M}W\rho\sigma_e^2$. Then, from (2.17), it follows that $\lim\limits_{x\to 0} G(y|x) = G_{|z|^2}(y) = 1 - e^{-\frac{y}{R_e}}$, and consequently,

$$\lim_{x\to 0} J_{\theta^*}(x,0) \quad = \quad (1-\tau)R_e e^{-\frac{\theta^*}{R_e}} - \theta^*\tau. \tag{A.12}$$

Thus, under the condition $R_e/\theta^* e^{-\frac{\theta^*}{R_e}} < \tau/(1-\tau)$, $\lim_{x\to 0} J_{\theta^*}(x,0) < 0$.

b) Using integration by parts, we can rewrite $J_{\theta^*}(x,r)$ as

$$J_{\theta^*}(x,0) = (1-\tau)\left(c_r x + R_e - \theta^* + \int_0^{\theta^*} G(u|x)du\right) - \theta^*\tau. \tag{A.13}$$

It follows that

$$q(x) = (c_r(1-\tau)-1)x + (1-\tau)R_e + (1-\tau)\int_0^{\theta^*} G(u|x)du,$$

and we can verify that

$$\lim_{x\to 0} q(x) = R_e(1-\tau) + (1-\tau)\theta^* > 0.$$

Furthermore, when $c_r < \frac{1}{1-\tau}$, it is clear that $\lim_{x\to\infty} q(x) = -\infty$. Since $G(y|x)$ is monotonically decreasing in x, we conclude that $q(x)$ is also monotonically decreasing in x.

A.4 Proof of Theorem 2.1

Let x_J and x_q be solutions to $J_{\theta^*}(x,0)=0$ and $q(x)=0$ respectively. From Lemma 2.2, we have

$$J_{\theta^*}(x,0)\begin{cases} <0 & \text{if } x<x_J \\ =0 & \text{if } x=x_J \\ >0 & \text{if } x>x_J \end{cases} \tag{A.14}$$

and

$$q(x)\begin{cases} <0 & \text{if } x>x_q \\ =0 & \text{if } x=x_q \\ >0 & \text{if } x<x_q \end{cases}. \tag{A.15}$$

Thus, one of the following two possibilities holds.

i. The case with $x_q \geq x_J$:

From the above discussions and the monotonicity properties of $J_{\theta^*}(\cdot,0)$ and $q(\cdot)$, it follows that

$$\max\left[x-\theta^*, J_{\theta^*}(x,0)\right]^+ = \begin{cases} x-\theta^* & \text{if } x>x_q \\ J_{\theta^*}(x,0) & \text{if } x\in[x_J,x_q] \\ 0 & \text{if } x<x_J \end{cases}. \tag{A.16}$$

Furthermore, from (A.16) and the optimality equation (2.21), we have that

$$\int_{x_J}^{x_q} J_{\theta^*}(u,0)\,dF(u) + \int_{x_q}^{\infty}(u-\theta^*)\,dF(u) = \frac{\theta^*\tau}{p_s}. \tag{A.17}$$

Consequently, it is clear that the optimal strategy is

$$\phi_n(R_n^{(1)}) = \begin{cases} 1 \text{ (transmit)} & \text{if } R_n^{(1)} > x_q \\ 2 \text{ (2-level)} & \text{if } R_n^{(1)} \in [x_J,x_q] \\ 0 \text{ (re-contend)} & \text{if } R_n^{(1)} < x_J \end{cases} \tag{A.18}$$

and when $\phi_n(R_n^{(1)})=2$, the strategy is

$$\psi_n(R_n^{(2)}) = \begin{cases} 1 \text{ (transmit)} & \text{if } R_n^{(2)} \geq \theta_A^* \\ 0 \text{ (re-contend)} & \text{if } R_n^{(2)} < \theta_A^* \end{cases} \tag{A.19}$$

where θ_A^* is the solution to (A.17). It can be seen that thresholds x_J and x_q are found as the solutions to $J_{\theta^*}(x,0)=0$ and $q(x)=0$ respectively. Thus, $\{x_J,x_q,\theta_A^*\}$

is the solution to the system (2.23). An illustration of Strategy A is depicted in Fig. 2.3.

ii. The case with $x_q < x_J$:

From (A.14) and (A.15), we have

$$\max\left[x-\theta^*, J_{\theta^*}(x,0)\right]^+ = \begin{cases} x-\theta^* & \text{if } x \geq \theta^* \\ 0 & \text{if } x < \theta^* \end{cases} \tag{A.20}$$

and $J_{\theta^*}(x,0) < \max\left[x-\theta^*, 0\right]$. Therefore, it is never optimal to perform second-level probing. From (A.20) and the optimality equation (2.21) we obtain

$$\int_{\theta^*}^{\infty} (x-\theta^*)\, dF(x) = \frac{\theta^*\tau}{p_s},$$

which is equivalent to (2.12). Thus from (A.20), the optimal strategy is

$$\phi(R_n^{(1)}) = \begin{cases} 1 \text{ (transmit)} & \text{if } R_n^{(1)} \geq \theta_B^* \\ 0 \text{ (re-contend)} & \text{if } R_n^{(1)} < \theta_B^*, \end{cases} \tag{A.21}$$

where the threshold θ_B^* is the solution to (2.12). An illustration of Strategy B is depicted in Fig. 2.4.

A.5 Proof of Theorem 2.2

Suppose $J_{\theta_A^*}(\theta_A^*, 0) \geq 0$. Then, this implies that $J_{\theta_A^*}(\theta_A^*, 0) \geq \max[x-\theta_A^*, 0]$ when $x = \theta_A^*$. Specifically, when $R_1^{(1)} = \theta_A^*$, performing second-level probing and using an optimal strategy thereafter yield an expected reward of $J_{\theta_A^*}(\theta_A^*, 0)$, which is at least as good as using Strategy B. Equivalently, we show that there exists at least one value of x (θ_A^* in this case) for which performing second-level probing is optimal. We conclude that Strategy A is optimal.

Next, we assume Strategy A is optimal and show that $J_{\theta_A^*}(\theta_A^*, 0) \geq 0$. Under such an assumption, there must exist some x_1 for which it is beneficial to demand additional information, i.e.

$$J_{\theta_A^*}(x_1, 0) \geq \max[x_1 - \theta_A^*, 0]. \tag{A.22}$$

We now investigate $J_{\theta_A^*}(\theta_A^*, 0)$ in two different cases, namely $\theta_A^* \geq x_1$ and $\theta_A^* < x_1$.

i. <u>The case with $\theta_A^* \geq x_1$:</u>

In this case,

$$J_{\theta_A^*}(\theta_A^*, 0) \geq J_{\theta_A^*}(x_1, 0) \geq \max[x_1 - \theta_A^*, 0] = 0, \tag{A.23}$$

where the first and second inequalities are due to the monotonicity of $J(\cdot, 0)$ and the assumed optimality of Strategy A, respectively.

ii. <u>The case with $\theta_A^* < x_1$:</u>

In this case,

$$J_{\theta_A^*}(\theta_A^*, 0) \geq J_{\theta_A^*}(x_1, 0) - x_1 + \theta_A^* \geq 0, \tag{A.24}$$

where the first inequality follows from the fact that $J_{\theta_A}(x, 0) - x + \theta$ is decreasing in x and the second inequality is due to (A.22).

Summarizing the above two cases, we conclude that $J_{\theta_A^*}(\theta_A^*, 0) \geq 0$ is a necessary condition for the optimality of Strategy A. Using contra position, we conclude that Strategy B is optimal if $J_{\theta_A^*}(\theta_A^*, 0) < 0$.

A.6 Proof of Lemma 2.5

It is clear that $V_{\gamma^*}(x, R_1)$ is monotonically increasing in x, and that

$$\lim_{x \to \infty} V_{\gamma^*}(x, R_1) = (1 - \tau)R_1 - \gamma^*.$$

Since $\gamma^* \leq \gamma_U$, it follows that $\lim\limits_{x \to \infty} V_{\gamma^*}(x, R_1) > 0$, provided that $\tau \leq 1 - p_s$. Furthermore, observe that

$$\begin{aligned}
\lim_{x \to 0} V_{\gamma^*}(x, R_1) &= (1 - \tau)(R_1 - \gamma^*)e^{-\frac{R_1}{R_e}} - \gamma^*\tau \\
&\leq \gamma^* \left((1 - \tau)(\frac{R_1}{\gamma_L} - 1)e^{-\frac{R_1}{R_e}} - \tau \right) \\
&= \tau\gamma^* \left((1 + \frac{1}{p_s})e^{\frac{R_1}{E[R^{(2)}]}} e^{-\frac{R_1}{R_e}} - 1 \right) \\
&\leq \tau\gamma^* \left((1 + \frac{1}{p_s})e^{-(1+2\tau)} - 1 \right),
\end{aligned}$$

where the last inequality follows due to the fact that $\frac{E[R^{(2)}]}{R_1} \leq 1$ and $\frac{E[R^{(2)}]}{R_e} \approx (1 + 2\tau)$. We conclude that

$$\lim_{x \to 0} V_{\gamma^*}(x, R_1) < 0, \quad \text{for } \tau \geq 0.5 \left(\ln \left(1 + \frac{1}{p_s} \right) - 1 \right).$$

The second part follows from the facts that for $x \geq R_1$,

$$q_{\gamma^*}(x, R_1) = (1-\tau)(1 - G(R_1|x))(\gamma^* - R_1) - \tau R_1 < 0,$$

and for $x < R_1$,

$$q_{\gamma^*}(x, R_1) = (1-\tau)R_1(1 - G(R_1|x)) + (1-\tau)G(R_1|x)\gamma^* \geq 0.$$

APPENDIX B

PROOFS OF RESULTS FROM CHAPTER 3

B.1 Proof of Theorem 3.1

Code construction: Pick a distribution

$$p(x_{pc1})p(x_{pc2})p(\mathbf{u}_{c1},\mathbf{u}_{c2})p(x_1,u_{pp,1}|\mathbf{u}_{c,1},\mathbf{u}_{c,2},x_{pc,1})p(x_2,u_{pp,2}|\mathbf{u}_{c,1},\mathbf{u}_{c,2},x_{pc,2}),$$

and let $R_i = R_{c,i} + R_{pc,i} + R_{pp,i},\ \ i = 1, 2$, be the given rate-split.

1. *Codewords for common messages:* For $i = 1, 2$, generate $2^{N(R_{c,i}+r_{c,i})}$, 2×1-vector codewords, $\mathbf{U}^N_{c,i}(\mathrm{k}_{c,i})$, for $\mathrm{k}_{c,i} \in \{1, \dots, 2^{N(R_{c,i}+r_{c,i})}\}$, where each symbol is drawn i.i.d. with $\mathbf{U}_{c,i} \sim P_{\mathbf{U}_{c,i}}(\cdot)$. Partition this set of codewords into $2^{NR_{c,i}}$ equal-sized bins, with $2^{Nr_{c,i}}$ codewords per bin. Observe that each bin corresponds to a common message index, and the bin corresponding to message $\mathrm{v}_{c,i} = k$ can be defined as, for $i = 1, 2$,

$$B^{(i)}_k = \left\{\mathbf{U}^N_{c,i}(\mathrm{k}_{c,i}) : \mathrm{k}_{c,i} \in \{(k-1)2^{Nr_{c,i}}+1, \dots, k2^{Nr_{c,i}}\}\right\},$$

For each pair of bin indices (k, l), define the product bin

$$\begin{aligned} B_{k,l} \quad = \quad & \Big\{(\mathbf{U}^N_{c,1}(\mathrm{k}_{c,1}), \mathbf{U}^N_{c,2}(\mathrm{k}_{c,2})) : \mathbf{U}^N_{c,1}(\mathrm{k}_{c,1}) \in B^{(1)}_k,\ \mathbf{U}^N_{c,2}(\mathrm{k}_{c,2}) \in B^{(2)}_l, \\ & \qquad (\mathbf{U}^N_{c,1}(\mathrm{k}_{c,1}), \mathbf{U}^N_{c,2}(\mathrm{k}_{c,2})) \in T^N_\epsilon(\mathbf{U}_{c,1}\mathbf{U}_{c,1})\Big\}. \end{aligned}$$

Declare an error if a typical codeword pair is not found in the bin.

2. *Codewords for public messages:* For $i = 1, 2$, generate $2^{NR_{pc,i}}$ codewords, $X^N_{pc,i}(\mathrm{v}_{pc,i}),\ \mathrm{v}_{pc,i} \in \mathcal{V}_{pc,i}$, where each symbol is drawn i.i.d. with $X_{pc,i} \sim P_{X_{pc,i}}(\cdot)$.

3. *Codewords for sub-private messages:* For $i = 1, 2$, for each $\mathbf{U}^N_{c,i}(\mathrm{k}_{c,i})$ and $X^N_{pc,i}(\mathrm{v}_{pc,i})$, generate $2^{N(R_{p,i}+r_{p,i})}$ codewords, $U^N_{pp,i}(\mathrm{k}_{c,i}, \mathrm{v}_{pc,i}, \mathrm{k}_{p,i})$, for $\mathrm{k}_{p,i} \in \{1, \dots, 2^{N(R_{pp,i}+r_{p,i})}\}$, where each symbol is drawn i.i.d. with

$$U_{pp,i} \sim P_{U_{pp,i}|\mathbf{U}_{c,i},X_{pc,i}}(\cdot|\mathbf{u}_{c,i}(\mathrm{k}_{c,i}), x_{pc,i}(\mathrm{v}_{pc,i})).$$

Partition this set of codewords into $2^{NR_{pp,i}}$ equal-sized bins, with $2^{Nr_{p,i}}$ codewords per bin. Observe that each bin corresponds to a sub-private message index, and the bin corresponding to message $\mathrm{v}_{pp,i} = k$ can be defined as

$$C_k(\mathrm{k}_{c,i}, \mathrm{v}_{pc,i}) = \left\{U^N_{pp,i}(\mathrm{k}_{c,i}, \mathrm{v}_{pc,i}, \mathrm{k}_{p,i}) : \mathrm{k}_{p,i} \in \{(k-1)2^{Nr_{p,i}} + 1, \dots, k2^{Nr_{p,i}}\}\right\}.$$

Further, $(X_{pc,i}, \mathbf{U}_{c,i})$ play the role of the cloud center while $U_{pp,i}$ is the satellite codeword. Codeword assignments are then revealed to both the encoders and the decoders.

Encoding and transmission:

1.*Joint binning for common messages:* For the given common messages, $(\mathrm{v}_{c,1}, \mathrm{v}_{c,2})$, the encoders try to find a pair of codewords $\mathbf{U}_{c,1}(\mathrm{k}_{c,1}), \mathbf{U}_{c,2}(\mathrm{k}_{c,2})$ from the product bin $B_{\mathrm{v}_{c,1},\mathrm{v}_{c,2}}$.

2. *Public message:* For the given message, $\mathrm{v}_{pc,1}$, encoder 1 picks $X^N_{pc,1}(\mathrm{v}_{pc,1})$.

3. *Binning for the sub-private message:* Given $\mathrm{v}_{pc,1}, \mathrm{k}_{c,1}$, and for given $\mathrm{v}_{pp,1}$, encoder 1 tries to find a codeword $U_{pp,1}(\mathrm{k}_{p,1})$ in the bin $C_{\mathrm{v}_{pp,1}}$, such that

$$\big(U^N_{pp,1}(\mathrm{k}_{c,1}, \mathrm{v}_{pc,1}, \mathrm{k}_{p,1}), \mathbf{U}^N_{c2}(\mathrm{k}_{c2})\big) \in T^N_\epsilon(U_{pp,1}\mathbf{U}_{c,2}|X_{pc,1}\mathbf{U}_{c,1}).$$

Encoder 1 declares an error otherwise.

4.*Codeword for transmission:* S_1 computes

$$X_1(\mathrm{v}_{pc,1}, \mathrm{k}_{c,1}, \mathrm{k}_{c,2}, \mathrm{k}_{p,1}) = f_1(X_{pc,1}, X_{c1}, X_{c2}, X_{pp,1}),$$

and transmits X^N_1. A similar procedure is performed at S_2 with index "1" and "2" exchanged.

Decoding: We illustrate the decoding procedure for the decoder of D_1. A similar procedure holds for the other decoder too. Given the received signal Y^N_1, decoder 1 performs the joint decoding of the common, public and sub-private messages. The decoder checks if there exists an unique tuple $(\hat{\mathrm{k}}_{c,1}, \hat{\mathrm{v}}_{pc,1}, \hat{\mathrm{k}}_{pp,1})$, and a $\hat{\mathrm{v}}_{pc,2}$ such that

$$\Big(\mathbf{U}^N_{c,1}(\hat{\mathrm{k}}_{c,1}), X^N_{pc,1}(\hat{\mathrm{v}}_{pc,1}), X^N_{pc,2}(\hat{\mathrm{v}}_{pc,2}), U^N_{pp,1}(\hat{\mathrm{v}}_{pc,1}, \hat{\mathrm{k}}_{c,1}, \hat{\mathrm{k}}_{pp,1}), Y^N_1\Big) \in T^N_\epsilon(\mathbf{U}_{c,1}X_{pc,1}X_{pc,2}U_{pp,1}Y_1),$$

and puts out $(\hat{\mathrm{v}}_{c,1}, \hat{\mathrm{v}}_{pc,1}, \hat{\mathrm{v}}_{pp,1})$ as the decoded message set, where $\hat{\mathrm{v}}_{c,1}$ and $\hat{\mathrm{v}}_{pp,1}$ are indices of the bins containing the respective typical codewords, i.e., $\mathbf{U}^N_{c,1}(\hat{\mathrm{k}}_{c,1}) \in B^{(1)}_{\hat{\mathrm{v}}_{c,1}}$ and $U^N_{pp,1}(\hat{\mathrm{k}}_{c,1}, \hat{\mathrm{v}}_{pc,1}, \hat{\mathrm{k}}_{p,1}) \in C_{\hat{\mathrm{v}}_{pp,1}}(\hat{\mathrm{k}}_{c,1}, \hat{\mathrm{v}}_{pc,1})$.

Analysis: By the symmetry of random code generation, the performance of the codes does not depend on the message set transmitted. Therefore, without loss of generality, we assume that message sets $\mathrm{v}_1 = (\mathrm{v}_{pc,1}, \mathrm{v}_{c,1}, \mathrm{v}_{pp,1}) = (1, 1, 1)$ and $\mathrm{v}_2 = (\mathrm{v}_{pc,2}, \mathrm{v}_{c,2}, \mathrm{v}_{pp,2}) = (1, 1, 1)$ are sent.

Encoding errors: The encoding errors can arise because of two events. The first corresponds to the binning of common messages in which encoders declare an error if

$$\left\{ \nexists (\mathbf{U}_{c,1}^N(\mathrm{k}_{c,1}), \mathbf{U}_{c,2}^N(\mathrm{k}_{c,2})) \in B_1^{(1)} \times B_1^{(2)} : (\mathbf{U}_{c,1}^N(\mathrm{k}_{c,1}), \mathbf{U}_{c,2}^N(\mathrm{k}_{c,2})) \in B_{1,1} \right\}.$$

Following [86], it can be shown that the probability of this event can be made arbitrarily small if N is very large, ϵ is small, and

$$r_{c1} + r_{c2} \geq I\left(\mathbf{U}_{c1};\ \mathbf{U}_{c2}\right), \tag{B.1}$$

which yields (3.3). The second encoding error corresponds to the binning of sub-private message against the common message codewords. Suppose it has been determined by both the encoders that $(\mathrm{k}_{c,1}, \mathrm{k}_{c,2}) = (1, 1)$. Then, encoder 1 declares an error if

$$\left\{ \nexists\ U_{pp,1}^N(1, 1, \mathrm{k}_{p,1}) \in C_1(1, 1) : \left(U_{pp,1}^N(1, 1, \mathrm{k}_{p,1}), \mathbf{U}_{c,2}^N(1)\right) \in T_\epsilon^N(U_{pp,1}\mathbf{U}_{c,2}|X_{pc,1}\mathbf{U}_{c,1}) \right\}.$$

Similarly, encoder 2 declares an error if

$$\left\{ \nexists\ U_{pp,2}^N(1, 1, \mathrm{k}_{p,2}) \in C_2(1, 1) : \left(U_{pp,2}^N(1, 1, \mathrm{k}_{p,2}), \mathbf{U}_{c,1}^N(1)\right) \in T_\epsilon^N(U_{pp,2}\mathbf{U}_{c,1}|X_{pc,2}\mathbf{U}_{c,2}) \right\}.$$

Based on [55, 56], it follows that the probability of the above error event can be made arbitrarily small if N is very large, ϵ is small, and

$$\begin{aligned} r_{p1} &\geq I\left(U_{pp1};\ \mathbf{U}_{c,2}|\mathbf{U}_{c,1}, X_{pc,1}\right) \\ r_{p2} &\geq I\left(U_{pp,2};\ \mathbf{U}_{c,1}|\mathbf{U}_{c,2}, X_{pc,2}\right), \end{aligned}$$

yielding (3.4) and (3.5).

Decoding errors: We consider decoding error analysis for only decoder 1. (By symmetry, similar analysis holds for decoder 2.) Define the event

$$E_{i,j,k,l} = \left\{ (X_{pc,1}^N(i), X_{pc,2}^N(j), \mathbf{U}_{c,1}^N(k), U_{pp,1}^N(i,k,l), Y_1^N) \in T_\epsilon^N(X_{pc,1}X_{pc,2}\mathbf{U}_{c,1}U_{pp,1}Y_1) \right\}.$$

Decoding errors occur if either the correct codewords are not typical with the received sequence or there is a pair of incorrect codewords that are typical with the received

i	j	k	l	Condition
*	*	*	*	$R_{pc1}+R_{pc2}+R_{c1}+r_{c1}+R_{pp1}+r_{p1} \leq I(Y;X_{pc1},X_{pc2},\mathbf{U}_{c1},U_{pp,1})$
*	*	1	*	$R_{pc1}+R_{pc2}+R_{pp1}+r_{p1} \leq I(Y;X_{pc1},X_{pc2},U_{pp,1}\|\mathbf{U}_{c1})$
*	1	*	*	$R_{pc1}+R_{c1}+r_{c1}+R_{pp1}+r_{p1} \leq I(Y;X_{pc1},\mathbf{U}_{c1},U_{pp,1}\|X_{pc2})$
*	1	1	*	$R_{pc1}+R_{pp1}+r_{p1} \leq I(Y;X_{pc1},U_{pp,1}\|\mathbf{U}_{c1},X_{pc2})$
1	*	*	*	$R_{pc2}+R_{c1}+r_{c1}+R_{pp1}+r_{p1} \leq I(Y;X_{pc2},\mathbf{U}_{c1},U_{pp,1}\|X_{pc1})$
1	*	1	*	$R_{pc2}+R_{pp1}+r_{p1} \leq I(Y;X_{pc2},U_{pp,1}\|X_{pc1},\mathbf{U}_{c1})$
1	1	*	*	$R_{c1}+r_{c1}+R_{pp1}+r_{p1} \leq I(Y;\mathbf{U}_{c1},U_{pp,1}\|X_{pc1},X_{pc2})$
1	1	1	*	$R_{pp1}+r_{p1} \leq I(Y;U_{pp,1}\|X_{pc1},\mathbf{U}_{c1},X_{pc2})$

Table B.1: Table depicting the error events ($i = *$ if $i \neq 1$)

.

sequence. That is, the first error corresponds to E^c_{1111}, and the latter corresponds to $\bigcup_{\{i,j,k,l\}\neq\{1,1,1,1\}} E_{ijkl}$. Therefore, we have the probability of error as

$$P_e = \mathbb{P}(E^c_{1111} \cup \bigcup_{\substack{\{i,j,k,l\} \\ \neq\{1,1,1,1\}}} E_{ijkl}) \leq \mathbb{P}(E^c_{1111}) + \sum_{\substack{\{i,j,k,l\} \\ \neq\{1,1,1,1\}}} \mathbb{P}(E_{ijkl}).$$

It is clear that, from the AEP [52, 53], $\mathbb{P}(E^c_{1111}) \to 0$. Rest of the error events are listed in Table B.1 along with the rate constraints to achieve the corresponding events with arbitrarily low error probability. Steps to evaluate these constraints are standard and are omitted here.

Note that some constraints are redundant due to the fact that the error-free decoding of $\mathrm{v}_{pp,i}$ is possible only if both $\mathrm{v}_{pc,1}$ and $\mathrm{v}_{c,1}$ are decoded correctly. Also, decoder 1 would not be interested in the error-free decoding of $\mathrm{v}_{pc,2}$, once $\mathrm{v}_{pc,1}, \mathrm{v}_{c,1}$ and $\mathrm{v}_{pp,1}$ are decoded without any error.

Summarizing, the error events at decoder 1 can be made to have arbitrarily small probability if the rate-tuple $(R_{pc,1}, R_{pc,2}, R_{c,1}, R_{pp,1})$, obey the following set of constraints:

$$R_{pp1} \leq I(Y; U_{pp,1} | X_{pc1}, \mathbf{U}_{c1}, X_{pc2}) - r_{p1}$$

$$R_{pc1} + R_{pp1} \leq I(Y; X_{pc1}, U_{pp,1} | \mathbf{U}_{c1}, X_{pc2}) - r_{p1}$$

$$R_{pc2} + R_{pp1} \leq I(Y; X_{pc2}, U_{pp,1} | X_{pc1}, \mathbf{U}_{c1}) - r_{p1}$$

$$R_{c1} + R_{pp1} \leq I(Y; \mathbf{U}_{c1}, U_{pp,1} | X_{pc1}, X_{pc2}) - r_{p1} - r_{c1}$$

$$R_{pc1} + R_{pc2} + R_{pp1} \leq I(Y; X_{pc1}, X_{pc2}, U_{pp,1} | \mathbf{U}_{c1}) - r_{p1}$$

$$R_{pc1} + R_{c1} + R_{pp1} \leq I(Y; X_{pc1}, \mathbf{U}_{c1}, U_{pp,1} | X_{pc2}) - r_{c1} - r_{p1}$$

$$R_{pc2} + R_{c1} + R_{pp1} \leq I(Y; X_{pc2}, \mathbf{U}_{c1}, U_{pp,1} | X_{pc1}) - r_{c1} - r_{p1}$$

$$R_{pc1} + R_{pc2} + R_{c1} + R_{pp1} \leq I(Y; X_{pc1}, X_{pc2}, \mathbf{U}_{c1}, U_{pp,1}) - r_{c1} - r_{p1}.$$

For convenience, we denote this set of constraints as $\mathcal{S}_1$. By symmetry, at decoder 2, the rate-tuple $(R_{pc,1}, R_{pc,2}, R_{c,2}, R_{pp,2})$ obey a set of constraints, $\mathcal{S}_2$, where $\mathcal{S}_2$ is same as $\mathcal{S}_1$, except that indices "1" and "2" interchanged everywhere in $\mathcal{S}_1$. Thus, for arbitrarily low probability of error, the rate-tuple $(R_{pc,1}, R_{pc,2}, R_{c,1}, R_{c,2}, R_{pp,1}, R_{pp,2})$ must belong to $\mathcal{S}_1 \bigcap \mathcal{S}_2$. Applying Fourier-Motzkin [87] elimination to the constraints $\mathcal{S}_1$ and $\mathcal{S}_2$, we eliminate $R_{pc,1}$ and $R_{pc,2}$ to obtain constraint sets $\mathcal{B}_1$ and $\mathcal{B}_2$.

B.2 Proof of Theorem 3.2

Code construction: First, pick a distribution

$$p(\mathbf{x}_{c1})p(\mathbf{x}_{c2})p(x_{pc1}|\mathbf{x}_{c2}, \mathbf{x}_{c1})p(x_{pc2}|\mathbf{x}_{c2}, \mathbf{x}_{c1})p(x_1|x_{pc1}, \mathbf{x}_{c2}, \mathbf{x}_{c1})p(x_1|x_{pc1}, \mathbf{x}_{c2}, \mathbf{x}_{c1}).$$

The private message $\mathrm{v}_{p,i}$ is further split into $(\mathrm{v}_{pp,i}, \mathrm{v}_{pc,i})$ with rates $(R_{pp,i}, R_{pc,i})$ such that $R_{p,i} = R_{pp,i} + R_{pc,i}$.

1. *Codewords for common messages:* For $i = 1, 2$, generate $2^{NR_{c,i}}$, 2×1-vector codewords, $\mathbf{X}_{c,i}^N(\mathrm{v}_{c,i})$, for $\mathrm{v}_{c,i} \in \mathcal{V}_{c,i}$, where each symbol is drawn i.i.d. with $\mathbf{X}_{c,i} \sim P_{\mathbf{X}_{c,i}}(\cdot)$.

2. *Codewords for public messages:* For $i = 1, 2$, for each common codeword pair, $(\mathbf{X}_{c,1}^N(\mathrm{v}_{c,1}), \mathbf{X}_{c,2}^N(\mathrm{v}_{c,2}))$, generate $2^{NR_{pc,i}}$ codewords, $X_{pc,i}^N(\mathrm{v}_{c,1}, \mathrm{v}_{c,2}, \mathrm{v}_{pc,i})$, $\mathrm{v}_{pc,i} \in \mathcal{V}_{pc,i}$, where

each symbol is drawn i.i.d. with $X_{pc,i} \sim P_{X_{pc,i}|\mathbf{X}_{c,1}\mathbf{X}_{c,2}}(\cdot)$.

3. *Codewords for sub-private messages:* At encoder i, $i = 1, 2$, for each set of common and public codewords, $(\mathbf{X}_{c,1}^N(\mathrm{v}_{c,1}), \mathbf{X}_{c,2}^N(\mathrm{v}_{c,2}), X_{pc,i}^N(\mathrm{v}_{c,1}, \mathrm{v}_{c,2}, \mathrm{v}_{pc,i}))$, generate $2^{NR_{pp,i}}$ codewords, $X_i^N(\mathrm{v}_{c,1}, \mathrm{v}_{c,2}, \mathrm{v}_{pc,i}, \mathrm{v}_{pp,i})$, for $\mathrm{v}_{pp,i} \in \mathcal{V}_{pp,i}$, where each symbol is drawn i.i.d. with $X_i \sim P_{X_i|\mathbf{X}_{c,1},\mathbf{X}_{c,1},X_{pc,i}}(\cdot|\mathbf{x}_{c,1}, \mathbf{x}_{c,2}, x_{pc,i})$.

Codeword assignments are revealed to both the encoders and the decoders.

Encoding and transmission:

Given $(\mathrm{v}_{c,1}, \mathrm{v}_{c2}, \mathrm{v}_{pc,1}, \mathrm{v}_{pp,1}) = (i, j, k, m)$, S_1 transmits $X_1^N(i, j, k, m)$. A similar procedure is adopted at S_2 with index "1" and "2" exchanged.

Decoding: We illustrate the decoding procedure for the decoder of D_1. A similar procedure holds for the other decoder too.

Given the received signal Y_1^N, the decoder checks if there exists an unique tuple $(\hat{i}, \hat{j}, \hat{k}, \hat{m})$ and a $\hat{l}$, such that

$$\left(\mathbf{X}_{c,1}^N(\hat{i}), \mathbf{X}_{c,2}^N(\hat{j}), X_{pc,1}^N(\hat{i}, \hat{j}, \hat{k}), X_{pc,2}^N(\hat{i}, \hat{j}, \hat{l}), X_1^N(\hat{i}, \hat{j}, \hat{k}, \hat{m}), Y_1^N\right) \in T_\epsilon^N(\mathbf{X}_{c,1}\mathbf{X}_{c,2}X_{pc,1}X_{pc,2}X_1Y_1)$$

and puts out $(\mathrm{v}_{c,1}, \mathrm{v}_{pc,1}, \mathrm{v}_{pp,1}) = (\hat{i}, \hat{j}, \hat{m})$ as the decoded message set. An error is declared if either such a tuple does not exist or the tuple is not unique.

Analysis: By the symmetry of random code generation, the performance of the codes does not depend on the message set transmitted. Therefore, without loss of generality, we assume that message sets $\mathrm{v}_1 = (\mathrm{v}_{pc,1}, \mathrm{v}_{c,1}, \mathrm{v}_{pp,1}) = (1, 1, 1)$ and $\mathrm{v}_1 = (\mathrm{v}_{pc,2}, \mathrm{v}_{c,2}, \mathrm{v}_{pp,2}) = (1, 1, 1)$ are sent. We consider decoding error analysis for only decoder 1. By symmetry, similar analysis holds for decoder 2. Define the event

$$E_{i,j,k,l,m} = \left\{(\mathbf{X}_{c,1}^N(i), \mathbf{X}_{c,2}^N(j), X_{pc,1}^N(i, j, k), X_{pc,2}^N(i, j, l), X_1^N(i, j, k, m), Y_1^N) \in T_\epsilon^N(\mathbf{X}_{c,1}\mathbf{X}_{c,2}X_{pc,1}X_{pc,2}X_1Y_1)\right\}.$$

i	j	k	l	m	Constraints
*	*	*	*	*	$R_{c1}+R_{c2}+R_{pc1}+R_{pc2}+R_{pp1} \leq I(Y_1;X_1,X_{pc2})$
*	1	*	*	*	$R_{c1}+R_{pc1}+R_{pc2}+R_{pp1} \leq I(Y_1;X_1,X_{pc2}\vert\mathbf{X}_{c,2})$
1	*	*	*	*	$R_{c2}+R_{pc1}+R_{pc2}+R_{pp1} \leq I(Y_1;X_1,X_{pc2}\vert\mathbf{X}_{c,1})$
1	1	*	*	*	$R_{pc1}+R_{pc2}+R_{pp1} \leq I(Y_1;X_1,X_{pc2}\vert\mathbf{X}_{c,1},\mathbf{X}_{c,2})$
1	1	*	1	*	$R_{pc1}+R_{pp1} \leq I(Y_1;X_1\vert X_{pc2}\mathbf{X}_{c,1},\mathbf{X}_{c,2})$
1	1	1	*	*	$R_{pc2}+R_{pp1} \leq I(Y_1;X_1,X_{pc2}\vert X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,2})$
1	1	1	1	*	$R_{pp1} \leq I(Y_1;X_1\vert X_{pc2},X_{pc1},\mathbf{X}_{c,1},\mathbf{X}_{c,1})$

Table B.2: Table depicting the error events ($i = *$ if $i \neq 1$)

.

Decoding error occurs if either the correct codewords are not typical with the received sequence or there is a set of incorrect codewords that are typical with the received sequence. That is, the first error corresponds to E^c_{11111}, and the latter corresponds to $\bigcup_{\{i,j,k,l,m\}\neq\{1,1,1,1,1\}} E_{ijklm}$. Therefore, we have the probability of error as

$$\begin{aligned} P_{e,1} &= \mathbb{P}(E^c_{11111} \cup \bigcup_{\substack{\{i,j,k,l,m\} \\ \neq\{1,1,1,1,1\}}} E_{ijklm}) \\ &\leq \mathbb{P}(E^c_{1111}) + \sum_{\substack{\{i,j,k,l,m\} \\ \neq\{1,1,1,1,1\}}} \mathbb{P}(E_{ijklm}). \end{aligned}$$

It is clear that, from the AEP [52], $\mathbb{P}(E^c_{11111}) \to 0$. Rest of the error events are listed in Table B.2 along with the rate constraints to achieve $P_{e,1} < \epsilon$ for any $\epsilon > 0$. Steps to evaluate these constraints are standard and are omitted here.

Note that constraints corresponding to some events are redundant due to the fact that the error-free decoding of $\mathrm{v}_{pp,1}$ is possible only if $\mathrm{v}_{c,1}, \mathrm{v}_{c,2}$ and $\mathrm{v}_{pc,1}$ are decoded correctly. Also, the fact that D_1 would not be interested in the error-free decoding of $\mathrm{v}_{pc,2}$, once $\mathrm{v}_{pc,1}, \mathrm{v}_{c,1}$ and $\mathrm{v}_{pp,1}$ are decoded without any error, makes the event $(1, 1, 1, l \neq 1, 1)$ irrelevant at D_1. The above analysis yields the following constraints on

$(R_{pc,1}, R_{pc,2}, R_{c,1}, R_{c,2}, R_{pp,1})$ at decoder 1.

$$R_{pp1} \leq I(Y_1; X_1 | X_{pc2}, X_{pc1}, \mathbf{X}_{c,1}, \mathbf{X}_{c,2})$$

$$R_{pc2} + R_{pp1} \leq I(Y_1; X_1, X_{pc2} | X_{pc1}, \mathbf{X}_{c,1}, \mathbf{X}_{c,2})$$

$$R_{pc1} + R_{pp1} \leq I(Y_1; X_1 | X_{pc2} \mathbf{X}_{c,1}, \mathbf{X}_{c,1})$$

$$R_{pc1} + R_{pc2} + R_{pp1} \leq I(Y_1; X_1, X_{pc2} | \mathbf{X}_{c,1}, \mathbf{X}_{c,2})$$

$$R_{c2} + R_{pc1} + R_{pc2} + R_{pp1} \leq I(Y_1; X_1, X_{pc2} | \mathbf{X}_{c,1})$$

$$R_{c1} + R_{pc1} + R_{pc2} + R_{pp1} \leq I(Y_1; X_1, X_{pc2} | \mathbf{X}_{c,2})$$

$$R_{c1} + R_{c2} + R_{pc1} + R_{pc2} + R_{pp1} \leq I(Y_1; X_1, X_{pc2}).$$

For convenience, we denote this set of constraints as $\mathcal{T}_1$. By symmetry, at decoder 2, the rate-tuple $(R_{pc,1}, R_{pc,2}, R_{c,1}, R_{c,2}, R_{pp,2})$ obey a set of constraints, $\mathcal{T}_2$, where $\mathcal{T}_2$ is same as $\mathcal{T}_1$, except that indices "1" and "2" interchanged everywhere in $\mathcal{T}_1$. Thus, for arbitrarily low probability of error, i.e., $P_e < \epsilon$ for any $\epsilon > 0$, the rate-tuple $(R_{pc,1}, R_{pc,2}, R_{c,1}, R_{c,2}, R_{pp,1}, R_{pp,2})$ must belong to $\mathcal{T}_1 \bigcap \mathcal{T}_2$. Applying Fourier-Motzkin [87] elimination to the constraints $\mathcal{T}_1$ and $\mathcal{T}_2$, we eliminate $R_{pc,1}$ and $R_{pc,2}$ to obtain constraint sets $\mathcal{C}_1$ and $\mathcal{C}_2$.

B.3 Proof of Proposition 3.1

For the MAC formed by $(U_{c,1}, U_{c,2}, Y_{d,1})$, we have the rate region as (see (3.14))

$$R_{c,1} \leq C\left(\frac{(\mathbf{a}^\dagger \mathbf{h}_1)^2}{\sigma_\zeta^2}\right) \tag{B.2}$$

$$R_{c,2} \leq C\left(\frac{(\mathbf{a}^\dagger \mathbf{h}_2)^2}{\sigma_\zeta^2}\right) \tag{B.3}$$

$$R_{c,1} + R_{c,2} \leq C\left(\frac{(\mathbf{a}^\dagger \mathbf{h}_1)^2 + (\mathbf{a}^\dagger \mathbf{h}_2)^2}{\sigma_\zeta^2}\right). \tag{B.4}$$

For the MAC formed by $(U_{c,1}, U_{c,2}, Y_{d,2})$, the rate region can be obtained from (B.2)-(B.4) by swapping $R_{c,1}$ and $R_{c,2}$ (see (3.15)). The intersection of the rate regions of these

MACs is given by

$$R_{c,1} \leq \min\left\{C\left(\frac{(\mathbf{a}^\dagger\mathbf{h}_2)^2}{\sigma_\zeta^2}\right), C\left(\frac{(\mathbf{a}^\dagger\mathbf{h}_1)^2}{\sigma_\zeta^2}\right)\right\} \tag{B.5}$$

$$R_{c,2} \leq \min\left\{C\left(\frac{(\mathbf{a}^\dagger\mathbf{h}_2)^2}{\sigma_\zeta^2}\right), C\left(\frac{(\mathbf{a}^\dagger\mathbf{h}_1)^2}{\sigma_\zeta^2}\right)\right\} \tag{B.6}$$

$$R_{c,1} + R_{c,2} \leq C\left(\frac{(\mathbf{a}^\dagger\mathbf{h}_1)^2 + (\mathbf{a}^\dagger\mathbf{h}_2)^2}{\sigma_\zeta^2}\right). \tag{B.7}$$

We first obtain the optimal power allocation policies for both the sum-rate and individual rate maximization, and then demonstrate that they are equal.

a. Maximizing the sum-rate:

We know that the sum-rate constraint is

$$R_{c,1} + R_{c,2} \quad \leq \quad C\left(\frac{(\mathbf{h}_1^\dagger\mathbf{a})^2 + (\mathbf{h}_2^\dagger\mathbf{a})^2}{\sigma_\zeta^2}\right). \tag{B.8}$$

The goal is to find the power allocation vector that maximizes the sum-rate given by the above equation. This problem can be posed as a constrained problem as

$$\text{Maximize } R_{c,1} + R_{c,2} \text{ subject to } \|\mathbf{a}\|^2 = P_c.$$

Since logarithm is increasing in its argument, the above problem is equivalent to

$$\max_{\mathbf{a}:\|\mathbf{a}\|^2 = P_c} (\mathbf{h}_1^\dagger\mathbf{a})^2 + (\mathbf{h}_2^\dagger\mathbf{a})^2.$$

Let $\mathbf{c} = \mathbf{a}/\sqrt{P_c}$, so that $\|\mathbf{c}\|^2 = \|\mathbf{a}\|^2/P_c$. Also, let

$$A = \mathbf{h}_1\mathbf{h}_1^\dagger + \mathbf{h}_2\mathbf{h}_2^\dagger = \begin{bmatrix} h^2+g^2 & 2hg \\ 2hg & h^2+g^2 \end{bmatrix}.$$

Then, the optimization problem simplifies to

$$\max_{\mathbf{c}:\|\mathbf{c}\|^2=1} \mathbf{c}^\dagger A\mathbf{c},$$

which is a standard eigenvalue problem. It turns out that the solution ($\mathbf{c}^*$) to the above problem is the eigenvector corresponding to the maximum eigenvalue (λ_{max}) of the matrix A. We can verify that $\lambda_{max} = (h+g)^2$, and

$$\mathbf{c}^* = \left[1/\sqrt{2}, 1/\sqrt{2}\right].$$

Thus, the optimal power allocation is $P_{cm} = P_{cu} = P_c/2$.

b. Maximizing the individual rates:

Since $P_{cm} = P_c - P_{cu}$, we can observe that, in the RHS of (B.5) or (B.6), the first term is decreasing in P_{cm}, while the second term is increasing in P_{cm}. Thus, the maxima for $R_{c,1}$ and $R_{c,2}$ are achieved when both the terms of RHS are equal. This is possible only when $P_{cm} = P_c - P_{cm}$. That is $P_{cm} = P_c/2$.

Substituting this into (B.5) and (B.6), and making use of the symmetry in the rates, we obtain

$$R_c = \min\left\{C\left(\frac{(h+g)^2 P_c}{2\sigma_\zeta^2}\right), \frac{1}{2}C\left(\frac{(h+g)^2 P_c}{\sigma_\zeta^2}\right)\right\}.$$

Setting $z = \frac{(h+g)^2 P_c}{\sigma_\zeta^2}$ and observing that $(1+\frac{z}{2})^2 - (1+z) = \frac{z^2}{4} \geq 0$, we conclude that

$$R_c = \frac{1}{2}C\left(\frac{(h+g)^2 P_c}{\sigma_\zeta^2}\right). \tag{B.9}$$

B.4 Determining DPC matrices and derivations of (3.11) and (3.12)

The derivation of optimal input covariance matrices applies the techniques in [58]. For the sake of completeness, we present the derivations in detail here. Considering the encoding strategy π_2, the aim is to design the DPC coefficients Σ_1', Σ_2' which maximize the sum rate given by

$$T_{\pi_2}(P_p, P_p) = \max_{\substack{\Sigma_1', \Sigma_2': \\ \mathrm{tr}(\Sigma_1' + \Sigma_2') \leq 2P_c'}} \frac{1}{2}C\left(\mathbf{h}_1^\dagger \Sigma_1' \mathbf{h}_1\right) + \frac{1}{2}C\left(\frac{\mathbf{h}_2^\dagger \Sigma_2' \mathbf{h}_2}{1 + \mathbf{h}_2^\dagger \Sigma_1' \mathbf{h}_2}\right),$$

where for convenience we have normalized the noise variance to unity by letting $\Sigma_1' = \frac{\Sigma_1}{\sigma_\zeta^2}, \Sigma_2' = \frac{\Sigma_2}{\sigma_\zeta^2}$ and $P_c' = \frac{P - P_p}{\sigma_\zeta^2}$. Using the MAC-BC transformation [58], we solve for Σ_1' and Σ_2'. To this end, let P_1, P_2 be the optimal powers allocated for the transmitters in the corresponding "dual-MAC". Then, for $j = 1, 2$, define

$$A_j = 1 + \mathbf{h}_j^\dagger \sum_{l=1}^{j-1} \Sigma_l' \mathbf{h}_j\,; \quad \mathbf{B}_j = \mathbf{I} + \sum_{l=j+1}^{2} P_l \mathbf{h}_l \mathbf{h}_l^\dagger.$$

Then the optimal covariance matrix is given by

$$\Sigma'_j \quad = \quad A_j P_j \cdot \mathbf{B}_j^{-\frac{1}{2}} G_j^\dagger F_j G_j F_j^\dagger \mathbf{B}_j^{-\frac{1}{2}},$$

where F_j, G_j are left and right singular vectors of $\mathbf{B}_j^{-\frac{1}{2}} \mathbf{h}_j A_j^{-\frac{1}{2}}$ obtained by the SVD $\mathbf{B}_j^{-\frac{1}{2}} \mathbf{h}_j A_j^{-\frac{1}{2}} = F_j \Lambda_j G_j$.

Consider the dual MAC problem. The dual uplink corresponds to a MAC having two single antenna users and a receiver with two antennas. We need to find optimal transmit powers P_1, P_2, that achieve the sum-capacity of the MAC, as the solution to the following optimization problem.

$$C_{MAC} = \max_{\substack{P_1, P_2: \\ P_1 + P_2 \leq 2P'_c}} \frac{1}{4} \log \left(\det \left[\mathbf{I} + P_1 \mathbf{h}_1 \mathbf{h}_1^\dagger + P_2 \mathbf{h}_2 \mathbf{h}_2^\dagger \right] \right).$$

By the symmetric structure in the dual uplink channels, it is clear that allocating equal power to each user maximizes the sum rate of the system. Thus sum-capacity is achieved with $P_1 = P_2 = P'_c$ and with any decoding order. Furthermore, we can verify that the sum-capacity is

$$C_{MAC} = \frac{1}{4} \log \left(1 + (h^2 - g^2)^2 (P'_c)^2 + 2P'_c (h^2 + g^2) \right). \tag{B.10}$$

Now we use the MAC-BC transformations to find the covariance matrices (Σ'_1, Σ'_2). On simplification, we obtain

$$A_1 = 1; \mathbf{B}_1 = \mathbf{I} + P'_c \mathbf{h}_2 \mathbf{h}_2^\dagger; G_1 = 1; F_1 = \frac{1}{\Lambda_1} \mathbf{B}_1^{-\frac{1}{2}} \cdot \mathbf{h}_1; \Lambda_1 = \sqrt{\mathbf{h}_1^\dagger \mathbf{B}_1^{-1} \cdot \mathbf{h}_1}$$

and the covariance matrix Σ'_1 is given by

$$\begin{aligned} \Sigma'_1 &= \frac{P'_c}{\Lambda_1^2} \mathbf{B}_1^{-1} \cdot \mathbf{h}_1 \mathbf{h}_1^\dagger \mathbf{B}_1^{-1} \\ &= \frac{P'_c}{\left(1 + P'_c \|\mathbf{h}\|^2\right)\left(\|\mathbf{h}\|^2 + P'_c h_\Delta^2\right)} \begin{bmatrix} h^2 \left(1 + P'_c h_\Delta\right)^2 & gh \left(1 - P'^2_c h_\Delta^2\right) \\ gh \left(1 - P'^2_c h_\Delta^2\right) & g^2 \left(1 - P'_c h_\Delta\right)^2 \end{bmatrix}, \end{aligned}$$

where $h_\Delta = h^2 - g^2$.

Similarly, for $j = 2$,

$$A_2 \quad = \quad \frac{\|\mathbf{h}\|^2 \left(1 + 2P'_c \|\mathbf{h}\|^2 + P'^2_c h_\Delta^2\right)}{\left(1 + P'_c \|\mathbf{h}\|^2\right)\left(\|\mathbf{h}\|^2 + P'_c h_\Delta^2\right)}, \mathbf{B}_2 = \mathbf{I}; \quad G_2 = 1; \quad F_2 = \frac{\mathbf{h}_2}{\|\mathbf{h}\|}; \Lambda_2 = \sqrt{A_2} \|\mathbf{h}_2\|$$

and the covariance matrix Σ_2' is given by

$$\Sigma_2' = P_c' A_2 \frac{\mathbf{h}_2 \mathbf{h}_2^\dagger}{\|\mathbf{h}\|^2} = \frac{P_c'\left(1 + 2P_c'\|\mathbf{h}\|^2 + P_c'^2 h_\Delta^2\right)}{\left(1 + P_c'\|\mathbf{h}\|^2\right)\left(\|\mathbf{h}\|^2 + P_c' h_\Delta^2\right)} \begin{bmatrix} g^2 & hg \\ hg & h^2 \end{bmatrix}.$$

Now one can easily verify that

$$P_{c,1}^{\pi_2} = [\Sigma_1 + \Sigma_2]_{1,1} = P - P_p; \quad P_{c,2}^{\pi_2} = [\Sigma_1 + \Sigma_2]_{2,2} = P - P_p.$$

Furthermore, we can verify that

$$T_{\pi_2}(P_p, P_p) = \frac{1}{4}\log\left(1 + (h^2 - g^2)^2 (P_c')^2 + 2P_c'(h^2 + g^2)\right), \tag{B.11}$$

which is equal to the sum-capacity of the dual-MAC (B.10).

arks: Note that the covariance matrices Σ_1 and Σ_2 are of single rank. Thus in our case, dirty paper coding involving scalar signals is sufficient to achieve the maximum sum rate. Indeed, by letting $\mathbf{b}_1 = \frac{\sqrt{P_c}}{\Lambda_1}\mathbf{B}^{-1}\mathbf{h}_1, \quad \mathbf{b}_2 = \sqrt{P_c A_2}\frac{\mathbf{h}_2}{\|\mathbf{h}_2\|}$, one can obtain vectors $\mathbf{X}_{c,1} = \mathbf{b}_1 X_{c,1}$ and $\mathbf{X}_{c,2} = \mathbf{b}_2 X_{c,2}$ where $X_{c,1}, X_{c,2}$ are independent Gaussian random variables of unit variance.