ABSTRACT

Title of dissertation: COOPERATIVE COMMUNICATION WITH WIRELESS NETWORK CODING

Wei Guan, Doctor of Philosophy, 2013

Dissertation directed by: Professor K. J. Ray Liu
Department of Electrical and Computer Engineering

Cooperative communication is a new communication paradigm that allows multiple transceivers to collaborate as a cluster for data transmission, and such clustering could greatly improve the transmission quality due to cooperative diversity. For conventional cooperation protocols, each cooperating device uses orthogonal channels to relay different messages for mitigating co-channel interference and avoiding transmission collision, but doing so would significantly reduce the bandwidth efficiency. One way to tackle this issue is to use wireless network coding, in which different messages are smartly combined at cooperating devices to save the channel use for data relaying.

Network coding has been widely used in wireline networks, but only until very recently was grafted onto the wireless networks. In the research community, it has been unknown for a long time whether network-coded cooperation is able to achieve the same diversity gain as the conventional diversity technique. On the industry side, how to efficiently apply network coding in the current wireless systems has also been an open design problem in the past few years. This thesis work aims to address

these important issues and challenges and provide some theoretical guidelines for real system design.

In the first part of this work, we study the fundamental diversity performance of uncoded cooperation systems with wireless network coding. It is demonstrated that network-coded cooperation generally cannot achieve the same diversity gain as the conventional diversity schemes; however, the diversity loss is usually very limited and occurs only under particular channel conditions. For example, for digital network coding we show that the error propagation issue would cause half of the total available diversity gain to be lost, and we develop several link adaptive schemes to mitigate the diversity loss. For analog network coding, we demonstrate that the associated co-channel interference may reduce the diversity as well, but such loss gradually diminishes as the transmitted power goes up. Finally for non-coherent network coding, we show that when the receivers do not know the channel state information, using blind signal detection would not hurt the dominant diversity gain, and the diversity loss occurs only at modest signal-to-noise ratio.

The second part of this work is focused on coded cooperation systems. The unique feature of coded systems is that the devices could somehow know the network dynamics such as the decoding status of a transmitted packet. We explore two transmission strategies that could efficiently exploit such information. For two-way relay channel, we propose a network-coded retransmission strategy, where wireless relaying is employed only when the direct link is in outage. To reduce the number of retransmissions, network coding is performed in a static or dynamic way to combine the to-be-retransmitted packets intended for different end terminals. We analyze

the throughput and develop power allocation scheme to maximize the throughput. We also develop a hybrid network coding scheme that can fully exploit the network coding gain in the multi-relay environment. Next for wireless uplink channel, we come up with a multi-user cooperation scheme based on node clustering. We develop inter-cluster cooperation strategy and intra-cluster transmit beamforming scheme to exploit the cooperative diversity gain. We demonstrate that there is a basic tradeoff between diversity gain and bandwidth efficiency, and different tradeoffs could be achieved by changing the formation of the clusters.

COOPERATIVE COMMUNICATION WITH WIRELESS NETWORK CODING

by

Wei Guan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor K. J. Ray Liu, Chair/Advisor
Professor Min Wu
Professor Adrian Papamarcou
Professor Gang Qu
Professor Lawrence C. Washington

Dedication

To my parents.

Acknowledgments

First of all, I would like to thank my advisor, Prof. K. J. Ray Liu, for his continuous inspiration, guidance and support on my thesis work. In the past few years, he constantly encouraged me to focus on those most important and challenging problems, and instructed me how to find a way toward innovative ideas and creative solutions. He also gave me a lot of freedom to choose the favorite research topics to strengthen my interest and enthusiasm in research. It is my great honor and pleasure to be able to finish my thesis work with a patient, kind and responsible advisor like Prof. Liu.

I would also like to thank all the members in Signal and Information Group. Because of their friendship, encouragement and helpful discussions, I gain the power to get through various hard times and no longer feel lonely along this long journey. Special thanks to Feng, Yang, Hui and Le, with whom I have had a lot of happy moments over the past years.

I also appreciate Prof. Min Wu, Prof. Adrian Papamarcou, Prof. Gang Qu and Prof. Lawrence C. Washington for serving on my committee and reviewing my thesis. Their comments and advices prove to be pretty helpful in improving the quality of this dissertation.

Lastly, I would like to express my deepest gratitude to my parents for their eternal love, support and comfort. Special gratitude to my grandpa, grandma and aunt, who keep encouraging me to pursue PhD degree but unfortunately leave me forever before the completion of my PhD study. To them I dedicate this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Channel Fading and Diversity

Nowadays, wireless applications such as Wifi, cellular phones and bluetooth have become an important part of daily life. But compared to the conventional wireline networks, wireless networks can only provide very limited data rate because the underlying channel is unreliable in nature. In practice, wireless channel is subject to fading, pathloss, shadowing and co-channel interference, and all these features would greatly degrade the quality of transmitted signals.

Channel fading is one of the major downside to wireless communication. Channel fading is caused by multipath propagation effect, which occurs when the reflectors surrounding the transmitter/receiver happen to create multiple propagation paths for the transmitted signals to traverse. Those multipath components may add constructively or destructively at the receiver side, thus making the amplitude of the received signal fluctuate randomly over time [1–3]. When the channel is in deep fading, the wireless link may totally get disconnected and no information can be delivered reliably.

Diversity techniques have been widely used to combat channel fading. Diversity is the capability to send the same signal repeatedly through independent channels. As the receiver is able to decode the source message as long as there exists at

least one good channel, the chance of link disconnection in cases of deep fading on all the channels could be reduced significantly. Mathematically, the diversity gain is defined as [1–3]

$$d = -\lim_{P\to\infty} \frac{\log P_e}{\log P}, \tag{1.1}$$

where P_e is the error rate and P is the signal-to-noise ratio (SNR). The diversity gain is a measure of the decay rate of transmission error in the high SNR regions. For conventional diversity systems, the scaling law of the error rate has a general form of $P_e = \mathrm{O}\left(\frac{1}{P^M}\right)$, where $f(x) = \mathrm{O}(g(x))$ means that $a \leq \lim_{x\to\infty} \frac{f(x)}{g(x)} \leq b$ for some positive constants a and b. The diversity gain M is equal to the number of independent paths that the transmitted signals traverse toward the receiver.

Conventionally, there are three generic types of diversity: time diversity, frequency diversity and spatial diversity [1–3]. Time diversity is to send the same signal copy in different time slots. To guarantee independent fading, the interval between adjacent transmissions must be greater than the channel coherence time, which would incur large processing delay especially when the channel is in slow fading. Frequency diversity is to send the same signal copy in sufficiently separated frequency bands that experience independent fading. However, frequency diversity is gained at a price of lower bandwidth efficiency, which is costly since frequency resource is pretty scarce.

Spatial diversity is a relatively new technique to address the drawbacks of time diversity and frequency diversity. Spatial diversity is achieved by deploying multiple antennas at the transmitter/receiver, such that there exists one independent propa-

gation path between each pair of transmitter antenna and receiver antenna. Multiple antenna technique has gained a lot of attention in recent years because it also provides an efficient way to improve bandwidth efficiency. For each channel mode in the eigen-space, it could carry one spatial stream without causing co-channel interference [4]. So The transmitters can choose to send multiple independent streams to increase the bandwidth efficiency, or send the same stream multiple times across different channel modes to improve the reliability, which is a fundamental tradeoff between diversity gain and multiplexing gain [5]. According to system design goals, different tradeoffs can be achieved by employing proper space-time coding schemes [6–8].

Theoretically, the spatial diversity gain and multiplexing gain could be arbitrarily high if it is possible to deploy infinitely many antennas at both the transmitter and receiver. But in practice, since the user devices are usually of very limited size and the adjacent antennas must be sufficiently separated to guarantee independency, it is pretty hard to equip too many antennas on any single user device. Those hardware constraints lead to a new concept of cooperative diversity. The main idea of cooperative diversity is to use distributed antennas instead of the co-located physical antennas, where the distributed antennas could be any independent relaying devices that may help to forward the source signals [3]. As each relay link is able to provide one additional diversity path, the available diversity gain could be quite remarkable in a dense wireless network where there are abundant relaying devices between the transmitter and receiver.

Although the research on cooperative communication dates back to late 1970s

[9], where the capacity of single-relay channels subject to additive white Gaussian noise (AWGN) was explored, only until recently has it gained a lot of interest in the research community. The performance gain of cooperative diversity in a two-user code-division multiplexing access (CDMA) system was first demonstrated in [10,11]. Since then, a bunch of cooperation protocols were developed and studied extensively.

Depending on the relay operations, all the cooperation protocols can be roughly divided into two broad categories: analog relaying and digital relaying [12]. In analog relaying protocols, each relay simply forwards the received signals to the receiver after performing some linear operations in the complex domain. As the additive noise is mixed with the signal component, it is amplified and forwarded to the intended receiver too. By contrast, in digital relaying protocols each relay needs to first decode the source message, re-encode it and then forward it to the receiver. So the relay node always forwards a "clean" message, although the message might be incorrect due to decoding errors. From an information theoretic view, simple digital relaying cannot achieve cooperative diversity; however, if the relay can somehow detect the decoding errors, then selectively forwarding the correct messages alone could recover the diversity loss [12].

For single-relay networks, the symbol error rate (SER) is studied in [13]. Both the exact SER and asymptotic SER are derived for analog relaying and digital relaying, respectively, based on which a set of optimum power allocation schemes are obtained. The outage probability and SER for multi-node parallel analog relaying networks are studied in [14] and [15], respectively. Parallel relaying has the disadvantage of low spectral efficiency, as each relay operates on orthogonal channels. To

address this issue, relay selection [16] and distributed space-time coding [17] could be used to coordinate multi-relay cooperation.

In practice, it is possible to exploit spatial diversity and cooperative diversity at the same time if all the devices are equipped with multiple antennas. For analog relaying, the system performances could be improved via relay precoding. The optimum precoding matrices for maximizing the achievable rate and minimizing the mean-squared errors were developed in [18] and [19], respectively. If power constraint is not stringent, the relay precoding matrix could also be optimized to achieve certain quality-of-service goals [20].

1.2 Wireless Network Coding

For cooperative diversity, the relays need to first acquire the source message before forwarding it to the receiver. However, practical devices are usually subject to half-duplex constraint, i.e., they cannot transmit and receive signals at the same time. As a result, the whole end-to-end data relaying is completed in two phases: data acquiring phase and data forwarding phase. Since an independent channel is required for each phase and only one message could be delivered across those two phases, it incurs a pre-log factor $\frac{1}{2}$ on the spectral efficiency [21]. For multi-relay systems, such rate loss is even larger if the intermediate relays operate on orthogonal channels [14].

To save channel use for data forwarding phase, the relay can choose to combine different source messages via network coding and forward a single mixed message

rather than forward the individual messages separately. Broadly speaking, network coding refers to arbitrary coding (i.e., mapping from input to output) at intermediate nodes [22,23]. But some pioneering literatures in this area focus only on wireline applications, where the physical channel is assumed to be error free and the contents of source messages are combined beyond the physical layer [23]. With these simplifications, it has been proved that network coding could achieve the min-cut max-flow throughput bound for multicast networks [22–24].

For mobile networks, it is very hard to connect the transmitter/receiver to the relay station by cable directly. So all the inter-node communications go through wireless links, and the underlying channel features play an important role in the design and analysis of network coding. As mentioned in Section 1.1, wireless channels suffer severe random fading that may result in serious transmission errors, and multiple transmitters would also cause co-channel interference. Consequently, the existing analytical results on wireline networks no longer hold for wireless applications, and new findings may rely on information theory and communication theory from a physical-layer view.

For wireless transmissions, the transmitted signal consists of the modulated symbols instead of the raw information bits. Depending on the way for mixing source messages, it gives rise to two different types of wireless network coding schemes. On one hand, the relay could choose to decode different source messages and then combine the bit-streams in the finite field. This is called digital network coding (DNC) and it is a legacy network coding scheme previously developed for wireline networks. Alternatively, the source signals could be combined symbol-wise in the complex field

directly to simplify relay operations, since the decoding could be omitted. This is a unique analog network coding (ANC) scheme dedicated for wireless applications, as the wireless devices usually have the capability of interference cancelation and multi-user detection [25]. In practice, DNC and ANC are suitable for digital relaying and analog relaying, respectively.

Thanks to the additive nature of wireless medium, wireless network coding could also save the channel use for data acquiring phase. In wireline networks, each cable defines a distinct channel between the connected terminals. If multiple transmitters send messages to a common intermediate node at the same time, the relay is able to obtain a "clean" message from each transmitter because there is no transmission collision. Those messages are then combined via network coding locally at the intermediate node. This scheme is conventionally referred to as link-layer network coding (LNC) [26]. For wireless applications, LNC is still applicable if the transmitters operate on different channels to avoid co-channel interference, but the bandwidth efficiency is low. By contrast, physical-layer network coding (PNC) allows all the transmitters operate on the same channel to reduce channel use [27]. Because of the additive nature of wireless medium, all the transmitted signals would be combined automatically in the air, which is a nature form of network coding. Then the relay only needs to amplify and forward such mixed signal to the intended receiver if ANC is employed [28], or map the mixed signal to the network-coded symbol if DNC is employed [27, 29].

1.3 Dissertation Outline

In this thesis, we aim to analyze and develop cooperative transmission strategies with wireless network coding. The whole thesis consists of two parts. In the first part from Chapter 2 to Chapter 5, we focus on the uncoded systems where there are no error detection/correction codes. We resolve a bunch of problems like transceiver design, power allocation, anti error propagation and space-time coding. Besides, we characterize the diversity performance and show that network-coded cooperation generally cannot achieve the same diversity gain as the conventional diversity schemes; however, the diversity loss is very limited and only occurs under certain channel conditions. The second part (i.e., Chapter 6 and Chapter 7) of this thesis is devoted to transmission strategy design for coded systems, in which the devices could somehow detect or correct the transmission errors. One key benefit of the coded systems is that the devices may learn the network dynamics such as the decoding status of a transmitted message and then choose the best response accordingly. We develop some network coding strategies for coded systems and characterize the performance in terms of error rate or throughput.

1.3.1 Error Performance of Two-Way Relay Channel with Digital Network Coding (Chapter 2)

Two-Way Relay Channel (TWRC) is one of the most important application scenarios of DNC. Many literatures [21, 30–32] have studied the achievable rate and reveal the tremendous gain over the conventional orthogonal relaying. Those

works are from an information theoretic view and assume perfect channel coding exists such that transmission errors are negligible. For practical system design, it is often more important to know the performance with given modulation schemes, and this is an area that very limited literatures [27, 33, 34] have ever touched. So far, no literatures have explicitly given the closed-form error rate and the achievable diversity gain in the fading channel, and we aim to fill this important gap.

To be specific, in this chapter we study the error performance of TWRC with differential binary phase-shift keying (DBPSK) modulation. We first design the maximum likelihood (ML) relay/receiver detectors for the general case with multiple parallel relays. As the exact ML relay detector is hard to manipulate, we approximate it as a multi-user detector (MLD) followed by a PNC encoder. For the single-relay case, we derive the closed-form end-to-end bit error rate (BER) and resolve the power allocation problem to minimize the average BER. We show that the optimal source power is inversely proportional to the square root of the channel gain of the source-relay channel, and the optimal relay power decreases with SNR. For the multi-relay case, the exact analysis is intractable and we develop upper bound and lower bound on BER and show that only half of the total available diversity gain could be exploited because the random relay detection errors could propagate to the end terminals [35, 36].

1.3.2 Mitigating Error Propagation for Wireless Uplink with Digital Network Coding (Chapter 3)

Error propagation would reduce the diversity gain of any digital relaying systems. For the conventional orthogonal relaying systems, many physical-layer techniques [37–42] have been developed to mitigate error propagation issue. However, those methods work only in the scenarios where there is a single source-destination pair, and they cannot apply to network-coded systems that deal with multiple users at the same time. Some literatures [43–46] also develop network coding schemes that rely on error detection/correction code, but those methods are not applicable for uncoded systems such as some sensor networks that have very limited processing capability. Very limited papers [47, 48] talk about anti error propagation for uncoded systems; however, global channel state information (CSI) is required in those methods and only receiver-side technique is considered, which largely limits their practical use.

Because of those concerns, we develop some practical anti error propagation methods for the uncoded two-user wireless uplink with DNC. We come up with some power scaling schemes and advanced detection schemes that require global CSI or only local CSI, respectively. For the soft power scaling scheme, we develop a virtual source-relay-destination channel model and demonstrate that the relay power should be such to balance the SNRs of the source-relay channel and relay-destination channel. For the hard power scaling scheme, we first design a decision rule based on total pairwise error probability (PEP), and then simplify it to the threshold-

based relaying strategy. At the receiver side, we show that the weighted minimum distance detection with the weight being determined by the relative link quality of source-relay channel and relay-destination channel can achieve full diversity if and only if global CSI is available, otherwise the maximum likelihood detection should be employed to achieve full diversity if the receiver only knows local CSI [49, 50].

1.3.3 Diversity Analysis of Wireless Uplink with Analog Network Coding (Chapter 4)

ANC is naturally immune to error propagation because the relay no longer needs to decode the source messages. Since different messages are combined in the complex field directly, they would become multi-user interference (MUI) to each other. For TWRC, ANC is essentially interference free as the end terminals are able to subtract the self-interference from the received signals [21,25,28,74]. By contrast, for uplink channel the receiver has no side information about any source messages and is unable to eliminate the co-channel interference. Many literatures [51–56] have studied the achievable rate region and space-time code design for uplink channel with ANC. However, the impact of MUI on the diversity gain remain unclear. In [57], the beamforming design when only the quantized CSI is available at the relays is studied and a generalized diversity measure is introduced to study the impact of MUI. However, this work is focused only on the instantaneous relay power constraints where instant CSI is available. The achievable diversity gain in the absence of instant CSI is still unclear.

So in this work, we provide a study on the diversity performance of a general K-user uplink channel with ANC. Depending on the relay power constraints, we investigate both the variable gain relaying (VGR) and the fixed gain relaying (FGR). We first study the single-relay networks, and show that full diversity can be achieved regardless of MUI. However, an logarithmic term would appear in the error rate expression and incur diversity loss at modest SNR. Several relaying schemes to achieve distributed spatial diversity when there are multiple relays are then explored. We first propose a relay selection strategy based on the principle of minimizing the maximum PEP and prove that full diversity can be achieved. Next, two distributed space-time coding (DSTC) schemes are studied. For distributed space-time block code (DSTBC), we show that DSTBC-FGR can always achieve full diversity, whereas the diversity of DSTBC-VGR is also bounded by the number of users. As the diversity of single-user DSTBC-VGR is limited by 2, we develop an adaptive relay power allocation scheme that can recover the diversity loss. Finally for diagonal distributed space-time coding (DDSTC), we show that both VGR and FGR can achieve full diversity, and the optimum code design criterion is to maximize the minimum product distance [58, 59].

1.3.4 Diversity Analysis of Wireless Uplink with Non-Coherent Network Coding (Chapter 5)

Perfect CSI is very important for the receiver to mitigate error propagation for DNC or suppress MUI for ANC. However, perfect CSI is not always available

due to various reasons such as high channel estimation overhead. To reduce the reliance on CSI, non-coherent modulation schemes have been widely used in real systems. Non-coherent orthogonal relaying strategies have received a lot of interest in the community [60–65], and some literatures [26, 34, 66–68] also discuss non-coherent transmissions for TWRC with network coding. Intuitively, using non-coherent modulation should decrease the system performance. For the traditional point-to-point channels, it is well known that it incurs 3dB SNR loss while the diversity gain remains the same [1]. But for the network-coded cooperation systems, very few literatures have ever explicitly discussed the performance loss, and this motivates our work.

To be specific, we study the two-user uplink channel with ANC or DNC, respectively. We consider FSK modulation to facilitate non-coherent detection. We first design the coherent and non-coherent ML receivers when global CSI and statistical CSI is available at the receivers, respectively. For ANC, as the non-coherent ML receiver has an intractable integral form, we develop two suboptimum receivers that are near optimum depending on the relative quality of source-relay channel and relay-destination channel. The PEP is then studied, and the scaling laws of different PEPs are derived at high SNRs. It is demonstrated that full dominant diversity is always achieved regardless of the CSI assumptions; however, using non-coherent detection would incur some diversity loss at modest SNRs such that the resulting error rates do not decrease that fast compared to coherent detection. Besides, we show that the performance loss of ANC is more serious due to the incapability to efficiently suppress MUI at the receiver [69, 70].

1.3.5 Network-Coded ARQ for Two-Way Relay Channel (Chapter 6)

In the previous chapters, we focus mainly on static relaying for uncoded systems, i.e., all the data transmission would go through the relay link regardless of the network dynamics. For coded systems, the decoding status of each packet could be known by performing cyclic redundancy check. If the transmission through direct link is already successful, wireless relaying could be omitted to save the channel use and transmitted power. Otherwise, the relay could help to retransmit the original source packet as a part of automatic repeat-request (ARQ) mechanism. Network coding could enhance the transmission efficiency of the conventional ARQ, since a couple of to-be-retransmitted packets could be combined to reduce the number of retransmissions. Network-coded ARQ has been studied for many applications, such as broadcast channels [71], wireless multicast [72] and multiple unicast flows [73].

In this work, we study the performance of network-coded ARQ for TWRC. The key distinction between our work and the existing literatures [31, 74–77] is that we take into account the maximum transmission constraint, which is a very practical concern. For single-relay networks, we derive the closed-from throughput when the retransmission is subject to per-hop constraint or end-to-end (E2E) constraint, respectively. We demonstrate that network coding can greatly improve the system throughput, but the throughput gain is upper bounded. Besides, we come up with a near-optimum power allocation scheme to maximize the throughput. For multi-relay networks, we show that successive relaying strategy suffers great throughput loss when the frame length is much smaller than the number of relays, and we develop

a hybrid network coding scheme to fully exploit the network coding gain [78, 79].

1.3.6 Clustering Based Space-Time Network Coding (Chapter 7)

When there is no dedicated relay in the systems, user devices have to help each other to enjoy the cooperative diversity gain. For a dense mobile network, how to coordinate the large number of user devices has been an open design problem for a long time. The existing strategies [80–82] tend to pursue the largest diversity gain, while the bandwidth efficiency is relatively low. So in this work, we aim to develop a new user cooperation strategy that can achieve better tradeoff between diversity gain and bandwidth efficiency. The core idea of our method is to divide the whole network into several small clusters, and different clusters help each other to relay the signals. The clusters send data successively in a time-division multiple access (TDMA) way. Each node in a certain cluster behaves as a digital relay to other clusters, and it uses linear coding to combine the local symbol and the relayed symbols. Linear decorrelator is used at the receivers to separate different source signals. We obtain both the exact SER and asymptotic SER of the M-ary phase-shift keying (PSK) signal. It is shown that different tradeoffs between diversity gain and bandwidth efficiency can be achieved by adjusting the formation of clusters [83].

Chapter 2

Error Performance of Two-Way Relay Channel with Digital Network Coding

For cellular systems, the uplink/downlink is a typical TWRC paradigm. Many literatures have discussed how DNC could improve the achievable rate against the conventional orthogonal relaying [21, 30–32]. However, those literatures are mainly from an information-theoretic view, which assumes perfect channel coding and suppose the transmission error could be arbitrarily small. But for real cellular systems, there are only a limited number of modulation schemes to choose, so the data rate usually belongs to a discrete set. On the engineering side, what is more important is the achievable error rate associated with each modulation scheme because it directly determines the network throughput. In the research community, very limited literatures [27, 33, 34] have ever discussed the error performance of TWRC with physical-layer DNC, and to the best of our knowledge, no literature has derived the closed-form error rate expression under fading channel.

So in this chapter, we study the error performance of TWRC where both source nodes use DBPSK modulation and physical-layer DNC is used at the relays. We first derive the relay denoising function and source detector based on ML principles, and then proceed to analyze the corresponding detection error at the end terminals. As it is hard to manipulate the ML denoising function directly, we approximate it

as a MUD followed by a PLNC encoder and obtain the closed-form relay detection error. For the single-relay case, we reveal the equivalence between the ML source detector and the typical DBPSK detector for the relay-source channel, based on which we obtain the exact end-to-end BER. We further investigate the power allocation problem for minimizing the average system BER by use of asymptotic analysis, and show that the optimal source power is inversely proportional to the square root of the channel gain of the source-relay channel, and the optimal relay power decreases with SNR. For the multi-relay networks with K parallel relay nodes, as the exact analysis is intractable, we develop upper bound and lower bound on BER and show that the diversity order is exactly equal to $\lceil \frac{K}{2} \rceil$.

Notations: Boldface lowercase letter $\mathbf{a}$ and boldface uppercase letter $\mathbf{A}$ represent vector in column form and matrix, respectively. $\|\mathbf{a}\|$ and $|\mathbf{A}|$ represent the Euclidean norm of a vector $\mathbf{a}$ and the determinant of a square matrix $\mathbf{A}$, respectively. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ stand for conjugate, transpose and conjugate transpose, respectively. We shall use abbreviation i.i.d. for independent and identically distributed, and denote $Z \sim \mathcal{CN}(\mu, \sigma^2)$ as a circularly symmetric complex Gaussian random variable Z. We define $sign(x)$=1 if x>0 and 0 otherwise. Finally, the probability of an event $\mathcal{A}$ and the probability density function (PDF) of a random variable Z are denoted by $\Pr(\mathcal{A})$ and $f(Z)$, respectively.

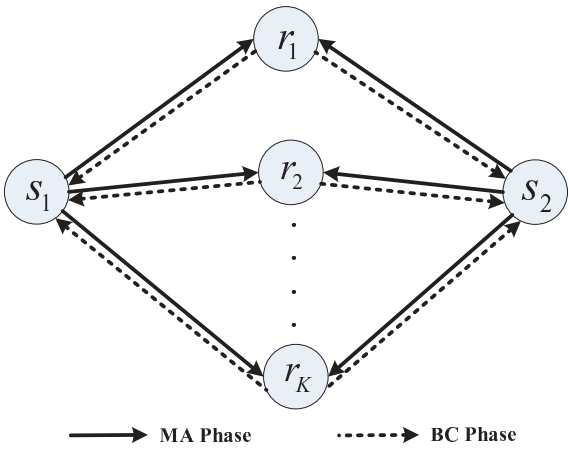


Figure 2.1: System model of network-coded TWRC.

2.1 System Model

In this chapter, we study the TWRC shown in Figure 2.1, where two sources S_1 and S_2 want to exchange information with the help of K parallel relays. The whole data transmission is completed in two phases: multiple-access (MA) phase and broadcasting (BC) phase. At the beginning of the MA phase, the source S_i first generates a sequence of i.i.d uncoded BPSK symbols $b_i(n) \in \{-1, 1\}$ of length L, where $n = 1, 2, \cdots, L$ is the symbol index. These raw symbols are then re-encoded through differential modulation, i.e., $x_i(n) = x_i(n-1) \times b_i(n)$ with $x_i(0)$=1 being the reference symbol. The two sources then send the whole block of differentially encoded symbols simultaneously to all the relays during MA phase. To facilitate demonstration, we define a sequence of network-coded symbols $b(n) = b_1(n) \times b_2(n) \in \{-1, 1\}$ for $n = 1, 2, \cdots, L$ to indicate whether the two source symbols of the same time index have the same sign or not. Note that because each source knows its own symbol, knowing the common information $b(n)$ is sufficient for both sources to decode

the symbol sent from the other end.

During the MA phase, the nth symbol received at the kth relay is given by

$$y_k(n) = \sqrt{P_{s_1}} h_{1,k}^{MA} x_1(n) + \sqrt{P_{s_2}} h_{2,k}^{MA} x_2(n) + w_k^{MA}(n) \tag{2.1}$$

for $k = 1, 2, \cdots, K$, where $P_{s_i} = \alpha_i P$ is the power of S_i, P is the total power and $\alpha_i \in [0, 1]$ stands for the corresponding source power ratio. $h_{i,k}^{MA} \sim \mathcal{CN}\left(0, \sigma_{i,k}^2\right)$ is the independent channel coefficient from S_i to the kth relay during MA phase, where $\sigma_{i,k}^2$ is the channel gain. Throughout this chapter, we assume that the channels remain unchanged within a block. Finally, $w_k^{MA}(n) \sim \mathcal{CN}(0, N_0)$ is the independent AWGN at the kth relay within the nth symbol interval during MA phase.

Suppose DNC is used at the relay node, the kth relay just maps the nth received symbol to another BPSK symbol $\hat{b}_{r_k}(n) \in \{-1, 1\}$ that can be used by both sources to uniquely determine the symbol transmitted from the other end, and this process is called denoising. Here $\hat{b}_{r_k}(n) \in \{-1, 1\}$ is an estimate of the network-coded symbol $b(n)$, so relay denoising is actually equivalent to detection for $b(n)$. According to [34], the single-symbol ML detector for $b(n)$ is given by

$$\hat{b}_{r_k}(n) = \arg \max_{b(n) \in \{-1,1\}} f\left(\mathbf{y}_k(n) \,|b(n)\right), \tag{2.2}$$

where $\mathbf{y}_k(n) = (y_k(n), y_k(n-1))^T$ is the vector of two consecutive received symbols. It is easy to show that given $b_1(n)$ and $b_2(n)$, $\mathbf{y}_k(n) \left|_{b_1(n), b_2(n)} \right. \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}_{b_1(n), b_2(n)}^k\right)$,

where the conditional covariance matrices are given by

$$\begin{cases} \mathbf{\Sigma}^k_{b_1(n)=1,b_2(n)=1} \triangleq \mathbf{\Sigma}_{1,r_k} = N_0\left(\gamma_{1,k}+\gamma_{2,k}+1\right)\mathbf{I}_2 + N_0\left(\gamma_{1,k}+\gamma_{2,k}\right)\hat{\mathbf{I}}_2 \\ \mathbf{\Sigma}^k_{b_1(n)=-1,b_2(n)=-1} \triangleq \mathbf{\Sigma}_{2,r_k} = N_0\left(\gamma_{1,k}+\gamma_{2,k}+1\right)\mathbf{I}_2 - N_0\left(\gamma_{1,k}+\gamma_{2,k}\right)\hat{\mathbf{I}}_2 \\ \mathbf{\Sigma}^k_{b_1(n)=1,b_2(n)=-1} \triangleq \mathbf{\Sigma}_{3,r_k} = N_0\left(\gamma_{1,k}+\gamma_{2,k}+1\right)\mathbf{I}_2 + N_0\left(\gamma_{1,k}-\gamma_{2,k}\right)\hat{\mathbf{I}}_2 \\ \mathbf{\Sigma}^k_{b_1(n)=-1,b_2(n)=1} \triangleq \mathbf{\Sigma}_{4,r_k} = N_0\left(\gamma_{1,k}+\gamma_{2,k}+1\right)\mathbf{I}_2 + N_0\left(\gamma_{2,k}-\gamma_{1,k}\right)\hat{\mathbf{I}}_2 \end{cases}. \quad (2.3)$$

Here $\gamma_{i,k} = \frac{P_{s_i}\sigma^2_{i,k}}{N_0} = \alpha_i\sigma^2_{i,k}\gamma$ is the received SNR, $\gamma = \frac{P}{N_0}$ is the reference system SNR, and

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \hat{\mathbf{I}}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

are two constant matrices. Based on the law of total probability, the conditional PDF of $\mathbf{y}_k(n)$ can be expressed as

$$f\left(\mathbf{y}_k(n)\,|b(n)\right) = \frac{1}{2}\sum_{b_1(n)\times b_2(n)=b(n)} f\left(\mathbf{y}_k(n)\,|b_1(n),b_2(n)\right). \quad (2.4)$$

After some manipulations, we can re-write the ML detector (2.2) as

$$\hat{b}_{r_k}(n) = sign\left(\ln\left(\mathrm{lrf}\left(\mathbf{y}_k(n)\,|b(n)\right)\right)\right), \quad (2.5)$$

where

$$\mathrm{lrf}\left(\mathbf{y}_k(n)\,|b(n)\right) = \frac{g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{1,r_k}\right)+g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{2,r_k}\right)}{g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{3,r_k}\right)+g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{4,r_k}\right)} \quad (2.6)$$

is the likelihood ratio function (LRF) of $\mathbf{y}_k(n)$, and

$$g\left(\mathbf{y},\mathbf{\Sigma}\right) = \frac{1}{\pi^2\left|\mathbf{\Sigma}\right|}\exp\left(-\mathbf{y}^H\mathbf{\Sigma}^{-1}\mathbf{y}\right) \quad (2.7)$$

is the PDF of $\mathbf{y} \sim \mathcal{CN}\left(\mathbf{0},\mathbf{\Sigma}\right)$. After detection, the kth relay needs to re-encode $\left\{\hat{b}_{r_k}(n)\right\}_{n=1}^{L}$ into $t_k(n) = t_k(n-1)\times\hat{b}_{r_k}(n)$ for $n = 1,2,\cdots,L$ through differential

modulation, where $t_k(0) = 0$ is the reference symbol. Note that due to random detection error, it is possible that $\hat{b}_{r_k}(n) \neq b(n)$.

During BC phase, all relays broadcast their own differentially re-encoded symbols together through a set of orthogonal channels. At S_i, the received signal from the kth relay is given by

$$r_{k,i}(n) = \sqrt{P_{r_k}} h_{k,i}^{BC} t_k(n) + w_{k,i}^{BC}(n),\ n = 0, 1, \cdots, L, \tag{2.8}$$

where $P_{r_k} = \beta_k P$ is the transmitted power of the kth relay and $\beta_k \in [0, 1]$ stands for the corresponding relay power ratio. $h_{k,i}^{BC} \sim \mathcal{CN}\left(0, \sigma_{i,k}^2\right)$ is the channel coefficient from the kth relay to the ith source during BC phase, and we assume $h_{k,i}^{BC}$ and $h_{i,k}^{MA}$ are independent but have the same channel gain that is determined by the distance between two terminals. Finally, $w_{k,i}^{BC}(n) \sim \mathcal{CN}(0, N_0)$ is the independent AWGN on the channel from the kth relay to the ith source within the nth symbol interval during BC phase.

As mentioned before, each source only needs to detect $b(n)$. For example, if the estimate of $b(n)$ at source S_1 is $\hat{b}_{s_1}(n) = 1$, then $b_2(n)$ can be detected as $\hat{b}_{2,s_1}(n) = b_1(n)$, otherwise $\hat{b}_{2,s_1}(n) = -b_1(n)$ if $\hat{b}_{s_1}(n) = -1$. Based on the observations $\{\mathbf{r}_{k,i}(n)\}_{k=1}^K$, the single-symbol ML detector for $b(n)$ at S_i is given by

$$\hat{b}_{s_i}(n) = \arg \max_{b(n) \in \{-1,1\}} f\left(\{\mathbf{r}_{k,i}(n)\}_{k=1}^K | b(n)\right), \tag{2.9}$$

where $\mathbf{r}_{k,i}(n) = (r_{k,i}(n), r_{k,i}(n-1))^T$ is the vector of two consecutive received symbols from the kth relay, and $\mathbf{r}_{k,i}(n)\left|_{\hat{b}_{r_k}(n)}\right. \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}^k_{\hat{b}_{r_k}(n), s_i}\right)$ where the conditional

covariance matrices are given by

$$\begin{cases} \boldsymbol{\Sigma}^k_{\hat{b}_{r_k}(n)=1,s_i} \triangleq \boldsymbol{\Sigma}^k_{1,s_i} = N_0\left(\bar{\gamma}_{k,i}+1\right)\mathbf{I}_2 + N_0\bar{\gamma}_{k,i}\hat{\mathbf{I}}_2 \\ \boldsymbol{\Sigma}^k_{\hat{b}_{r_k}(n)=-1,s_i} \triangleq \boldsymbol{\Sigma}^k_{2,s_i} = N_0\left(\bar{\gamma}_{k,i}+1\right)\mathbf{I}_2 - N_0\bar{\gamma}_{k,i}\hat{\mathbf{I}}_2 \end{cases}, \tag{2.10}$$

where $\bar{\gamma}_{k,i} = \frac{P_{r_k}\sigma^2_{i,k}}{N_0} = \beta_k\sigma^2_{i,k}\gamma$ is the received SNR of the kth relay-source channel. As the signals from different relays are conditionally independent given $b(n)$, we can rewrite the joint PDF in (2.9) as

$$f\left(\{\mathbf{r}_{k,i}(n)\}_{k=1}^K\,|b(n)\right) = \prod_{k=1}^K \sum_{\hat{b}_{r_k}(n)\in\{-1,1\}} f\left(\mathbf{r}_{k,i}(n)\,\middle|\hat{b}_{r_k}(n)\right) P\left(\hat{b}_{r_k}(n)\,|b(n)\right), \tag{2.11}$$

where we use the law of total probability and the fact $\mathbf{r}_{k,i}(n)$ is conditionally independent of $b(n)$ given $\hat{b}_{r_k}(n)$. From (2.11), the ML source detector (2.9) can be simplified to

$$\hat{b}_{s_i}(n) = sign\left(\sum_{k=1}^K \ln\left(\mathrm{lrf}\left(\mathbf{r}_{k,i}(n)\,|b(n)\right)\right)\right), \tag{2.12}$$

where

$$\mathrm{lrf}\left(\mathbf{r}_{k,i}(n)\,|b(n)\right) = \frac{g\left(\mathbf{r}_{k,i}(n),\boldsymbol{\Sigma}^k_{1,s_i}\right)\left(1-P_{M,r_k}\right) + g\left(\mathbf{r}_{k,i}(n),\boldsymbol{\Sigma}^k_{2,s_i}\right)P_{M,r_k}}{g\left(\mathbf{r}_{k,i}(n),\boldsymbol{\Sigma}^k_{1,s_i}\right)P_{F,r_k} + g\left(\mathbf{r}_{k,i}(n),\boldsymbol{\Sigma}^k_{2,s_i}\right)\left(1-P_{F,r_k}\right)} \tag{2.13}$$

is the LRF of $\mathbf{r}_{k,i}(n)$, and

$$P_{M,r_k} = \Pr\left(\hat{b}_{r_k}(n)=-1\,|b(n)=1\right) = \Pr\left(\mathrm{lrf}\left(\mathbf{y}_k(n)\,|b(n)\right)\le 1\,|b(n)=1\right), \tag{2.14}$$

$$P_{F,r_k} = \Pr\left(\hat{b}_{r_k}(n)=1\,|b(n)=-1\right) = \Pr\left(\mathrm{lrf}\left(\mathbf{y}_k(n)\,|b(n)\right)>1\,|b(n)=-1\right) \tag{2.15}$$

are two kinds of conditional detection errors at the kth relay. The calculation of those two terms is postponed to the next section.

2.2 Performance Analysis: Single-Relay Case

In this section, we examine the detection error rate for the single-relay case. Without loss of generality, we assume only a single relay (i.e., the kth relay) is activated to assist the information exchange between two sources. To optimize the end-to-end error performance, we also investigate the power allocation problem.

2.2.1 Relay Detection Error

Using the law of total probability, we can write the relay detection error as

$$\Pr\left(\hat{b}_{r_k}(n) \neq b(n)\right) \triangleq P_{e,r_k} = \frac{P_{M,r_k} + P_{F,r_k}}{2}, \tag{2.16}$$

where P_{M,r_k} and P_{F,r_k} are two kinds of conditional detection errors defined in (2.14) and (2.15), and both of them are related to $\mathrm{lrf}\left(\mathbf{y}_k(n)\,|b(n)\right)$. After substituting (2.7) into (2.6) and doing some manipulations, we can obtain

$$\begin{aligned}\mathrm{lrf}\left(\mathbf{y}_k(n)\,|b(n)\right) = \frac{|\mathbf{\Sigma}_{3,r_k}|}{|\mathbf{\Sigma}_{1,r_k}|} &\frac{\cosh\left(\frac{N_0\left(\gamma_{1,k}+\gamma_{2,k}\right)}{|\mathbf{\Sigma}_{1,r_k}|}\mathbf{y}_k^H(n)\hat{\mathbf{I}}_2\mathbf{y}_k(n)\right)}{\cosh\left(\frac{N_0\left(\gamma_{1,k}-\gamma_{2,k}\right)}{|\mathbf{\Sigma}_{3,r_k}|}\mathbf{y}_k^H(n)\hat{\mathbf{I}}_2\mathbf{y}_k(n)\right)} \\ &\times \exp\left(\frac{|\mathbf{\Sigma}_{1,r_k}| - |\mathbf{\Sigma}_{3,r_k}|}{|\mathbf{\Sigma}_{1,r_k}|\,|\mathbf{\Sigma}_{3,r_k}|} N_0\left(\gamma_{1,k}+\gamma_{2,k}+1\right)\|\mathbf{y}_k(n)\|^2\right)\end{aligned} \tag{2.17}$$

where $\cosh(x) = \frac{e^x+e^{-x}}{2}$ is the hyperbolic cosine function. As it is really hard to analyze the error rate based on the above LRF directly, we use the following approximation to simplify the analysis

$$\cosh(x) \approx \frac{\max\left(e^x, e^{-x}\right)}{2} = \frac{e^{|x|}}{2}, \tag{2.18}$$

which is quite tight when $|x|$ is not too small. After such approximation, only exponential terms are left with the exponent being a quadratic form of $\mathbf{y}_k(n)$, which

is analytically tractable.

After substituting (2.18) back into (2.17), we have

$$\mathrm{lrf}\left(\mathbf{y}_k(n)\,|b(n)\,\right) \approx \frac{\max\left(g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{1,r_1}\right),g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{2,r_k}\right)\right)}{\max\left(g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{3,r_k}\right),g\left(\mathbf{y}_k(n),\mathbf{\Sigma}_{4,r_k}\right)\right)}. \tag{2.19}$$

Now if we use (2.19) instead in (2.5), it is easy to see that this suboptimal detector is actually a MUD

$$\left(\hat{b}_{1,r_k}(n),\hat{b}_{2,r_k}(n)\right) = \arg\max_{b_i(n)\in\{-1,1\},i=1,2} f\left(\mathbf{y}_k(n)\,|b_1(n),b_2(n)\,\right) \tag{2.20}$$

followed by a PLNC encoder $\hat{b}_{r_k}(n) = \hat{b}_{1,r_k}(n)\times\hat{b}_{2,r_k}(n)$. That is, the relay first jointly detects the BPSK symbols $b_1(n)$ and $b_2(n)$, and then maps the detected symbols to a single BPSK symbol $\hat{b}_{r_k}(n)$ as an estimate of the network-coded symbol $b(n)$. As we shall see in simulations, this suboptimal relay detector is almost as good as the ML detector (2.5) in all cases. The reason is that the two conditional PDFs of $\mathbf{y}_k(n)$ given $b(n)$ are very well separated. As a result, the ML region on $b(n)$ is very close to the direct union of the individual ML regions on $(b_1(n),b_2(n))$, which leads to the max operation in (2.20).

To characterize the error performance, let us first calculate P_{M,r_k}. After substituting (2.19) into (2.14) and making some manipulations, we have

$$\begin{aligned} P_{M,r_k} = \frac{1}{2}\sum_{b_1(n)\times b_2(n)=1} \Pr\Big(&(a_k+b_k)\,|\hat{y}_{k,1}(n)|^2 + (a_k-b_k)\,|\hat{y}_{k,2}(n)|^2 \;\le \ln\gamma_{th}^k \;\&\\ &(a_k-b_k)\,|\hat{y}_{k,1}(n)|^2 + (a_k+b_k)\,|\hat{y}_{k,2}(n)|^2 \;\le \ln\gamma_{th}^k\,|b_1(n),b_2(n)\Big), \end{aligned} \tag{2.21}$$

where

$$a_k = -\frac{4\gamma_{1,k}\gamma_{2,k}\left(\gamma_{1,k}+\gamma_{2,k}+1\right)}{N_0\left(2\gamma_{1,k}+2\gamma_{2,k}+1\right)\left(2\gamma_{1,k}+1\right)\left(2\gamma_{2,k}+1\right)}, \tag{2.22}$$

$$b_k = \frac{4\gamma_{1,k}\gamma_{2,k}\left(\gamma_{1,k}+\gamma_{2,k}\right)+2\min\left(\gamma_{1,k},\gamma_{2,k}\right)\left(2\gamma_{1,k}+2\gamma_{2,k}+1\right)}{N_0\left(2\gamma_{1,k}+2\gamma_{2,k}+1\right)\left(2\gamma_{1,k}+1\right)\left(2\gamma_{2,k}+1\right)}, \tag{2.23}$$

$$\gamma_{th}^{k} = \frac{\left(2\gamma_{1,k}+2\gamma_{2,k}+1\right)}{\left(2\gamma_{1,k}+1\right)\left(2\gamma_{2,k}+1\right)}, \tag{2.24}$$

and we define

$$\hat{\mathbf{y}}_k(n) = \left(\hat{y}_{k,1}(n), \hat{y}_{k,2}(n)\right)^T = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\mathbf{y}_k(n) \tag{2.25}$$

as an auxiliary signal vector. Since $\mathbf{y}_k(n)\big|_{b_1(n),b_2(n)} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}^k_{b_1(n),b_2(n)}\right)$, $\left|\hat{y}_{k,1}(n)\right|^2$ and $\left|\hat{y}_{k,2}(n)\right|^2$ are conditionally independent exponential random variables given $b_1(n)$ and $b_2(n)$. Therefore, (2.21) can be easily evaluated as

$$P_{M,r_k} = h\left(u_{1,k}, u_{2,k}, a_k, b_k, \gamma_{th}^{k}\right), \tag{2.26}$$

where $h\left(t_1, t_2, a, b, \gamma\right)$ is a function with five parameters and it is given by

$$h\left(t_1, t_2, a, b, \gamma\right) = \frac{4abt_1t_2}{a^2\left(t_1-t_2\right)^2 - b^2\left(t_1+t_2\right)^2}\exp\left(-\frac{t_1+t_2}{2a}\ln\gamma\right), \tag{2.27}$$

and the two constants are given by

$$u_{1,k} = \frac{1}{N_0\left(2\gamma_{1,k}+2\gamma_{2,k}+1\right)},\ u_{2,k} = \frac{1}{N_0}. \tag{2.28}$$

In a similar manner, we can show that

$$P_{F,r_k} = 1 - h\left(u_{3,k}, u_{4,k}, a_k, b_k, \gamma_{th}^{k}\right), \tag{2.29}$$

where

$$u_{3,k} = \frac{1}{N_0\left(2\gamma_{1,k}+1\right)},\ u_{4,k} = \frac{1}{N_0\left(2\gamma_{2,k}+1\right)}. \tag{2.30}$$

Finally, plugging (2.26) and (2.29) back into (2.16) leads to the closed-form relay detection error.

2.2.2 Source Detection Error

When there is only one active relay in the system, the source detector (2.12) is reduced to

$$\begin{aligned}\hat{b}_{s_i}(n) &= sign\left(\ln\left(\mathrm{lrf}\left(\mathbf{r}_{k,i}(n)\,|b(n)\right)\right)\right) \\ &= sign\left(\ln\left(\mathrm{lrf}\left(\mathbf{r}_{k,i}(n)\,\middle|\hat{b}_{r_k}(n)\right)\right)\right) \triangleq \hat{b}_{r_k,s_i}(n), \end{aligned} \tag{2.31}$$

where

$$\mathrm{lrf}\left(\mathbf{r}_{k,i}(n)\,\middle|\hat{b}_{r_k}(n)\right) = \frac{g\left(\mathbf{r}_{k,i}(n), \mathbf{\Sigma}^k_{1,s_i}\right)}{g\left(\mathbf{r}_{k,i}(n), \mathbf{\Sigma}^k_{2,s_i}\right)}. \tag{2.32}$$

Note that the detector on the second line of (2.31) is actually a typical non-coherent DBPSK detector [1, Eqn. (14-4-23)] for the point-to-point channel from the kth relay to the ith source. Consequently, detection for the true network-coded symbol $b(n)$ is equivalent to detection for the transmitted symbol $\hat{b}_{r_k}(n)$ at the kth relay. Using this property, we can write the source detection error as

$$\Pr\left(\hat{b}_{s_i}(n) \neq b(n)\right) = \Pr\left(\hat{b}_{r_k,s_i}(n) \neq b(n)\right) \triangleq P^k_{e,s_i} = \frac{1}{2}\left(P^k_{M,s_i} + P^k_{F,s_i}\right), \tag{2.33}$$

where

$$P^k_{M,s_i} = \Pr\left(\hat{b}_{s_i}(n) = -1\,|b(n) = 1\right) = \Pr\left(\hat{b}_{r_k,s_i}(n) = -1\,|b(n) = 1\right), \tag{2.34}$$

$$P^k_{F,s_i} = \Pr\left(\hat{b}_{s_i}(n) = 1\,|b(n) = -1\right) = \Pr\left(\hat{b}_{r_k,s_i}(n) = 1\,|b(n) = -1\right) \tag{2.35}$$

are two kinds of conditional detection error at the ith source, and we have used the relation $\hat{b}_{s_i}(n) = \hat{b}_{r_k,s_i}(n)$ in (2.33)–(2.35). After expanding (2.34) by use of the law

of total probability, we have

$$\begin{aligned} P_{M,s_i}^k &= \sum_{\hat{b}_{r_k}(n)\in\{-1,1\}} \Pr\left(\hat{b}_{r_k,s_i}(n) = -1 \left| \hat{b}_{r_k}(n), b(n) = 1 \right.\right) P\left(\hat{b}_{r_k}(n) \left| b(n) = 1 \right.\right) \\ &\overset{(a)}{=} \sum_{\hat{b}_{r_k}(n)\in\{-1,1\}} \Pr\left(\hat{b}_{r_k,s_i}(n) = -1 \left| \hat{b}_{r_k}(n) \right.\right) P\left(\hat{b}_{r_k}(n) \left| b(n) = 1 \right.\right) \\ &\overset{(b)}{=} P_{D,s_i}^k \left(1 - P_{M,r_k}\right) + \left(1 - P_{D,s_i}^k\right) P_{M,r_k} \end{aligned} \tag{2.36}$$

where we use in (a) the fact that $\hat{b}_{r_k,s_i}(n)$ is conditionally independent of $b(n)$ given $\hat{b}_{r_k}(n)$, and in (b) we use the fact that the two kinds of conditional detection errors of a typical non-coherent DBPSK detector are equal and given by [1, Eqn. (14-4-26)]

$$\begin{aligned} \Pr\left(\hat{b}_{r_k,s_i}(n) = 1 \left| \hat{b}_{r_k}(n) = -1 \right.\right) &= \Pr\left(\hat{b}_{r_k,s_i}(n) = -1 \left| \hat{b}_{r_k}(n) = 1 \right.\right) \\ &\triangleq P_{D,s_i}^k = \frac{1}{2\left(\bar{\gamma}_{k,i} + 1\right)}. \end{aligned} \tag{2.37}$$

In a similar way, we can obtain

$$P_{F,s_i}^k = \left(1 - P_{D,s_i}^k\right) P_{F,r_k} + P_{D,s_i}^k \left(1 - P_{F,r_k}\right). \tag{2.38}$$

Plugging (2.36) and (2.38) back into (2.33) we have

$$P_{e,s_i}^k = \left(1 - P_{D,s_i}^k\right) P_{e,r_k} + P_{D,s_i}^k \left(1 - P_{e,r_k}\right), \tag{2.39}$$

which is the end-to-end BER at the ith source.

2.2.3 Power Allocation

So far, we have obtained the end-to-end BER as a function of the transmitted power. To minimize the average BER, the power could be smartly allocated among

the terminals. The optimum power allocation problem is formulated as

$$\begin{aligned} \min \ & P_e^k = \frac{1}{2}\left(P_{e,s_1}^k + P_{e,s_2}^k\right) \\ \text{s.t.} \ & \alpha_1 + \alpha_2 + \beta_k = 1, \\ & 0 \le \alpha_1, \alpha_2, \beta_k \le 1. \end{aligned} \tag{2.40}$$

However, it is very hard to manipulate the exact BER expression directly, and the optimal power level can be derived only through exhaustive search. In order to obtain a simple closed-form solution, we choose to use the asymptotic error rate at high SNRs (i.e., $\gamma \gg 1$). After some approximations, we can obtain

$$\begin{cases} P_{M,r_k} \approx \frac{c_{M,r_k}}{\gamma}, \ c_{M,r_k} = \frac{1}{2\min\left(\alpha_1\sigma_{1,k}^2, \alpha_2\sigma_{2,k}^2\right)} \\ P_{F,r_k} \approx \frac{d_{F,r_k}}{\gamma} \ln \frac{\gamma}{d_{F,r_k}}, \ d_{F,r_k} = \frac{\alpha_1\sigma_{1,k}^2 + \alpha_2\sigma_{2,k}^2}{2\alpha_1\alpha_2\sigma_{1,k}^2\sigma_{2,k}^2} \\ P_{D,s_i}^k \approx \frac{q_{D,s_i}^k}{\gamma}, \ q_{D,s_i}^k = \frac{1}{2\beta_k\sigma_{i,k}^2}, i = 1, 2 \end{cases} . \tag{2.41}$$

After plugging these approximations back into (2.39), we can obtain the asymptotic end-to-end BER given by

$$P_e^k \approx \frac{1}{2\gamma}\left(c_{M,r_k} + d_{F,r_k} \ln \frac{\gamma}{d_{F,r_k}} + q_{D,s_1}^k + q_{D,s_2}^k\right), \tag{2.42}$$

where we neglect the higher-order terms. There are several important observations. Firstly, it is observed that BER is dominated by P_{F,r_k}, which scales as $\gamma^{-1}\ln\gamma$ at high SNRs. Therefore, more power should be allocated to the sources in order to reduce the relay detection error. Secondly, for point-to-point channels the BER of non-coherent DBPSK modulation scales as γ^{-1} [1, Eqn. (14-4-28)], which decreases faster than the dominant error term P_{F,r_k}. As a result, wireless relaying has no advantage over direct transmission at high SNRs. Finally, it can be observed that

$P_{F,r_k} > P_{M,r_k}$ when source power is fixed and SNR is sufficiently high. This is because it is relatively easier to detect $b(n)$ when the two source symbols have the same sign, in which case the two consecutive observations $y_k(n)$ and $y_k(n-1)$ would have similar envelopes at high SNRs.

Now let us proceed to solve (2.40) by use of the asymptotic expression (2.42). Note that the first two terms in (2.42) depend only on source power ratio α_1 and α_2 while the last two terms depend only on β_k. So the optimization problem (2.40) can be resolved by two steps. In the first step, we fix β_k and seek to find the optimal source power, i.e.,

$$\begin{aligned} \min \ & \frac{c_{M,r_k} + d_{F,r_k} \ln \frac{\gamma}{d_{F,r_k}}}{2\gamma} \approx \frac{d_{F,r_k}}{2\gamma} \ln \frac{\gamma}{d_{F,r_k}} \\ \text{s.t.} \ & \alpha_1 + \alpha_2 = 1 - \beta_k, \\ & 0 \le \alpha_1, \alpha_2 \le 1 - \beta_k. \end{aligned} \tag{2.43}$$

where we neglect the term c_{M,r_k} because it is much smaller than $\ln \gamma$ at high SNRs. Note that the function $\phi(x) = x \ln x$ is increasing when $x < e^{-1}$, which is the case for sufficiently large γ. Therefore, it is equivalent to minimizing d_{F,r_k} instead in (2.43), and the optimizer is

$$\begin{cases} \alpha_1^{opt} = (1-\beta_k) \frac{\sigma_{2,k}}{\sigma_{1,k}+\sigma_{2,k}} \\ \alpha_2^{opt} = (1-\beta_k) \frac{\sigma_{1,k}}{\sigma_{1,k}+\sigma_{2,k}} \end{cases} . \tag{2.44}$$

Clearly, the optimal source power is inversely proportional to the square root of the channel gain of the corresponding source-relay channel. That is, more power should be allocated to the source that is far away from the relay, otherwise its signal would be shadowed by that from the other end during MA phase, which increases the

relay detection error. Therefore, the above source power allocation scheme actually provides an elegant way to resolve the near-far issue. Next, if we plug (2.44) into (2.42), it leads to an objective function that involves the relay power coefficient β_k. After some manipulations, the second optimization problem can be formulated as

$$\min \ \frac{\eta_{1,k}}{1-\beta_k} + \frac{\eta_{2,k}}{\beta_k}, \ \text{s.t.} \ 0 \leq \beta_k \leq 1, \tag{2.45}$$

where

$$\eta_{1,k} = \frac{\sigma_{1,k}+\sigma_{2,k}}{4\gamma\sigma_{1,k}\sigma_{2,k}\min\left(\sigma_{1,k},\sigma_{2,k}\right)} + \frac{\left(\sigma_{1,k}+\sigma_{2,k}\right)^2}{4\gamma\sigma_{1,k}^2\sigma_{2,k}^2}\ln\frac{2\gamma\sigma_{1,k}^2\sigma_{2,k}^2}{\left(\sigma_{1,k}+\sigma_{2,k}\right)^2}, \tag{2.46}$$

$$\eta_{2,k} = \frac{\sigma_{1,k}^2+\sigma_{2,k}^2}{4\gamma\sigma_{1,k}^2\sigma_{2,k}^2}. \tag{2.47}$$

Note that we neglect the term $(1-\beta_k)$ within the logarithmic function in (2.46) when deriving the objective function in (2.45), as it is generally much smaller than γ at high SNRs. The optimizer of (2.45) can be easily derived as

$$\beta_k^{opt} = \frac{\sqrt{\eta_{2,k}}}{\sqrt{\eta_{1,k}}+\sqrt{\eta_{2,k}}}. \tag{2.48}$$

It can be shown that β_k^{opt} is a decreasing function of SNR, which coincides with our previous analysis that more power should be allocated to the sources as SNR increases. Another observation is that the power allocation coefficients depend only on the channel gains and system SNR, which are static if the inter-node distances are fixed.

2.3 Performance Analysis: Multi-Relay Case

In this section, we turn our attention to the multi-relay case. However, the exact end-to-end BER analysis based on the ML source detector (2.9) is intractable

due to the non-linearity of the decision metric. Alternatively, we seek to characterize the diversity gain at high SNRs, which reveals how the system performances improve as the number of relays increases. Following is the main conclusion of this section.

Proposition 2.1. *When there are K orthogonal relay links, the diversity gain is*

$$d(K) = -\lim_{\gamma\to\infty} \frac{\log P_{e,s_i}}{\log \gamma} = \begin{cases} \frac{K+1}{2}, K \text{ is odd} \\ \frac{K}{2}, K \text{ is even} \end{cases} = \left\lceil \frac{K}{2} \right\rceil, \tag{2.49}$$

where P_{e,s_i} is the detection error at the ith source.

The above result is somewhat counter-intuitive, as the diversity gain is only about half of the number of relays. Such performance penalty is due to error propagation, as the relays are assumed to forward whatever they detect without any error correction. To prove this result, we seek to find an upper bound and a lower bound on BER and show that they indicate the same diversity gain as (2.49).

2.3.1 BER Upper Bound

In this section, we would derive an upper bound on BER, the diversity gain of which provides a lower bound on $d(K)$ in (2.49). Note that the ML source detector (2.12) is optimum in the sense of minimizing the detection error, thus any suboptimal source detector would lead to a strictly higher BER. So we simply investigate a post-combining detector, where the ith source first performs the single-relay detection (2.31) on each relay-source channel and obtains a set of K estimates $\left\{\hat{b}_{r_k,s_i}(n)\right\}_{k=1}^{K}$,

and then feeds these estimates into a combiner whose output is

$$\hat{b}_{s_i}^U(n) = \begin{cases} 1\ , \text{ if } \left|D_{s_i}^U\right| > \left|\bar{D}_{s_i}^U\right| \\ -1, \text{ if } \left|D_{s_i}^U\right| \leq \left|\bar{D}_{s_i}^U\right| \end{cases}, \tag{2.50}$$

where $D_{s_i}^U = \left\{m \left| \hat{b}_{r_m,s_i}(n) = 1 \right.\right\}$ with the complement set being $\bar{D}_{s_i}^U$.

When K is odd, the decision rule (2.50) is equivalent to $\hat{b}_{s_i}^U(n) = 1$ if $\left|D_{s_i}^U\right| \geq \frac{K+1}{2}$. So the detection error at the ith source can be written in a similar way as (2.33), which is given by

$$\begin{aligned} P_{e,s_i}^U &= \frac{1}{2}\left\{P\left(\hat{b}_{s_i}^U(n) = -1 \left| b\left(n\right) = 1\right.\right) + P\left(\hat{b}_{s_i}^U(n) = 1 \left| b\left(n\right) = -1\right.\right)\right\} \\ &= \frac{1}{2}\left\{P\left(\left|\bar{D}_{s_i}^U\right| \geq \frac{K+1}{2} \left| b\left(n\right) = 1\right.\right) + P\left(\left|D_{s_i}^U\right| \geq \frac{K+1}{2} \left| b\left(n\right) = -1\right.\right)\right\}. \end{aligned} \tag{2.51}$$

Note that the detections on different relay-source channels are independent, and at high SNRs the conditional detection errors on the kth branch can be derived from (2.36), (2.38) and (2.41) and are given by

$$\Pr\left(\hat{b}_{r_k,s_i}(n) = -1 \left| b\left(n\right) = 1\right.\right) \approx P_{M,r_k} + P_{D,s_i}^k \approx \gamma^{-1}\left(c_{M,r_k} + q_{D,s_i}^k\right), \tag{2.52}$$

$$\Pr\left(\hat{b}_{r_k,s_i}(n) = 1 \left| b\left(n\right) = -1\right.\right) \approx P_{F,r_k} + P_{D,s_i}^k \approx \gamma^{-1}\left(q_{D,s_i}^k + d_{F,r_k} \ln\frac{\gamma}{d_{F,r_k}}\right). \tag{2.53}$$

Therefore, we have

$$\Pr\left(\left|\bar{D}_{s_i}^U\right| \geq \frac{K+1}{2} \left| b\left(n\right) = 1\right.\right) \approx \gamma^{-\frac{K+1}{2}} \sum_{\left|\bar{D}_{s_i}^U\right| = \frac{K+1}{2}} \prod_{l \in \bar{D}_{s_i}^U} \left(q_{D,s_i}^l + c_{M,r_l}\right), \tag{2.54}$$

$$\Pr\left(\left|D_{s_i}^U\right| \geq \frac{K+1}{2} \left| b\left(n\right) = -1\right.\right) \approx \gamma^{-\frac{K+1}{2}} \sum_{\left|D_{s_i}^U\right| = \frac{K+1}{2}} \prod_{m \in D_{s_i}^U} \left(q_{D,s_i}^m + d_{F,r_m} \ln\frac{\gamma}{d_{F,r_m}}\right), \tag{2.55}$$

where we neglect the higher-order terms. Clearly, P_{e,s_i}^U has a diversity gain of $\frac{K+1}{2}$ as both of the two components have the same diversity gain.

The case when K is even can be characterized in a similar way. Now the decision rule (2.50) is reduced to $\hat{b}_{s_i}^U(n) = 1$ if $\left|D_{s_i}^U\right| \geq \frac{K}{2} + 1$, and the detection error is given by

$$P_{e,s_i}^U = \frac{1}{2}\left\{P\left(\left|\bar{D}_{s_i}^U\right| \geq \frac{K}{2} \left| b\left(n\right) = 1\right.\right) + P\left(\left|D_{s_i}^U\right| \geq \frac{K}{2} + 1 \left| b\left(n\right) = -1\right.\right)\right\}. \tag{2.56}$$

From (2.52) and (2.53), we can obtain the two kinds of conditional detection errors given by

$$\Pr\left(\left|\bar{D}_{s_i}^U\right| \geq \frac{K}{2} \left| b\left(n\right) = 1\right.\right) \approx \gamma^{-\frac{K}{2}} \sum_{\left|\bar{D}_{s_i}^U\right| = \frac{K}{2}} \prod_{l \in \bar{D}_{s_i}^U} \left(q_{D,s_i}^l + c_{M,r_l}\right) \tag{2.57}$$

$$\Pr\left(\left|D_{s_i}^U\right| \geq \frac{K}{2} + 1 \left| b\left(n\right) = -1\right.\right) \approx \gamma^{-\left(\frac{K}{2}+1\right)} \sum_{\left|D_{s_i}^U\right| = \frac{K}{2}+1} \prod_{m \in D_{s_i}^U} \left(q_{D,s_i}^m + d_{F,r_m} \ln \frac{\gamma}{d_{F,r_m}}\right) \tag{2.58}$$

As the aggregate detection error is dominated by (2.57), the diversity gain is equal to $\frac{K}{2}$. After combining these two cases, we observe that the diversity gain of the BER lower bound agrees with $d\left(K\right)$ in (2.49).

2.3.2 BER Lower Bound

In this section, we would instead derive a lower bound on BER, the diversity gain of which provides an upper bound on $d\left(K\right)$ in (2.49). Here we use a similar technique proposed in [84]. Specifically, we shall make the following two ideal assumptions, i.e.,

(1) The relay-source channel is distortion free, i.e., $r_{k,i}(n) = t_k(n)$, such that both sources can observe the relay symbols $\left\{\hat{b}_{r_k}(n)\right\}_{k=1}^K$ directly;

(2) All relays have the same detection capability as the best relay, i.e.,

$$P_M = \min_{k\in\{1,2,\cdots,K\}} P_{M,r_k} \overset{\gamma\gg 1}{\approx} \frac{c_M}{\gamma},$$

$$P_F = \min_{k\in\{1,2,\cdots,K\}} P_{F,r_k} \overset{\gamma\gg 1}{\approx} \frac{d_F}{\gamma}\ln\frac{\gamma}{d_F},$$

where $c_M = \min_{k\in\{1,2,\cdots,K\}} c_{M,r_k}$ and $d_F = \min_{k\in\{1,2,\cdots,K\}} d_{F,r_k}$.

Note that both assumptions would bring positive effects on system performances, therefore helping to lower the BER. Like (2.9), the single-symbol ML detector at the ith source can be written as

$$\begin{aligned}\hat{b}_{s_i}^L(n) &= \arg\max_{b(n)\in\{-1,1\}} f\left(\left\{\hat{b}_{r_k}(n)\right\}_{k=1}^K | b(n)\right) \\ &= sign\left(\left|D_{s_i}^L\right|\ln\frac{1-P_M}{P_F} + \left|\bar{D}_{s_i}^L\right|\ln\frac{P_M}{1-P_F}\right)\end{aligned} \tag{2.59}$$

where $D_{s_i}^L = \left\{m \left| \hat{b}_{r_m}(n) = 1\right.\right\}$ with the complement set being $\bar{D}_{s_i}^L$. At high SNRs, both P_M and P_F approach 0 and $\frac{\ln P_M}{\ln P_F} \overset{\gamma\gg 1}{\approx} 1$, so the above decision rule is reduced to

$$\hat{b}_{s_i}^L(n) = \begin{cases} 1\ , \text{ if } \left|D_{s_i}^L\right| > \left|\bar{D}_{s_i}^L\right| \\ -1, \text{ if } \left|D_{s_i}^L\right| \le \left|\bar{D}_{s_i}^L\right| \end{cases}, \tag{2.60}$$

which is similar to (2.50). So the error analysis can be done in the same way as we did in the last sub-section, and we skip some tedious intermediate steps and give the final results directly. In sum, when K is odd the detection error at the ith source is given by

$$P_{e,s_i}^U = \frac{1}{2}\left\{P\left(\left|\bar{D}_{s_i}^L\right| \ge \frac{K+1}{2} | b(n) = 1\right) + P\left(\left|D_{s_i}^L\right| \ge \frac{K+1}{2} | b(n) = -1\right)\right\}, \tag{2.61}$$

$$P\left(\left|\bar{D}_{s_i}^L\right| \ge \frac{K+1}{2} | b(n) = 1\right) \approx \gamma^{-\frac{K+1}{2}} \begin{pmatrix} K \\ \frac{K+1}{2} \end{pmatrix} c_M^{\frac{K+1}{2}}, \tag{2.62}$$

$$P\left(\left|D_{s_i}^L\right| \geq \frac{K+1}{2} \left| b\left(n\right) = -1 \right.\right) \approx \gamma^{-\frac{K+1}{2}} \left(\begin{array}{c} K \\ \\ \frac{K+1}{2} \end{array} \right) \left(d_F \ln \frac{\gamma}{d_F} \right)^{\frac{K+1}{2}}. \quad (2.63)$$

When K is even, the detection error at the ith source is given by

$$P_{e,s_i}^L = \frac{1}{2} \left\{ P\left(\left|\bar{D}_{s_i}^L\right| \geq \frac{K}{2} \left| b\left(n\right) = 1 \right.\right) + P\left(\left|D_{s_i}^L\right| \geq \frac{K}{2} + 1 \left| b\left(n\right) = -1 \right.\right) \right\}, \quad (2.64)$$

$$P\left(\left|\bar{D}_{s_i}^L\right| \geq \frac{K}{2} \left| b\left(n\right) = 1 \right.\right) \approx \gamma^{-\frac{K}{2}} \left(\begin{array}{c} K \\ \\ \frac{K}{2} \end{array} \right) c_M^{\frac{K}{2}}, \quad (2.65)$$

$$P\left(\left|D_{s_i}^L\right| \geq \frac{K}{2} + 1 \left| b\left(n\right) = -1 \right.\right) \approx \gamma^{-\left(\frac{K}{2}+1\right)} \left(\begin{array}{c} K \\ \\ \frac{K}{2}+1 \end{array} \right) \left(d_F \ln \frac{\gamma}{d_F} \right)^{\frac{K}{2}+1}. \quad (2.66)$$

We can observe that the BER upper bound also have the same diversity gain indicated by (2.49), thus completing the proof.

2.4 Simulations

In this section, we present simulation results to verify the analytical results. Throughout simulations, we use the path loss model $\sigma^2 = d^{-4}$, where σ^2 is the channel gain and d is the distance between two terminals. For simplicity, we normalize the distance between two sources to 1, and we always place the relays on the line connecting two sources. In all cases, BER refers to the average detection error at both sources. Without special mention, the transmit power is always split equally among all terminals.

We first examine the performance of the single-relay systems, where $d_{1,r}$ and $d_{2,r}$ are the distances between the relay and two sources, respectively. In Figure 2.2,

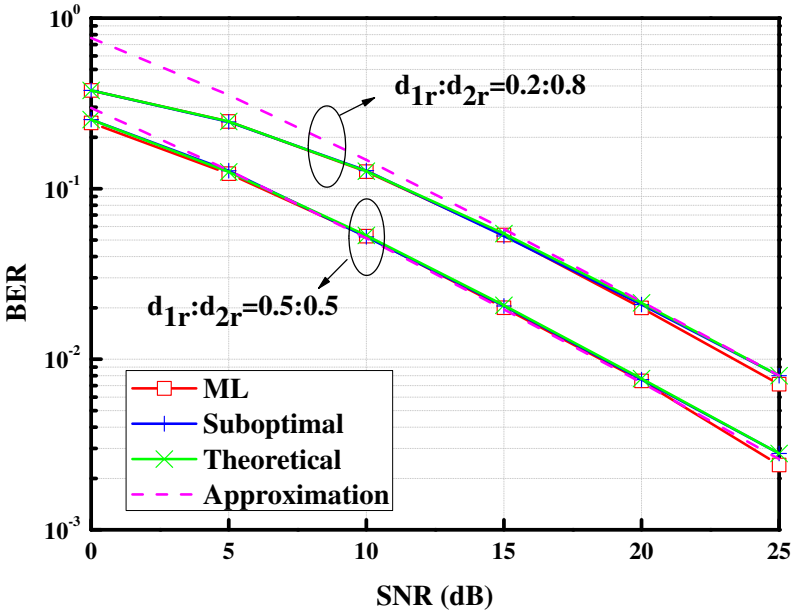


Figure 2.2: BER performances versus SNR.

we compare the BER of different relay detectors with the theoretical results. The suboptimal relay detector refers to the MUD followed by a PLNC encoder. It is observed that there is almost no difference between the ML detector and the suboptimal one, and both of them coincide with the theoretical results (2.39). Besides, the asymptotic BER (2.42) is tight when SNR is sufficiently high, e.g., when $\gamma \geq$ 15dB for $d_{1,r} : d_{2,r} = 0.2 : 0.8$ and when $\gamma \geq$ 5dB for $d_{1,r} : d_{2,r} = 0.5 : 0.5$. The tightness for the latter case is due to the high channel gains of both the source-relay channels, which make it easier to satisfy the high SNR assumption.

Then in Figure 2.3 and Figure 2.4, we proceed to study the benefits of power allocation. The optimal scheme is found through exhaustive search, and the suboptimal one refers to that given by (2.44) and (2.48) derived through asymptotic analysis. It is observed that the suboptimal scheme performs almost as well as the

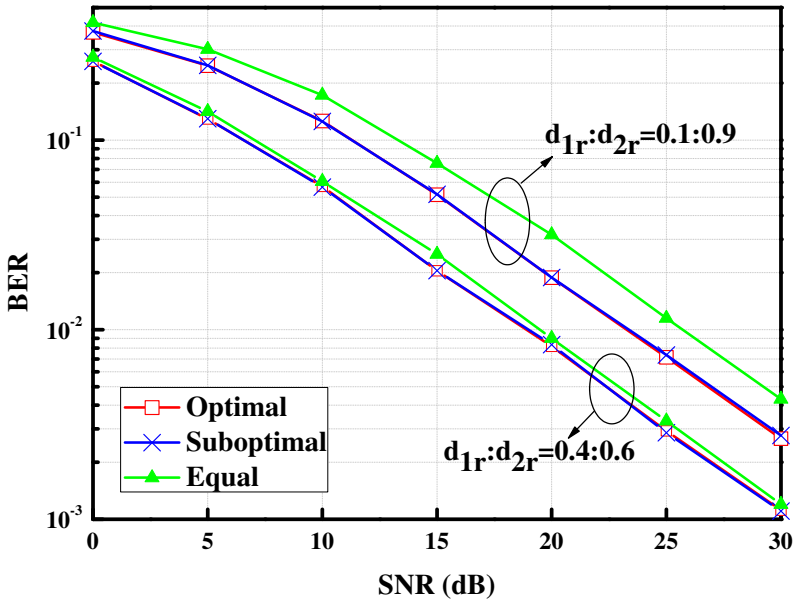


Figure 2.3: BER performances with power allocation versus SNR.

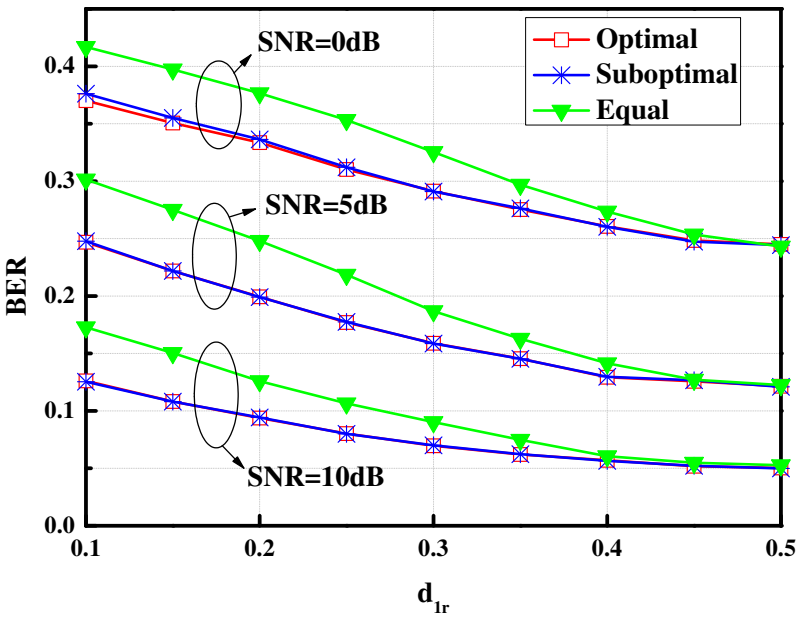


Figure 2.4: BER performances with power allocation versus relay placement.

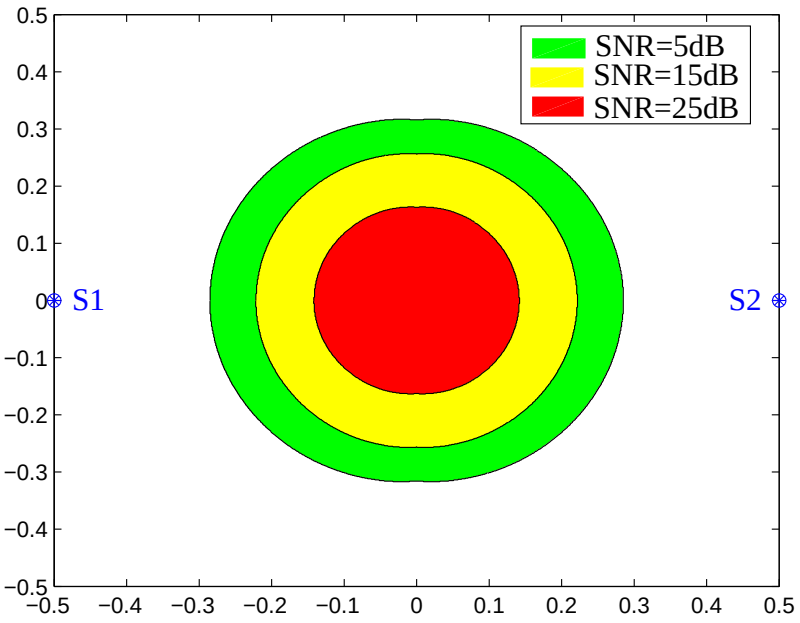


Figure 2.5: Comparison of wireless relaying and direct transmission. Colored areas correspond to the places where wireless relaying can achieve better BER.

optimal scheme in most cases. From Figure 2.4, we can observe some slight performance degradation when the SNR is low and the relay is far from source 2. This is because the channel SNR from source 2 to the relay is so low that the high SNR assumption is not fully effective on that channel. Compared with equal power allocation, about 2dB SNR gain can be observed in Figure 2.3 when $d_{1,r} : d_{2,r} = 0.1 : 0.9$. Such performance gain is diminishing as the relay moves to the halfway between two sources, in which case the equal power allocation is near-optimal.

We also compare wireless relaying with direct transmission using the same modulation scheme in Figure 2.5. To this end, we place the two sources at $(-0.5, 0)$ and $(0.5, 0)$, respectively. We then compare the BER of these two systems at each grid on a square plane, and the colored areas correspond to places where wireless

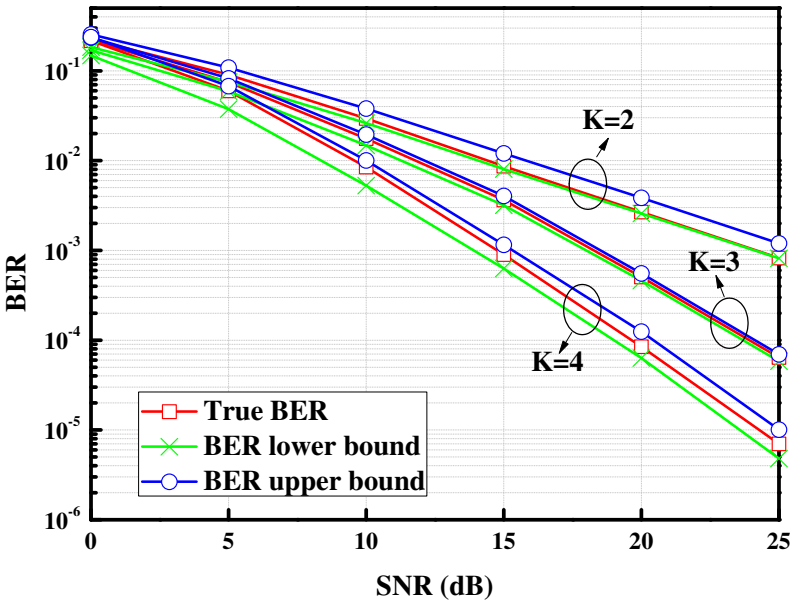


Figure 2.6: BER performances with multiple relays – all relays are at halfway between two sources.

relaying achieves better BER. To fairly compare the performance, we split the power equally between two sources for the direct transmission; as for wireless relaying, we use a mixed power allocation scheme that first determines the source power ratio by (2.44) and then finds the optimal relay power through one-dimensional search in order to reduce the time complexity. It is observed that the preferred relay locations are always concentrated around the halfway between two sources, otherwise wireless relaying cannot benefit from the high channel gain as a result of the shorter source-relay distances. Another observation is that the preferred relay locations actually shrink as SNR is increasing. This coincides with our analysis that direct transmission is more preferable at high SNRs.

Finally in Figure 2.6 and Figure 2.7 we investigate the multi-relay scenario. We first locate all relays at halfway between two sources, in which case they should

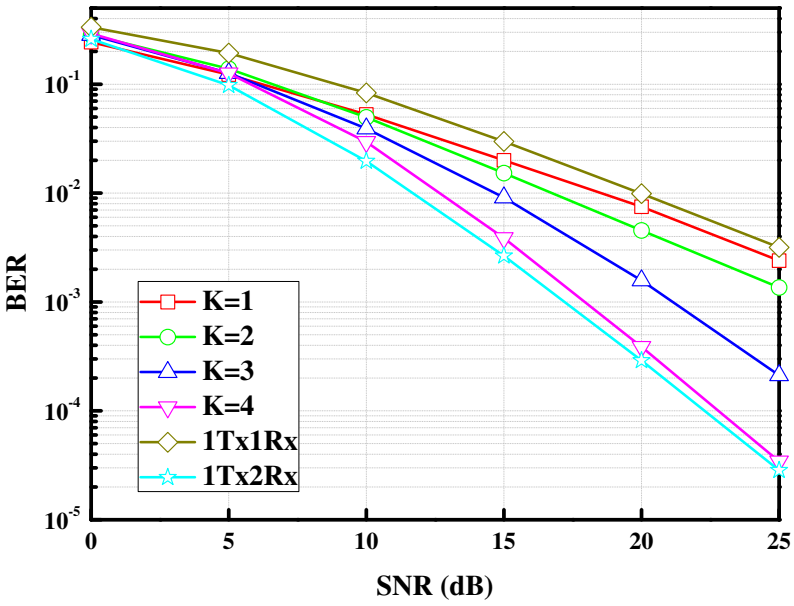


Figure 2.7: BER performances with multiple relays – all relays are equispaced between two sources.

have the same detection ability. We observe in Figure 2.6 that both of the BER bounds are tight in all cases, and they have the same slopes as we showed before. In Figure 2.7, we further compare with the typical received diversity system using one transmit antenna and K receive antenna (1TxKRx), which is well known to have a diversity gain of K [1, Eqn. (14-4-28)]. It is clear that the diversity gain of the system having 1 relay or 2 relays is 1 as 1Tx1Rx, and the system having 3 relays or 4 relays has a diversity gain of 2 as 1Tx2Rx, which validates our proposition (2.49). It should be mentioned that as all relays operate on orthogonal channels, adding more relays would reduce the spectral efficiency. Since the diversity gain is achieved along with a double loss of spectral efficiency, it is better to deploy only a small number of relays in practical systems to achieve better tradeoff between these two performance measures.

2.5 Conclusions

In this work, we studied the error performance of TWRC using DBPSK modulation. For single-relay case, we obtained the closed-form BER and developed a near-optimal power allocation scheme. We demonstrated that more power should be allocated to the sources as SNR increases, and the source associated with the weaker link should use more power to mitigate the near-far effect. For multi-relay case, we showed that around half of the total diversity gain is lost due to error propagation, as the relays may detect wrong symbols and forward such errors to the end terminals unknowingly. As a result, cooperative diversity cannot always achieve the same diversity gain as the conventional diversity schemes having the same spatial degree of freedom.

Chapter 3

Mitigating Error Propagation for Wireless Uplink with Digital Network Coding

In Chapter 2, we have demonstrated that error propagation would reduce the diversity gain of digital relaying, and the diversity loss is very severe. In practice, the best way to overcome error propagation is through error detection code [43–46], which requires the devices to be able to distinguish the correct data from the incorrect data. However, error detection mechanism requires extra overhead and is thus not available for certain networks such as censor networks, where the devices have very limited power and the redundant processing should be omitted as much as possible to extend the network lifetime. Very limited papers [47, 48] have discussed the anti error propagation strategies for uncoded systems with network coding; however, those methods require global CSI that would incur large channel estimation overhead and is thus hard to acquire.

So in this chapter, we develop some practical anti error propagation methods with reasonable CSI requirement. To be specific, we study a two-user single-relay uplink channel using DNC. We first show that due to error propagation, no diversity gain could be achieved by using the conventional transmission protocol. To address this issue, we propose to properly scale the contribution of relay link, either through power scaling at the relay side or through weighted combining at the receiver side.

Both global-CSI based methods and local-CSI based methods are explored. For soft power scaling scheme, we first develop a virtual channel model for the relay branch and demonstrate that the relay power should be such to balance the signal-to-noise ratios of the source-relay channel and relay-destination channel. For hard power scaling scheme, we first design a decision rule based on total PEP, and then simplifies it to the threshold-based relaying strategy. At the receiver side, we show that link adaptive combining with the weight being determined by the relative link quality of source-relay channel and relay-destination channel can achieve full diversity once global CSI is available, otherwise the maximum likelihood detection should be employed to achieve full diversity if the receiver only knows local CSI.

Notations: $|\cdot|$ and $(\cdot)^*$ stand for absolute value and conjugate, respectively. We shall use abbreviation i.i.d. for independent and identically distributed, and denote $Z \sim \mathcal{CN}(\mu, \sigma^2)$ as a circularly symmetric complex Gaussian random variable. The probability of an event $\mathcal{A}$ and the PDF of a random variable Z are denoted by $\Pr(\mathcal{A})$ and $f(Z)$, respectively. We define $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$ as the Q-function, and denote $g(z, \sigma^2) = \frac{1}{\pi\sigma^2} \mathrm{e}^{-\frac{|z|^2}{\sigma^2}}$ as the PDF of $Z \sim \mathcal{CN}(0, \sigma^2)$. Finally, we say $h(x) = \mathrm{O}(g(x))$ if $a \leq \lim_{x\to\infty} \frac{h(x)}{g(x)} \leq b$ for some positive constants a and b.

3.1 System Model

Consider a wireless uplink channel where two source nodes send data to a single destination, as shown in Figure 3.1. The whole data transmission is completed in 3 phases. In the kth phase for $k = 1, 2$, the kth source broadcasts its message to the

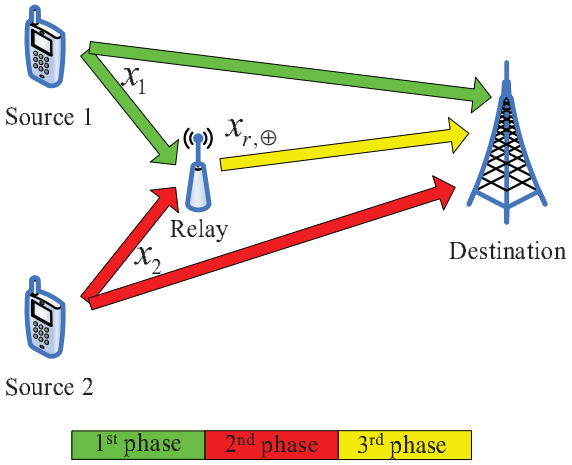


Figure 3.1: System model of the network-coded uplink.

relay and destination. The received signal can be represented as

$$y_{kt} = h_{kt}\sqrt{P}x_k + n_{kt} = \bar{h}_{kt}x_k + n_{kt} \tag{3.1}$$

for $t \in \{r, d\}$ and $k = 1, 2$. Here y_{kt} is the received signal at node t from source k, $n_{kt} \sim \mathcal{CN}(0, N_0)$ is the additive noise, $h_{kt} \sim CN(0, \lambda_{kt})$ is the Rayleigh fading channel coefficient with λ_{kt} being the channel gain, P is the transmitted power, and x_k is the source symbol with normalized power, i.e., $E|x_k|^2 = 1$. To facilitate the following analysis, we assume BPSK signal is used by the two sources, i.e., $x_k \in \Omega = \{1, -1\}$. The extension to higher-order modulations shall be discussed in later sections. The network-coded symbol as $x_\oplus = x_1 \oplus x_2 = -x_1x_2$. Note that $x_\oplus \in \{-1, 1\}$ is also BPSK signal. Besides, we define $\bar{h}_{kt} = \sqrt{P}h_{kt}$ as the equivalent channel, and define $\gamma_{kt} = |h_{kt}|^2\Gamma$ as the instantaneous channel SNR with $\Gamma = \frac{P}{N_0}$ being the reference system SNR. It is easy to show that γ_{kt} is an exponential random variable with mean $\Gamma_{kt} = \lambda_{kt}\Gamma$.

As the source symbols are randomly picked from the constellation with equal

probability, ML detection at the relay node is equivalent to minimum distance detection given by

$$x_{r,k} = \arg\min_{\hat{x}_k \in \Omega} \left| y_{kr} - \bar{h}_{kr}\hat{x}_k \right|^2 \tag{3.2}$$

for $k = 1, 2$. Then, the detected source messages are mixed through network coding, and the re-encoded message is $x_{r,\oplus} = x_{r,1} \oplus x_{r,2} = -x_{r,1}x_{r,2}$. Note that due to random detection error, it is possible that $x_{r,\oplus} \neq x_{\oplus}$. Finally in the third phase, the relay node forwards the network-coded message $x_{r,\oplus}$ to the destination, and the received signal is

$$y_{rd} = h_{rd}\sqrt{\alpha P}x_{r,\oplus} + n_{rd} = \bar{h}_{rd}\sqrt{\alpha}x_{r,\oplus} + n_{rd}. \tag{3.3}$$

Here $n_{rd} \sim \mathcal{CN}(0, N_0)$ is the additive noise, and $h_{rd} \sim CN(0, \lambda_{rd})$ is the Rayleigh fading channel coefficient with λ_{rd} being the channel gain. Besides, we define $\bar{h}_{rd} = \sqrt{P}h_{rd}$ as the equivalent relay-destination channel, and define $\gamma_{rd} = |h_{rd}|^2\Gamma$ as the corresponding channel SNR that follows exponential distribution with mean $\Gamma_{rd} = \lambda_{rd}\Gamma$. Without loss of generality, we assume the additive noises and channel coefficients of different channels are all independent. Note that the power scaling coefficient α ($0 \leq \alpha \leq 1$) in (3.3) could be adaptive to channel conditions, as will be clear later.

We consider uncoded systems throughout this chapter. As a result, neither the relay node nor the destination knows the detection status of $x_{r,\oplus}$, i.e., whether $x_{r,\oplus} = x_{\oplus}$ or not. The weighted minimum distance combining can be employed at the destination to jointly detect the two source symbols based on the observations

y_{1d}, y_{2d} and y_{rd}, i.e.,

$$\mathbf{x}_d \triangleq (x_{d,1}, x_{d,2}) = \arg \min_{\hat{x}_1, \hat{x}_2 \in \Omega} \left(\sum_{k=1}^{2} \left| y_{kd} - \bar{h}_{kd} \hat{x}_k \right|^2 + w \left| y_{rd} - \bar{h}_{rd} \sqrt{\alpha} \hat{x}_{\oplus} \right|^2 \right), \quad (3.4)$$

where the combining weight w can be leveraged to account for the possible relay detection error, as will be clear later.

Depending on how much CSI is known at each node, we consider two distinct cases. For local-CSI based methods, we assume that the receiver of each channel knows the corresponding instantaneous channel coefficient (or equivalently, the instantaneous channel SNR). Specifically, $\bar{h}_{kr}$ ($\bar{\gamma}_{kr}$) are known at the relay node for $k = 1, 2$, and $\bar{h}_{kd}$ (γ_{kd}) and $\bar{h}_{rd}$ (γ_{rd}) are known at the destination for $k = 1, 2$. For global-CSI based methods, we further assume that the relay node knows $\bar{h}_{rd}$ (γ_{rd}) and the destination knows $\bar{h}_{kr}$ (γ_{kr}) for $k = 1, 2$ besides local CSI. As the average channel SNRs are second-order statistics that stay stationary over a long time, we assume that they are available to all nodes with trivial feedback overhead.

3.2 Performance Analysis

In this work, the diversity gain is defined as

$$d = - \lim_{\Gamma \to \infty} \frac{\log \Pr \left(\mathbf{x_d} \neq \mathbf{x} \right)}{\log \Gamma}, \quad (3.5)$$

where $\mathbf{x} = (x_1, x_2)$ is the source symbol vector. Note that the maximum diversity gain is 2 because each source symbol can reach the receiver through two independent channels, i.e., the individual direct link and the common relay branch, as the network-coded symbol provides information for both sources. Unfortunately, the

exact error analysis is intractable due to the complexity in deriving the closed-form decision regions of (3.4). Alternatively, we study PEP which provides a tight bound on the real error rate.

Using the law of total probability, we can express the PEP as

$$\Pr\left(\mathbf{x} \to \hat{\mathbf{x}}\right) = \Pr\left(\mathbf{x} \to \hat{\mathbf{x}}, \Phi_{prop}, \Phi_{on}\right) + \Pr\left(\mathbf{x} \to \hat{\mathbf{x}}, \Phi_{free}, \Phi_{on}\right) + \Pr\left(\mathbf{x} \to \hat{\mathbf{x}}, \Phi_{off}\right). \tag{3.6}$$

Here Φ_{on} and Φ_{off} are the event that the relay node does forward the message (i.e., $\alpha \neq 0$) and stays idle (i.e., $\alpha = 0$), respectively. In the case of $\alpha = 0$, the weight w in (3.4) should be set to 0 too as there is no information sent from the relay node at all. On the other hand, Φ_{free} is the event that the relay node obtains the correct network-coded symbol (i.e., $x_{r,\oplus} = x_{\oplus}$), and Φ_{prop} means $x_{r,\oplus} \neq x_{\oplus}$. According to the definition of network-coded symbol, we have

$$\begin{aligned}\Pr\left(\Phi_{prop}\right) &= \Pr\left(x_{r,1} = x_1\right)\Pr\left(x_{r,2} \neq x_2\right) + \Pr\left(x_{r,1} \neq x_1\right)\Pr\left(x_{r,2} = x_2\right) \\ &= \frac{1}{2}\left(1 - \sqrt{\frac{\Gamma_{1r}}{1+\Gamma_{1r}}}\sqrt{\frac{\Gamma_{2r}}{1+\Gamma_{2r}}}\right) \overset{\Gamma\to\infty}{\approx} \frac{\lambda_{1r}+\lambda_{2r}}{4\lambda_{1r}\lambda_{2r}}\Gamma^{-1},\end{aligned} \tag{3.7}$$

and $\Pr\left(\Phi_{free}\right) = 1 - \Pr\left(\Phi_{prop}\right)$.

After some manipulations, it is also straightforward to show that[1]

$$\begin{cases} \Pr\left(\mathbf{x} \to -\mathbf{x} \,|h\right) = Q\left(\sqrt{2\sum\limits_{k\in\{1,2\}} \gamma_{kd}}\right), & \text{(3.8a)} \\ \Pr\left(\mathbf{x} \to (-x_1, x_2) \,|\Phi_{off}, h\right) = Q\left(\sqrt{2\gamma_{1d}}\right), & \text{(3.8b)} \\ \Pr\left(\mathbf{x} \to (-x_1, x_2) \,|\Phi_{free}\,, \Phi_{on}, h\right) = Q\left(\dfrac{\sqrt{2}\left(\gamma_{1d} + \alpha w \gamma_{rd}\right)}{\sqrt{\gamma_{1d} + \alpha w^2 \gamma_{rd}}}\right), & \text{(3.8c)} \\ \Pr\left(\mathbf{x} \to (-x_1, x_2) \,|\Phi_{prop}\,, \Phi_{on}, h\right) = Q\left(\dfrac{\sqrt{2}\left(\gamma_{1d} - \alpha w \gamma_{rd}\right)}{\sqrt{\gamma_{1d} + \alpha w^2 \gamma_{rd}}}\right). & \text{(3.8d)} \end{cases}$$

By using the integral representation of Q-function [85]

$$Q\left(x\right) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta \tag{3.9}$$

and averaging (3.8a) and (3.8b) over channel distributions, we can further obtain

$$\begin{cases} \Pr\left(\mathbf{x} \to -\mathbf{x}\right) = \dfrac{1}{\pi}\displaystyle\int_0^{\pi/2} \prod_{k\in\{1,2\}} \left(1 + \frac{\Gamma_{kd}}{\sin^2\theta}\right)^{-1} d\theta \overset{\Gamma\to\infty}{\approx} \frac{3}{16\lambda_{1d}\lambda_{2d}}\Gamma^{-2}, & \text{(3.10a)} \\ \Pr\left(\mathbf{x} \to (-x_1, x_2) \,|\Phi_{off}\right) = \dfrac{1}{\pi}\displaystyle\int_0^{\pi/2} \left(1 + \frac{\Gamma_{1d}}{\sin^2\theta}\right)^{-1} d\theta \overset{\Gamma\to\infty}{\approx} \frac{1}{4\lambda_{1d}}\Gamma^{-1}. & \text{(3.10b)} \end{cases}$$

From (3.8a), it is observed that the error event that neither of the two source symbols is detected correctly at the receiver has the same conditional probability regardless of the relay detection status, and the corresponding diversity gain is equal to 2. Therefore, the dominant error event occurs when only one of the source symbols flips at the receiver, which determines the overall diversity performance.

For conventional transmission protocol with full power relaying (i.e., $\alpha = 1$) and maximal-ratio combining (i.e., $w = 1$), it is easy to show that

$$\Pr\left(\mathbf{x} \to (-x_1, x_2)\right) \geq \Pr\left(\mathbf{x} \to (-x_1, x_2)\,, \Phi_{prop}, \Phi_{on}\right) = \mathrm{O}\left(\Gamma^{-1}\right). \tag{3.11}$$

[1]The symbol h means the probability is conditional on the related channels. Same convention is used throughout this work.

That is, the diversity gain is only 1 due to error propagation at the relay. By contrast, for genie-aided relaying where the relay detection error could be perfectly detected, the best response is to drop the incorrect symbol and forward only the correct symbol with full power, i.e., $\alpha = \mathbb{1}\{x_{r,\oplus} = x_{\oplus}\}$ where $\mathbb{1}\{\cdot\}$ is the indicator function. In this case, it is easy to show that full diversity gain of 2 could be achieved. However, for uncoded systems it is pretty hard to perform perfect error detection. So in the sequel, we would design some power scaling schemes and detection schemes that can achieve full diversity without using error detection.

3.3 Relay-Side Schemes

In this section, we develop two power scaling schemes at the relay side. For both methods, the combining weight w in (3.4) is set to 1, i.e., the regular equal-weight minimum distance combining is employed at the receiver. We demonstrate that full diversity can be achieved by smartly designing the power scaling coefficient α according to channel conditions.

3.3.1 Soft Power Scaling

Soft power scaling was first proposed in [40] to mitigate error propagation for the orthogonal relaying systems, and it is also called link adaptive relaying (LAR). The idea is to adapt the relay power to the channel conditions so as to limit the interference of relay detection error. However, LAR was mainly developed in the context of single-source communication, and it cannot be employed directly in the

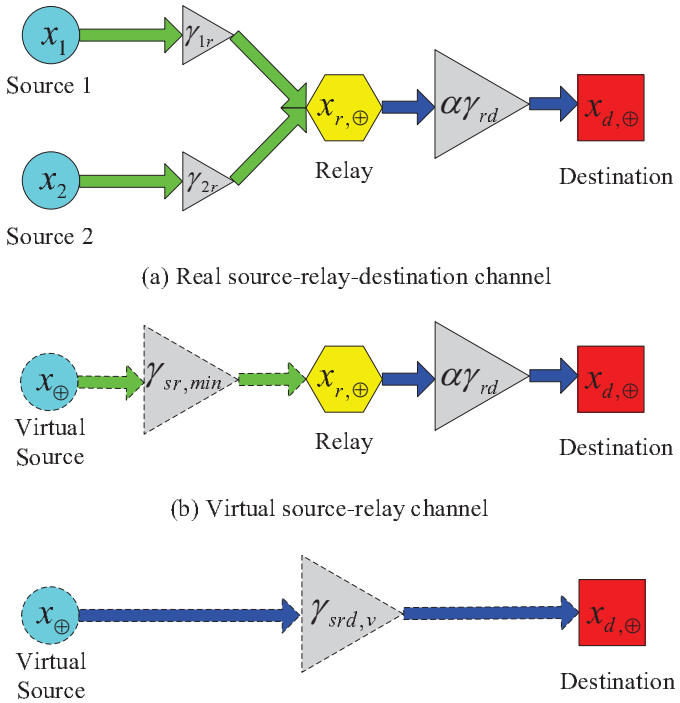


Figure 3.2: Virtual channel model for the relay branch.

network-coded uplink which accommodates multiple source nodes simultaneously.

To extend the spirit of LAR, we develop a virtual channel model for the relay branch, as shown in Figure 3.2. For the real link in Figure 3.2(a), the relay node simply forwards an estimate $x_{r,\oplus}$ of $x_\oplus$ to the destination, which is not totally reliable but still provides some information for both sources. Suppose now the destination just detects $x_\oplus$ as $x_{d,\oplus}$ based on the observation $y_{r,d}$, then the end-to-end BER $\Pr\left(x_{d,\oplus} \neq x_\oplus\right)$ is a good measure of the reliability of this two-hop relay branch. To

this end, we approximate the conditional relay detection error as

$$\begin{aligned}
&\Pr\left(x_{r,\oplus} \neq x_{\oplus} \middle| h\right) \\
&\quad = \Pr\left(x_{r,1} = x_1 \middle| h\right) \Pr\left(x_{r,2} \neq x_2 \middle| h\right) + \Pr\left(x_{r,1} \neq x_1 \middle| h\right) \Pr\left(x_{r,2} = x_2 \middle| h\right) \\
&\quad = Q\left(\sqrt{2\gamma_{1r}}\right) + Q\left(\sqrt{2\gamma_{2r}}\right) - 2Q\left(\sqrt{2\gamma_{1r}}\right) Q\left(\sqrt{2\gamma_{2r}}\right) \\
&\quad \approx Q\left(\sqrt{2\gamma_{1r}}\right) + Q\left(\sqrt{2\gamma_{2r}}\right) \approx Q\left(\sqrt{2\gamma_{sr,min}}\right), \qquad (3.12)
\end{aligned}$$

where $\gamma_{sr,min} = \min\left(\gamma_{1r}, \gamma_{2r}\right)$ represents the SNR of the worse source-relay channel. As γ_{1r} and γ_{2r} are independent exponential random variables, $\gamma_{sr,min}$ is also an exponential random variable with mean $\Gamma_{sr,min} = \lambda_{sr,min}\Gamma$, where $\lambda_{sr,min} = \frac{\lambda_{1r}\lambda_{2r}}{\lambda_{1r}+\lambda_{2r}}$. Such approximation is quite tight when γ_{1r}, γ_{2r} and their difference are reasonably large, as the Q-function $Q\left(x\right)$ decays really fast with the argument x. The above approximation shows that the multiple-input single-output source-relay channel can be accurately characterized by a single-input single-output virtual channel with the channel input being the true network-coded symbol $x_{\oplus}$ and the channel SNR being $\gamma_{sr,min}$, as shown in Figure 3.2(b). This virtual channel model can be justified by observing that the conditional BER $\Pr\left(x_{r,\oplus} \neq x_{\oplus} \middle| h\right)$ over the virtual source-relay channel, which happens to be $Q\left(\sqrt{2\gamma_{sr,min}}\right)$, is approximately the same as that over the real one. In a similar way, the end-to-end BER of this two-hop relay branch can

be approximated as

$$\begin{aligned}
&\Pr\left(x_{d,\oplus} \neq x_{\oplus} \middle| h\right) \\
&\quad = \Pr\left(x_{d,\oplus} \neq x_{r,\oplus} \middle| h\right) \Pr\left(x_{r,\oplus} = x_{\oplus} \middle| h\right) + \Pr\left(x_{d,\oplus} = x_{r,\oplus} \middle| h\right) \Pr\left(x_{r,\oplus} \neq x_{\oplus} \middle| h\right) \\
&\quad \approx Q\left(\sqrt{2\gamma_{sr,min}}\right) + Q\left(\sqrt{2\alpha\gamma_{rd}}\right) - 2Q\left(\sqrt{2\gamma_{sr,min}}\right) Q\left(\sqrt{2\alpha\gamma_{rd}}\right) \\
&\quad \approx Q\left(\sqrt{2\gamma_{srd,v}}\right), \qquad (3.13)
\end{aligned}$$

where $\gamma_{srd,v} = \min\left(\gamma_{sr,min}, \alpha\gamma_{rd}\right)$. Using the same arguments, we can further model this two-hop branch as a point-to-point virtual link with the channel input being $x_{\oplus}$ and the equivalent channel SNR being $\gamma_{srd,v}$, as shown in Figure 3.2(c). Note that the link quality is uniquely characterized by this virtual SNR, which is independent of the relay detection error patterns. Clearly, when $\gamma_{sr,min} \leq \gamma_{rd}$, the source-relay channel becomes the bottleneck, so increasing α beyond $\frac{\gamma_{sr,min}}{\gamma_{rd}}$ makes no sense as $\gamma_{srd,v} \equiv \gamma_{sr,min}$. On the other hand, if $\gamma_{sr,min} \geq \gamma_{rd}$, then the relay-destination channel becomes the bottleneck and the relay node should forward the message with full power. With the above observation, we can design the power scaling coefficient α as

$$\alpha = \begin{cases} \min\left(\dfrac{\gamma_{sr,min}}{\gamma_{rd}}, 1\right), \text{global CSI} & (3.14a) \\ \min\left(\dfrac{\gamma_{sr,min}}{\Gamma_{rd}}, 1\right), \text{local CSI} & (3.14b) \end{cases}$$

Note that γ_{rd} is unknown to the relay node when only local CSI is available, so we have used its mean Γ_{rd} in (3.14b) as a blind estimate. In some sense, the relay node behaves like a link coordinator that strives to balance the channel SNRs of the two hops, as the worse hop limits the whole link quality. As for the diversity performances, we have the following proposition.

Proposition 3.1. *Both global-CSI and local-CSI based soft power scaling can achieve a diversity gain of 2.*

Proof. We prove the first part using our virtual channel model. As the relay branch is modeled as a point-to-point link with the channel input being $x_{\oplus}$, the whole transmitted codeword now becomes $(x_1, x_2, x_{\oplus})$, where each symbol is delivered in different time slots. After some manipulations, it is easy to show that given any power scaling coefficient α employed at the relay side and any combining weight w employed at the receiver side, the PEP can be in general approximated as

$$\Pr(\mathbf{x} \to \hat{\mathbf{x}}) \approx E\left[Q\left(\frac{\sum\limits_{k\in\{1,2\}} \gamma_{kd}|x_k - \hat{x}_k|^2 + w\alpha\gamma_{rd}|x_{\oplus} - \hat{x}_{\oplus}|^2}{\sqrt{2\sum\limits_{k\in\{1,2\}} \gamma_{kd}|x_k - \hat{x}_k|^2 + 2\alpha^2 w^2 \gamma_{rd}^2 |x_{\oplus} - \hat{x}_{\oplus}|^2 / \gamma_{srd,v}}}\right)\right]. \tag{3.15}$$

For global-CSI based soft power scaling (3.14a), the virtual channel SNR is $\gamma_{srd,v} = \min(\gamma_{sr,min}, \gamma_{rd}) \triangleq \gamma_{srd,min}$. Here $\gamma_{srd,min}$ follows exponential distribution with mean $\Gamma_{srd,min} = \lambda_{srd,min}\Gamma$, where $\lambda_{srd,min} = \frac{\lambda_{1r}\lambda_{2r}\lambda_{rd}}{\lambda_{1r}\lambda_{2r}+\lambda_{1r}\lambda_{rd}+\lambda_{2r}\lambda_{rd}}$. After applying the Chernoff bound [1] on the Q-function and plugging in $w = 1$, we can further obtain

$$\begin{aligned}\Pr(\mathbf{x} \to \hat{\mathbf{x}}) &\leq E\left[\frac{1}{2}\exp\left(-\frac{\sum\limits_{k\in\{1,2\}} \gamma_{kd}|x_k - \hat{x}_k|^2 + \gamma_{srd,min}|x_{\oplus} - \hat{x}_{\oplus}|^2}{4}\right)\right] \\ &\overset{\Gamma\to\infty}{\approx} \frac{1}{2}\Big(\prod\nolimits_{i=1}^{r} \Lambda_i\Big)^{-1} \Gamma^{-r},\end{aligned} \tag{3.16}$$

where Λ_i and r is the ith non-zero eigenvalue and the rank of the diagonal matrix

$$\begin{pmatrix} \frac{\lambda_{1d}|x_1-\hat{x}_1|^2}{4} & 0 & 0 \\ 0 & \frac{\lambda_{2d}|x_2-\hat{x}_2|^2}{4} & 0 \\ 0 & 0 & \frac{\lambda_{srd,min}|x_{\oplus}-\hat{x}_{\oplus}|^2}{4} \end{pmatrix},$$

respectively. Because at least two diagonal elements are non-zero when an error event happens, we have $\max_{\hat{\mathbf{x}}\neq\mathbf{x}} \Pr(\mathbf{x}\to\hat{\mathbf{x}}) = \mathrm{O}(\Gamma^{-2})$, which completes the proof of the first part. For the second part, the PEP can be alternatively bounded using (3.6) after plugging in (3.8c), (3.8d), (3.12) and $\Pr(\Phi_{off}) = 0$ as

$$\begin{aligned}\Pr(\mathbf{x}\to(-x_1,x_2)) \leq\ & E\left[\left(Q\left(\frac{\sqrt{2}\,(\gamma_{1d}-\alpha\gamma_{rd})}{\sqrt{\gamma_{1d}+\alpha\gamma_{rd}}}\right)Q\left(\sqrt{2\gamma_{sr,min}}\right)\right)\right] \\ & + E\left[\left(Q\left(\sqrt{2\,(\gamma_{1d}+\alpha\gamma_{rd})}\right)\right)\right]. \end{aligned} \tag{3.17}$$

Note that the two source-relay channels have been put into one virtual channel with the virtual channel SNR being $\gamma_{sr,min}$, which still follows exponential distribution. As a result, we can follow the similar steps in [40] to show that both terms in (3.17) scale as $\mathrm{O}(\Gamma^{-2})$ at high SNRs. □

3.3.2 Hard Power Scaling

For soft power scaling, the relay node has to constantly change its power level and let the destination know its transmitted power for performing coherent detection. To save the extra overhead, the relay could instead apply hard power scaling (i.e., $\alpha \in \{0,1\}$). As the total PEP upper bounds the real detection error rate at the receiver, we propose to turn on the relay node (i.e., $\alpha = 1$) when

$$\sum_{\hat{\mathbf{x}}\neq\mathbf{x}} \Pr(\mathbf{x}\to\hat{\mathbf{x}}|\,\Phi_{on}, h_{1r}, h_{2r}) \leq \sum_{\hat{\mathbf{x}}\neq\mathbf{x}} \Pr(\mathbf{x}\to\hat{\mathbf{x}}|\,\Phi_{off}). \tag{3.18}$$

That is, the relay node always chooses the action that promises smaller total PEP. If (3.18) is otherwise false, then the relay node should stay idle by letting $\alpha = 0$.

After some manipulations, we can show that the above decision rule is equivalent to

$$\Pr\left(\Phi_{prop}\,|h_{1r},h_{2r}\right)\leq\frac{\sum\limits_{\hat{\mathbf{x}}\neq\mathbf{x}}\left(\Pr\left(\mathbf{x}\rightarrow\hat{\mathbf{x}}\,|\Phi_{off}\right)-\Pr\left(\mathbf{x}\rightarrow\hat{\mathbf{x}}\,|\Phi_{free}\,,\Phi_{on}\right)\right)}{\sum\limits_{\hat{\mathbf{x}}\neq\mathbf{x}}\left(\Pr\left(\mathbf{x}\rightarrow\hat{\mathbf{x}}\,|\Phi_{prop}\,,\Phi_{on}\right)-\Pr\left(\mathbf{x}\rightarrow\hat{\mathbf{x}}\,|\Phi_{free}\,,\Phi_{on}\right)\right)},\tag{3.19}$$

where $\Pr\left(\Phi_{prop}\,|h_{1r},h_{2r}\right)$ is given by (3.12), $\Pr\left(\mathbf{x}\rightarrow-\mathbf{x}\right)$ and $\Pr\left(\mathbf{x}\rightarrow(-x_1,x_2)\,|\Phi_{off}\right)$ is shown in (3.10a) and (3.10b), respectively. After plugging $\alpha=w=1$ back into (3.8c) and (3.8d) and averaging over channel distribution, we have

$$\begin{cases}\Pr\left(\mathbf{x}\rightarrow(-x_1,x_2)\,|\Phi_{free},\Phi_{on}\right)\overset{\Gamma\rightarrow\infty}{\approx}\dfrac{3}{16\lambda_{1d}\lambda_{rd}}\Gamma^{-2}, & (3.20a)\\ \Pr\left(\mathbf{x}\rightarrow(-x_1,x_2)\,|\Phi_{prop},\Phi_{on}\right)\overset{\Gamma\rightarrow\infty}{\approx}\dfrac{\lambda_{rd}}{\lambda_{1d}+\lambda_{rd}}, & (3.20b)\end{cases}$$

where the high-SNR approximation in (3.20b) has been proved in [38]. Note that the exact decision rule (3.19) is somewhat intractable, as the average of Q-function over channel distribution is hard to manipulate. Alternatively, we choose to use the high-SNR approximations to simplify the right-hand side of (3.19), i.e.,

$$\begin{cases}\sum\limits_{\hat{\mathbf{x}}\neq\mathbf{x}}\Pr\left(\mathbf{x}\rightarrow\hat{\mathbf{x}}\,|\Phi_{off}\right)\overset{\Gamma\rightarrow\infty}{\approx}\dfrac{\lambda_{1d}+\lambda_{2d}}{4\lambda_{1d}\lambda_{2d}}\Gamma^{-1}, & (3.21a)\\ \sum\limits_{\hat{\mathbf{x}}\neq\mathbf{x}}\Pr\left(\mathbf{x}\rightarrow\hat{\mathbf{x}}\,|\Phi_{prop}\,,\Phi_{on}\right)\overset{\Gamma\rightarrow\infty}{\approx}\dfrac{\lambda_{rd}}{\lambda_{2d}+\lambda_{rd}}+\dfrac{\lambda_{rd}}{\lambda_{1d}+\lambda_{rd}}, & (3.21b)\\ \sum\limits_{\hat{\mathbf{x}}\neq\mathbf{x}}\Pr\left(\mathbf{x}\rightarrow\hat{\mathbf{x}}\,|\Phi_{free}\,,\Phi_{on}\right)\overset{\Gamma\rightarrow\infty}{\approx}\dfrac{\lambda_{1d}+\lambda_{2d}+\lambda_{rd}}{\lambda_{1d}\lambda_{2d}\lambda_{rd}}\dfrac{3}{16}\Gamma^{-2}. & (3.21c)\end{cases}$$

By using the virtual source-relay channel model in Figure 3.2(b) and applying the Chernoff bound [1], we can further simplify the decision rule (3.19) as

$$Q\left(\sqrt{2\gamma_{sr,min}}\right)\leq\frac{1}{2}e^{-\gamma_{sr,min}}\leq\frac{1}{2\lambda_T}\Gamma^{-1},\tag{3.22}$$

or equivalently,

$$\gamma_{sr,min}\geq\log\lambda_T\Gamma,\tag{3.23}$$

where

$$\lambda_T = \frac{2\lambda_{1d}\lambda_{2d}\lambda_{rd}\left(\lambda_{1d}+\lambda_{2d}+2\lambda_{rd}\right)}{\left(\lambda_{1d}+\lambda_{2d}\right)\left(\lambda_{1d}+\lambda_{rd}\right)\left(\lambda_{2d}+\lambda_{rd}\right)} \tag{3.24}$$

is a constant determined by the second-order statistics. Consequently, the complex decision rule (3.19) is simplified to the threshold-based relaying strategy. We observe that the two source-relay channels have to meet the same SNR threshold, as the relay detection error is bounded by the worse channel as shown in (3.12). We also observe that imposing any threshold of the form $\log\left(\lambda_{T,k}\Gamma\right)$ on γ_k for $\lambda_{T,k} > 0$ and $k = 1, 2$ would lead to the same diversity performance, since $\log \lambda_{T,k}\Gamma \overset{\Gamma\to\infty}{\approx} \log\Gamma$. The special λ_T given in (3.24) can be justified by the following proposition.

Proposition 3.2. *For all the hard power scaling strategies with*

$$\Phi_{on} = \left\{\gamma_{kr} \geq \log \lambda_{T,k}\Gamma,\ k = 1, 2\right\}, \tag{3.25}$$

where $\lambda_{T,k}$ is a positive constant, a diversity gain of 2 can be achieved. Besides, $\lambda_{T,1} = \lambda_{T,2} = \lambda_T$ is optimum in the sense of minimizing the total end-to-end PEP.

Proof. From the decision rule (3.25), we can show that

$$\begin{aligned}\Pr\left(\Phi_{on}\right) &= \prod_{k\in\{1,2\}} \Pr\left(\gamma_{kr} \geq \log \lambda_{T,k}\Gamma\right) = \exp\left(-\sum_{k\in\{1,2\}} \frac{\log \lambda_{T,k}\Gamma}{\lambda_{kr}\Gamma}\right) \\ &\overset{\Gamma\to\infty}{\approx} 1 - \sum_{k\in\{1,2\}} \frac{\log \lambda_{T,k}\Gamma}{\lambda_{kr}\Gamma} \overset{\Gamma\to\infty}{\approx} 1,\end{aligned} \tag{3.26}$$

and $\Pr\left(\Phi_{off}\right) = 1 - \Pr\left(\Phi_{on}\right) \overset{\Gamma\to\infty}{\approx} \sum_{k\in 1,2} \frac{\log \lambda_{T,k}\Gamma}{\lambda_{kr}\Gamma}$. The conditional PDF of γ_{kr} given Φ_{on} is

$$f\left(\gamma_{kr}\left|\Phi_{on}\right.\right) = \frac{f\left(\gamma_{kr}\right)}{\Pr\left(\gamma_{kr} \geq \log \lambda_{T,k}\Gamma\right)} = \frac{1}{\lambda_{kr}\Gamma}\exp\left(-\frac{\gamma_{kr} - \log \lambda_{T,k}\Gamma}{\lambda_{kr}\Gamma}\right) \tag{3.27}$$

for $\gamma_{kr} \geq \log \lambda_{T,k}\Gamma$ and $k = 1, 2$. Now we can obtain

$$\begin{aligned}\Pr\left(x_{r,k} \neq x_k \,|\Phi_{on}\right) &= \int_{\log \lambda_{T,k}\Gamma}^{\infty} Q\left(\sqrt{2\gamma_{kr}}\right) \frac{1}{\lambda_{kr}\Gamma} \exp\left(-\frac{\gamma_{kr} - \log \lambda_{T,k}\Gamma}{\lambda_{kr}\Gamma}\right) d\gamma_{kr} \\ &\leq \frac{1}{2\lambda_{kr}\Gamma} \int_{\log \lambda_{T,k}\Gamma}^{\infty} \exp\left(-\gamma_{kr}\right) d\gamma_{kr} = \frac{1}{2\lambda_{kr}\lambda_{T,k}}\Gamma^{-2}, \end{aligned} \tag{3.28}$$

which leads to

$$\Pr\left(\Phi_{prop} \,|\Phi_{on}\right) \leq \sum_{k\in\{1,2\}} \Pr\left(x_{r,k} \neq x_k \,|\Phi_{on}\right) \leq \frac{\lambda_{1r}\lambda_{T,1} + \lambda_{2r}\lambda_{T,2}}{2\lambda_{1r}\lambda_{2r}\lambda_{T,1}\lambda_{T,2}}\Gamma^{-2} = \mathrm{O}\left(\Gamma^{-2}\right), \tag{3.29}$$

and $\Pr\left(\Phi_{free} \,|\Phi_{on}\right) = 1 - \Pr\left(\Phi_{prop} \,|\Phi_{on}\right) \overset{\Gamma\to\infty}{\approx} 1$. After plugging (3.21) and the above results back into (3.6), we have

$$\begin{aligned}\sum_{\hat{\mathbf{x}}\neq\mathbf{x}} \Pr\left(\mathbf{x} \to \hat{\mathbf{x}}\right) \overset{\Gamma\to\infty}{\leq} &\sum_{k\in\{1,2\}} \frac{\log \lambda_{T,k}\Gamma}{\lambda_{kr}} \frac{\lambda_{1d} + \lambda_{2d}}{4\lambda_{1d}\lambda_{2d}}\Gamma^{-2} + \frac{\lambda_{1d} + \lambda_{2d} + \lambda_{rd}}{\lambda_{1d}\lambda_{2d}\lambda_{rd}} \frac{3}{16}\Gamma^{-2} \\ &+ \sum_{k\in\{1,2\}} \frac{\lambda_{rd}}{\lambda_{kd} + \lambda_{rd}} \frac{\lambda_{1r}\lambda_{T,1} + \lambda_{2r}\lambda_{T,2}}{2\lambda_{1r}\lambda_{2r}\lambda_{T,1}\lambda_{T,2}}\Gamma^{-2}. \end{aligned} \tag{3.30}$$

Therefore, a diversity gain of 2 is achieved. To prove the second part, we need to find the optimum $\lambda_{T,k}$ for $k = 1, 2$ to minimize the above bound, i.e.,

$$\lambda_{T,k}^{*} = \arg\min_{\lambda_{T,k}} \left(\log \lambda_{T,k} \frac{\lambda_{1d} + \lambda_{2d}}{2\lambda_{1d}\lambda_{2d}} + \frac{1}{\lambda_{T,k}} \sum_{k\in\{1,2\}} \frac{\lambda_{rd}}{\lambda_{kd} + \lambda_{rd}} \right). \tag{3.31}$$

It is easy to check that $\lambda_{T,1}^{*} = \lambda_{T,2}^{*} = \lambda_T$ as given in (3.24). □

3.4 Receiver-Side Schemes

So far we have shown that the error propagation issue could be efficiently addressed at the relay side. Alternatively, we show in this section that full diversity can also be achieved through receiver-side processing even when there is no power scaling at the relay side. Throughout this section, we assume the relay node always forwards message using full power (i.e., $\alpha = 1$).

3.4.1 Link Adaptive Combining

One way to properly scale the contribution of the relay link at the receiver side is by adopting a proper combining weight w in (3.4). Basically, the combining weight w is a kind of confidence measure that reflects how reliable the relay branch is. When the relay detection error is very likely to occur, the destination should adaptively lower the combining weight to heavily discount the contribution of the relay branch.

Before describing our choice of w, let us first revisit the virtual relay branch shown in Figure 3.2(c) to gain more insights. As mentioned before, this virtual channel has the BPSK input $x_{\oplus}$ and the channel SNR is $\gamma_{srd,v}$. As the relay-destination channel coefficient is h_{rd}, we can approximate the real received signal y_{rd} in (3.3) as

$$\tilde{y}_{rd} = \bar{h}_{rd}\sqrt{\alpha}x_{\oplus} + \tilde{n}_{rd}, \tag{3.32}$$

where $\tilde{n}_{rd} \sim \mathcal{CN}(0, \frac{\gamma_{rd}\alpha}{\gamma_{srd,v}})$ is the virtual channel noise, and the noise power is such that the SNR of this virtual signal is exactly $\gamma_{srd,v}$. With the above signal model, it is easy to show that the ML detection based on the observations y_{1d}, y_{2d} and $\tilde{y}_{rd}$ is

$$\begin{aligned}\mathbf{x}_d \stackrel{\Delta}{=} (x_{d,1}, x_{d,2}) &= \arg \max_{\hat{x}_1,\hat{x}_2\in\Omega} f\left(y_{1d}, y_{2d}, \tilde{y}_{rd} | \, \hat{x}_1, \hat{x}_2\right) \\ &= \arg \max_{\hat{x}_1,\hat{x}_2\in\Omega} g\left(\tilde{y}_{rd} - \bar{h}_{rd}\sqrt{\alpha}\hat{x}_{\oplus}, \frac{\gamma_{rd}\alpha}{\gamma_{srd,v}}\right) \prod_{k\in\{1,2\}} g\left(y_{kd} - \bar{h}_{kd}\hat{x}_k, N_0\right), \end{aligned} \tag{3.33}$$

where we exploit the independence of the three received signals. As $\alpha = 1$, we can show that the above ML detector is actually equivalent to the weighted minimum

distance combiner (3.4) by letting

$$w = \frac{\gamma_{srd,v}}{\gamma_{rd}} = \min\left(\frac{\gamma_{sr,min}}{\gamma_{rd}}, 1\right). \tag{3.34}$$

We remark that our design is asymptotically the same as that proposed in [47]; however, unlike [47] which directly extends the scheme in [41] in a heuristic way, we justify such design using our virtual channel model, which clearly shows that the adaptive weight should be such to equalize the power of virtual channel noise before entering the combiner. To be specific, when $\gamma_{sr,min} \leq \gamma_{rd}$, the virtual noise power is $\frac{\gamma_{rd}}{\gamma_{sr,min}} \geq 1$, which reflects the fact that the delivered symbol $x_{\oplus}$ is unreliable as the source-relay channel is the system bottleneck. On the other hand if $\gamma_{sr,min} > \gamma_{rd}$, the relay-destination channel becomes the bottleneck, then the virtual noise has unit power and the relay branch is given full credit as the other two source-destination channels.

By comparing global-CSI based soft power scaling (3.14a) and global-CSI based link adaptive combining scheme (3.34), we observe that

$$\alpha w \gamma_{rd} = \gamma_{srd,v} = \min\left(\gamma_{sr,min}, \gamma_{rd}\right) \tag{3.35}$$

in both schemes. This factor can be regarded as the aggregate scaling coefficient effective on $x_{\oplus}$ to mitigate the impact of relay decoding error. So basically, the two schemes are following the same principle to address the error propagation issue. Due to such relation, one may guess that when only local CSI is available, we can replace $\gamma_{sr,min}$ by its average $\Gamma_{sr,min}$ in (3.34), i.e., let

$$w = \min\left(\frac{\Gamma_{sr,min}}{\gamma_{rd}}, 1\right) \tag{3.36}$$

and still achieve a diversity gain of 2 as is the case of local-CSI based soft power scaling scheme. However, this is not true as we show in the following proposition.

Proposition 3.3. *For global-CSI based link adaptive combining scheme (3.34), a diversity gain of 2 can be achieved. However, the diversity gain of local-CSI based link adaptive combining scheme (3.36) is only 1.*

Proof. The first part is easy to prove, as the PEP upper bound is the same as (3.16) after plugging (3.35) into (3.15). So let us focus on the local-CSI based link adaptive combining. As $\alpha = 1$ and $\Pr(\Phi_{off}) = 0$, we derive from (3.6)

$$\Pr(\mathbf{x} \to (-x_1, x_2)) \geq \Pr(\mathbf{x} \to (-x_1, x_2) | \Phi_{prop}, \Phi_{on}) \Pr(\Phi_{prop}). \tag{3.37}$$

After plugging $\alpha = 1$ and (3.36) back into (3.8d), we have

$$\begin{aligned}
&\Pr(\mathbf{x} \to (-x_1, x_2) | \Phi_{prop}, \Phi_{on}) \\
&\quad = E\left[Q\left(\frac{\sqrt{2}(\gamma_{1d} - \min(\Gamma_{sr,min}, \gamma_{rd}))}{\sqrt{\gamma_{1d} + \min\left(\Gamma_{sr,min}^2, \gamma_{rd}^2\right)/\gamma_{rd}}}\right)\right] \\
&\quad \geq E_{\gamma_{1d} < \Gamma_{sr,\min} \leq \gamma_{rd}}\left[Q\left(\frac{\sqrt{2}(\gamma_{1d} - \min(\Gamma_{sr,min}, \gamma_{rd}))}{\sqrt{\gamma_{1d} + \min\left(\Gamma_{sr,min}^2, \gamma_{rd}^2\right)/\gamma_{rd}}}\right)\right] \\
&\quad \geq \frac{1}{2}\Pr(\gamma_{rd} > \Gamma_{sr,min}, \gamma_{1d} < \Gamma_{sr,min}) \\
&\quad = \frac{1}{2}e^{-\frac{\lambda_{sr,min}}{\lambda_{rd}}}\left(1 - e^{-\frac{\lambda_{sr,min}}{\lambda_{1d}}}\right) = \mathrm{O}(1),
\end{aligned} \tag{3.38}$$

where in the last inequality we use the fact $Q(x) \geq \frac{1}{2}$ for $x \leq 0$. Recall that $\Pr(\Phi_{prop}) = \mathrm{O}(\Gamma^{-1})$ as shown in (3.7), we conclude that $\Pr(\mathbf{x} \to (-x_1, x_2)) = \mathrm{O}(\Gamma^{-1})$. □

3.4.2 Maximum Likelihood Detection

So far, we have focused on linear combining at the destination; however, full diversity cannot be achieved when the receiver only knows local CSI. The problem is that link adaptive combining is based on the virtual channel model, so the linear detector is strictly suboptimum. So in this subsection, we study the diversity performance of the exact ML detection, which is optimum in the sense of minimizing detection errors.

The exact ML detector based on real observations y_{1d}, y_{2d} and y_{rd} is given by

$$\mathbf{x}_d \triangleq (x_{d,1}, x_{d,2}) = \arg \max_{\hat{x}_1, \hat{x}_2 \in \Omega} f\left(y_{rd} \middle| \hat{x}_1, \hat{x}_2\right) \prod_{k \in \{1,2\}} g\left(y_{kd} - \bar{h}_{kd} \hat{x}_k, N_0\right), \tag{3.39}$$

where

$$f\left(y_{rd} \middle| x_1, x_2\right) = g\left(y_{rd} + \bar{h}_{rd} x_{\oplus}, N_0\right) \Pr\left(\Phi_{prop}\right) + g\left(y_{rd} - \bar{h}_{rd} x_{\oplus}, N_0\right) \Pr\left(\Phi_{free}\right) \tag{3.40}$$

is the conditional PDF of y_{rd} given the two source symbols x_1 and x_2, and (3.12) and (3.7) should be plugged in to replace the term $\Pr\left(\Phi_{prop}\right)$ for global-CSI based ML detection and local-CSI based ML detection, respectively. Our main result is summarized below.

Proposition 3.4. *Both global-CSI based ML detection and local-CSI based ML detection can achieve a diversity gain of 2.*

Proof. We only prove the case for local-CSI based ML detection, which is dominated by the global-CSI based ML detection. As $f\left(y_{rd} \middle| x_1, x_2\right) = f\left(y_{rd} \middle| -x_1, -x_2\right)$, it is easy to show that $\Pr\left(\mathbf{x} \to -\mathbf{x}\right) = \mathrm{O}\left(\Gamma^{-2}\right)$. Next we investigate the PEP

$\Pr(\mathbf{x} \to (-x_1, x_2))$. After some manipulations, we have

$$\Pr(\mathbf{x} \to (-x_1, x_2)) = \Pr\left(\frac{4\mathrm{Re}\left(y_{1d}\bar{h}_{1d}^* x_1\right)}{N_0} \leq q\left(\Pr(\Phi_{prop}), \frac{4\mathrm{Re}\left(y_{rd}\bar{h}_{rd}^* x_1 x_2\right)}{N_0}\right)\right), \tag{3.41}$$

where

$$q(\varepsilon, t) = \log\frac{\varepsilon + (1-\varepsilon)e^t}{\varepsilon e^t + (1-\varepsilon)} \approx \begin{cases} \log\frac{1-\varepsilon}{\varepsilon}, & t \geq \log\frac{1-\varepsilon}{\varepsilon} \\ t, & \log\frac{\varepsilon}{1-\varepsilon} \leq t \leq \log\frac{1-\varepsilon}{\varepsilon} \\ \log\frac{\varepsilon}{1-\varepsilon}, & t \leq \log\frac{\varepsilon}{1-\varepsilon} \end{cases}. \tag{3.42}$$

The last piece-wise linear approximation in (3.42) is proved in [84]. Define $Z = \frac{4\mathrm{Re}\left(y_{1d}\bar{h}_{1d}^* x_1\right)}{N_0}$, which can be rewritten as the quadratic form of two independent complex Gaussian random variables. According to [86], the PDF of Z is

$$f(z) = \begin{cases} \frac{ab}{a+b}e^{-az}, z > 0 \\ \frac{ab}{a+b}e^{bz}, \quad z \leq 0 \end{cases}, \tag{3.43}$$

where

$$\begin{cases} a = \sqrt{1+\Gamma_{1d}^{-1}} - 1 \overset{\Gamma\to\infty}{\approx} \frac{1}{2\lambda_{1d}}\Gamma^{-1} \\ b = \sqrt{1+\Gamma_{1d}^{-1}} + 1 \overset{\Gamma\to\infty}{\approx} 2 \end{cases}. \tag{3.44}$$

With the above PDF, it is easy to show that $\Pr(Z \leq -\log\Gamma) \overset{\Gamma\to\infty}{\approx} \frac{1}{4\lambda_{1d}}\Gamma^{-3}$ and $\Pr(Z \leq \log\Gamma) = \frac{1}{2\lambda_{1d}}\frac{\log\Gamma}{\Gamma}$. Likewise, we can show that the conditional PDFs of $T = \frac{4\mathrm{Re}\left(y_{rd}\bar{h}_{rd}^* x_1 x_2\right)}{N_0}$ are

$$f(t|\,\Phi_{prop}) = \begin{cases} \frac{cd}{c+d}e^{-ct}, t > 0 \\ \frac{cd}{c+d}e^{dt}, \quad t \leq 0 \end{cases} \tag{3.45}$$

and

$$f(t|\,\Phi_{free}) = \begin{cases} \frac{cd}{c+d}e^{-dt}, \quad t > 0 \\ \frac{cd}{c+d}e^{ct}, \quad t \leq 0 \end{cases}, \tag{3.46}$$

respectively, where

$$\begin{cases} c = \sqrt{1+\Gamma_{rd}^{-1}} - 1 \overset{\Gamma\to\infty}{\approx} \frac{1}{2\lambda_{rd}}\Gamma^{-1} \\ d = \sqrt{1+\Gamma_{rd}^{-1}} + 1 \overset{\Gamma\to\infty}{\approx} 2 \end{cases} . \tag{3.47}$$

Again it is easy to show that $\Pr\left(Z \le T \middle| \Phi_{free}\right) \overset{\Gamma\to\infty}{\approx} \frac{3}{16\lambda_{1d}\lambda_{rd}}\Gamma^{-2} = \mathrm{O}\left(\Gamma^{-2}\right)$. With the above results, we can obtain

$$\begin{aligned} \Pr\left(\mathbf{x} \to (-x_1, x_2) \middle| \Phi_{prop}\right) &= \Pr\left(Z \le q\left(\Pr\left(\Phi_{prop}\right), T\right) \middle| \Phi_{prop}\right) \\ &\le \Pr\left(Z \le \nu\right) \overset{\Gamma\to\infty}{\approx} \Pr\left(Z \le \log\Gamma\right) = \mathrm{O}\left(\frac{\log\Gamma}{\Gamma}\right), \end{aligned} \tag{3.48}$$

where

$$\nu = \log\frac{1-\Pr\left(\Phi_{prop}\right)}{\Pr\left(\Phi_{prop}\right)} \overset{\Gamma\to\infty}{\approx} \log\Gamma \tag{3.49}$$

according to the high-SNR approximation of $\Pr\left(\Phi_{prop}\right)$ in (3.7). Besides, it is easy to show that

$$\begin{aligned} &\Pr\left(\mathbf{x} \to (-x_1, x_2) \middle| \Phi_{free}\right) \\ &\quad\approx \Pr\left(Z \le \nu \le T \middle| \Phi_{free}\right) + \Pr\left(Z \le -\nu, T \le -\nu \middle| \Phi_{free}\right) \\ &\qquad + \Pr\left(Z \le T, -\nu \le T \le \nu \middle| \Phi_{free}\right) \\ &\quad\le \Pr\left(Z \le T \middle| \Phi_{free}\right) + \Pr\left(Z \le -\nu\right) = \mathrm{O}\left(\Gamma^{-2}\right), \end{aligned} \tag{3.50}$$

where we use the piece-wise linear approximation (3.42). After plugging (3.7), (3.48) and (3.50) back into (3.6), we have $\Pr\left(\mathbf{x} \to (-x_1, x_2)\right) = \mathrm{O}\left(\frac{\log\Gamma}{\Gamma^2}\right)$. Consequently, a diversity gain of 2 is achieved. □

3.5 More Discussions

In this section, we compare the aforementioned schemes in terms of relay power consumption and signalling overhead. We also briefly discuss the extension to higher-order modulations.

3.5.1 Relay Power Consumption Ratio

For the relay-side schemes, the relay power is adaptively scaled by the coefficient α. To compare the relay power consumption of different schemes, we define the relay power consumption ratio as $\bar{\alpha} = E(\alpha)$. Under this definition, the relay power consumption ratio of the receiver-side schemes is 1 as the relay node always sends message with full power (i.e., $\alpha = 1$). After some manipulations, we can show that

$$\begin{cases} \bar{\alpha}_{hard} \approx \Pr\left(\gamma_{sr,min} \geq \log \lambda_T \Gamma\right) = \exp\left(-\dfrac{\log \lambda_T \Gamma}{\lambda_{sr,min} \Gamma}\right) \overset{\Gamma \to \infty}{\approx} 1, & \text{(3.51a)} \\ \bar{\alpha}_{soft,local} = \dfrac{\lambda_{sr,min}}{\lambda_{rd}} \left(1 - e^{-\frac{\lambda_{rd}}{\lambda_{sr,min}}}\right), & \text{(3.51b)} \\ \bar{\alpha}_{soft,global} = \dfrac{\lambda_{sr,min}}{\lambda_{rd}} \log\left(\dfrac{\lambda_{rd}}{\lambda_{sr,min}} + 1\right). & \text{(3.51c)} \end{cases}$$

Clearly, for hard power scaling, the relay power consumption increases as the relay detection error probability reduces with SNR. On the contrary, the relay power consumption ratio is independent of SNR for soft power scaling, as it is adaptive to the relative quality of source-relay channel and relay-destination channel. When the source-relay channel is much better than the relay-destination channel, we have $\frac{\lambda_{sr,min}}{\lambda_{rd}} \to \infty$ and $\bar{\alpha}_{soft} \to 1$. On the contrary, if the source-relay channel is the bottleneck, we have $\frac{\lambda_{sr,min}}{\lambda_{rd}} \to 0$ and $\bar{\alpha}_{soft} \to 0$, in which case the network-coded

uplink reduces to the conventional TDMA without node cooperation.

3.5.2 Signalling Overhead

The CSI assumptions directly determine the signalling overhead of the whole system. For hard power scaling, as local CSI is exploited and $\alpha \in \{0, 1\}$, the relay node only needs to send 1 bit indicating ON or OFF to the destination. As for global-CSI based soft power scaling, the destination has to feed γ_{rd} back to the relay node, which then sends back the calculated power scaling coefficient α. So the signalling overhead depends largely on the quantization accuracy of α and γ_{rd}. The story is totally different for local-CSI based soft power scaling. Indeed, after the relay node estimates the source-relay channel coefficients, it can compute the power scaling coefficient α and effect it on the training sequence sent to the destination. After that, the destination can obtain the equivalent channel coefficient $\bar{h}_{rd}\sqrt{\alpha}$ that is needed for minimum distance combining. Consequently, there is no additional signalling overhead for local-CSI based soft power scaling, which is also the case for local-CSI based ML detection as the destination only exploits the average source-relay channel gain. Finally for global-CSI based link adaptive combining and global-CSI based ML detection, the relay node needs to report the source-relay channel SNR to the destination.

3.5.3 Extension To Higher-Order Modulations

Although we focus primarily on BPSK signals so far, the aforementioned schemes can also achieve full diversity for higher-order modulations. For soft power scaling and link adaptive combining, we observe that the power scaling coefficient α and combining weight w are independent of the underlying modulation scheme. Through some straightforward algebra, it is easy to show that our virtual channel model still fits for higher-order modulations, i.e., the quality of relay branch is approximately characterized by the worst channel inside. Therefore, full diversity can be achieved by following the same proof in the binary case.

As for hard power scaling, the decision rule (3.19) depends directly on the error probability at the relay node, which is hard to manipulate. Alternatively, we choose to extend the spirit of the threshold-based relaying (3.23) in a heuristic way. Recall that for hard power scaling, the relay node is actually striving to prevent the error propagation by setting a stringent SNR threshold, such that the conditional error rate when the relay node passes the threshold test scales like $\Pr\left(\Phi_{prop}\,|\Phi_{on}\right) = \mathrm{O}\left(\Gamma^{-2}\right)$. Intuitively, if the same scaling law is preserved for higher-order modulations, we can expect to achieve full diversity as well. As an example, we propose the following design for M-ary PSK signals.

Proposition 3.5. *For M-ary PSK signals, if we adopt the following decision rule for hard power scaling*

$$\Phi_{on} = \left\{\gamma_{sr,min} \geq g_{psk}^{-1} \log \Gamma\right\}, \tag{3.52}$$

where $g_{psk} = \sin^2\left(\frac{\pi}{M}\right)$*, then the conditional error rate is* $\Pr\left(\Phi_{prop}\,|\Phi_{on}\right) = \mathrm{O}\left(\Gamma^{-2}\right)$.

Proof. The conditional PDF of γ_{kr} given Φ_{on} is

$$f\left(\gamma_{kr}\right|\Phi_{on}) = \frac{f\left(\gamma_{kr}\right)}{\Pr\left(\gamma_{kr} \geq g_{psk}^{-1}\log\Gamma\right)} = \frac{1}{\lambda_{kr}\Gamma}\exp\left(-\frac{\gamma_{kr} - g_{psk}^{-1}\log\Gamma}{\lambda_{kr}\Gamma}\right) \tag{3.53}$$

for $\gamma_{kr} \geq g_{psk}^{-1}\log\Gamma$ and $k = 1, 2$. The conditional SER for M-ary PSK signal is [85]

$$\mathrm{P}_e\left(\gamma\right) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\exp\left(-\frac{g_{psk}\gamma}{\sin^2\theta}\right)d\theta. \tag{3.54}$$

Averaging the above probability over the conditional PDF of γ_{kr} leads to

$$\begin{aligned}
&\Pr\left(x_{r,k} \neq x_k\right|\Phi_{on}) \\
&= \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\int_{g_{psk}^{-1}\log\Gamma}^{\infty}\exp\left(-\frac{g_{psk}\gamma}{\sin^2\theta}\right)\frac{1}{\lambda_{kr}\Gamma}\exp\left(-\frac{\gamma - g_{psk}^{-1}\log\Gamma}{\lambda_{kr}\Gamma}\right)d\gamma d\theta \\
&= \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\frac{\sin^2\theta}{g_{psk}\lambda_{kr}\Gamma + \sin^2\theta}\exp\left(-\frac{\log\Gamma}{\sin^2\theta}\right)d\theta \\
&\leq \frac{1}{\pi g_{psk}\lambda_{kr}\Gamma}\int_0^{\frac{M-1}{M}\pi}\exp\left(-\frac{\log\Gamma}{\sin^2\theta}\right)d\theta \\
&\leq \frac{1}{\pi g_{psk}\lambda_{kr}\Gamma}\exp\left(-\log\Gamma\right)\frac{M-1}{M}\pi = \frac{M-1}{M g_{psk}\lambda_{kr}}\Gamma^{-2}.
\end{aligned} \tag{3.55}$$

Now we can conclude that $\Pr\left(\Phi_{prop}\right|\Phi_{on}) \leq \sum_{k\in\{1,2\}}\Pr\left(x_{r,k} \neq x_k\right|\Phi_{on}) = \mathrm{O}\left(\Gamma^{-2}\right)$.

□

3.6 Simulations

In this section, we present simulation results to validate our diversity analysis. In simulation results, SPS, HPS, LAC and MLD are short for soft power scaling, hard power scaling, link adaptive combining and ML detection, respectively. Throughout simulations, we use the path loss model $\lambda = D^{-3}$, where λ is the channel gain and D is the distance between two terminals. Pair error probability is used as the performance metric, i.e., the probability that at least one of the source symbols is

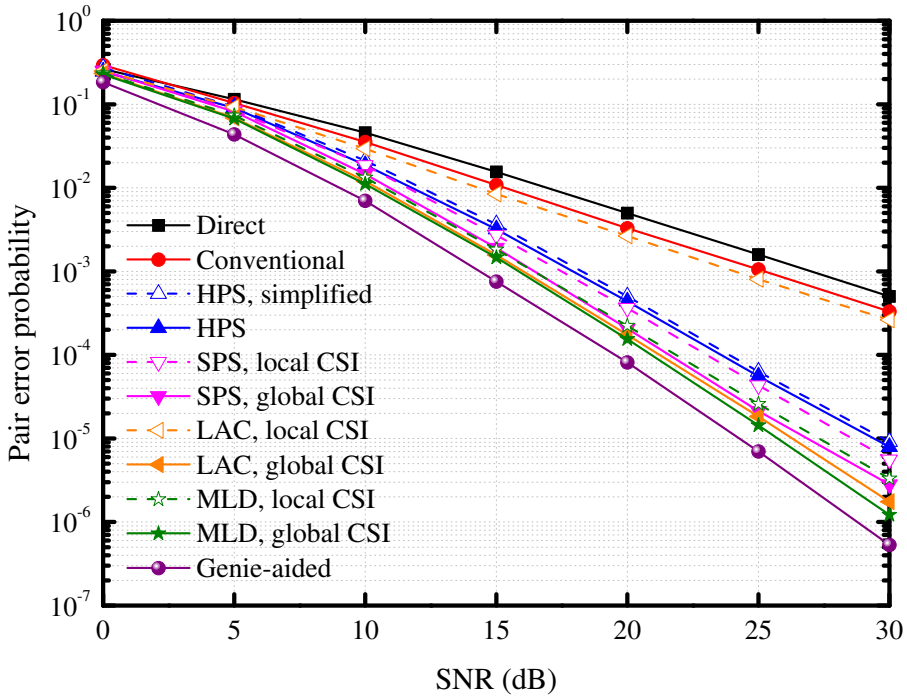


Figure 3.3: Error performances of BPSK signal in a symmetric network.

detected incorrectly at the destination. We also simulate the genie-aided relaying and the conventional transmission protocol that uses full power relaying and equal weight combining as the baseline schemes.

Figure 3.3 shows the error performances in a symmetric network, where the distance between any two nodes is normalized. We observe that local-CSI based link adaptive combining only achieves a diversity gain of 1 as direct transmission and conventional scheme, while all the other schemes achieve a diversity gain of 2. The genie-aided relaying is the benchmark for all the practical schemes, thus having the best error performances. It is also observed that the performance of simplified hard power scaling (3.23) is very close to that based on the exact decision rule (3.19) at all SNRs. Local-CSI based soft power scaling is slightly better than hard power scaling.

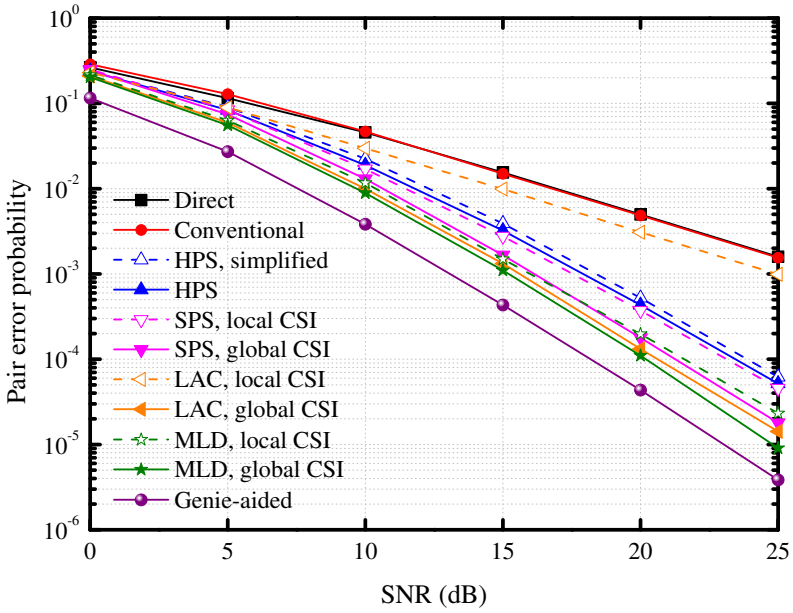


Figure 3.4: Error performances of BPSK signal in an asymmetric network with strong relay-destination channel.

The performances of three global-CSI based methods are very close. Comparatively, ML detection is the best scheme among all, but it performs nearly the same as link adaptive combining which enjoys lower detection complexity.

Then in Figure 3.4 and Figure 3.5 we present the error performances for two asymmetric networks. For the network with strong relay-destination channel and with strong source-relay channel, we set $D_{rd} = 0.4$ and $D_{sr} = 0.4$ respectively while normalizing the other distances. In the former scenario, the source-relay channel is the system bottleneck. As the relay decoding is unreliable, the conventional scheme performs almost the same as direct transmission. Besides, the performance gap between genie-aided relaying and all other schemes expands compared to the symmetric scenario, which reflects the importance of preventing error propagation. As

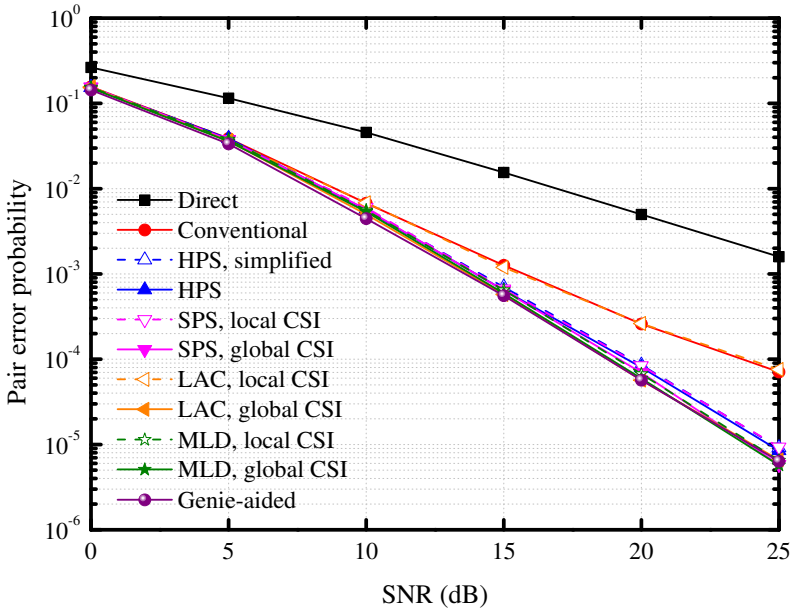


Figure 3.5: Error performances of BPSK signal in an asymmetric network with strong source-relay channel.

for the network where the relay-destination channel is worse, the error propagation issue is comparatively mitigated. We observe in Figure 3.5 that the conventional scheme and local-CSI based link adaptive combining now have huge coding gain against direct transmission; however, the diversity gain is still 1. For all the remaining schemes, the performances are almost the same, and a diversity gain of 2 is achieved.

Next we investigate the error performances with different relay positions. For the network topology, we place the destination at $(0, 0)$, and locate the two source nodes at $(\frac{\sqrt{3}}{2}, \pm\frac{1}{2})$, respectively. The relay node shall move along the x-axis from $(0.2, 0)$ to $(2, 0)$. The error performance is shown in Figure 3.6. It is observed that for all the schemes, the best performance is attained when the relay node is

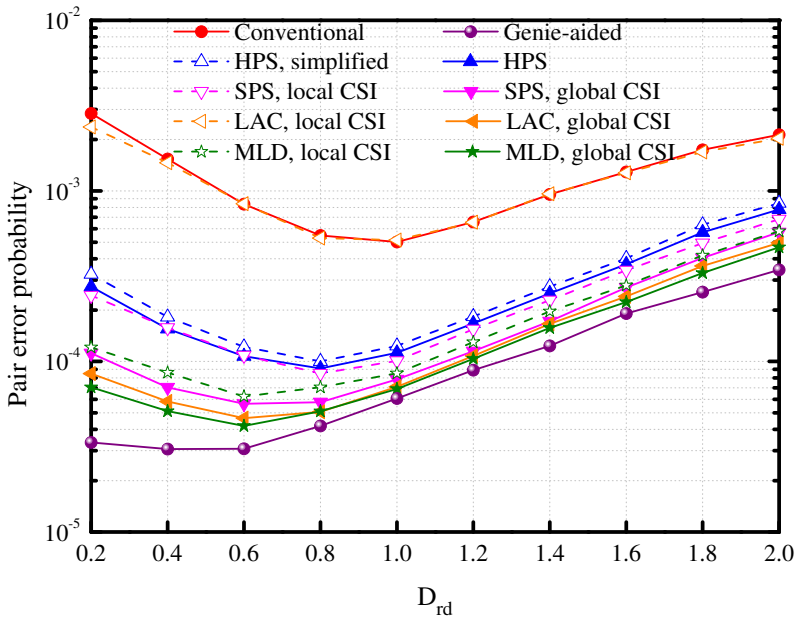


Figure 3.6: Error performances of BPSK signal with $\Gamma = 20$dB and different relay positions.

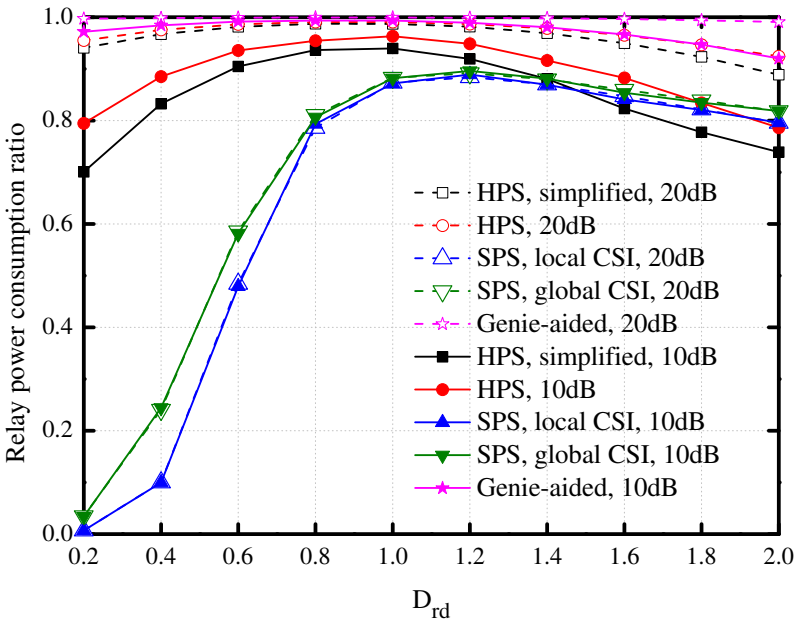


Figure 3.7: Relay power consumption ratio with different relay positions.

close to the source nodes, as the relay detection error dominates the overall system performance. In all cases, global-CSI based schemes perform much better than their local-CSI based counterparts at a price of higher signalling overhead.

For the same network, we also plot the relay power consumption ratio in Figure 3.7. The simulation results are consistent with our analysis, i.e., the relay power consumption of hard power scaling increases with SNR, while for soft power scaling it is independent of SNR. For soft power scaling, we observe that the power consumption is really low when the relay is close to the destination, since the source-relay link is comparatively unreliable; as the relay node moves far away from the destination, the relay node gradually increases its power until the relay-destination channel becomes the bottleneck. For hard power scaling and genie-aided relaying, the relay power consumption maximizes when the relay node is close to the source, in which case the relay decoding is really reliable and the chance of forwarding the message is large. We also observe that soft power scaling is much more power efficient than hard power scaling in most cases. This is because for hard power scaling, the relay node is always very conservative in forwarding the message so as to keep the conditional error rate low. Note that although the relay node always uses full power in the receiver-side schemes, better performances are also achieved compared to the relay-side schemes.

Finally, we study the error performances using higher-order modulations in Figure 3.8 and Figure 3.9. Clearly, a diversity gain of 2 is achieved by all the schemes except local-CSI based link adaptive combining, which justifies our analysis.

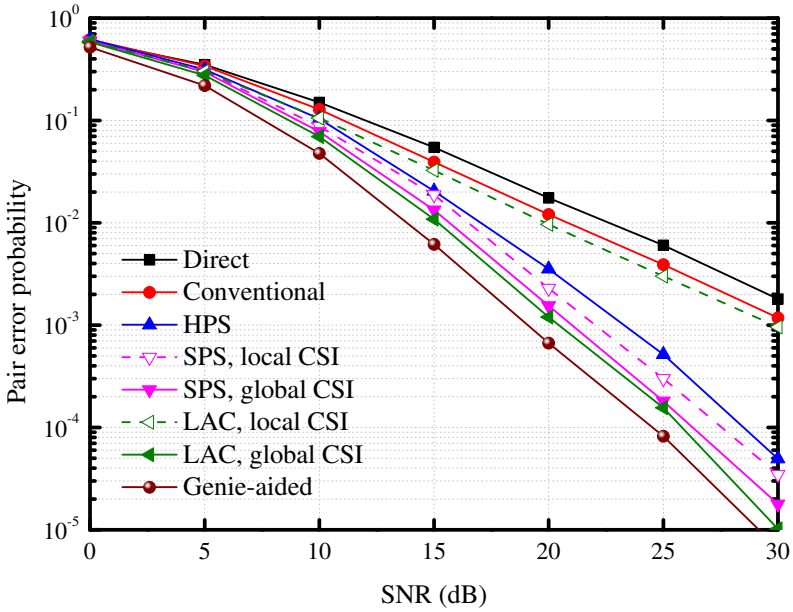


Figure 3.8: Error performances of QPSK signal in a symmetric network.

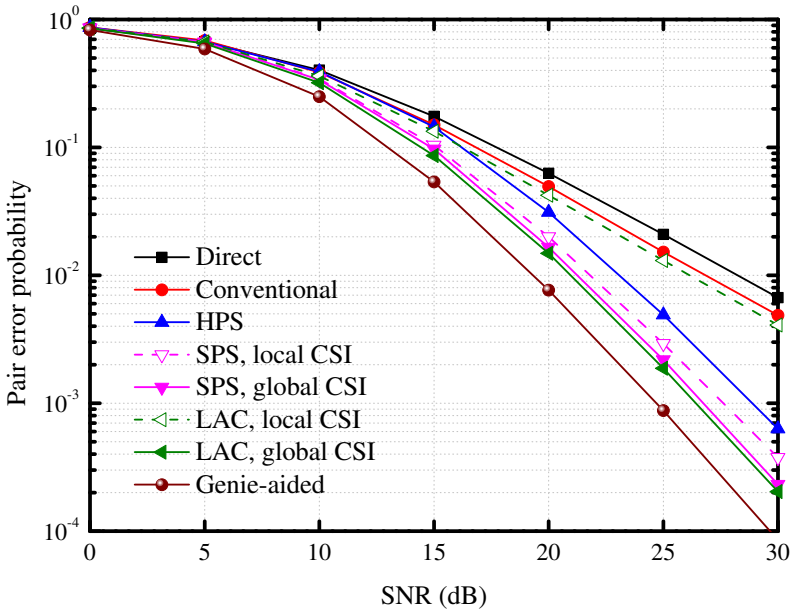


Figure 3.9: Error performances of 8PSK signal in a symmetric network.

3.7 Conclusions

In this work, we proposed two power scaling schemes at the relay side and two detection schemes at the receiver side, respectively, that can mitigate error propagation and thus achieve full diversity for the wireless network-coded uplink. We showed that the receiver-side schemes generally has better error performances, whereas the relay-side schemes are more power efficient. We also demonstrated that there is a basic tradeoff between the error performance and signalling overhead to acquire CSI. We remark that the error propagation issue is addressed either at the relay side or at the receiver side in this work to achieve full diversity. One interesting issue for possible future consideration is how to jointly optimize the relaying scheme and detection scheme so as to improve the coding gain.

Chapter 4

Diversity Analysis of Wireless Uplink with Analog Network Coding

Different from DNC that suffers error propagation issue, ANC is naturally immune to relay detection error. So ANC is widely accepted as a good substitute for DNC that could always achieve full diversity gain. However, there are also disadvantages associated with ANC. On one hand, as the relay does not make a hard decision on the source message, the noise component could not be perfectly removed from the received signal, and such noise would be amplified and forwarded to the intended received along with the desired messages. On the other hand, because different source messages are directly combined in the complex field, it incurs co-channel interference. If the receiver is unable to eliminate the interference, MUD has to be used to separate different source messages. Many literatures [21, 25, 28, 74] have studied the performance of ANC for TWRC, where self-interference could be elegantly removed and the impact of MUI is largely neglected. Very limited literatures [57] have ever studied ANC in the presence of MUI. However, only the special case where the relay knows the instant CSI is discussed, and the story on the other side is still unknown.

So in this work, we provide a comprehensive study on how MUI would impact the diversity gain of ANC. Both VGR and FGR that require instant/statistical CSI at the relay nodes are properly discussed. We first study the error performance of

single-relay uplink and show that MUI would incur diversity loss at modest SNRs. Then for multi-relay networks, we develop a Min-Max relay selection strategy and prove that full diversity can be achieved. Next we investigate the diversity performance of two DSTC schemes. For DSTBC, we show that full diversity can be achieved by FGR but not VGR, and we propose a selective DSTBC-VGR scheme to recover the diversity loss by adaptively allocating the relay power when there is only one user. Finally for DDSTC, we show that both FGR and VGR can achieve full diversity, and the optimum code design criterion is to maximize the minimum product distance.

Notations: $|\cdot|$, $(\cdot)^T$ and $(\cdot)^H$ stand for absolute value, transpose and conjugate transpose, respectively. The boldface lowercase letter $\mathbf{a}$ and the boldface uppercase letter $\mathbf{A}$ represent vector in column form and matrix, respectively. $\|\mathbf{a}\|$ and $\det \mathbf{A}$ denote the Euclidean norm of a vector $\mathbf{a}$ and the determinant of a square matrix $\mathbf{A}$, respectively. $\mathbb{Z}$ and $\mathbb{C}$ stand for the set of integers and the set of complex numbers, respectively. We shall use the abbreviation i.i.d. for independent and identically distributed, and denote $Z \sim \mathcal{CN}(\mu, \sigma^2)$ as a circularly symmetric complex Gaussian random variable. The probability of an event $\mathcal{A}$ is denoted by $\Pr(\mathcal{A})$. The cumulative distribution function (CDF) and PDF of a random variable Z are denoted by $F_Z(z)$ and $f_Z(z)$, respectively. We define the Q-function as $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$. Finally, we say $h(x) = \mathrm{O}(g(x))$ if $a \leq \lim_{x\to\infty} \frac{h(x)}{g(x)} \leq b$ for some positive constants a and b.

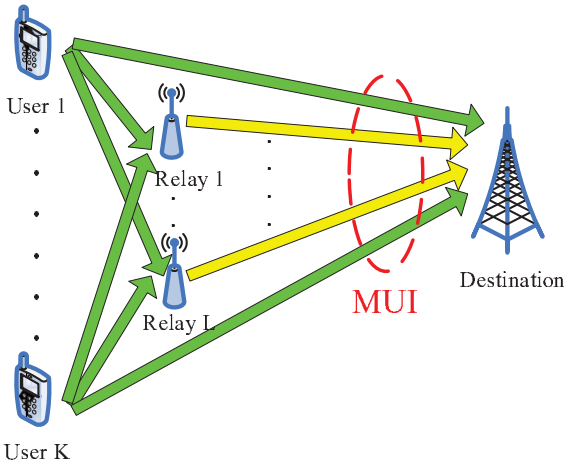


Figure 4.1: System model of a multi-user multi-relay uplink channel.

4.1 Multi-User Single-Relay Systems

In this section, we first study the single-relay uplink channel. The analytical results obtained here will be used repeatedly in later sections when we consider the multi-relay network.

4.1.1 System Model

Consider a uplink channel where K users send data to a single destination with the help of a single relay node, as shown in Figure 4.1 with $L = 1$. Let $f_k \sim \mathcal{CN}(0, 1)$ be the channel coefficient from the kth user to the relay, $h_k \sim \mathcal{CN}(0, 1)$ be the channel coefficient from the kth user to the destination, and $g \sim \mathcal{CN}(0, 1)$ be the channel coefficient from the relay to the destination, respectively. All the channel coefficients are independent, and the additive noises on different channels are also independently distributed as $\mathcal{CN}(0, 1)$. Without loss of generality, throughout this work we focus only on the symmetric networks, where all the user-relay channels

have the same path-loss coefficient λ_{sr}, and all the user-destination channels have the same path-loss coefficient λ_{sd}. The path-loss coefficient of the relay-destination channel is denoted by λ_{rd}. We remark that such assumption is just to simplify the notations, and our analysis can be easily extended to any asymmetric networks. As will be seen later, these path-loss coefficients are only related to the coding gain but have nothing to do with the diversity gain, which is the main concern of this work.

Due to half-duplex constraint, the whole data transmission is completed in two phases. In the first phase, all the users broadcast their data simultaneously, and the received signal at the relay and destination can be respectively represented as[1]

$$\begin{cases} y_{sr} = \sqrt{P\lambda_{sr}} \sum_{k=1}^{K} f_k s_k + n_{sr}, & \text{(4.1a)} \\ y_{sd} = \sqrt{P\lambda_{sd}} \sum_{k=1}^{K} h_k s_k + n_{sd}. & \text{(4.1b)} \end{cases}$$

Here P is the transmitted power, n_{sd} and n_{sr} are the additive noises, and s_k is the transmitted symbol of the kth user, which is picked from some constellation Ω with normalized power, i.e., $E|s_k|^2 = 1$. The transmitted signal of the relay node is $x_r = \sqrt{\alpha P} y_{sr}$, where α is the amplification factor to normalize the relay power. In this work, we consider two different ways to normalize the relay power. For VGR, the amplification factor is chosen in such a way that the relay power is limited to P at any time instant, i.e., $E\left(|x_r|^2 \middle| \mathbf{f}\right) = P$ with $\mathbf{f} = (f_1, f_2, \cdots, f_K)^T$. This requires the relay node to adjust the amplification factor according to the real-time channel

[1]In this work, we assume all the transmitters are perfectly synchronized. The effect of synchronization errors is beyond the scope of this work.

conditions, thus α_{VGR} keeps changing all the time and is given by

$$\alpha_{VGR} = \frac{1}{P\lambda_{sr}\sum_{k=1}^{K}|f_k|^2 + 1}. \tag{4.2}$$

Alternatively, the relay node could also use a constant amplification factor such that the average relay power is normalized to P in the long run, i.e., $E|x_r|^2 = P$, which is referred to as FGR. The resulting amplification factor α_{FGR} is given by

$$\alpha_{FGR} = \frac{1}{KP\lambda_{sr} + 1}. \tag{4.3}$$

Note that in this scheme, the amplification factor α_{FGR} is a constant depending only on the second-order statistics of channel distributions, so the relay node needs not to know the instantaneous channel conditions. It should also be pointed out that the relay power of FGR may momentarily exceed the maximum load of the power amplifier. However, such power saturation issue is not considered in this work to simplify the analysis.

After proper power scaling, the relay node then forwards the amplified signal to the destination in the second phase. The received signal is

$$y_{rd} = \sqrt{\lambda_{rd}}gx_r + n_{rd} = \sqrt{\alpha P^2\lambda_{sr}\lambda_{rd}}g\sum_{k=1}^{K}f_k s_k + \tilde{n}_{rd}, \tag{4.4}$$

where $\tilde{n}_{rd} \triangleq \sqrt{\alpha P\lambda_{rd}}gn_{sr} + n_{rd} \sim \mathcal{CN}\left(0, \alpha P\lambda_{rd}|g|^2 + 1\right)$ is the equivalent additive noise. Upon observing the signals y_{sd} and y_{rd}, the destination performs ML detection to jointly detect the K user symbols as

$$\mathbf{s}_d = \arg\min_{\hat{s}_k\in\Omega}\left|y_{sd} - \sqrt{P\lambda_{sd}}\sum_{k=1}^{K}h_k\hat{s}_k\right|^2 + \frac{\left|y_{rd} - \sqrt{\alpha P^2\lambda_{sr}\lambda_{rd}}g\sum_{k=1}^{K}f_k\hat{s}_k\right|^2}{\alpha P\lambda_{rd}|g|^2 + 1}, \tag{4.5}$$

where $\mathbf{s}_d = (s_{d,1}, s_{d,2}, \cdots, s_{d,K})^T$ is the decoded symbol vector, and different channel noises are assumed to be independent. Note that as all the source symbols reach the destination through 2 independent paths (i.e., one through direct link and the other through relay branch), the maximum diversity gain is equal to 2.

4.1.2 Performance Analysis

In this subsection, we study the PEPs of the aforementioned system. PEP is defined as the probability that a transmitted symbol $\mathbf{x}$ is mistaken by a different symbol $\hat{\mathbf{x}}$, which provides a tight bound on the error rates. According to (4.5), the PEP of mistaking $\mathbf{s}$ by $\hat{\mathbf{s}}$ is given by

$$\Pr(\mathbf{s} \to \hat{\mathbf{s}}) = E\left[Q\left(\sqrt{2(W_d + W_r)}\right)\right] \leq E\left[\exp\left(-(W_d + W_r)\right)\right], \tag{4.6}$$

where $W_d = \frac{1}{4}P\lambda_{sd}\left|\mathbf{h}^T\Delta\mathbf{s}\right|^2$, $W_r = \frac{\alpha P^2\lambda_{sr}\lambda_{rd}|g|^2\left|\mathbf{f}^T\Delta\mathbf{s}\right|^2}{4\left(\alpha P\lambda_{rd}|g|^2+1\right)}$, $\mathbf{h} = (h_1, h_2, \cdots, h_K)^T$, $\Delta\mathbf{s} = \mathbf{s} - \hat{\mathbf{s}}$, and we have applied Chernoff bound [1] in the inequality. As W_d follows the exponential distribution, we have

$$E\left[\exp\left(-W_d\right)\right] = \frac{4}{P\lambda_{sd}\|\Delta\mathbf{s}\|^2 + 4} \overset{P\to\infty}{\approx} \frac{4}{\lambda_{sd}\|\Delta\mathbf{s}\|^2}P^{-1}. \tag{4.7}$$

To evaluate the expectation of the second term in (4.6), we first prove the following resutls.

Proposition 4.1. *Let* $W_1 = \frac{abXY}{aX+bY+cZ+1}$ *and* $W_2 = \frac{abXY}{aX+c}$, *where* a, b *and* c *are constants,* X *and* Y *are independent exponential random variables with unit mean, and* Z *is independent of* X *and* Y *and has the gamma distribution* $f(z) =$

$\frac{1}{\Gamma(N)} z^{N-1} e^{-z}$, *then for* $w \geq 0$ *the CDF of* W_1 *and* W_2 *are respectively given by*

$$F_{W_1}(w) \leq 1 - \exp\left(-\frac{a+b}{ab}w\right) 2\sqrt{\frac{w+w^2}{ab}} K_1\left(2\sqrt{\frac{w+w^2}{ab}}\right) + \frac{Ncw}{ab} \exp\left(-\frac{a+b-c}{ab}w\right) E_1\left(\frac{cw}{ab}\right) \overset{w \ll 1}{\approx} \frac{Nc+1}{ab} w \log\frac{1}{w}, \quad (4.8)$$

$$F_{W_2}(w) = 1 - \exp\left(-\frac{w}{b}\right) \sqrt{\frac{4wc}{ab}} K_1\left(\sqrt{\frac{4wc}{ab}}\right) \overset{w \ll 1}{\approx} \frac{c}{ab} w \log\frac{1}{w}, \quad (4.9)$$

where $E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt$ *is the exponential integral function [87, 5.1.1], and* $K_1(x)$ *is the first-order modified Bessel function of the second kind [87, 9.6.1].*

Proof. For W_2, it is easy to show that

$$F_{W_2}(w) = \Pr(bY - w < 0) + \Pr\left(X \leq \frac{wc}{(bY-w)a}, bY - w > 0\right). \quad (4.10)$$

The first term is given by $\Pr(bY - w < 0) = 1 - \exp\left(-\frac{w}{b}\right)$. For the second term, we have

$$\Pr\left(X \leq \frac{wc}{(bY-w)a}, bY - w > 0\right) = \exp\left(-\frac{w}{b}\right) - \exp\left(-\frac{w}{b}\right) \sqrt{\frac{4wc}{ab}} K_1\left(\sqrt{\frac{4wc}{ab}}\right), \quad (4.11)$$

where $K_1(x)$ is the first-order modified Bessel function of the second kind [87, 9.6.1], and we use [88, 3.478.4] in the equality. Combining these two terms leads to the first part of (4.9), which is consistent with a previous result derived using different algebra [89]. Next we examine the asymptotic behaviors of $F_{W_2}(w)$ when $w \ll 1$. Using [87, 9.6.11], we have $K_1(z) \overset{z \ll 1}{\approx} z^{-1} + \log\left(\frac{1}{2}z\right) I_1(z)$, where $I_1(z)$ is the first-order modified Bessel function of the first kind [87, 9.6.1], which can be further approximated as $I_1(z) \overset{z \ll 1}{\approx} \frac{z}{2}$ [87, 9.6.7]. Therefore we have

$$zK_1(z) \overset{z \ll 1}{\approx} 1 + \frac{1}{2}z^2 \log z. \quad (4.12)$$

Using the above approximation, we can obtain

$$F_{W_2}(w) \overset{w\ll 1}{\approx} 1-\left(1-\frac{w}{b}\right)\left(1+\frac{2wc}{ab}\log\left(\sqrt{\frac{4wc}{ab}}\right)\right) \overset{w\ll 1}{\approx} \frac{c}{ab}w\log\frac{1}{w}. \tag{4.13}$$

Next we study $F_{W_1}(w)$. Let $T(w)=abXY-awX-bwY$, then

$$F_{W_1}(w)=\Pr\left(T(w)\le w\right)+\Pr\left(w\le T(w)\le cwZ+w\right). \tag{4.14}$$

For any $t\ge 0$, we have

$$F_T(t)=\Pr\left(aX-w\le 0\right)+\Pr\left(Y\le\frac{t+awX}{(aX-w)b},aX-w\ge 0\right). \tag{4.15}$$

The first term is given by $\Pr\left(aX-w\le 0\right)=1-\exp\left(-\frac{w}{a}\right)$. For the second term, we have

$$\begin{aligned}&\Pr\left(Y\le\frac{t+awX}{(aX-w)b},aX-w\ge 0\right)\\&\quad=\exp\left(-\frac{w}{a}\right)-\exp\left(-\frac{a+b}{ab}w\right)2\sqrt{\frac{t+w^2}{ab}}K_1\left(2\sqrt{\frac{t+w^2}{ab}}\right),\end{aligned} \tag{4.16}$$

where we use [88, 3.478.4] again. Consequently,

$$F_T(t)=1-\exp\left(-\frac{a+b}{ab}w\right)2\sqrt{\frac{t+w^2}{ab}}K_1\left(2\sqrt{\frac{t+w^2}{ab}}\right),\ t\ge 0. \tag{4.17}$$

Using the relation $K_\nu'(x)=-K_{\nu-1}(x)-\frac{\nu}{x}K_\nu(x)$ [87, 9.6.26], we have

$$\left(xK_1(x)\right)'=K_1(x)+xK_1'(x)=K_1(x)+x\left(-K_0(x)-\frac{1}{x}K_1(x)\right)=-xK_0(x), \tag{4.18}$$

where $K_0(x)$ is the zeroth-order modified Bessel function of the second kind [87, 9.6.1]. Thus

$$f_T(t)=\frac{2}{ab}\exp\left(-\frac{a+b}{ab}w\right)K_0\left(2\sqrt{\frac{t+w^2}{ab}}\right),\ t\ge 0. \tag{4.19}$$

Note that

$$1 - F_Z(z) = \frac{1}{\Gamma(N)} \int_z^\infty t^{N-1} e^{-t} dt = \frac{\Gamma(N,z)}{\Gamma(N)} = e^{-z} \sum_{k=0}^{N-1} \frac{z^k}{k!}, \tag{4.20}$$

where $\Gamma(s,x)$ is the upper incomplete gamma function [87, 6.5.3], we can obtain

$$\begin{aligned} &\Pr(w \le T(w) \le cwZ + w) \\ &\quad \le \frac{2cw}{ab} \exp\left(-\frac{a+b}{ab} w\right) \sum_{k=0}^{N-1} \frac{1}{k!} \int_0^\infty t^k e^{-t} K_0\left(2\sqrt{\frac{cw}{ab}}\sqrt{t}\right) dt, \end{aligned} \tag{4.21}$$

where we use the fact that $K_0(x)$ is a decreasing function in the inequality. To evaluate the integral within the summation, we use [88, 6.643.3] and obtain

$$\int_0^\infty t^k e^{-t} K_0\left(2\sqrt{\frac{cw}{ab}}\sqrt{t}\right) dt = \frac{\Gamma^2(k+1)}{2\sqrt{\frac{cw}{ab}}} \exp\left(\frac{cw}{2ab}\right) W_{-(k+\frac{1}{2}),0}\left(\frac{cw}{ab}\right). \tag{4.22}$$

Here $W_{\kappa,\mu}(z) = e^{-\frac{1}{2}z} z^{\frac{1}{2}+\mu} U\left(\frac{1}{2}+\mu-\kappa, 1+2\mu, z\right)$ is the Whittaker's function [87, 13.1.33], and $U(a,b,z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1} dt$ is the Kummer's function [87, 13.2.5]. Now we have

$$\begin{aligned} &\int_0^\infty t^k e^{-t} K_0\left(2\sqrt{\frac{cw}{ab}}\sqrt{t}\right) dt = \frac{1}{2}\Gamma(k+1) \int_0^\infty e^{-\frac{cw}{ab}t} \frac{t^k}{(1+t)^{k+1}} dt \\ &\quad \le \frac{1}{2}\Gamma(k+1) \int_0^\infty \frac{\exp\left(-\frac{cw}{ab}t\right)}{1+t} dt = \frac{1}{2}\Gamma(k+1) \exp\left(\frac{cw}{ab}\right) E_1\left(\frac{cw}{ab}\right), \end{aligned} \tag{4.23}$$

where $E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt$ is the exponential integral function [87, 5.1.1]. Finally we have

$$\Pr(w \le T(w) \le cwZ + w) \le \frac{Ncw}{ab} \exp\left(-\frac{a+b-c}{ab} w\right) E_1\left(\frac{cw}{ab}\right). \tag{4.24}$$

Plugging (4.17) and (4.24) back into (4.14) leads to the first part of (4.8). Using (4.12), we have

$$F_T(w) \overset{w \ll 1}{\approx} 1 - \left(1 - \frac{a+b}{ab} w\right)\left(1 + \frac{w}{ab} \log\left(\frac{4w}{ab}\right)\right) \overset{w \ll 1}{\approx} \frac{1}{ab} w \log\frac{1}{w}. \tag{4.25}$$

Using the inequality $e^x E_1(x) \leq \log\left(1+\frac{1}{x}\right)$ [87, 5.1.20], we have

$$\Pr\left(w \leq T(w) \leq cwZ + w\right) \leq \frac{Ncw}{ab}\exp\left(-\frac{a+b}{ab}w\right)\log\left(1+\frac{ab}{cw}\right) \overset{w\ll 1}{\approx} \frac{Nc}{ab}w\log\frac{1}{w}. \tag{4.26}$$

Plugging (4.25) and (4.26) back into (4.14) leads to the second part of (4.8). □

Now we proceed to study the PEPs of FGR and VGR, respectively. For FGR, after plugging (4.3) in W_r, we have $W_{r,FGR} = \frac{P^2\lambda_{sr}\lambda_{rd}|g|^2\left|\mathbf{f}^T\Delta\mathbf{s}\right|^2}{4\left(P\lambda_{rd}|g|^2+KP\lambda_{sr}+1\right)}$. According to Proposition 4.1, the CDF of $W_{r,FGR}$ can be obtained by substituting $a_1 = P\lambda_{rd}$, $b_1 = \frac{1}{4}P\lambda_{sr}\|\Delta\mathbf{s}\|^2$, and $c_1 = KP\lambda_{sr}+1$ in (4.9), and we have

$$\begin{aligned} E\left[\exp\left(-W_{r,FGR}\right)\right] &\overset{(a)}{=} \frac{1}{b_1+1} + \frac{b_1c_1}{a_1(b_1+1)^2}\exp\left(\frac{c_1}{a_1(b_1+1)}\right)E_1\left(\frac{c_1}{a_1(b_1+1)}\right) \\ &\overset{(b)}{\leq} \frac{1}{b_1+1} + \frac{b_1c_1}{a_1(b_1+1)^2}\log\left(1+\frac{a_1(b_1+1)}{c_1}\right) \\ &\overset{P\to\infty}{\approx} \frac{4K}{\lambda_{rd}\|\Delta\mathbf{s}\|^2}\frac{\log P}{P}, \end{aligned} \tag{4.27}$$

where we use [88, 6.643.3], [87, 13.1.33] and [87, 13.2.5] in (a), and the inequality $E_1(z) < e^{-z}\log\left(1+\frac{1}{z}\right)$ [87, 5.1.20] in (b). Combining (4.7) with (4.27) leads to

$$\Pr\left(\mathbf{s}\to\hat{\mathbf{s}}|\,\mathrm{FGR}\right) \overset{P\to\infty}{\approx} \frac{16K}{\lambda_{rd}\lambda_{sd}\|\Delta\mathbf{s}\|^4}\frac{\log P}{P^2} \leq \frac{16K}{\lambda_{rd}\lambda_{sd}d_{\min}^4}\frac{\log P}{P^2}, \tag{4.28}$$

where $d_{\min} = \min\limits_{s,\hat{s}\in\Omega,s\neq\hat{s}}|s-\hat{s}|$ is the minimum distance of any two distinct points in the set Ω.

Next we study the PEPs of VGR. After plugging (4.2) in W_r, we have $W_{r,VGR} = \frac{P^2\lambda_{sr}\lambda_{rd}|g|^2\left|\mathbf{f}^T\Delta\mathbf{s}\right|^2}{4\left(P\lambda_{rd}|g|^2+P\lambda_{sr}\sum\limits_{k=1}^{K}|f_k|^2+1\right)}$. For any error vector $\Delta\mathbf{s}\neq 0$, we can always find a $K\times K$ unitary matrix $\mathbf{U}$ with the first row being $\frac{\Delta\mathbf{s}^T}{\|\Delta\mathbf{s}\|}$. Define a new vector $\tilde{\mathbf{f}} = \mathbf{U}\mathbf{f}$, then we have $W_{r,VGR} = \frac{\|\Delta\mathbf{s}\|^2}{4}\tilde{W}_{r,VGR}$ with $\tilde{W}_{r,VGR} = \frac{P^2\lambda_{sr}\lambda_{rd}|g|^2\left|\tilde{f}_1\right|^2}{P\lambda_{rd}|g|^2+P\lambda_{sr}\left|\tilde{f}_1\right|^2+P\lambda_{sr}\sum\limits_{k=2}^{K}\left|\tilde{f}_k\right|^2+1}$. As

$\tilde{\mathbf{f}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, $\left|\tilde{f}_1\right|^2$ is independent of $\sum_{k=2}^{K} \left|\tilde{f}_k\right|^2$, which has the gamma distribution. According to *Lemma 1*, the CDF of $\tilde{W}_{r,VGR}$ can be obtained after plugging $a_2 = P\lambda_{rd}$, $b_2 = c_2 = P\lambda_{sr}$ and $N = K-1$ in (4.8), and we have

$$\begin{aligned} E\left[\exp\left(-W_{r,VGR}\right)\right] &\overset{(a)}{\leq} 1 - \frac{\|\Delta\mathbf{s}\|^2}{4} \exp\left(-\frac{1}{\sqrt{a_2 b_2}}\right) \left(\frac{\|\Delta\mathbf{s}\|^2}{4} + \frac{\left(\sqrt{a_2}+\sqrt{b_2}\right)^2}{a_2 b_2}\right)^{-1} \\ &+ \frac{(K-1)\, a_2 \|\Delta\mathbf{s}\|^2}{4} \left(\frac{16}{\|\Delta\mathbf{s}\|^4 a_2^2} \log\left(1 + \frac{\|\Delta\mathbf{s}\|^2 a_2}{4}\right) - \frac{1}{\frac{\|\Delta\mathbf{s}\|^2 a_2}{4}\left(1 + \frac{\|\Delta\mathbf{s}\|^2 a_2}{4}\right)}\right) \\ &\overset{P\to\infty}{\approx} \frac{4\,(K-1)}{\|\Delta\mathbf{s}\|^2 \lambda_{rd}} \frac{\log P}{P}, \end{aligned} \tag{4.29}$$

where we use [88, 6.227.1] and the inequality $xK_1(x) \geq \exp(-x)$ [57] in (a). Combining (4.7) with (4.29) leads to

$$\Pr\left(\mathbf{s} \to \hat{\mathbf{s}} | \text{ VGR}\right) \overset{P\to\infty}{\approx} \frac{16\,(K-1)}{\lambda_{rd}\lambda_{sd}\|\Delta\mathbf{s}\|^4} \frac{\log P}{P^2} \leq \frac{16\,(K-1)}{\lambda_{rd}\lambda_{sd} d_{\min}^4} \frac{\log P}{P^2}. \tag{4.30}$$

4.1.3 Discussions

It is observed that the error rate of ANC-FGR and ANC-VGR has a scaling law of $\mathrm{O}\left(\frac{\log P}{P^2}\right)$ at high SNRs. According to the definition of diversity gain, we have

$$d = -\lim_{P\to\infty} \frac{\log \Pr\left(\mathbf{s}_d \neq \mathbf{s}\right)}{\log P} = 2 - \lim_{P\to\infty} \frac{\log\log P}{\log P} = 2.$$

Consequently, we conclude that ANC is able to achieve full dominant diversity gain even in the presence of MUI. However, we observe that there is also another logarithmic-term (i.e., $\log P$) in the numerator of the error rate expression. Although its impact on the diversity gain vanishes at extremely high SNRs (i.e., $\lim_{P\to\infty} \frac{\log\log P}{\log P} = 0$), this $\log P$ term would introduce some diversity loss at modest

SNRs. For example, when $P \leq 30$dB we have $\frac{\log\log P}{\log P} \geq 0.28$. As a result, ANC cannot achieve the same diversity gain as the conventional diversity schemes discussed in Section 1.1 that do not suffer such log-term loss, but the diversity loss is very limited and only occurs at modest SNRs.

Recall that for the conventional single-user analog relaying, the error rate scales as $\mathrm{O}\left(\frac{1}{P^2}\right)$ [12, 13] and $\mathrm{O}\left(\frac{\log P}{P^2}\right)$ [17, 90] for VGR and FGR, respectively. So such logarithmic-term loss could be avoided if assigning orthogonal relaying channels to different users, and the relay must normalize its instant power. To explain this phenomenon, let us revisit the relay signal model (4.4), where the signal component of the kth user is given by $\sqrt{\alpha P^2 \lambda_{sr}\lambda_{rd}} g f_k s_k$. For the single-user VGR, the amplification factor is given by $\alpha_{VGR} = \left(P\lambda_{sr}|f_k|^2 + 1\right)^{-1} \overset{P\to\infty}{\approx} \frac{1}{P\lambda_{sr}|f_k|^2}$, where the approximation holds with probability 1. Consequently, the signal component becomes $h_{eq,k}s_k$ at high SNRs, and the equivalent channel $h_{eq,k} \overset{\Delta}{=} \sqrt{P\lambda_{rd}} g e^{j\varphi(f_k)}$ still follows Rayleigh fading, where $\varphi(f_k)$ is the phase of f_k. On the contrary, for ANC and single-user FGR the effective channel is proportional to $f_k g$, which follows double-Rayleigh fading [91] and it introduces the logarithmic-term in the error rate expression.

From (4.28) and (4.30), we also observe that increasing the user number K beyond 2 would not degrade the diversity gain further. However, since the dominant PEPs are proportional to K, there is some linear loss in coding gain as the number of users increases. Relatively speaking, VGR performs slightly better than FGR. This is because the instantaneous output power at the relay node is always normalized to P in VGR, which helps to mitigate the extent of channel fading. Finally,

it is also noteworthy that the dominant PEPs are inversely proportional to λ_{rd} but is independent of λ_{sr}. This is because the received signal power of y_{sr} is approximately proportional to λ_{sr}, so the path-loss effects of source-relay channels would be counteracted during power normalization at the relay node. Consequently, the quality of relay-destination channel dominates the overall error performance, and the best relay position should be closer to the destination.

4.2 Relay Selection Strategy

From now on, we consider the more general multi-relay networks with L relays. Given that there are totally $L+1$ independent diversity paths from each user to the destination, the design objective is to achieve full diversity gain $L+1$. To this end, we develop a relay selection strategy based on the Min-Max criterion in this section. More sophisticated DSTC schemes will be discussed in later sections.

The system model is a natural extension of the single-relay model discussed in the last section. Suppose there are now L parallel relays. Let f_{kl} be the channel coefficient from the kth user to the lth relay, and g_l be the channel from the lth relay to the destination. We still consider the symmetric networks, where all the user-relay channels have the same path-loss coefficient λ_{sr}, and all the relay-destination channels have the same path-loss coefficient λ_{rd}. As there are multiple parallel relays now, the relaying strategy becomes much more flexible. An intuitive scheme is to let each relay node forwards the uncoded data one after another during the second phase, and the destination then performs ML detection by constructively

combining all the received signals. Although this scheme does achieve full diversity, the bandwidth efficiency is really low, as L independent channels are required for orthogonal relaying.

To address this issue, we propose to select only one good relay (i.e., the qth relay) each time to help forward the source messages. The resulting signal model is basically the same as that in the single-relay networks after properly modifying the subscripts in (4.1a) and (4.4). To be specific, all the users still broadcast concurrently in the first phase. The received signal at the qth relay is

$$y_{sr_q} = \sqrt{P\lambda_{sr}} \sum_{k=1}^{K} f_{kq} s_k + n_{sr_q},$$

and the received signal at the destination is still given by (4.1b). Then in the second phase, the qth relay will amplify and forward the data using either $\alpha_{q,VGR} = \left(P\lambda_{sr} \sum_{k=1}^{K} |f_{kq}|^2 + 1 \right)^{-1}$ or $\alpha_{q,FGR} = (KPG\lambda_{sr} + 1)^{-1}$. Note that $\alpha_{q,FGR}$ is slightly different from (4.3) after introducing the additional factor G. This is because after relay selection, the incoming channels are generally under very good conditions and the channel distributions would greatly change due to order statistics. As it is very hard to obtain the exact distribution functions, we simply assume that $\sum_{k=1}^{K} E|f_{k,q}|^2 = KG$, where G (>1) is a bounded constant that can be obtained through computer simulation, and this factor would not influence the diversity performance. The received signal at the destination during the second phase is

$$y_{r_q d} = \sqrt{\alpha P^2 \lambda_{sr} \lambda_{rd}} g_q \sum_{k=1}^{K} f_{kq} s_k + \tilde{n}_{r_q d},$$

where $\tilde{n}_{r_q d} \sim \mathcal{CN}\left(0, \alpha P \lambda_{rd} |g_q|^2 + 1\right)$. Finally, ML detection similar to (4.5) is performed to detect all the user symbols based on the observations y_{sd} and $y_{r_q d}$, and

the conditional PEP is given by

$$\Pr\left(\mathbf{s}\to\hat{\mathbf{s}}|\,\mathbf{h},\mathbf{f}_{sq},g_q\right)=Q\left(\sqrt{2\left(W_d+W_{r_q}\left(\Delta\mathbf{s}\right)\right)}\right)\leq\exp\left(-W_d-W_{r_q}\left(\Delta\mathbf{s}\right)\right),\tag{4.31}$$

where $W_d=\frac{1}{4}P\lambda_{sd}\left|\mathbf{h}^T\Delta\mathbf{s}\right|^2$, $W_{r_q}\left(\Delta\mathbf{s}\right)=\frac{\alpha P^2\lambda_{sr}\lambda_{rd}|g_q|^2\left|\mathbf{f}_{sq}^T\Delta\mathbf{s}\right|^2}{4\left(\alpha P\lambda_{rd}|g_q|^2+1\right)}$ and $\mathbf{f}_{sq}=\left(f_{1q},f_{2q},\cdots,f_{Kq}\right)^T$. It is observed that the quality of the qth relay branch is uniquely characterized by $W_{r_q}\left(\Delta\mathbf{s}\right)$, which itself depends on the error vector $\Delta\mathbf{s}$. As the real error probability is lower bounded by any PEP, we propose to select the relay branch that can minimize the maximum PEP. Since the exponential function is monotonically decreasing, the above Min-Max relay selection strategy can be equivalently formulated as

$$q=\arg\min_{l=1,2,\cdots,L}\max_{\hat{\mathbf{s}}\neq\mathbf{s}}\Pr\left(\mathbf{s}\to\hat{\mathbf{s}}|\,\mathbf{h},\mathbf{f}_{sl},g_l\right)=\arg\max_{l=1,2,\cdots,L}\min_{\Delta\mathbf{s}\neq\mathbf{0}}W_{r_l}\left(\Delta\mathbf{s}\right).\tag{4.32}$$

We remark that the above Min-Max criterion is also independently studied in [57] for VGR without considering the direct link. In the sequel, we obtain the diversity gain of VGR and FGR by using a much simpler approach.

Proposition 4.2. *For the Min-Max relay selection strategy (4.32), the scaling law of the error rate is* $\mathrm{O}\left(\frac{(\log P)^L}{P^{L+1}}\right)$ *for both VGR and FGR.*

Proof. Let $W_{r_l}^*=\min\limits_{\Delta\mathbf{s}\neq\mathbf{0}}W_{r_l}\left(\Delta\mathbf{s}\right)$ and $W_{r_q}^*=\max\limits_{l=1,2,\cdots,L}W_{r_l}^*$, then we have

$$F_{W_{r_q}(\Delta\mathbf{s})}\left(w\right)\leq\Pr\left(W_{r_q}^*\leq w\right)=\prod_{l=1}^{L}\Pr\left(W_{r_l}^*\leq w\right)\leq\prod_{l=1}^{L}\sum_{\Delta\mathbf{s}\neq\mathbf{0}}F_{W_{r_l}(\Delta\mathbf{s})}\left(w\right).\tag{4.33}$$

For VGR, we have

$$W_{r_l,FGR}\left(\Delta\mathbf{s}\right)=\frac{P^2\lambda_{sr}\lambda_{rd}|g_l|^2\left|\mathbf{f}_{sl}^T\Delta\mathbf{s}\right|^2}{4\left(P\lambda_{rd}|g_l|^2+KPG\lambda_{sr}+1\right)}\overset{P\to\infty}{\geq}\frac{P\lambda_{sr}\lambda_{rd}|g_l|^2\left|\mathbf{f}_{sl}^T\Delta\mathbf{s}\right|^2}{4\left(\lambda_{rd}|g_l|^2+KG\lambda_{sr}+1\right)}.$$

The CDF of $W_{r_l,FGR}(\Delta\mathbf{s})$ is thus given by

$$F_{W_{r_l,FGR}(\Delta\mathbf{s})}(w) \overset{P\to\infty}{\leq} \Pr\left(\frac{\lambda_{sr}\lambda_{rd}|g_l|^2\left|\mathbf{f}_{sl}^T\Delta\mathbf{s}\right|^2}{4\left(\lambda_{rd}|g_l|^2+KG\lambda_{sr}+1\right)}\leq\frac{w}{P}\right) \overset{P\to\infty}{\approx} \frac{c_3}{a_3b_3}\frac{w}{P}\log\left(\frac{P}{w}\right), \tag{4.34}$$

where $a_3=\lambda_{rd}$, $b_3=\frac{\lambda_{sr}\|\Delta\mathbf{s}\|^2}{4}$, $c_3=KG\lambda_{sr}+1$, and the approximation is due to (4.9). Likewise, for VGR we have

$$\begin{aligned} W_{r_l,VGR}(\Delta\mathbf{s}) &= \frac{P^2\lambda_{sr}\lambda_{rd}|g_l|^2\left|\mathbf{f}_{sl}^T\Delta\mathbf{s}\right|^2}{4\left(P\lambda_{rd}|g_l|^2+P\lambda_{sr}\sum_{k=1}^{K}|f_{kl}|^2+1\right)} \\ &\overset{P\to\infty}{\geq} \frac{P\lambda_{sr}\lambda_{rd}|g_l|^2\left|\mathbf{f}_{sl}^T\Delta\mathbf{s}\right|^2}{4\left(\lambda_{rd}|g_l|^2+\lambda_{sr}\sum_{k=1}^{K}|f_{kl}|^2+1\right)}. \end{aligned}$$

The CDF of $W_{r_l,VGR}(\Delta\mathbf{s})$ is given by

$$\begin{aligned} F_{W_{r_l,VGR}(\Delta\mathbf{s})}(w) &\overset{P\to\infty}{\leq} \Pr\left(\frac{\lambda_{sr}\lambda_{rd}|g_l|^2\left|\mathbf{f}_{sl}^T\Delta\mathbf{s}\right|^2}{4\left(\lambda_{rd}|g_l|^2+\lambda_{sr}\sum_{k=1}^{K}|f_{kl}|^2+1\right)}\leq\frac{w}{P}\right) \\ &\overset{P\to\infty}{\approx} \frac{(K-1)c_4+1}{a_4b_4}\frac{w}{P}\log\left(\frac{P}{w}\right), \end{aligned} \tag{4.35}$$

where $a_4=\lambda_{rd}$, $b_4=\frac{\lambda_{sr}\|\Delta\mathbf{s}\|^2}{4}$, $c_4=\lambda_{sr}$, and we use the small value approximation in (4.8). Plugging (4.34) and (4.35) back into (4.33) leads to

$$F_{W_{r_q}(\Delta\mathbf{s})}(w) \overset{P\to\infty}{\leq} C\left[\frac{w}{P}\log\left(\frac{P}{w}\right)\right]^L,$$

where C is some constant. Therefore, for both the VGR and FGR we can obtain

$$\begin{aligned} E\left[\exp\left(-W_{r_q}\left(\Delta\mathbf{s}\right)\right)\right] &\overset{P\to\infty}{\leq} CP(-1)^L \int_0^\infty \exp\left(-Pw\right) w^L (\log w)^L dw \\ &\overset{(a)}{=} CP(-1)^L \frac{\partial^L}{\partial w^L}\left\{P^{-w}\Gamma\left(w\right)\right\}\Big|_{w=L+1} \\ &= CP(-1)^L \left\{\sum_{k=0}^{L}\binom{L}{k}\frac{\partial^k P^{-w}}{\partial w^k}\frac{\partial^{L-k}\Gamma\left(w\right)}{\partial w^{L-k}}\right\}\Bigg|_{w=L+1} \\ &\overset{P\to\infty}{\approx} CP(-1)^L \left\{\frac{\partial^L P^{-w}}{\partial w^L}\Gamma\left(w\right)\right\}\Big|_{w=L+1} \\ &= C\Gamma\left(L+1\right)\frac{(\log P)^L}{P^L}, \end{aligned} \tag{4.36}$$

where we use [88, 4.358.5] in (a), and $\Gamma\left(x\right)$ is the Gamma function [88, 8.310.1]. Combining the above result with (4.7) completes the proof. □

From Proposition 4.2, we conclude that relay selection could achieve full dominant diversity gain $L+1$ at extremely high SNRs; however, the logarithmic-term loss is also proportional to the number of relays.

4.3 Distributed Space-Time Block Coding

Although relay selection can achieve full diversity, it would induce some loss in coding gain as each time there is only one relay node helping forward data. To fully exploit the spatial diversity, we investigate DSTBC in this section, where all the relay nodes participate in data relaying using some linear coding on the received signals.

4.3.1 Signal Model of DSTBC

The whole data transmission is still completed in two phases. In the first phase, all the users simultaneously broadcast a block of data $\mathbf{s}_k = (s_{k1}, s_{k2}, \cdots, s_{kT})^T$ containing T ($>L$) symbols. Suppose the channel is quasi-static, i.e., the channel coefficients stay constant during a block interval, then the received signal vector at the lth relay and at the destination can be respectively expressed as

$$\begin{cases} \mathbf{y}_{sr_l} = \sqrt{P\lambda_{sr}} \sum_{k=1}^{K} f_{kl}\mathbf{s}_k + \mathbf{n}_{sr_l}, & \text{(4.37a)} \\ \mathbf{y}_{sd} = \sqrt{P\lambda_{sd}} \sum_{k=1}^{K} h_k\mathbf{s}_k + \mathbf{n}_{sd}, & \text{(4.37b)} \end{cases}$$

where $\mathbf{n}_{sr_l}, \mathbf{n}_{sd} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Then each relay node performs linear coding on the received signal, and the transmitted signal at the lth relay node is $\mathbf{x}_{r_l} = \sqrt{\alpha_l P}\mathbf{A}_l\mathbf{y}_{sr_l}$. To simplify the analysis, we assume the $T \times T$ coding matrices $\{\mathbf{A}_l\}$ are unitary, i.e., $\mathbf{A}_l\mathbf{A}_l^H = \mathbf{I}$. The amplification factor α_l at the lth relay node is still given by (4.2) for VGR and by (4.3) for FGR, respectively. Then in the second phase, all the relay nodes forward their signals simultaneously to the destination while the source nodes stay silent. The received signal at the destination in the second phase is

$$\mathbf{y}_{rd} = \sqrt{\lambda_{rd}} \sum_{l=1}^{L} g_l\mathbf{x}_{r_l} + \mathbf{n}_{rd} = P\sqrt{\lambda_{rd}\lambda_{sr}} \sum_{k=1}^{K} \mathbf{M}_k\mathbf{v}_k + \tilde{\mathbf{n}}_{rd}, \tag{4.38}$$

where $\tilde{\mathbf{n}}_{rd} \sim \mathcal{CN}\left(\mathbf{0}, \left(\sum_{l=1}^{L} \alpha_l P\lambda_{rd}|g_l|^2 + 1\right)\mathbf{I}\right)$, $\mathbf{v}_k = \left(\sqrt{\alpha_1}f_{k1}g_1, \cdots, \sqrt{\alpha_L}f_{kL}g_L\right)^T$ is the equivalent channel vector, and $\mathbf{M}_k = (\mathbf{A}_1\mathbf{s}_k, \mathbf{A}_2\mathbf{s}_k, \cdots, \mathbf{A}_L\mathbf{s}_k)$ is the codeword of the kth user. Clearly, $\mathbf{M}_k$ plays the role of the space-time code in the multiple antenna systems. In the following, we assume that $\Delta\mathbf{M}_k = (\mathbf{A}_1\Delta\mathbf{s}_k, \mathbf{A}_2\Delta\mathbf{s}_k, \cdots, \mathbf{A}_L\Delta\mathbf{s}_k)$ always have full rank for any $\Delta\mathbf{s}_k \neq \mathbf{0}$, which is the sufficient condition to achieve

full diversity for the single-user DSTBC-FGR systems [17, 90].

4.3.2 Error Performance Analysis

Using the ML detection similar to (4.5), the PEPs are still given by (4.6) after redefining $W_d = \frac{P\lambda_{sd}}{4}\|\Delta\mathbf{Sh}\|^2$ and $W_r = \frac{P^2\lambda_{rd}\lambda_{sr}\left\|\sum_{k=1}^{K}\Delta\mathbf{M}_k\mathbf{v}_k\right\|^2}{4\left(P\lambda_{rd}\sum_{l=1}^{L}\alpha_l|g_l|^2+1\right)}$. After some manipulations, it is easy to show that

$$E\left[\exp\left(-W_d\right)\right] = \det{}^{-1}\left(\mathbf{I}+\frac{1}{4}P\lambda_{sd}\Delta\mathbf{S}^H\Delta\mathbf{S}\right) \overset{P\to\infty}{\approx} \frac{4^r}{\lambda_{sd}^r\prod_{i=1}^{r}\tau_i}P^{-r} \le \frac{4}{P\lambda_{sd}\tau}, \quad (4.39)$$

where $\Delta\mathbf{S} = (\Delta\mathbf{s}_1, \Delta\mathbf{s}_2, \cdots, \Delta\mathbf{s}_K)$, and τ_i and r are the ith eigen-value and the rank of the matrix $\Delta\mathbf{S}^H\Delta\mathbf{S}$, respectively. Note that the error performance is lower bounded by the worst-case PEP, which occurs when $r = 1$ and leads to the last inequality, where τ is the minimum of the eigen-values of all the matrices $\Delta\mathbf{S}^H\Delta\mathbf{S}$ for any $\Delta\mathbf{S} \neq \mathbf{0}$. Next we study the term $E\left[\exp\left(-W_r\right)\right]$. For FGR, we have $W_{r,FGR} = \frac{P^2\lambda_{rd}\lambda_{sr}\left\|\sum_{k=1}^{K}\Delta\mathbf{M}_k\mathbf{D}(\mathbf{g})\mathbf{f}_{kr}\right\|^2}{4\left(P\lambda_{rd}\sum_{l=1}^{L}|g_l|^2+KP\lambda_{sr}+1\right)}$, where $\mathbf{f}_{kr} = (f_{k1}, f_{k2}, \cdots, f_{kL})^T$ and $\mathbf{D}\left(\mathbf{g}\right) = diag\left(g_1, g_2, \cdots, g_L\right)$. It is easy to show that the expectation of $\exp\left(-W_{r,FGR}\right)$ is

$$\begin{aligned} E_{\{\mathbf{f}_{kr}\}}\left[\exp\left(-W_{r,FGR}\right)\right] &= \det{}^{-1}\left(\mathbf{I}+\frac{P^2\lambda_{rd}\lambda_{sr}\sum_{k=1}^{K}\Delta\mathbf{M}_k\mathbf{D}\left(\mathbf{g}\right)\mathbf{D}^H\left(\mathbf{g}\right)\Delta\mathbf{M}_k^H}{4\left(P\lambda_{rd}\sum_{l=1}^{L}|g_l|^2+KP\lambda_{sr}+1\right)}\right) \\ &\le \prod_{k=1}^{L}\left(1+\frac{P\lambda_{rd}\lambda_{sr}\eta|g_k|^2}{4\left(\lambda_{rd}\sum_{l=1}^{L}|g_l|^2+K\lambda_{sr}+1\right)}\right)^{-1}, \end{aligned} \quad (4.40)$$

where we use the inequality $\det\left(\mathbf{A}+\mathbf{B}\right) \ge \det\mathbf{A} + \det\mathbf{B}$ for any $\mathbf{A}, \mathbf{B} \ge \mathbf{0}$, and η is the minimum of the eigen-values of all the matrices $\Delta\mathbf{M}^H\Delta\mathbf{M}$ for any $\Delta\mathbf{M} \neq \mathbf{0}$. When there is a large number of relay nodes, we can simplify the above expression by

using the approximation $\sum_{k=1}^{L} |g_l|^2 \overset{L\gg 1}{\approx} L$ without affecting the diversity performances [17, 90]. Now we can obtain

$$E\left[\exp\left(-W_{r,FGR}\right)\right] \leq \prod_{k=1}^{L} E_{g_k}\left[\left(1+P\delta|g_k|^2\right)^{-1}\right] = \prod_{k=1}^{L}\left[\frac{1}{P\delta}\exp\left(\frac{1}{P\delta}\right)E_1\left(\frac{1}{P\delta}\right)\right]$$
$$\leq \prod_{k=1}^{L}\left[\frac{1}{P\delta}\log\left(1+P\delta\right)\right] \overset{P\to\infty}{\approx} \frac{1}{\delta^L}\left(\frac{\log P}{P}\right)^L, \tag{4.41}$$

where $\delta = \frac{\lambda_{rd}\lambda_{sr}\eta}{4(L\lambda_{rd}+K\lambda_{sr}+1)}$, and we use $e^x E_1(x) \leq \log\left(1+\frac{1}{x}\right)$ [87, 5.1.20] in the last inequality. Plugging (4.39) and (4.41) back into (4.6), we have

$$\Pr\left(\mathbf{s}\to\hat{\mathbf{s}}|\,\text{DSTBC-FGR}\right) \overset{P\to\infty}{\approx} \frac{4}{\delta^L\lambda_{sd}\tau}\frac{(\log P)^L}{P^{L+1}}. \tag{4.42}$$

It is observed that the dominant diversity gain of DSTBC-FGR is $L+1$ and it is independent of the number of users; however, there is some loss in coding gain compared to the single-user case as δ is a decreasing function of K.

Next we study DSTBC-VGR. Consider a specific error pair where $\Delta\mathbf{s}_k = \mathbf{0}$ and $\Delta\mathbf{M}_k = \mathbf{0}$ for $k = 2, 3, \cdots, K$. Then we have $W_d = \frac{P\lambda_{sd}\|\Delta\mathbf{s}_1\|^2}{4}|h_1|^2$. Let ρ be the maximum eigen-value of $\Delta\mathbf{M}_1^H\Delta\mathbf{M}_1$, and without loss of generality, suppose $|g_1|^2 \geq |g_2|^2 \geq \cdots \geq |g_L|^2$, then we can obtain

$$W_{r,VGR} = \frac{P^2\lambda_{rd}\lambda_{sr}\|\Delta\mathbf{M}\mathbf{v}_1\|^2}{4\left(P\lambda_{rd}\sum_{l=1}^{L}\alpha_{l,VGR}|g_l|^2+1\right)} \overset{(a)}{\leq} \frac{\rho LP\lambda_{rd}|g_1|^2\left(P\lambda_{sr}\sum_{k=1}^{K}|f_{k1}|^2+1\right)}{4\left(P\lambda_{rd}|g_1|^2+P\lambda_{sr}\sum_{k=1}^{K}|f_{k1}|^2+1\right)}$$
$$\leq \frac{\rho L}{4}\min\left(P\lambda_{rd}|g_1|^2, P\lambda_{sr}\sum_{k=1}^{K}|f_{k1}|^2+1\right), \tag{4.43}$$

where we use $P\lambda_{sr}\alpha_{l,VGR}|f_{1l}|^2 \leq 1$, $\sum_{l=1}^{L}|g_l|^2 \leq L|g_1|^2$ and $\sum_{l=1}^{L}\alpha_{l,VGR}|g_l|^2 \geq \alpha_{1,VGR}|g_1|^2$ in (a). Now we can lower bound the PEP by

$$\Pr\left(\mathbf{s}\to\hat{\mathbf{s}}|\,\text{DSTBC-VGR}\right) \geq \max\left(F_1, F_2\right),$$

where

$$F_1 = E\left[Q\left(\sqrt{\mu P\left(|h_1|^2+|g_1|^2\right)}\right)\right] \overset{P\to\infty}{\approx} \frac{(2L+1)!}{(L+1)!}\left(\frac{8}{\mu}\right)^{L+1} P^{-(L+1)}, \tag{4.44}$$

$$F_2 = E\left[Q\left(\sqrt{\mu\left(P|h_1|^2+P\sum_{k=1}^{K}|f_{k1}|^2+1\right)}\right)\right] \overset{P\to\infty}{\approx} C\left(\frac{2}{\mu}\right)^{K+1} P^{-(K+1)}, \tag{4.45}$$

and $\mu = \frac{1}{4}\max\left(\lambda_{sd}\|\Delta\mathbf{s}_1\|^2, \rho L\lambda_{rd}, \rho L\lambda_{sr}, \rho L\right)$, and

$$C = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{\mu}{2\sin^2\theta}\right)(\sin\theta)^{2(K+1)}d\theta$$

is some constant. Consequently, we have

$$\Pr\left(\mathbf{s}\to\hat{\mathbf{s}}|\,\text{DSTBC-VGR}\right) \geq \mathrm{O}\left(P^{-\min(K+1,L+1)}\right). \tag{4.46}$$

So the diversity gain of DSTBC-VGR is upper bounded by $\min\left(K+1, L+1\right)$.

Comparing (4.42) and (4.46), we can observe that the diversity performances of DSTBC-FGR and DSTBC-VGR are very distinct. To be specific, full diversity is always achieved by DSTBC-FGR, whereas the diversity of DSTBC-VGR is also upper bounded by the number of users. This phenomenon is caused by the noise amplification effect of VGR, as can be seen by studying the noise power (denoted by σ^2) of the relay-destination signal (4.38). For DSTBC-FGR, we have $\sigma^2_{FGR} = \sum_{l=1}^{L}\frac{P\lambda_{rd}|g_l|^2}{KP\lambda_{sr}+1}+1 \overset{P\to\infty}{\approx} \sum_{l=1}^{L}\frac{\lambda_{rd}|g_l|^2}{K\lambda_{sr}}+1$, which is independent of the transmitted power P and the user-relay channels. On the contrary, for DSTBC-VGR we have $\sigma^2_{VGR} = \sum_{l=1}^{L}\frac{P\lambda_{rd}|g_l|^2}{P\lambda_{sr}\sum_{k=1}^{K}|f_{kl}|^2+1}+1$. As a result, the noise power would be comparable to the transmitted power P whenever $P\lambda_{sr}\sum_{k=1}^{K}|f_{kl}|^2 = \mathrm{O}\left(1\right)$, in which case the detection error is likely to occur at the receiver with very high probability. It is easy to show that the probability of such dominant error events is $\mathrm{O}\left(P^{-K}\right)$, thus the overall diversity of DSTBC-VGR is also upper bounded by $K+1$.

4.3.3 Selective DSTBC-VGR for Single-User Networks

So far, we have shown that the diversity of DSTBC-VGR is upper bounded by the number of users. For the conventional single-user analog relaying (i.e., $K = 1$), this implies that the diversity gain is limited by 2 regardless of the number of relay nodes. We have also seen that the noise amplification effect is the main cause of the diversity loss, which happens when all the channels from the users to a certain relay node experience deep fading. Note that we have implicitly assumed that the relay nodes are using full power all the time. Intuitively, if the input channels are in bad conditions, the relay nodes should lower its transmitted power or even stay idle to mitigate the noise enhancement effect. In the following, we shall adopt this idea and develop the selective DSTBC-VGR to recover the diversity loss when there is only one user (i.e., $K = 1$).

Since the quality of the relay branch is uniquely characterized by $W_{r,\mathrm{VGR}}$, we first rewrite this term as $W_{r,VGR} = \frac{\kappa P \lambda_{sr} \sum_{l=1}^{L} a_l |f_{1l}|^2 \nu_l}{4\left(\sum_{l=1}^{L} a_l \nu_l + 1\right)}$, where $\kappa \in [\eta, \rho]$ is a constant, $\nu_l = \frac{P\lambda_{rd}|g_l|^2}{P\lambda_{sr}|f_{1l}|^2+1}$ and $a_l \leq 1$ is the power scaling coefficient. Clearly, the best power allocation scheme is to maximize $W_{r,VGR}$, i.e.,

$$\max g(\mathbf{a}) \triangleq \frac{\sum_{l=1}^{L} a_l |f_{1l}|^2 \nu_l}{\sum_{l=1}^{L} a_l \nu_l + 1}, \text{ s.t. } a_l \leq 1 \text{ for } l = 1, 2, \cdots, L, \tag{4.47}$$

where $\mathbf{a} = (a_1, a_2, \cdots, a_L)^T$. Although the above optimization problem is also studied in [92], the authors only propose an iterative algorithm to search for the optimizer and no performance analysis is performed there. Unlike that work, we give the closed-form solution to the above problem and prove that full diversity can be

achieved. Without loss of generality, we assume $|f_{11}|^2 > |f_{12}|^2 > \cdots > |f_{1L}|^2$ and define $f_{1(L+1)} = 0$.

Proposition 4.3. *The solution to the problem (4.47) is*

$$a_l^* = \begin{cases} 1, \ l = 1, 2, \cdots, l_0 \\ \\ 0, \qquad o.w. \end{cases}, \tag{4.48}$$

where $l_0 \in \{1, 2, \cdots, L\}$ *is the smallest index satisfying*

$$\sum_{l=1}^{l_0} \left(|f_{1l}|^2 - \left| f_{1(l_0+1)} \right|^2 \right) \nu_l \geq \left| f_{1(l_0+1)} \right|^2.$$

Besides, the error rate has a scaling law of $\mathrm{O}\left(\frac{1}{P^{L+1}}\right)$*, and the diversity gain of selective DSTBC-VGR is* $L + 1$*.*

Proof. In [92], it has been proved that $a_l^* \in \{0, 1\}$. We now prove by contradiction that if $a_{l_1}^* = 1$, then $a_{l_2}^* = 1$ for any $l_2 \leq l_1$. Suppose now there exists an $a_{l_2}^* = 0$ with $l_2 < l_1$. Then we can always find a sufficiently small quantity $\delta \ll 1$ such that $\frac{\delta \nu_{l_1}}{\nu_{l_2}} \leq 1$. Consequently, there exists another feasible solution $\hat{\mathbf{a}}$ with $\hat{a}_{l_1} = a_{l_1}^* - \delta$ and $\hat{a}_{l_2} = \frac{\delta \nu_{l_1}}{\nu_{l_2}}$, and the other elements of $\hat{\mathbf{a}}$ are the same as $\mathbf{a}^*$. It is easy to check that $g(\hat{\mathbf{a}}) - g(\mathbf{a}^*) = \frac{\delta \nu_{l_1} \left(|f_{1l_2}|^2 - |f_{1l_1}|^2 \right)}{\sum_{l=1}^{L} a_l^* \nu_l + 1} > 0$, which contradicts the optimality of $\mathbf{a}^*$. Now the optimizer can be limited to the finite set $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_L\}$, where $\mathbf{a}_l = \left(\mathbf{1}_{l \times 1}, \mathbf{0}_{(L-l) \times 1} \right)$. After some manipulations, we can show that $g(\mathbf{a}_k) \geq g(\mathbf{a}_{k+1})$ is equivalent to the condition $\sum_{l=1}^{k} \left(|f_{1l}|^2 - \left| f_{1(k+1)} \right|^2 \right) \nu_l \geq \left| f_{1(k+1)} \right|^2$, where the left-hand side is increasing with k and the right-hand side is decreasing with k. Consequently, we have $g(\mathbf{a}_1) \leq g(\mathbf{a}_2) \leq \cdots \leq g(\mathbf{a}_{l_0})$ and $g(\mathbf{a}_{l_0}) \geq g(\mathbf{a}_{l_0+1}) \geq \cdots \geq g(\mathbf{a}_{L+1})$, which completes the proof of the first part. To show the diversity

performance, we note that the vectors $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_L\}$ also belong to the feasible set, where $\mathbf{e}_l$ is a $L \times 1$ vector with the kth element being 1 and the other elements being 0. Note that each $\mathbf{e}_l$ corresponds to the case when only the lth relay is selected to forward data. Consequently, the selective DSTBC-VGR performs strictly better than the best relay selection scheme, the error rate of which has a scaling law of $\mathrm{O}\left(\frac{1}{P^{L+1}}\right)$ [16]. □

Basically, the above results show that each relay node either stays idle or forwards data with full power, and the relays with better user-relay channel conditions have the priority to be selected. These facts indicate an easy implementation of the proposed selective DSTBC strategy, i.e., the destination can first compute the active relay set and then feed back a single threshold. The relays whose input channel conditions are better than the threshold then stay active in the second phase. On the contrary, for the iterative algorithm proposed in [92], the destination has to feed back the whole active relay set, and the feedback overhead is prohibitive when there is a large number of relays.

4.4 Diagonal Distributed Space-Time Coding

In this section, we study DDSTC that can achieve full diversity for both VGR and FGR. Different from DSTBC which employs unitary coding matrices at the relay nodes, DDSTC has a diagonal structure by letting only one relay forward data at each time instant.

The whole data transmission is still completed in two phases. In the first

phase, all the users broadcast a set of L symbols. The received signals at the lth relay node and at the destination are given by (4.37a) and by (4.37b), respectively. Then each relay node performs linear transformation on the received signal vector, and the transmitted signal at the lth relay node is $x_{r_l} = \sqrt{\alpha_l LP}\mathbf{t}_l^T \mathbf{y}_{sr_l}$. Here $\mathbf{t}_l$ is a $L \times 1$ coding vector having unit norm (i.e., $\|\mathbf{t}_l\|^2 = 1$), the design criterion of which will be clear later. The constant L is introduced to normalize the total transmitted power, as each relay node only forwards data in one time slot in the second phase. The amplification factor α_l at the lth relay node is still given by (4.2) for VGR and by (4.3) for FGR, respectively. Then in the second phase, all the relay nodes take turns to forward data, and the received signal at the destination during the lth time slot is

$$y_{r_l d} = \sqrt{\lambda_{rd}} g_l x_{r_l} + n_{r_l d} = P\sqrt{L\lambda_{sr}\lambda_{rd}\alpha_l} g_l \sum_{k=1}^{K} f_{kl}\mathbf{t}_l^T \mathbf{s}_k + \tilde{n}_{r_l d}, \tag{4.49}$$

where $\tilde{n}_{r_l d} \sim \mathcal{CN}\left(0, LP\lambda_{rd}\alpha_l |g_l|^2 + 1\right)$ is the equivalent additive noise. Upon observing the signals $\mathbf{y}_{sd}$ and $\{y_{r_l d}\}$, the destination performs ML detection to jointly detect the K user symbols as

$$\mathbf{s}_d = \arg \min_{\hat{\mathbf{s}}_k \in \Omega^L} \left\| \mathbf{y}_{sd} - \sqrt{P\lambda_{sd}} \sum_{k=1}^{K} h_k \hat{\mathbf{s}}_k \right\|^2 + \sum_{l=1}^{L} \frac{\left\| y_{r_l d} - P\sqrt{L\lambda_{sr}\lambda_{rd}\alpha_l} g_l \sum_{k=1}^{K} f_{kl}\mathbf{t}_l^T \hat{\mathbf{s}}_k \right\|^2}{LP\lambda_{rd}\alpha_l |g_l|^2 + 1}, \tag{4.50}$$

and the PEP is given by

$$\Pr(\mathbf{s} \to \hat{\mathbf{s}}) = E\left[Q\left(\sqrt{2\left(W_d + \sum_{l=1}^{L} W_{r_l} \right)} \right) \right] \leq E\left[\exp\left(-W_d - \sum_{l=1}^{L} W_{r_l} \right) \right], \tag{4.51}$$

where $W_d = \frac{P\lambda_{sd}}{4} \|\Delta\mathbf{S}\mathbf{h}\|^2$, $W_{r_l} = \frac{\alpha_l P^2 L \lambda_{sr}\lambda_{rd}|g_l|^2 \left|\mathbf{f}_{sl}^T \Delta\mathbf{S}^T \mathbf{t}_l\right|^2}{4\left(LP\lambda_{rd}\alpha_l|g_l|^2 + 1\right)}$, $\mathbf{f}_{sl} = (f_{1l}, f_{2l}, \cdots, f_{Kl})^T$ and $\Delta\mathbf{S} = (\Delta\mathbf{s}_1, \Delta\mathbf{s}_2, \cdots, \Delta\mathbf{s}_K)$. The first term $E\left[\exp\left(-W_d\right)\right]$ has been given by

(4.39). As for the second term, we observe that each W_{r_l} has the similar form as W_r in the single-relay case discussed before. By following the same steps, we can obtain

$$E\left[\exp\left(-W_{r_l}\right)\right] \overset{P\to\infty}{\leq} \frac{4K}{L\lambda_{rd}\|\Delta\mathbf{S}^T\mathbf{t}_l\|^2}\frac{\log P}{P} \tag{4.52}$$

for both VGR and FGR. Plugging (4.39) and (4.52) back into (4.51) leads to

$$\Pr\left(\mathbf{s}\to\hat{\mathbf{s}}\right) \overset{P\to\infty}{\leq} \frac{4^{L+1}}{\lambda_{sd}\tau}\left(\frac{K}{L\lambda_{rd}}\right)^L\left(\prod_{l=1}^{L}\left|\Delta\mathbf{s}_k^T\mathbf{t}_l\right|^2\right)^{-1}\frac{(\log P)^L}{P^{L+1}}, \tag{4.53}$$

where we use $\prod_{l=1}^{L}\left\|\Delta\mathbf{S}^T\mathbf{t}_l\right\|^2 \geq \prod_{l=1}^{L}\left|\Delta\mathbf{s}_k^T\mathbf{t}_l\right|^2$ in the inequality, i.e., the dominant error events occur when only one user symbol vector is decoded incorrectly. Unlike DSTBC, we observe that both of DDSTC-VGR and DDSTC-FGR can achieve full diversity gain $L+1$. This is because the diagonal structure of DDSTC can efficiently mitigate the noise enhancement effect, as the output noises of the relay nodes would not be combined concurrently at the destination. Besides, we conclude that the best code design criterion is to maximize the minimum product distance, i.e.,

$$\max_{\mathbf{t}_l\in\mathbb{C}^{L\times 1},\|\mathbf{t}_l\|^2=1}\min_{\Delta\mathbf{s}\neq\mathbf{0}}\prod_{l=1}^{L}\left|\Delta\mathbf{s}^T\mathbf{t}_l\right|^2, \tag{4.54}$$

which has been well studied in the literatures. For example, when $L=2^s$ $(s\geq 1)$ and the signal constellation Ω has the form $\mathbb{Z}[j]=\left\{a+jb|\, a,b\in\mathbb{Z}, j=\sqrt{-1}\right\}$, the optimum coding matrix is given by [93] $\mathbf{T}_{\text{opt}} = \frac{1}{\sqrt{L}}\mathbf{V}\left(\theta_1,\theta_2,\cdots,\theta_L\right)$, where $\mathbf{T}=$

$(\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_L)^T$, $\theta_l = e^{j\frac{4l-3}{2L}\pi}$ for $l = 1, 2, \cdots, L$, and

$$\mathbf{V}(\theta_1, \theta_2, ..., \theta_L) = \begin{pmatrix} 1 & \theta_1 & \cdots & \theta_1^{L-1} \\ 1 & \theta_2 & \cdots & \theta_2^{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \theta_L & \cdots & \theta_L^{L-1} \end{pmatrix} \tag{4.55}$$

is the $L \times L$ Vandermonde matrix with parameters $\theta_1, \theta_2, \cdots, \theta_L$. For more code designs, please refer to [94] and the references therein.

4.5 Simulations

In this section, we present some simulation results to validate our analysis. We use the path loss model $\lambda = D^{-3}$, where λ is the path-loss coefficient and D is the distance between two terminals. Pair error probability is used as the performance metric, i.e., the probability that at least one of the user symbols is decoded incorrectly at the destination. To simplify the simulation settings, only symmetric networks with one or two users are considered, and D_{sd} is always normalized to 1.

Figure 4.2 shows the error performances with different channel conditions, where the two users use QPSK signals and there is only one relay node. Compared with direct transmission (DT), a diversity gain of 2 is achieved due to node cooperation. We observe that VGR has about 1dB SNR gain over FGR in all cases. It is also observed that the error performances almost remain unchanged after improving the qualities of user-relay channels, whereas about 3dB SNR gain is achieved when the relay-destination channels become better, which is consistent with our analysis

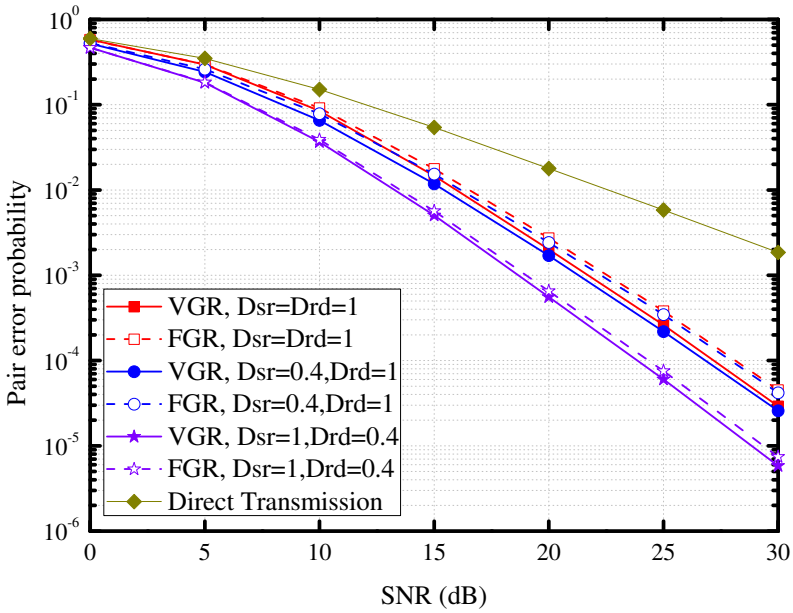


Figure 4.2: Error performances of a two-user network with different channel conditions.

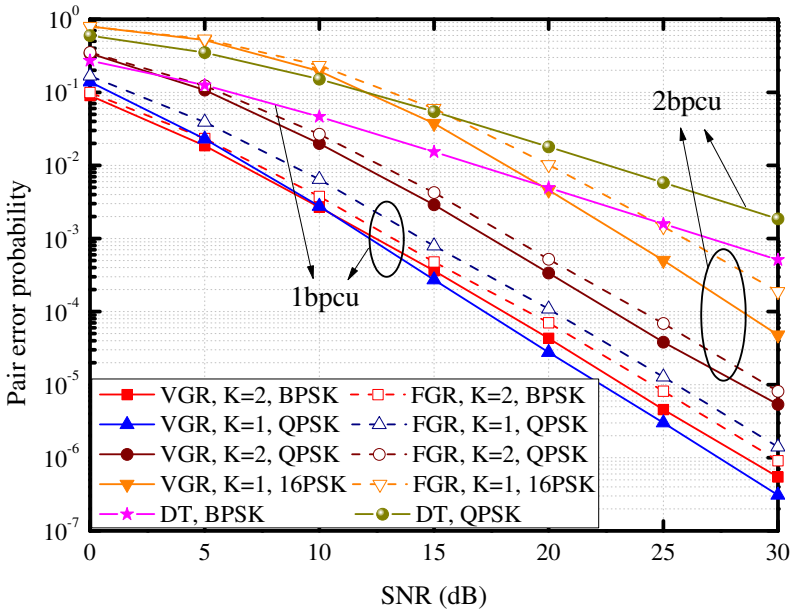


Figure 4.3: Comparison of two-user and single-user network with different data rate.

that the relay-destination channel conditions dominate the error performances.

In Figure 4.3, we compare with the conventional single-user (i.e., $K = 1$) amplify-and-forward relaying [12], where all the users are served separately in a TDMA manner. The transmitted power and the total data rate have been properly normalized, and for network topology we let $D_{sr} = D_{rd} = 0.5$. We observe that full diversity is achieved in all cases. However, the single-user VGR is superior to all the other schemes when the total data rate is only 1 bit per channel use (bpcu) because there is no logarithmic-term loss at modest SNRs. When the data rate is 2bpcu, although the error probability of single-user VGR still decreases faster, the spectral efficiency dominates the overall performances and thus the two-user systems show huge performance gain.

Next we study the error performances of relay selection in Figure 4.4, where QPSK signal is employed. As the reference, we also simulate Ding's scheme [54] for VGR, where the qth relay is selected if

$$q = \arg \max_{l=1,2,\cdots,L} \frac{P\lambda_{rd}|g_l|^2 \left(P\lambda_{sr} \sum_{k=1}^{K} |f_{kl}|^2 + 1 \right)}{P\lambda_{rd}|g_l|^2 + P\lambda_{sr} \sum_{k=1}^{K} |f_{kl}|^2 + 1}.$$

Clearly, our Min-Max scheme can achieve full diversity for both VGR and FGR, whereas the diversity of Ding's scheme is bounded by 2. This is because in Ding's scheme, the channel phases have nothing to do with relay selection. However, the channel phases actually have tremendous effects on PEPs, since the source messages are randomly mixed in the air and the MUI depends largely on the orthogonality of the instant channel coefficients. Although Ding's scheme does achieve full diversity in terms of outage capacity [54], where successive interference cancelation is em-

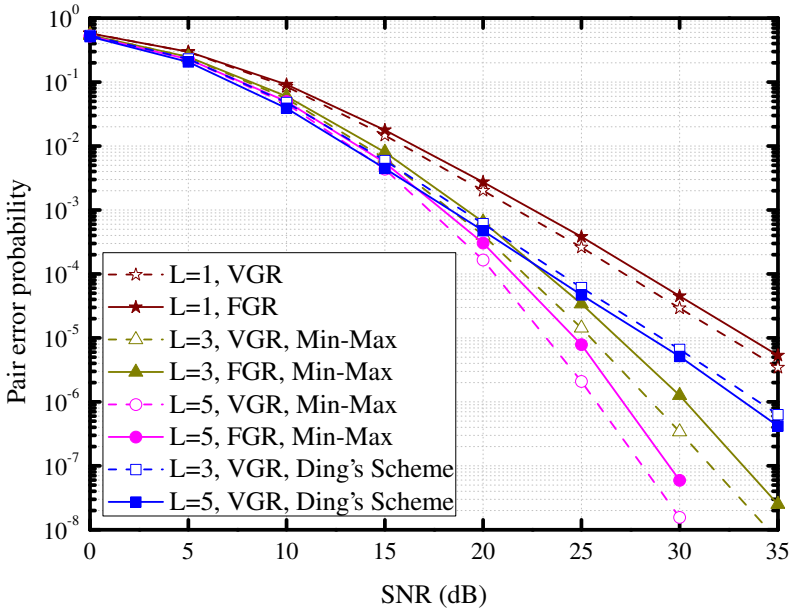


Figure 4.4: Error performances of a two-user network with relay selection.

ployed at the destination and error-free decoding is assumed, our Min-Max strategy is more practical for uncoded systems.

In Figure 4.5, we show the performances of DSTBC-FGR and DDSTC with two users, where QPSK signal is employed. We observe that both coding schemes can achieve full diversity with multiple relays. Comparatively, DSTBC-FGR performs slightly better than DDSTC-FGR, since the diagonal structure of DDSTC limits the minimum distances of the codeword. About 1dB SNR gain is achieved by using DDSTC-VGR against DDSTC-FGR regardless of the number of relays. Comparing Figure 4.4 and Figure 4.5, we observe that DSTC can achieve much higher coding gain than relay selection does, as all the relay nodes are contributing to forward data.

Finally we study the performances of DSTBC-VGR using QPSK signal. When

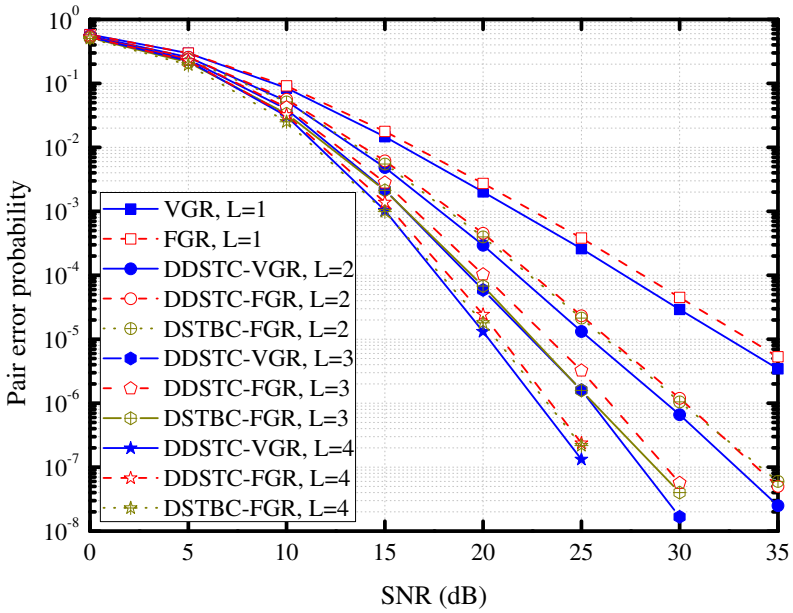


Figure 4.5: Error performances of a two-user network with DDSTC and DSTBC-FGR.

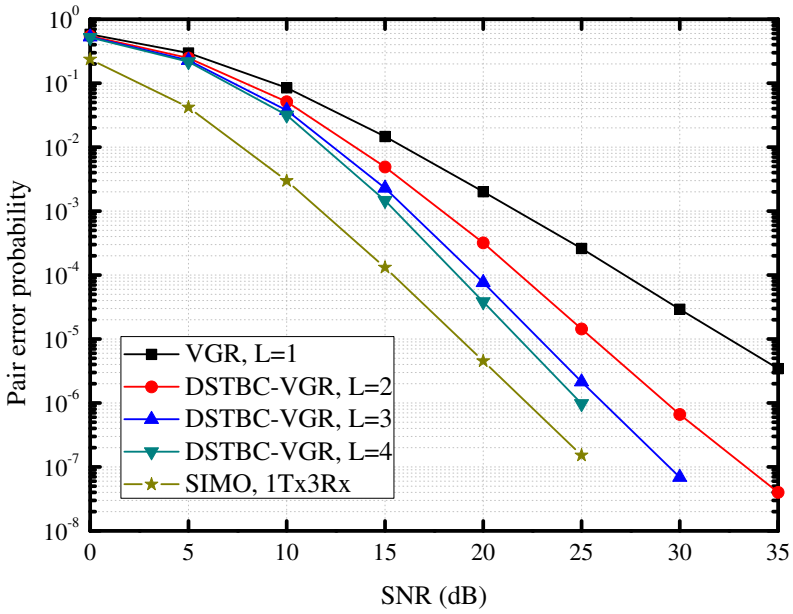


Figure 4.6: Error performances of a two-user network with DSTBC-VGR.

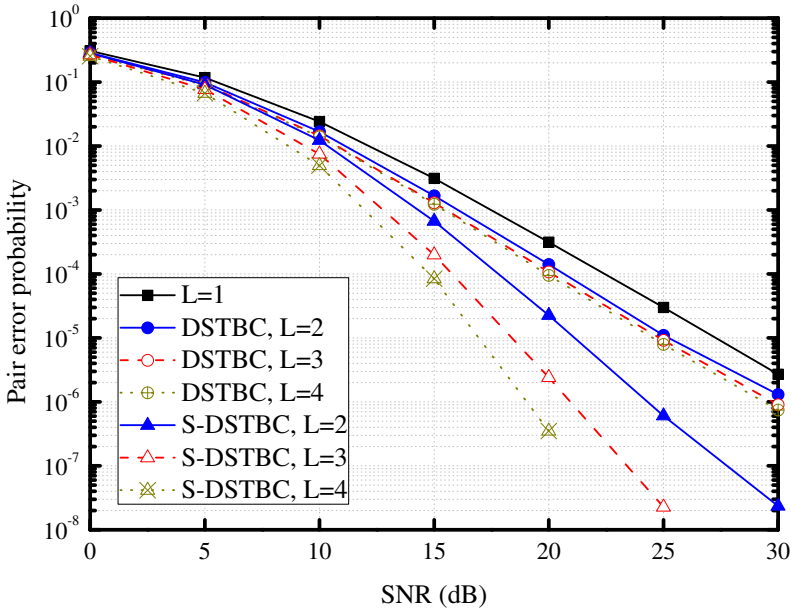


Figure 4.7: Error performances of a single-user network with DSTBC-VGR and selective DSTBC-VGR.

there are two users, we observe in Figure 4.6 that the performance of DSTBC-VGR is bounded by that of the single-input multiple-output systems with one transmitted antenna and three received antennas (1Tx3Rx), which is well known to have a diversity gain of 3. As for the single-user systems shown in Figure 4.7, it is observed that the diversity gain is always 2, and the marginal coding gain is very trivial by increasing the number of relay nodes beyond 2. Clearly, our selective DSTBC (S-DSTBC) can fully recover the diversity loss with very small signalling overhead.

4.6 Conclusions

In this work, we studied the diversity gain of ANC uplink. We showed that full dominant diversity can be achieved through relay selection and DSTC, but the

logarithmic term in the error rate expression would degrade the diversity performance at modest SNRs. For DSTBC, the diversity gain would also be bounded by the number of users due to noise amplification effect of VGR. For future work, one may study the code design for DSTBC and extend our selective DSTBC scheme to the multi-user networks. One may also investigate other distributed beamforming schemes to improve the coding gain.

Chapter 5

Diversity Analysis of Wireless Uplink with Non-Coherent Network Coding

In the previous chapters, we assume that instantaneous CSI is known to the whole network, and we studied the diversity performance of ANC and DNC with coherent transmission. Although perfect CSI is very important for the receiver to mitigate error propagation for DNC or suppress MUI for ANC, it is not always available in practice. This could occur when the channels experience fast fading such that it is really hard to track the real-time channel variations at the receiver. Channel estimation overhead is another concern. For the multi-user relay channel, the channel estimation overhead increases linearly with the product of the number of users and the number of relays, which may become formidable and even outweigh the network coding gain. Besides, channel estimation requires additional wireless resources such as power and dedicated channels that could be otherwise used to transmit the data stream, so the bandwidth efficiency is also reduced to some extent. Because of these concerns, non-coherent transmission that do not require perfect CSI are of more practical interests under these circumstances.

Non-coherent transmission would reduce the performance. For the traditional point-to-point channel, it is well known that non-coherent transmission would incur 3dB SNR loss, while the diversity gain remains the same [1]. However, as for the

network-coded cooperation systems very few literatures have explicitly discussed the performance loss. So in this work, we study the non-coherent network-coded uplink and explore the impact on diversity gain. Depending on the available CSI, we first develop the coherent and non-coherent ML receivers. For ANC, as the non-coherent receivers have no closed form, we develop two suboptimum receivers according to the average link qualities. Next we study the error rates, and show that full diversity can always be achieved at extremely high SNRs regardless of the CSI assumptions. However, the lack of perfect CSI would incur some diversity loss at modest SNRs. Besides, the performance loss of ANC is more serious due to the incapability to efficiently suppress multi-user interferences at the receivers. Extensive simulations are performed to verify our analytical results.

Notations: $|\cdot|$, $(\cdot)^T$ and $(\cdot)^H$ stand for absolute value, transpose and conjugate transpose, respectively. The boldface lowercase letter $\mathbf{a}$ and the boldface uppercase letter $\mathbf{A}$ represent vector in column form and matrix, respectively. $\|\mathbf{a}\|$ and $\det \mathbf{A}$ denote the Euclidean norm of a vector $\mathbf{a}$ and the determinant of a square matrix $\mathbf{A}$, respectively. We shall use abbreviation i.i.d. for independent and identically distributed. We denote $Z \sim \mathcal{CN}(\mathbf{u}, \mathbf{\Sigma})$ as a circularly symmetric complex Gaussian random variable vector with mean $\mathbf{u}$ and covariance matrix $\mathbf{\Sigma}$, $Z \sim \exp(\mu)$ as an exponential random variable with mean μ, and $Z \sim \chi_k^2$ as a chi-square random variable with the degree of freedom being k. The probability of an event $\mathcal{A}$ is denoted by $\Pr(\mathcal{A})$. The CDF and PDF of a random variable Z are denoted by $F_Z(z)$ and $f_Z(z)$, respectively. Finally, we say $h(x) = \mathrm{O}(g(x))$ if $a \leq \lim_{x\to\infty} \frac{h(x)}{g(x)} \leq b$ for some positive constants a and b.

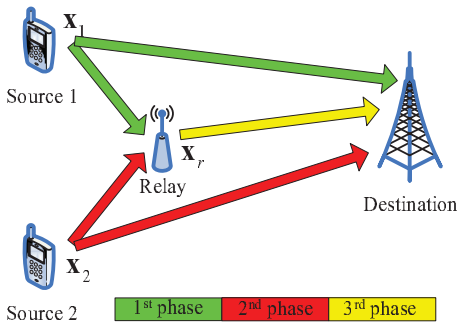


Figure 5.1: System model of the two-user network-coded uplink.

5.1 System Model

Consider the wireless uplink channel where two source nodes send data to a common destination with the help of a single relay node, as shown in Figure 5.1. Let $f_k \sim \mathcal{CN}(0,1)$ and $h_k \sim \mathcal{CN}(0,1)$ for $k = 1, 2$ be the channel coefficients from the kth source to the relay and to the destination, respectively, and $g \sim \mathcal{CN}(0,1)$ be the channel coefficient from the relay to the destination. All the channel coefficients are independent, and the additive noises on different channels are also i.i.d. $\mathcal{CN}(0,1)$. The channel gains are denoted by λ_{sr}, λ_{sd} and λ_{rd} for source-relay channels, source-destination channels and relay-destination channel, respectively. The channel gains are some constants that are determined by the distances and the path-loss exponents.

Only uncoded systems are considered throughout this work, i.e., there is no error detection/correction code. The two source nodes are supposed to use the M-ary FSK modulations. The symbol set is denoted by $\Omega = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_M\}$, where $\mathbf{e}_l$ is the lth unit vector with the lth element being 1 and the other elements being 0. The whole data transmission is completed in three phases. In the kth phase for $k = 1, 2$, the kth source node broadcasts its own signals while the other source

node keeps silent. The received signals at the relay node and at the destination are denoted by

$$\mathbf{y}_{kr} = \sqrt{P\lambda_{sr}} f_k \mathbf{x}_k + \mathbf{n}_{kr}, \tag{5.1}$$

$$\mathbf{y}_{kd} = \sqrt{P\lambda_{sd}} h_k \mathbf{x}_k + \mathbf{n}_{kd}, \tag{5.2}$$

respectively. Here P is the transmitted power, $\mathbf{x}_k$ is the kth source symbol with $\mathbf{x}_k \in \Omega$, $\mathbf{n}_{kr}$ and $\mathbf{n}_{kd}$ are the corresponding additive noises. As the noise power has been normalized, the transmitted power P is also treated as the system SNR throughout this work. The relay operation is dependent upon the network coding schemes that would be detailed later. At this moment, we simply denote the relay symbol by $\mathbf{x}_r$ and assume it has unit power, i.e., $E\|\mathbf{x}_r\|^2 = 1$. Then in the third phase, the relay node forwards its symbol to the destination while the two source nodes remain silent. The received signal is given by

$$\mathbf{y}_{rd} = \sqrt{P\lambda_{rd}} g \mathbf{x}_\mathbf{r} + \mathbf{n}_{rd}, \tag{5.3}$$

where $\mathbf{n}_{rd}$ is the additive noises. Upon observing $\mathbf{y}_{rd}$ and $\mathbf{y}_{kd}$ for $k = 1, 2$, the destination performs ML detection to jointly detect the two source symbols as

$$(\mathbf{x}_{d,1}, \mathbf{x}_{d,2}) = \arg \max_{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \in \Omega} L\left(\mathbf{y}_{rd} | \, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi\right) \times \prod_{k=1}^{2} L\left(\mathbf{y}_{kd} | \, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi\right), \tag{5.4}$$

where $L(\cdot)$ is the corresponding likelihood function, and Ψ is the set of instantaneous CSI available at the destination. For coherent detection, we have $\Psi_F = \{f_1, f_2, h_1, h_2, g\}$, i.e., global CSI is supposed to be known at the destination. Alternatively, the destination can choose to perform non-coherent detection with Ψ_S being the empty set, such that channel estimation is no longer required and all the

real-time channel variations are blind to the whole network. To facilitate explanations, we also define the partial coherent detection, where the destination only knows the relay-destination channel, i.e., $\Psi_P = \{g\}$. As the channel gains (i.e., λ) are only second-order statistics that remain unchanged over a long time, we assume these coefficients are known to all the nodes in the network. According to these definitions, we have

$$L\left(\mathbf{y}_{kd}|\,\hat{\mathbf{x}}_k, \Psi\right) = \begin{cases} p\left(\mathbf{y}_{kd} - \sqrt{P\lambda_{sd}} h_k \hat{\mathbf{x}}_k, \mathbf{I}\right), \ \Psi = \Psi_F & \text{(5.5a)} \\ p\left(\mathbf{y}_{kd}, P\lambda_{sd}\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^H + \mathbf{I}\right), \ \Psi \in \{\Psi_P, \Psi_S\} & \text{(5.5b)} \end{cases}$$

for $k = 1, 2$, where $\mathbf{I}$ is the identity matrix and $p\left(\mathbf{y}, \mathbf{\Sigma}\right) = \frac{1}{\pi^M |\mathbf{\Sigma}|} \exp\left(-\mathbf{y}^H \mathbf{\Sigma}^{-1} \mathbf{y}\right)$ is the PDF of $\mathbf{y} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}\right)$. The likelihood function of the relay link signal $\mathbf{y}_{rd}$ is related to the network coding schemes and would be detailed later.

5.2 Transceiver Design

In this section, we study the relay operation and the ML detection at the destination under different network coding schemes.

5.2.1 Analog Network Coding

If ANC is used at the relay node, the two received signals $\mathbf{y}_{kr}$ for $k = 1, 2$ are combined in the complex field directly by $\mathbf{x}_r = \sqrt{\alpha}\left(\mathbf{y}_{1r} + \mathbf{y}_{2r}\right)$, where

$$\alpha = \frac{1}{2\left(P\lambda_{sr} + M\right)} \tag{5.6}$$

is the amplification factor to normalize the relay power. From (5.3), the received signal $\mathbf{y}_{rd}$ can then be rewritten as

$$\mathbf{y}_{rd} = \sqrt{\alpha P \lambda_{rd}} g \sum_{k=1}^{2} \left(\sqrt{P\lambda_{sr}} f_k \mathbf{x}_k + \mathbf{n}_{kr} \right) + \mathbf{n}_{rd} = \sqrt{\alpha P^2 \lambda_{rd} \lambda_{sr}} g \sum_{k=1}^{2} f_k \mathbf{x}_k + \tilde{\mathbf{n}}_{rd}, \tag{5.7}$$

where $\tilde{\mathbf{n}}_{rd} = \sqrt{\alpha P \lambda_{rd}} g \sum_{k=1}^{2} \mathbf{n}_{kr} + \mathbf{n}_{rd}$ is the equivalent noises. The likelihood function of the relay link signal $\mathbf{y}_{rd}$ is then given by

$$L\left(\mathbf{y}_{rd} | \, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi_F\right) = p\left(\mathbf{y}_{rd} - \sqrt{\alpha P^2 \lambda_{rd} \lambda_{sr}} g \sum_{k=1}^{2} f_k \hat{\mathbf{x}}_k, \left(2\alpha P \lambda_{rd} |g|^2 + 1\right) \mathbf{I} \right), \tag{5.8}$$

$$L\left(\mathbf{y}_{rd} | \, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi_P\right) = p\left(\mathbf{y}_{rd}, \alpha P^2 \lambda_{rd} \lambda_{sr} |g|^2 \sum_{k=1}^{2} \hat{\mathbf{x}}_k \hat{\mathbf{x}}_k^H + \left(2\alpha P \lambda_{rd} |g|^2 + 1\right) \mathbf{I} \right) \tag{5.9}$$

for coherent and partial coherent detection, respectively. For non-coherent detection, the likelihood function can be obtained by averaging (5.9) over the distributions of g, i.e.,

$$L\left(\mathbf{y}_{rd} | \, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi_S\right) = E_g \left[L\left(\mathbf{y}_{rd} | \, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi_P\right) \right]. \tag{5.10}$$

Unfortunately, the above expression is an integral form that has no closed-form solution, which would greatly complicate the implementation. To simplify the receiver design, we develop two suboptimum yet efficient receivers in the sequel. To that end, let us first revisit the signal model (5.7). The aggregate scaling coefficient effective on the signal component from the input of the relay node to the destination is given by

$$\sqrt{\alpha P \lambda_{rd}} g = \sqrt{\frac{P\lambda_{rd}}{2\left(P\lambda_{sr} + M\right)}} g \overset{P \gg 1}{\approx} \sqrt{\frac{\lambda_{rd}}{2\lambda_{sr}}} g. \tag{5.11}$$

When the source-relay channel is much better than the relay-destination channel (i.e., $\lambda_{rd} \ll \lambda_{sr}$), the above scaling coefficient remains small with large probability,

whereas the noise power of $\mathbf{n}_{rd}$ is a constant. As a result, we can approximate g by its mean and obtain

$$L\left(\mathbf{y}_{rd}|\,\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_2,\Psi_S\right)\approx L\left(\mathbf{y}_{rd}|\,\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_2,\Psi_P,g=1\right), \tag{5.12}$$

which is called *fading elimination approximation* (FEA). The error performance is expected to remain similar because the channel fading only brings very limited effects when the scaling coefficient is small on average. On the other hand, if the source-relay channel is much worse than the relay-destination channel (i.e., $\lambda_{rd}\gg\lambda_{sr}$), then the noises $\mathbf{n}_{rd}$ would have much lower power than the scaled noises $\sqrt{\alpha P\lambda_{rd}}g\mathbf{n}_{kr}$. Consequently, we can intentionally neglect $\mathbf{n}_{rd}$ and obtain

$$\mathbf{y}_{rd}\approx\tilde{\mathbf{y}}_{rd}=\sqrt{\alpha P\lambda_{rd}}g\sum_{k=1}^{2}\left(\sqrt{P\lambda_{sr}}f_k\mathbf{x}_k+\mathbf{n}_{kr}\right). \tag{5.13}$$

To obtain the likelihood function of $\tilde{\mathbf{y}}_{rd}$, we first prove the following lemma.

Proposition 5.1. *Suppose* $\mathbf{v}\sim\mathcal{CN}\left(\mathbf{0},diag\left\{\sigma_{v_i}^2\right\}_{i=1}^{M}\right)$ *and* $u\sim\mathcal{CN}\left(0,\sigma_u^2\right)$ *are independent random variables, then the PDF of* $\mathbf{z}=u\mathbf{v}$ *is*

$$f\left(\mathbf{z}\right)=\left(\prod_{i=1}^{M}\frac{2}{\pi\sigma_u^2\sigma_{v_i}^2}\right)q\left(\frac{4}{\sigma_u^2}\sum_{i=1}^{M}\frac{|z_i|^2}{\sigma_{v_i}^2}\right), \tag{5.14}$$

where $q\left(x\right)=x^{-\frac{M-1}{2}}K_{M-1}\left(\sqrt{x}\right)$ *and* $K_M\left(x\right)$ *is the* M*th-order modified Bessel function of the second kind [87, 9.6.1].*

Proof. Denote $z_i=r_ie^{j\theta_i}$ for $i=1,2,\cdots,M$, then it is easy to show that the phases $\{\theta_i\}$ are independent of the amplitudes $\{r_i\}$, and $\{\theta_i\}$ are i.i.d. and uniformly distributed on $[0,2\pi)$. Therefore,

$$f\left(\mathbf{z}\right)=\frac{1}{|J|}f\left(\mathbf{r},\boldsymbol{\theta}\right)=\prod_{i=1}^{M}\left(2\pi r_i\right)^{-1}f\left(\mathbf{r}\right), \tag{5.15}$$

where $|J| = \prod_{i=1}^{M} r_i$ is the Jacobian determinant. The CDF of $\mathbf{r}$ is given by

$$F(\mathbf{r}) = \int_0^{\infty} \frac{1}{\sigma_u^2} \exp\left(-\frac{x}{\sigma_u^2}\right) \prod_{i=1}^{M} \left(1 - \exp\left(-\frac{r_i^2}{\sigma_{v_i}^2 x}\right)\right) dx. \tag{5.16}$$

After taking derivatives, we can obtain

$$\begin{aligned} f(\mathbf{r}) &= \frac{1}{\sigma_u^2} \prod_{i=1}^{M} \frac{2r_i}{\sigma_{v_i}^2} \int_0^{\infty} x^{-M} \exp\left(-\frac{x}{\sigma_u^2} - \frac{1}{x} \sum_{i=1}^{M} \frac{r_i^2}{\sigma_{v_i}^2}\right) dx \\ &= \left(\prod_{i=1}^{M} \frac{4r_i}{\sigma_u^2 \sigma_{v_i}^2}\right) q\left(\frac{4}{\sigma_u^2} \sum_{i=1}^{M} \frac{r_i^2}{\sigma_{v_i}^2}\right), \end{aligned} \tag{5.17}$$

where we use [88, 3.478.4] in the last equality. Plugging (5.17) back into (5.15) completes the proof. □

According to Proposition 5.1, the likelihood function of $\tilde{\mathbf{y}}_{rd}$ can be obtained after redefining the parameters in (5.14). To be specific, for $\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2 = \mathbf{e}_l$ we have $\sigma_u^2 = \alpha P \lambda_{rd}$, $\sigma_{v_l}^2 = 2(P\lambda_{sr} + 1)$ and $\sigma_{v_i}^2 = 2$ for $i \neq l$, whereas for $\hat{\mathbf{x}}_1 = \mathbf{e}_k, \hat{\mathbf{x}}_2 = \mathbf{e}_l$ with $k \neq l$ we have $\sigma_u^2 = \alpha P \lambda_{rd}$, $\sigma_{v_k}^2 = \sigma_{v_l}^2 = P\lambda_{sr} + 2$ and $\sigma_{v_i}^2 = 2$ for $i \neq l, k$. In later sections, this scheme is referred to as *noise elimination approximation* (NEA).

5.2.2 Digital Network Coding

For DNC, the two source symbols are first individually detected at the relay node as

$$\mathbf{x}_{r,k} = \begin{cases} \arg\max\limits_{\hat{\mathbf{x}}_k \in \Omega} p\left(\mathbf{y}_{kr} - \sqrt{P\lambda_{sr}} f_k \hat{\mathbf{x}}_k, \mathbf{I}\right), \Psi = \Psi_F & \text{(5.18a)} \\ \arg\max\limits_{\hat{\mathbf{x}}_k \in \Omega} p\left(\mathbf{y}_{kr}, P\lambda_{sr} \hat{\mathbf{x}}_k \hat{\mathbf{x}}_k^H + \mathbf{I}\right), \Psi = \Psi_S & \text{(5.18b)} \end{cases}$$

where the relay node is assumed to know the same type of CSI as the destination. Next, the detected source symbols are combined in the finite field, and the relay

symbol is given by $\mathbf{x}_r = \mathbf{x}_{r,1} \oplus \mathbf{x}_{r,2}$. We remark that the conventional DNC is performed bitwise through exclusive-or operations. Here to simplify the notations we omit the mapping between the bits and symbols and express the network-coded output in the symbol space directly. It is also noteworthy that $\mathbf{x}_r$ belongs to the M-ary FSK symbol set Ω as well, but it maybe different from the true network-coded symbol $\mathbf{x}_\oplus = \mathbf{x}_1 \oplus \mathbf{x}_2$ due to random detection errors.

The likelihood function of the relay link signal $\mathbf{y}_{rd}$ is

$$L\left(\mathbf{y}_{rd}|\,\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi\right) = \sum_{k=1}^{M} \Pr\left(\hat{\mathbf{x}}_r = \mathbf{e}_k|\,\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi\right) L\left(\mathbf{y}_{rd}|\,\hat{\mathbf{x}}_r = \mathbf{e}_k, \Psi\right), \tag{5.19}$$

where the second term within the summation has the same form as (5.5) after properly modifying the subscripts, and the first term is given by

$$\Pr\left(\hat{\mathbf{x}}_r = \mathbf{e}_k|\,\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \Psi\right) = \sum_{\hat{\mathbf{x}}_{r,1} \oplus \hat{\mathbf{x}}_{r,2} = \mathbf{e}_k} \Pr\left(\hat{\mathbf{x}}_{r,k}|\,\hat{\mathbf{x}}_1, \Psi\right) \Pr\left(\hat{\mathbf{x}}_{r,2}|\,\hat{\mathbf{x}}_2, \Psi\right). \tag{5.20}$$

Note that each term within the summation stands for the transition probability from the trial source symbol $\hat{\mathbf{x}}_k$ to the trial relay detected symbol $\hat{\mathbf{x}}_{r,k}$. For FSK modulations, the error rates at the relay node can be obtained from [1, 5.2.19] and [1, 5.4.42] after properly accommodating the channel fading, which are given by

$$P_{r,k} = \begin{cases} 1 - \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} \left(1 - Q\left(x\right)\right)^{M-1} e^{-\frac{1}{2}\left(x - \sqrt{2P\lambda_{sr}|f_k|^2}\right)^2} dx, \ \Psi = \Psi_F & \text{(5.21a)} \\ \displaystyle\sum_{k=1}^{M-1} \binom{M-1}{k} \frac{(-1)^{k+1}}{1 + k\left(1 + P\lambda_{sr}\right)}, \ \Psi = \Psi_S & \text{(5.21b)} \end{cases}$$

where $Q\left(x\right)$ is Q-function. Due to symmetry, the detection errors of FSK modulations are uniformly distributed in the error symbol set, so each transition probability

is given by

$$\Pr\left(\hat{\mathbf{x}}_{r,k}|\,\hat{\mathbf{x}}_k,\Psi\right)=\begin{cases}\dfrac{1}{M-1}P_{r,k},\ \hat{\mathbf{x}}_{r,k}\neq\hat{\mathbf{x}}_k & \text{(5.22a)}\\ 1-P_{r,k},\ \hat{\mathbf{x}}_{r,k}=\hat{\mathbf{x}}_k & \text{(5.22b)}\end{cases}$$

for $k=1,2$. Note that for coherent detection, the destination knows the instantaneous relay detection errors, whereas for non-coherent detection only the average relay detection errors are supposed to be known at the destination.

5.3 Error Performance Analysis

In this section, we study the error rates of the aforementioned systems. Unfortunately, the exact error rates are analytically intractable due to the complex ML decision regions. Instead, we study the PEPs, which are defined as the probability of mistaking the true symbols $(\mathbf{x}_1,\mathbf{x}_2)$ by another trial symbols $(\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_2)$. It is well known that the real error rates can be approximately characterized by the dominant PEPs. Without loss of generality, we focus only on the binary FSK modulations throughout this section. Due to symmetry, there are four types of PEPs as listed in Table 5.1.

Table 5.1: Four Types of PEP

Notations	**True Symbols**	**Trial Symbols**
P_1	$(\mathbf{e}_1,\mathbf{e}_1)$	$(\mathbf{e}_2,\mathbf{e}_2)$
P_2	$(\mathbf{e}_1,\mathbf{e}_2)$	$(\mathbf{e}_2,\mathbf{e}_1)$
P_3	$(\mathbf{e}_1,\mathbf{e}_1)$	$(\mathbf{e}_1,\mathbf{e}_2)$
P_4	$(\mathbf{e}_1,\mathbf{e}_2)$	$(\mathbf{e}_1,\mathbf{e}_1)$

5.3.1 Coherent Analog Network Coding

After some manipulations, we can obtain the conditional PEPs given by

$$\mathrm{PEP}|_{\Psi_F} = Q\left(\sqrt{\frac{1}{2}\left(P\lambda_{sd}\sum_{k=1}^{2}|h_k|^2\|\mathbf{x}_k-\hat{\mathbf{x}}_k\|^2+\frac{\alpha P^2\lambda_{rd}\lambda_{sr}|g|^2}{1+2\alpha P\lambda_{rd}|g|^2}\left\|\sum_{k=1}^{2}f_k\left(\mathbf{x}_k-\hat{\mathbf{x}}_k\right)\right\|^2\right)}\right). \tag{5.23}$$

To obtain the average PEPs, we need to average the above expression over the channel distributions, which is analytically intractable. Instead, we seek to develop some bounds that have the same scaling laws as the true PEPs at high SNRs.

To obtain the PEP lower bound, we use the integral representation of Q-function given by [85]

$$Q\left(x\right) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\exp\left(-\frac{x^2}{2\sin^2\theta}\right)d\theta. \tag{5.24}$$

After replacing the Q-function in (5.23) with the above expression and averaging over the channel distributions by use of (4.9), we can obtain

$$\begin{aligned}\mathrm{PEP} \geq & \frac{(2n+1)!!}{(2n+2)!!}\frac{2^{n-1}}{\left(2+P\lambda_{sd}\right)^n} \\ & \times\left(\frac{4}{4+nP\lambda_{sr}}+\frac{n\lambda_{sr}}{\alpha\lambda_{rd}(4+nP\lambda_{sr})^2}\log\left(1+4\alpha P\lambda_{rd}+\alpha nP^2\lambda_{rd}\lambda_{sr}\right)\right),\end{aligned} \tag{5.25}$$

where $(\cdot)!!$ is double factorial, $n = \mathbb{1}\{\mathbf{x}_1 \neq \hat{\mathbf{x}}_1\} + \mathbb{1}\{\mathbf{x}_2 \neq \hat{\mathbf{x}}_2\}$ with $\mathbb{1}\{\cdot\}$ being indicator function, and we use $\frac{1}{2}\log\left(1+\frac{2}{x}\right) < e^x E_1\left(x\right)$ [87, 5.1.20] in the inequality. To obtain the PEP upper bound, we use the Chernoff bound [1] of Q-function given by $Q\left(x\right) \leq \frac{1}{2}\exp\left(-\frac{1}{2}x^2\right)$. In a similar way, we have

$$\begin{aligned}\mathrm{PEP} \leq & \frac{2^{n-1}}{\left(2+P\lambda_{sd}\right)^n} \\ & \times\left(\frac{4}{4+nP\lambda_{sr}}+\frac{2nP\lambda_{sr}}{\alpha P\lambda_{rd}(4+nP\lambda_{sr})^2}\log\left(1+2\alpha P\lambda_{rd}+\frac{1}{2}\alpha nP^2\lambda_{rd}\lambda_{sr}\right)\right),\end{aligned} \tag{5.26}$$

where we use $e^x E_1(x) < \log\left(1+\frac{1}{x}\right)$ [87, 5.1.20] in the inequality. From (5.23), it is observed that Type-1 PEP and Type-2 PEP have the same conditional distributions, and Type-3 PEP and Type-4 PEP also have the same conditional distributions. That is why the PEP bounds (5.25) and (5.26) depend only on n but not the true and trial symbol pairs. At high SNRs (i.e. $P \gg 1$), we can show that both the upper bound and lower bound scale as $\mathrm{O}\left(\frac{\log P}{P^{n+1}}\right)$. Clearly, the dominant error events occur when only one of the two source symbols is detected incorrectly at the destination (i.e., $n=1$), and the error rates of coherent ANC scale as $\mathrm{O}\left(\frac{\log P}{P^2}\right)$ at high SNRs.

5.3.2 Non-Coherent Analog Network Coding

Since the non-coherent ML receiver (5.10) has an intractable integral form, we study the partial coherent receiver (5.9). As will be shown in simulations, the performances of these two receivers are very close to each other.

For Type-1 PEP, we have

$$\begin{aligned} P_1 &= \Pr\left\{\frac{\lambda_{sd}P}{1+\lambda_{sd}P}\left(U_2^1+V_2^1\right)+\frac{\lambda_{sr}PQ}{1+\lambda_{sr}PQ}W_2^1 \geq \lambda_{sd}P\left(U_1^1+V_1^1\right)+\lambda_{sr}PQW_1^1\right\} \\ &\leq \Pr\left\{U_2^1+V_2^1+W_2^1 \geq PY\right\}, \end{aligned} \tag{5.27}$$

where $W_1^1 = \frac{\left|y_{rd,1}\right|^2}{1+2\alpha P\lambda_{rd}|g|^2+2\alpha P^2\lambda_{rd}\lambda_{sr}|g|^2}$, $U_1^1 = \frac{\left|y_{1d,1}\right|^2}{1+\lambda_{sd}P}$, $V_1^1 = \frac{\left|y_{2d,1}\right|^2}{1+\lambda_{sd}P}$, $W_2^1 = \frac{\left|y_{rd,2}\right|^2}{1+2\alpha P\lambda_{rd}|g|^2}$, $U_2^1 = |y_{1d,2}|^2$, $V_2^1 = |y_{2d,2}|^2$ are i.i.d. exponential random variables with unit mean, and

$$Q = \frac{2\alpha P\lambda_{rd}|g|^2}{1+2\alpha P\lambda_{rd}|g|^2}. \tag{5.28}$$

Since $U_2^1+V_2^1+W_2^1 \sim \frac{1}{2}\chi_6^2$, the conditional PEP given $Y \triangleq \lambda_{sd}\left(U_1^1+V_1^1\right)+$

$\lambda_{sr} Q W_1^1 = y$ is

$$P_1|_{Y=y} \leq \left(1 + Py + \frac{1}{2}P^2 y^2\right) \exp\left(-Py\right). \tag{5.29}$$

At high SNRs, the conditional PEP decreases really fast with y. Therefore, the average PEP is roughly determined by the behavior of the PDF of $Y \ll 1$. Denoting $T \triangleq Q W_1^1$, we can obtain

$$F_Y\left(y\right) \overset{y \ll 1}{\approx} \int_0^{\frac{y}{\lambda_{sr}}} f_T\left(t\right) \frac{\left(y - \lambda_{sr} t\right)^2}{\lambda_{sd}^2} dt, \tag{5.30}$$

Taking derivative with respect to y leads to

$$f_Y\left(y\right) \overset{y \ll 1}{\approx} \frac{2\lambda_{sr}}{\lambda_{sd}^2} \int_0^{\frac{y}{\lambda_{sr}}} F_T\left(t\right) dt = -\frac{y^2}{2\alpha P \lambda_{sr} \lambda_{rd} \lambda_{sd}^2} \left(\log \frac{y}{\lambda_{sr}} - \frac{1}{2}\right), \tag{5.31}$$

where we use $F_T\left(t\right) \overset{t \ll 1}{\approx} -\frac{1}{2\alpha P \lambda_{rd}} t \log t$ from (4.9) in the second equality. Using the above PDF to average the conditional PEP in (5.29), we obtain

$$P_1 \leq \frac{20 \log\left(\sqrt{e} P \lambda_{sr}\right) - 39 + 20\gamma}{2\alpha P^4 \lambda_{sr} \lambda_{rd} \lambda_{sd}^2} = \mathrm{O}\left(\frac{\log P}{P^3}\right), \tag{5.32}$$

where γ is Euler constant [88, 4.352.2].

For Type-2 PEP, the likelihood function of $\mathbf{y}_{rd}$ remains the same under both hypotheses, which greatly simplifies the computations. After some manipulations, we can obtain

$$P_2 = \Pr\left\{\left|y_{1d,2}\right|^2 + \left|y_{2d,1}\right|^2 \geq \left|y_{1d,1}\right|^2 + \left|y_{2d,2}\right|^2\right\} = \frac{4 + 3\lambda_{sd} P}{\left(2 + \lambda_{sd} P\right)^3} = \mathrm{O}\left(\frac{1}{P^2}\right). \tag{5.33}$$

For Type-3 PEP, we have

$$\begin{aligned} P_3 = \Pr\Bigg\{ & \frac{\lambda_{sd} P}{1 + \lambda_{sd} P} V_2^3 + \frac{\lambda_{sr} PQ}{2 + \lambda_{sr} PQ} W_2^3 \\ & \geq \lambda_{sd} P V_1^3 + \frac{\lambda_{sr} PQ}{2 + \lambda_{sr} PQ} W_1^3 + \log \frac{\left(2 + \lambda_{sr} PQ\right)^2}{4\left(1 + \lambda_{sr} PQ\right)} \Bigg\} \\ \leq \Pr & \left\{W_2^3 + V_2^3 \geq \lambda_{sd} P V_1^3 + \log Z\right\}, \end{aligned} \tag{5.34}$$

where $W_1^3 = \frac{\left|y_{rd,1}\right|^2}{1+2\alpha P\lambda_{rd}|g|^2+2\alpha P^2\lambda_{rd}\lambda_{sr}|g|^2}$, $V_1^3 = \frac{\left|y_{2d,1}\right|^2}{1+\lambda_{sd}P}$, $W_2^3 = \frac{\left|y_{rd,2}\right|^2}{1+2\alpha P\lambda_{rd}|g|^2}$, and $V_2^3 = |y_{2d,2}|^2$ are i.i.d. exponential random variables with unit mean, and $Z = \frac{2+\lambda_{sr}PQ}{4} \leq \frac{1}{2} + \frac{P\lambda_{sr}}{4}$. The conditional PEP given $Z = z$ is given by

$$P_3|_{Z=z} \leq \frac{(1+2\lambda_{sd}P) + (1+\lambda_{sd}P)\log\left(\frac{1}{2} + \frac{P\lambda_{sr}}{4}\right)}{(1+\lambda_{sd}P)^2 z}. \tag{5.35}$$

By defining $\eta_1 = \alpha P\lambda_{rd}(2+\lambda_{sr}P)$, we can further obtain

$$E\left(\frac{1}{Z}\right) = \frac{1}{\eta_1}\left(4\alpha\lambda_{rd}P + \frac{2\lambda_{sr}P}{2+\lambda_{sr}P}\exp\left(\frac{1}{\eta_1}\right)E_1\left(\frac{1}{\eta_1}\right)\right). \tag{5.36}$$

After combining (5.35) and (5.36) and using the inequality $e^x E_1(x) \leq \log\left(1+\frac{1}{x}\right)$ [87, 5.1.20], we obtain $P_3 \leq \mathrm{O}\left(\frac{\log^2 P}{P^2}\right)$

Finally for Type-4 PEP, we have

$$P_4 = \Pr\left\{\frac{\lambda_{sd}P}{1+\lambda_{sd}P}V_1^4 + \frac{\lambda_{sr}PQ}{2(1+\lambda_{sr}PQ)}W_1^4 + \log\frac{(2+\lambda_{sr}PQ)^2}{4(1+\lambda_{sr}PQ)} \geq \lambda_{sd}PV_2^4 + \frac{1}{2}\lambda_{sr}PQW_2^4\right\}, \tag{5.37}$$

where $W_1^4 = \frac{\left|y_{rd,1}\right|^2}{1+2\alpha P\lambda_{rd}|g|^2+\alpha P^2\lambda_{rd}\lambda_{sr}|g|^2}$, $V_1^4 = |y_{2d,1}|^2$, $W_2^4 = \frac{\left|y_{rd,2}\right|^2}{1+2\alpha P\lambda_{rd}|g|^2+\alpha P^2\lambda_{rd}\lambda_{sr}|g|^2}$, $V_2^4 = \frac{\left|y_{2d,2}\right|^2}{1+\lambda_{sd}P}$ are i.i.d. exponential random variables with unit mean. As the logarithmic term is upper bounded by $\log\frac{(2+\lambda_{sr}P)^2}{4(1+\lambda_{sr}P)} \triangleq \eta_2$, we have

$$P_4 \leq \underbrace{\Pr\left\{W_1^{IV} + V_1^{IV} \geq \max\left\{\frac{1}{2}\lambda_{sd}PV_2^{IV}, \eta_2\right\}\right\}}_{\triangleq P_4^{U_1}} + \underbrace{\Pr\left\{2\eta_2 \geq \lambda_{sd}PV_2^{IV} + \frac{1}{2}\lambda_{sr}PQW_2^{IV}\right\}}_{\triangleq P_4^{U_2}}. \tag{5.38}$$

After some lengthy algebra, we have

$$\begin{aligned} P_4^{U_1} &= (1+\eta_2)\, e^{-\eta_2} - \left(\frac{\lambda_{sd}P}{2+\lambda_{sd}P}\right)^2 \left(1+\frac{2+\lambda_{sd}P}{\lambda_{sd}P}\eta_2\right) e^{-\frac{2+\lambda_{sd}P}{\lambda_{sd}P}\eta_2} \\ &= \mathrm{O}\left(\frac{\log^2 P}{P^2}\right), \end{aligned} \tag{5.39}$$

$$P_4^{U_2} \overset{P\gg 1}{\approx} \frac{\eta_2^2}{\alpha P^3 \lambda_{sr}\lambda_{sd}\lambda_{rd}} \left(1 - 2\log\left(\frac{4\eta_2}{\lambda_{sr}P}\right)\right) = \mathrm{O}\left(\frac{\log^3 P}{P^2}\right). \tag{5.40}$$

Consequently, we have $P_4 \leq \mathrm{O}\left(\frac{\log^3 P}{P^2}\right)$, which appears to dominate all the PEPs. To make the argument rigorous, yet we still need to show that this is the exact scaling law of Type-4 PEP. To that end, we develop a lower bound on P_4 by neglecting the first two terms on the left-hand side of the inequality in (5.37), which leads to

$$P_4 \geq P_4^L = \Pr\left\{\log\frac{2+\lambda_{sr}PQ}{4} \geq \lambda_{sd}PV_2^4 + \frac{1}{2}\lambda_{sr}PQW_2^4\right\}. \tag{5.41}$$

The conditional probability of P_4^L given $Q=q$ is given by

$$\begin{aligned} P_4^L\big|_{Q=q} = 1 &- \frac{2\lambda_{sd}}{2\lambda_{sd}-q\lambda_{sr}} \exp\left(-\frac{1}{\lambda_{sd}P}\log\frac{2+q\lambda_{sr}P}{4}\right) \\ &+ \frac{q\lambda_{sr}}{2\lambda_{sd}-q\lambda_{sr}} \exp\left(-\frac{2}{q\lambda_{sr}P}\log\frac{2+q\lambda_{sr}P}{4}\right). \end{aligned} \tag{5.42}$$

When $q \geq P^{-\beta}$ for any constant $\beta \in (0,1)$, we have $qP \geq P^{1-\beta} \gg 1$ for $P \gg 1$, thus

$$P_4^L\big|_{Q=q} \overset{P\gg 1}{\approx} \frac{1}{q\lambda_{sr}\lambda_{sd}P^2}\left(\log\frac{2+q\lambda_{sr}P}{4}\right)^2. \tag{5.43}$$

The final step is to average the above expression over the distribution of Q. The PDF of Q is given by

$$f_Q(q) = \frac{1}{2\alpha P\lambda_{rd}(1-q)^2} \exp\left(-\frac{q}{2\alpha P\lambda_{rd}(1-q)}\right),\ 0 \leq q \leq 1 \tag{5.44}$$

It is easy to see that $f_Q(q)$ is a continuous function with $f_Q(0) = \frac{1}{2\alpha P\lambda_{rd}}$ and $f_Q(1) = 0$, therefore it is lower bounded by some constant C on the region $q \in [0, \eta_3]$

with $\eta_3 < 1$ being some fixed number. Using the above facts, we can obtain

$$\begin{aligned} P_4^L &\geq \int_{P^{-\beta}}^{\eta_3} \frac{1}{q\lambda_{sr}\lambda_{sd}P^2}\left(\log\frac{2+q\lambda_{sr}P}{4}\right)^2 f_Q(q)\,dq \\ &\geq \frac{C\log\left(\eta_3 P^{\beta}\right)}{\lambda_{sr}\lambda_{sd}P^2}\left(\log\frac{2+\lambda_{sr}P^{1-\beta}}{4}\right)^2 = \mathrm{O}\left(\frac{\log^3 P}{P^2}\right). \end{aligned} \tag{5.45}$$

As the upper bound and lower bound on P_4 have exactly the same scaling laws, we conclude that the error rates of partial coherent ANC scale as $\mathrm{O}\left(\frac{\log^3 P}{P^2}\right)$ at high SNRs.

5.3.3 Digital Network Coding

For DNC, we assume $\mathbf{x}_1 \oplus \mathbf{x}_2 = \mathbf{e}_1$ for $\mathbf{x}_1 = \mathbf{x}_2$ and $\mathbf{x}_1 \oplus \mathbf{x}_2 = \mathbf{e}_2$ for $\mathbf{x}_1 \neq \mathbf{x}_2$. The relay symbol $\mathbf{x}_r = \mathbf{x}_{r,1} \oplus \mathbf{x}_{r,2}$ is a single binary FSK symbol that carries information for both sources. Thus the likelihood function of the relay link signal $\mathbf{y}_{rd}$ depends only on the true network-coded symbol $\mathbf{x}_{\oplus}$ rather than the source symbol pair $(\mathbf{x}_1, \mathbf{x}_2)$. From Table I, we can observe that regardless of the type of receiver, Type-1 PEP and Type-2 PEP are always the same, and Type-3 PEP and Type-4 PEP are the same as well. So in the sequel, we only study Type-1 PEP and Type-3 PEP, in which cases the true source symbols are $(\mathbf{e}_1, \mathbf{e}_1)$ and the true network-coded symbol is $\mathbf{x}_{\oplus} = \mathbf{e}_1 \oplus \mathbf{e}_1 = \mathbf{e}_1$.

To proceed, we first define

$$t_{kd} = \begin{cases} 2\sqrt{P\lambda_{sd}}\mathrm{Re}\left\{h_k^*\left(y_{kd,1} - y_{kd,2}\right)\right\}, \ \Psi = \Psi_F & \text{(5.46a)} \\ \dfrac{P\lambda_{sd}}{1+P\lambda_{sd}}\left(|y_{kd,1}|^2 - |y_{kd,2}|^2\right), \ \Psi = \Psi_S & \text{(5.46b)} \end{cases}$$

for $k = 1, 2$ and

$$t_{rd} = \begin{cases} 2\sqrt{P\lambda_{rd}}\mathrm{Re}\left\{g^*\left(y_{rd,1} - y_{rd,2}\right)\right\}, \ \Psi = \Psi_F & (5.47a) \\ \dfrac{P\lambda_{rd}}{1+P\lambda_{rd}}\left(|y_{rd,1}|^2 - |y_{rd,2}|^2\right), \ \Psi = \Psi_S & (5.47b) \end{cases}$$

It can be shown that the above metrics can be rewritten as the quadratic forms of some independent complex Gaussian random variables. According to [86], the PDF of t_{kd} is $f_{t_{kd}}(t) = f(a_{sd}, b_{sd}, t)$ for $k = 1, 2$, the conditional PDF of t_{rd} given $\mathbf{x}_r = \mathbf{x}_\oplus$ is $f_{t_{rd}}(t)|_{\mathbf{x}_r = \mathbf{x}_\oplus} = f(a_{rd}, b_{rd}, t)$, and the conditional PDF of t_{rd} given $\mathbf{x}_r \neq \mathbf{x}_\oplus$ is $f_{t_{rd}}(t)|_{\mathbf{x}_r \neq \mathbf{x}_\oplus} = f(b_{rd}, a_{rd}, t)$, where

$$f(a, b, t) = \frac{ab}{a+b}\left(\exp(bt)\,\mathbb{1}\{t < 0\} + \exp(-at)\,\mathbb{1}\{t \geq 0\}\right), \tag{5.48}$$

and for $k \in \{s, r\}$, $b_{kd} = a_{kd} + 1$ and

$$a_{kd} = \begin{cases} \dfrac{1}{2}\left(\sqrt{1 + \dfrac{2}{P\lambda_{kd}}} - 1\right), \ \Psi = \Psi_F & (5.49a) \\ \dfrac{1}{P\lambda_{kd}}, \ \Psi = \Psi_S & (5.49b) \end{cases}$$

After some algebra, we can show that Type-1 PEP is given by

$$P_1 = \Pr\left(\sum_{k=1}^{2} t_{kd} \leq 0\right) = \frac{a_{sd}^2\left(a_{sd} + 3b_{sd}\right)}{\left(a_{sd} + b_{sd}\right)^3}. \tag{5.50}$$

At high SNRs, Type-1 PEP can be approximated by

$$P_1 \overset{P \gg 1}{\approx} \begin{cases} \dfrac{3}{4P^2\lambda_{sd}^2}, \ \text{coherent} & (5.51a) \\ \dfrac{3}{P^2\lambda_{sd}^2}, \ \text{non-coherent} & (5.51b) \end{cases}$$

and the scaling law is $P_1 = \mathrm{O}\left(\frac{1}{P^2}\right)$ regardless of the type of the receiver. As for Type-3 PEP, we have

$$P_3 = \Pr\left(t_{2d} + \log\frac{(1 - P_r)\,\mathrm{e}^{t_{rd}} + P_r}{P_r\mathrm{e}^{t_{rd}} + (1 - P_r)} \leq 0\right), \tag{5.52}$$

where P_r is the relay detection error rate of the network-coded symbol $\mathbf{x}_{\oplus}$. For coherent detection and non-coherent detection, P_r are respectively given by

$$P_r = \prod_{k=1}^{2} Q\left(\sqrt{P\lambda_{sr}|f_k|^2}\right)\left(1 - Q\left(\sqrt{P\lambda_{sr}\left|f_{\{1,2\}\backslash\{k\}}\right|^2}\right)\right) \triangleq P_{r,F}, \tag{5.53}$$

$$P_r = \frac{2\left(1+P\lambda_{sr}\right)}{\left(2+P\lambda_{sr}\right)^2} \triangleq P_{r,S}. \tag{5.54}$$

As the exact analysis is intractable, we use the piecewise linear approximation [84] given by

$$\log\frac{\left(1-P_r\right)\mathrm{e}^{t_{rd}}+P_r}{P_r\mathrm{e}^{t_{rd}}+\left(1-P_r\right)} \approx \begin{cases} T_r, t_{rd} \geq T_r \\ t_{rd}, -T_{r,S} \leq t_{rd} \leq T_r \\ -T_r, t_{rd} \leq -T_r \end{cases} \tag{5.55}$$

where $T_r = \log\frac{1-P_r}{P_r}$. After applying the above approximation in (5.52) and averaging over the channel distributions of $\{h_1, h_2, g\}$, we can obtain the conditional Type-3 PEP given $\{f_1, f_2\}$ as

$$\begin{aligned} P_3|_{\{f_1,f_2\}} &= \left(1-P_r\right)h_0\left(a_{rd},b_{rd},a_{sd},b_{sd}\right) + P_r h_0\left(b_{rd},a_{rd},a_{sd},b_{sd}\right) \\ &-\frac{a_{sd}}{b_{rd}+a_{sd}}\left(1-P_r\right)h_1\left(b_{rd},a_{rd},a_{sd},b_{sd},T_r\right) - \frac{a_{sd}}{a_{rd}+a_{sd}}P_r h_1\left(a_{rd},b_{rd},a_{sd},b_{sd},T_r\right) \\ &+\frac{b_{sd}}{a_{rd}+b_{sd}}\left(1-P_r\right)h_1\left(a_{rd},b_{rd},b_{sd},a_{sd},T_r\right) \\ &+\frac{b_{sd}}{b_{rd}+b_{sd}}P_r h_1\left(b_{rd},a_{rd},b_{sd},a_{sd},T_r\right), \end{aligned} \tag{5.56}$$

where

$$h_0\left(a,b,c,d\right) = \frac{a}{a+b}\left\{1+\frac{bc}{\left(c+d\right)\left(a+d\right)} - \frac{bd}{\left(c+d\right)\left(b+c\right)}\right\}, \tag{5.57}$$

$$h_1\left(a,b,c,d,t\right) = \frac{b}{a+b}\frac{d}{c+d}\exp\left(-\left(a+c\right)t\right). \tag{5.58}$$

For non-coherent detection, (5.56) is exactly the average Type-3 PEP, because $P_{r,S}$ in (5.54) is independent of the source-relay channels $\{f_1, f_2\}$. At high SNRs, it can

be approximated by

$$P_3 \overset{P\gg 1}{\approx} \frac{4+2\log\left(\frac{P\lambda_{sr}}{2}\right)}{P^2\lambda_{sr}\lambda_{sd}} + \frac{3}{P^2\lambda_{rd}\lambda_{sd}} = \mathrm{O}\left(\frac{\log P}{P^2}\right). \tag{5.59}$$

By comparing (5.51b) and (5.59), we observe that Type-3 PEP is the dominant error term, thus the error rates of non-coherent DNC scale as $\mathrm{O}\left(\frac{\log P}{P^2}\right)$ at high SNRs.

Next we study the average Type-3 PEP of coherent DNC, which can be obtained after averaging (5.56) over the channel distributions of f_1 and f_2. Since the exact analysis is intractable, we shall make some approximations. As the first step, we approximate $P_{r,F}$ in (5.53) by

$$P_{r,F} \approx Q\left(\sqrt{\frac{1}{2}P\lambda_{sr}|f|^2}\right) \approx h_2\left(\sqrt{\frac{1}{2}P\lambda_{sr}|f|^2}\right), \tag{5.60}$$

where $|f|^2 = 2\min\left(|f_1|^2, |f_2|^2\right) \sim \exp(1)$, and

$$h_2(x) = \frac{1}{B\sqrt{2\pi}x}\left(1-\exp\left(-\frac{A}{\sqrt{2}}x\right)\right)\exp\left(-\frac{1}{2}x^2\right), A = 1.98, B = 1.135. \tag{5.61}$$

The first approximation is quite tight from modest to high SNRs, as the relay detection performance is roughly determined by the worse source-relay channel. The advantage of such approximation is that the conditional PEP (5.56) now relies only on the single variable $|f|$. As for the second approximation, we use $Q(x) \approx h_2(x)$ [95]. Next, we observe that the average PEP depends largely on the small values of $|f|$, as the conditional PEP decreases very fast with the variable $|f|$ at high SNRs. As a result, we can approximate $T_{r,F}$ by

$$T_{r,F} = \log\frac{1-P_{r,F}}{P_{r,F}} \overset{|f|\ll 1}{\approx} \sqrt{\frac{4}{\pi}P\lambda_{sr}|f|^2}. \tag{5.62}$$

After plugging (5.60) and (5.62) back into (5.56) and doing some lengthy algebra,

we can obtain

$$\begin{aligned} P_3 \;\approx\; & h_0\left(a_{rd},b_{rd},a_{sd},b_{sd}\right)+h_5\left(0,1,0,1,\frac{1}{2}P\lambda_{sr}\right)h_0\left(b_{rd},a_{rd},a_{sd},b_{sd}\right) \\ & -h_5\left(0,1,0,1,\frac{1}{2}P\lambda_{sr}\right)h_0\left(a_{rd},b_{rd},a_{sd},b_{sd}\right) \\ & -\frac{a_{sd}}{b_{rd}+a_{sd}}\left(h_4\left(b_{rd},a_{rd},a_{sd},b_{sd},\frac{1}{2}P\lambda_{sr}\right)-h_5\left(b_{rd},a_{rd},a_{sd},b_{sd},\frac{1}{2}P\lambda_{sr}\right)\right) \\ & +\frac{b_{sd}}{a_{rd}+b_{sd}}\left(h_4\left(a_{rd},b_{rd},b_{sd},a_{sd},\frac{1}{2}P\lambda_{sr}\right)-h_5\left(a_{rd},b_{rd},b_{sd},a_{sd},\frac{1}{2}P\lambda_{sr}\right)\right) \\ & -\frac{a_{sd}}{a_{rd}+a_{sd}}h_5\left(a_{rd},b_{rd},a_{sd},b_{sd},\frac{1}{2}P\lambda_{sr}\right) \\ & +\frac{b_{sd}}{b_{rd}+b_{sd}}h_5\left(b_{rd},a_{rd},b_{sd},a_{sd},\frac{1}{2}P\lambda_{sr}\right) \qquad (5.63) \\ \overset{P\gg 1}{\approx}\; & \frac{3}{4P^2\lambda_{rd}\lambda_{sd}}+\frac{2}{P^2\lambda_{sr}\lambda_{sd}}=\mathrm{O}\left(\frac{1}{P^2}\right), \qquad (5.64) \end{aligned}$$

where

$$h_4\left(a,b,c,d,t\right)=\frac{b}{a+b}\frac{d}{c+d}\left(1-\sqrt{8(a+c)^2t}h_3\left(\sqrt{\frac{4(a+c)^2t}{\pi}}\right)\right), \qquad (5.65)$$

$$\begin{aligned} h_5\left(a,b,c,d,t\right)=&\frac{b}{a+b}\frac{d}{c+d}\frac{2}{B\sqrt{t\left(2+t\right)}} \\ &\times\left(h_3\left(\sqrt{\frac{8(a+c)^2t}{\pi\left(2+t\right)}}\right)-h_3\left(\sqrt{\frac{t}{2\left(2+t\right)}}\left(\frac{4\left(a+c\right)}{\sqrt{\pi}}+A\right)\right)\right), \qquad (5.66) \end{aligned}$$

and $h_3\left(x\right)=\exp\left(\frac{x^2}{2}\right)Q\left(x\right)$. From (5.51a) and (5.64), we conclude that the error rates of coherent DNC scale as $\mathrm{O}\left(\frac{1}{P^2}\right)$ at high SNRs.

Table 5.2: Scaling Laws of The Error Rates

	ANC	**DNC**
Coherent	$\mathrm{O}\left(\frac{\log P}{P^2}\right)$	$\mathrm{O}\left(\frac{1}{P^2}\right)$
Partial/Non-coherent	$\mathrm{O}\left(\frac{\log^3 P}{P^2}\right)$	$\mathrm{O}\left(\frac{\log P}{P^2}\right)$

5.3.4 Discussions

We summarize the scaling laws of the error rates in Table 5.2. It is observed that the error rates can be put in a general form of $P_e = \mathrm{O}\left(\frac{\log^{d_1} P}{P^{d_2}}\right)$. The diversity gain is defined as

$$d = -\frac{\log P_e}{\log P} = d_2 - d_1 \frac{\log \log P}{\log P}. \tag{5.67}$$

At extremely high SNRs (i.e., $P \to \infty$), the second term vanishes and the dominant diversity gain is d_2. But at modest SNRs, the logarithmic term would incur noticeable diversity loss and make the error curves decrease very slowly with SNRs. The extent of such diversity loss can be measured by d_1.

From Table 5.2, we can observe that the dominant diversity gain is 2 regardless of the network coding scheme and the type of receiver. The diversity gain comes from node cooperation, as each source symbol can reach the destination through two independent paths, i.e., direct link and relay link. However, the logarithmic term loss is totally different for these schemes. For coherent detection, the logarithmic term loss of ANC has an order of 1, whereas there is no logarithmic term loss for DNC. This is because for ANC, different source signals are randomly combined in the complex field through linear addition, so MUI would appear at the destination and degrade the performances of ML detection. On the contrary, for DNC the source signals are combined in the finite field through bit operation, thus the relay symbol is a single network-coded FSK symbol without any MUI.

For partial/non-coherent detection, the lack of perfect CSI would incur more logarithmic term loss for both ANC and DNC. Specifically, the order of the loga-

rithmic term loss are 1 and 3 respectively for coherent ANC and partial coherent ANC, while for coherent DNC and non-coherent DNC the order of the logarithmic term loss are 0 and 1, respectively. As a result, the performance loss of ANC is more serious, and the reasons are given in the sequel. For DNC, the relay node may detect the wrong symbols and propagate these errors to the destination. So the relay link signal is properly scaled through nonlinear operations in the ML combiner according to the estimated relay detection error rates at the destination. For non-coherent detection, the real-time relay detection states are blind to the destination, and the scaling is based on the average relay detection error rates. Although such scaling is accurate on average, it cannot capture the instantaneous network dynamics and thus incurs some performance loss. The situation becomes worse for ANC, as the formation of MUI is totally blind to the destination when perfect CSI is unknown. Consequently, the destination cannot efficiently suppress MUI to clearly separate the two source signals. That is why ANC suffers more serious performance loss than DNC when perfect CSI is unavailable at the destination.

Lastly, we would like to point out that the non-coherent detection may behave very differently for the single-user channels and the multi-user channels. For the single-user point-to-point channel, it is well known that non-coherent detection only brings 3dB SNR loss compared to coherent detection, but the diversity performances remain the same [1]. As for the multi-user network-coded uplink channel, we have demonstrated that non-coherent detection would incur diversity loss at modest SNRs, but the dominant diversity gains remain the same at extremely high SNRs.

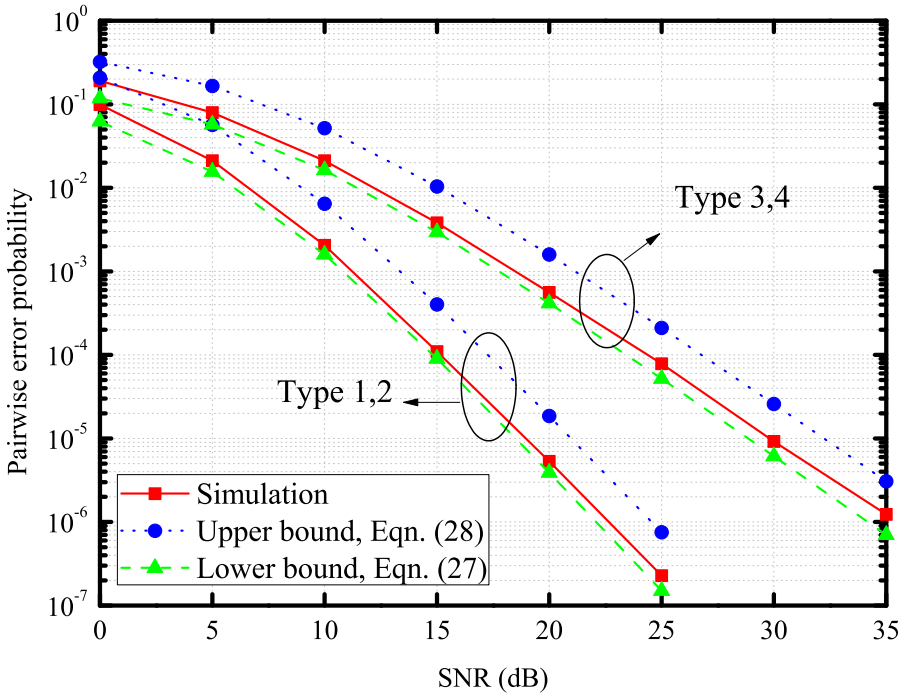


Figure 5.2: PEP of coherent ANC.

5.4 Simulations

In this section, we present some simulation results to validate our analysis. The path loss model is $\lambda = D^{-3}$, where λ is the channel gain and D is the distance between two terminals. Pair error probability is used as the performance metric, i.e., the probability that at least one of the source symbols is detected incorrectly at the destination. To simplify the simulation settings, only binary FSK modulation is studied and D_{sd} is always normalized to 1. In the simulations, direct transmission means the two source nodes take turns to deliver their own information to the destination without the help of the relay node. The transmitted power for direct transmission is normalized in such a way that the total energy consumption of each source node is identical to that for network coding.

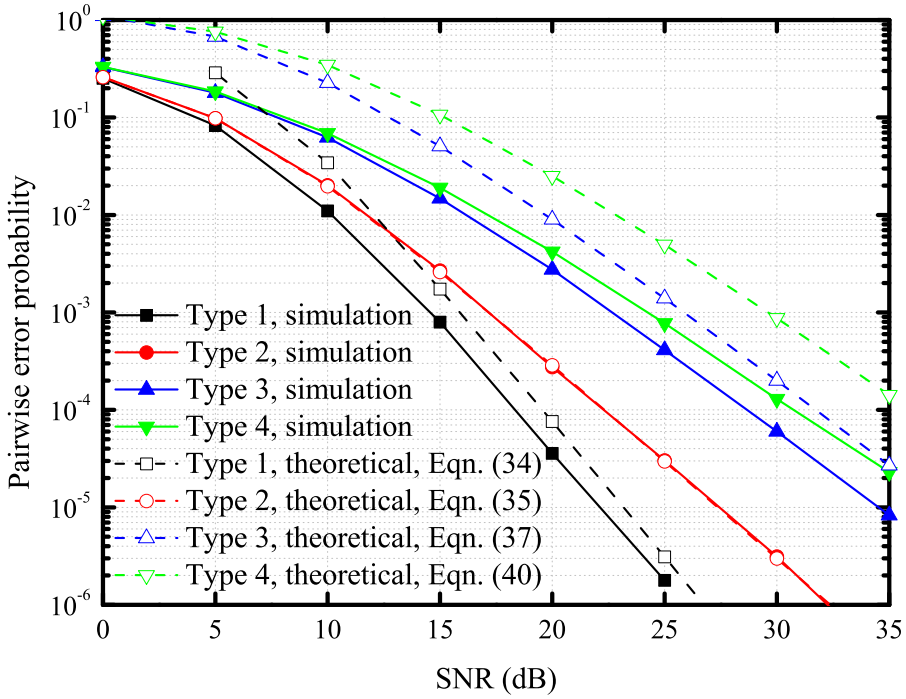


Figure 5.3: PEP of partial coherent ANC.

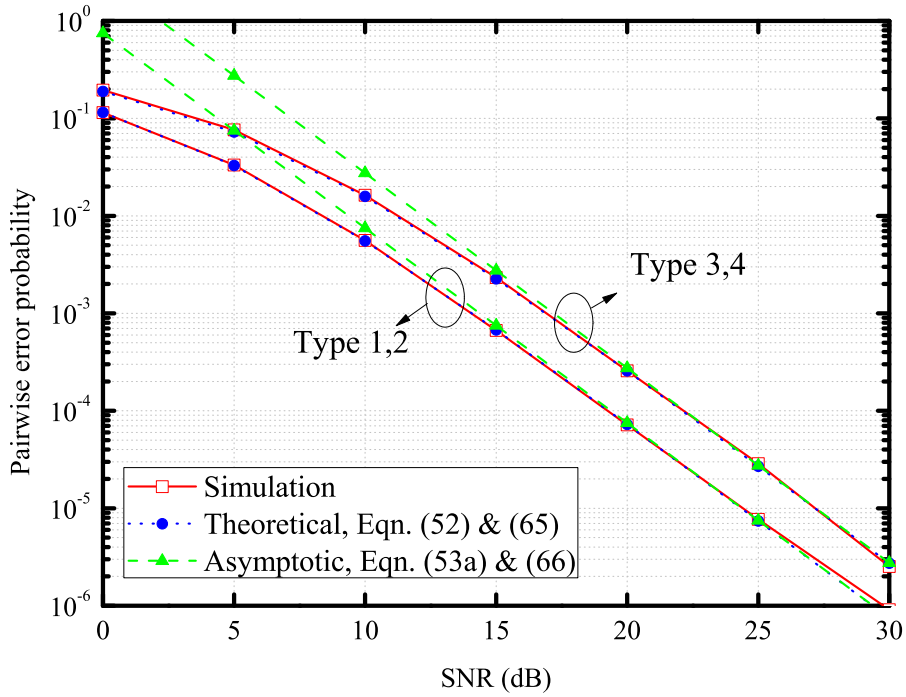


Figure 5.4: PEP of coherent DNC.

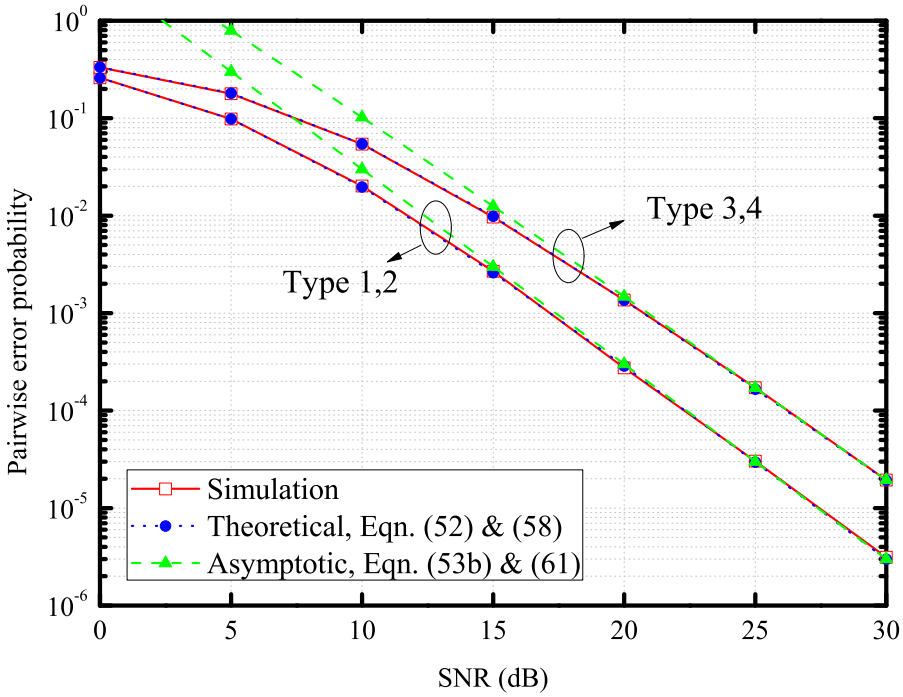


Figure 5.5: PEP of non-coherent DNC.

First we compare the simulated PEPs with our theoretical results in the symmetric networks, where all the inter-node distances are normalized to 1. For coherent ANC, we observe in Figure 5.2 that the error curves are always between the two bounds. Besides, the two bounds have the same slopes at high SNRs, since their scaling laws are the same. Next in Figure 5.3, we plot the four types of PEPs of partial coherent ANC, where for Type-4 PEP only the upper bound is given. It is observed that all the error curves have the same scaling behaviors as our theoretical results at high SNRs. Although Type-2, Type-3 and Type-4 PEPs have the same dominant diversity order of 2, the PEPs with larger logarithmic term loss decrease much more slowly at moderate SNRs. Then in Figure 5.4 and Figure 5.5, we show the PEPs of coherent and non-coherent DNC. It is observed that the two error curves

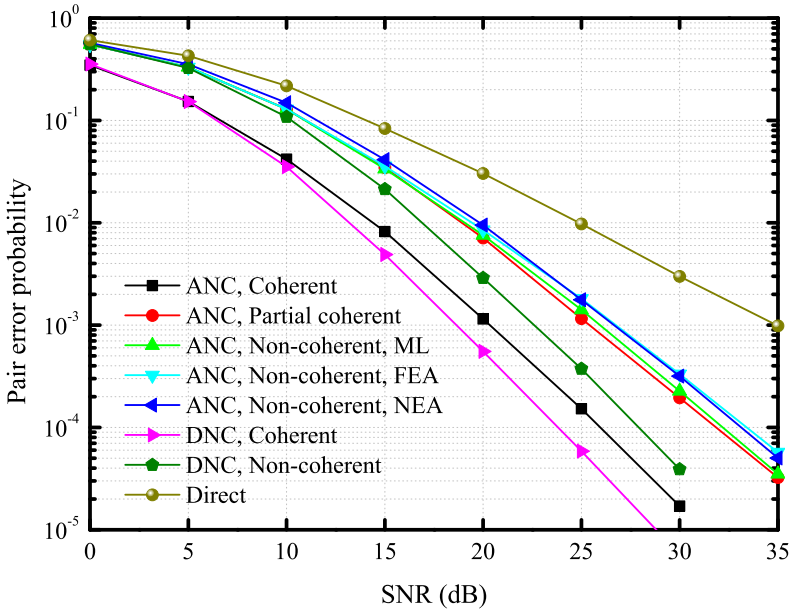


Figure 5.6: Error rates of the symmetric networks.

of coherent DNC have the same slopes at high SNRs, whereas for non-coherent DNC Type-3 PEP suffers some diversity loss at modest SNRs. In all cases, our theoretical results match perfectly well with the simulation results.

Then in Figure 5.6, we compare the real error rates in the symmetric networks. Compared to direct transmission without user cooperation, both coherent and non-coherent cooperation can provide a diversity order of 2. For ANC, it is observed that the partial coherent receivers perform almost the same as non-coherent receivers. Besides, the two suboptimum non-coherent receivers also perform reasonably well with about 1dB SNR loss compared to the ML receivers. Compared to coherent receivers, the partial/non-coherent receivers suffer huge performance loss due to the logarithmic term loss. For example, the non-coherent receivers have about 6dB SNR loss when the error rate is 10^{-4}. For DNC, we can observe that the coherent receivers

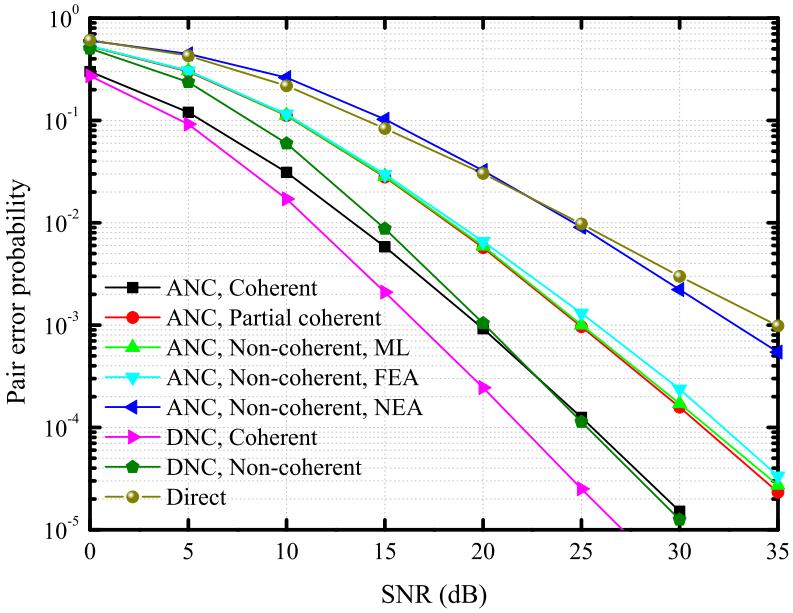


Figure 5.7: Error rates of the asymmetric networks with better source-relay channels.

perform much better than the non-coherent receivers, but the performance gap is not as large as that of ANC. In all cases, DNC performs better than ANC using the same type of receivers. This is because the extent of the logarithmic term loss is smaller for DNC, and there is no MUI within the relay branch. Finally, we observe that the non-coherent DNC and coherent ANC have the same scaling behaviors at high SNRs, which is consistent with our analysis.

We also compare the error rates in the asymmetric networks. In Figure 5.7, we study the networks with better source-relay channels, where $D_{sr} = 0.4$ and all the other distances are normalized. By comparing Figure 5.6 and Figure 5.7, we observe that the performances of DNC have greatly improved, since better source-relay channel conditions would lead to reduced relay detection errors. On the contrary, the performances of ANC almost remain the same due to the power normalization

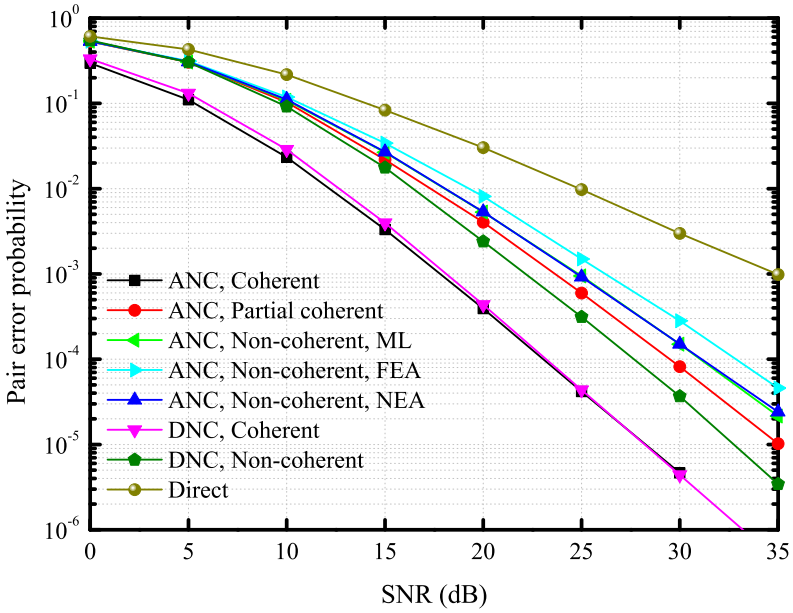


Figure 5.8: Error rates of the asymmetric networks with better relay-destination channels.

effects at the relay node. For non-coherent ANC, we observe that the FEA receivers perform very close to the ML receivers, whereas the NEA receivers suffer great performance loss, since neglecting the dominant noise terms on the second hop would greatly distort the likelihood function. Next in Fig. 8, we study the networks with better relay-destination channels, where $D_{rd} = 0.4$ and all the other distances are normalized. We observe that improving the relay-destination channel conditions would bring huge coding gain to ANC. The performances of coherent ANC and DNC are almost indistinguishable, whereas for non-coherent receivers the performance gap is greatly reduced. For non-coherent ANC, the NEA receivers perform almost the same as the ML receivers, whereas the FEA receivers suffer about 1.5dB SNR loss.

Finally we study the error rates with different relay positions in Figure 5.9,

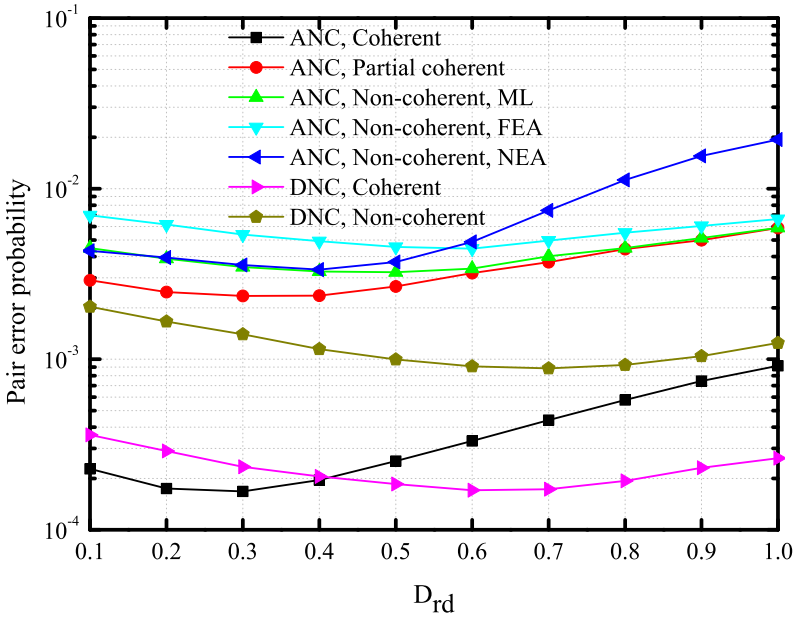


Figure 5.9: Error rates with different relay positions.

where the system SNR is 20dB. For the network topology, we place the destination at $(0, 0)$, and locate the two source nodes at $(\frac{\sqrt{3}}{2}, \pm\frac{1}{2})$. The relay node moves along the x-axis from $(0.1, 0)$ to $(1, 0)$. The simulation results clearly show that the best performance of ANC is achieved when the relay node is close to the destination, whereas for DNC the best performance is achieved when the relay node is close to the sources. In all cases, non-coherent DNC performs better than the corresponding ANC due to the smaller logarithmic term loss. For non-coherent ANC, we observe that the NEA receivers are nearly optimum when the relay node is close to the destination, whereas the performances of the FEA receivers get closer to the performances of the ML receivers as the relay node moves to the two source nodes. Throughout simulations, we have observed that the partial coherent ANC performs very close to the non-coherent ANC. So we conjecture that these two schemes ac-

tually have the same scaling behaviors. A rigorous proof is deferred to our future work.

5.5 Conclusions

In this chapter, we studied the diversity performance of non-coherent network coding. We developed the optimum/sub-optimum transceivers under different CSI conditions. It is demonstrated that non-coherent transmission would incur diversity loss at modest SNRs, but the dominant diversity gain remains the same at extremely high SNRs. It is also demonstrated that non-coherent ANC would incur more serious performance loss than non-coherent DNC, as the destination cannot efficiently suppress MUI.

Chapter 6

Network-Coded ARQ for Two-Way Relay Channel

So far, we have studied the diversity gain of uncoded cooperation systems. Since this chapter, we would shift attention to the coded systems. One key benefit of the coded systems is that the network could somehow know the decoding status of the messages by use of error detection code. Depending on the network dynamics, the relay may choose to perform network coding when necessary or just stay idle if the decoding fails. Note that the error propagation issue associated with DNC discussed in Chapter 2 and Chapter 3 is no longer a big concern, which would bring huge performance gain over the uncoded systems. How to design a sound transmission strategy to exploit such benefit would be our main focus since this chapter.

In this work, we focus particularly on the application of network-coded ARQ for coded TWRC. Some related studies can be found in [31,74–77] and the references therein. The novelty of our work is summarized below:

1) The direct link between the source nodes is neglected in the previous studies. As a result, all data flows have to go through relay link regardless of the network dynamics. In contrast, we incorporate the direct link in our system model, and wireless relaying is used if and only if the direct link is in outage and the source packets have been correctly decoded at relay nodes. Such incremental relaying

scheme can save the channel use and transmitted power that could be used to relay new packets whenever possible.

2) In previous studies, it is assumed that any packet can be retransmitted infinitely many times until successfully received, so packet loss is neglected. However, for most practical applications there is always some maximum transmission constraint to limit the transmission delay, and the packet will be dropped after several attempts. In this work, we carefully incorporate such constraints, where the per-hop transmission or E2E transmission of the same packet cannot exceed certain times. We derive the closed-form throughput under these constraints after considering the packet loss.

3) The transmitted power is supposed to be fixed in the previous work regardless of the network topology. In contrast, we obtain a closed-form near-optimum power control scheme to maximize the system throughput. We show that power control is extremely important for asymmetric networks, where the terminal of the bottleneck link should use more power to compensate for the larger packet loss rate.

4) While the previous studies focus only on the single-relay networks, we also study multi-relay networks in this work. We show that using localized DNC alone may not fully leverage the network coding gain, especially when the frame length is much smaller than the number of relays. To address this issue, we propose a hybrid network coding schemes in which both DNC and ANC may be used within the relay array.

Notations: The abbreviation i.i.d. stands for independent and identically distributed. The probability of an event $\mathcal{A}$ is denoted by $\Pr(\mathcal{A})$. The CDF and the

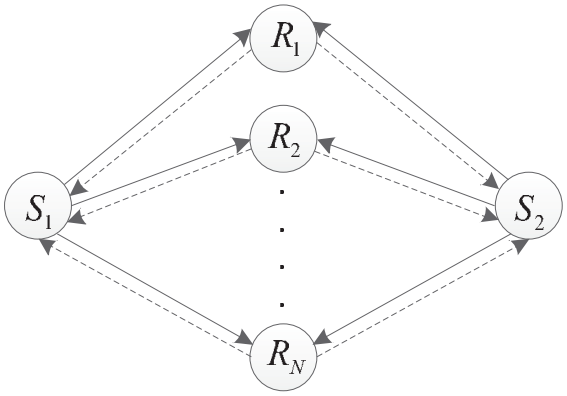


Figure 6.1: System model of two-way relay channel.

probability mass function (PMF) of a random variable Z are denoted by $F_Z(z)$ and $f_Z(z)$, respectively. We shall use $Z \sim \mathcal{CN}(\mathbf{u}, \mathbf{\Sigma})$ as a circularly symmetric complex Gaussian random variable vector with mean $\mathbf{u}$ and covariance matrix $\mathbf{\Sigma}$. $Z \sim \text{Bin}(p, n)$ stands for binomial distribution with parameter p and n. $Z \sim \text{Geom}(p)$ stands for geometric distribution with parameter p, i.e., $f_Z(k) = (1-p)\,p^{k-1}$ for $k = 1, 2, \cdots$.

6.1 System Model

Consider the bidirectional relay channel shown in Figure 6.1, where the two source nodes S_1 and S_2 want to exchange data with the help of a set of relay nodes R_k for $k = 1, 2, \cdots, N$. Suppose the source data are sent in a frame-by-frame manner, and each frame consists of K packets. The kth packet of S_i is denoted by $X_i(k)$ for $i = 1, 2$. For simplicity, we assume all the packets are of the same length, and the data rate is fixed and is denoted by r. The time that is needed to deliver one packet through any point-to-point channel is called one time unit.

The signal model for data transmission from node i to node j is given by

$$y_{ij} = h_{ij}\sqrt{P_i}x_i + n_{ij}. \tag{6.1}$$

Here P_i is the transmitted power of node i, x_i and y_{ij} are the transmitted signal and the received signal, respectively, $n_{ij} \sim \mathcal{CN}(0,1)$ is the additive white Gaussian noise, $h_{ij} \sim \mathcal{CN}(0,\lambda_{ij})$ is the Rayleigh fading channel coefficient, where λ_{ij} is the path loss coefficient determined by the distance between the transmitter and receiver. All the channel coefficients and additive noises are independent for different channels.

Due to random channel fading, the receiver may be unable to decode the transmitted packet correctly. Usually, some parity bits are added to the raw data, based on which the receiver can perform cyclic redundant check to tell whether the decoding is successful. In this work, we assume such error detection is perfect. The packet error rate is denoted by q_{ij}, and it is approximately equal to the channel outage rate given by

$$q_{ij} \approx \Pr\left(\log_2\left(1+|h_{ij}|^2 P_i\right) \le r\right) = 1 - \exp\left(-\frac{2^r-1}{P_i\lambda_{ij}}\right). \tag{6.2}$$

Depending on the decoding status of the packet, the receiver needs to feed back the ACK/NACK signal to inform the transmitter that the decoding is successful/failed. In this work, we assume that the feedback channels are perfect, and the channel use of such flag bits are negligible compared to that of the information bits.

Suppose ARQ is used to guarantee the quality-of-service of data transmission. That is, the current erroneous packet is dropped and the receiver will feed back the NACK signal if the decoding is failed; afterwards, the transmitter may retransmit the same packet until the decoding is successful at the receiver or the current packet

transmission expires when it exceeds a certain period. In this work, we investigate two types of maximum transmission constraints, i.e., per-hop constraint and E2E constraint. For per-hop constraint, packet transmissions on different point-to-point channels are treated as independent sessions, and each packet transmission session cannot exceed L times. In contrast, for E2E constraint the transmission of the same packet from the original source to the final sink is treated as one whole session and cannot exceed L times, regardless of the transmission routes. To give an example, suppose there is packet transmission from S_1 to S_2 and the transmission route is $S_1 \rightarrow R_1 \rightarrow S_2$. Then for per-hop constraint, S_1 can transmit the packet up to L times and if R_1 can successfully decode this packet within L times, it can also attempt to deliver the same packet to S_2 up to L times. In contrast, for E2E constraint the total transmission of this packet cannot exceed L times, regardless it is sent from S_1 or R_1.

As the performance measure, the effective throughput is defined as the average number of successfully delivered packets per time unit within each frame, and is given by

$$\eta = \frac{E[M]}{E[T]}, \tag{6.3}$$

where M and T are the total number of successively delivered packets and the total times to exchange K packets between S_1 and S_2. In this work, we assume all nodes are subject to half-duplex constraint such that they cannot transmit and receive at the same time. As a result, the effective throughput is always bounded by $0 \leq \eta \leq 1$. The exact throughput is dependent on the transmission strategies,

which will be detailed in the next section.

6.1.1 Some Preliminaries

Before leaving this section, we first present some mathematical results that appear to be pretty useful later. For notational convenience, we define the bounded geometric distribution (BGeom) as $Z = \min(X, L) \sim \text{BGeom}(p, L)$, where $X \sim \text{Geom}(p)$ for $0 \leq p \leq 1$ and $L > 0$ is some integer. After some simple algebra, it is easy to show that the PMF and CDF of Z are respectively given by

$$f_Z(k) = \begin{cases} (1-p)\, p^{k-1}, k = 1, 2, \cdots, L-1 \\ p^{L-1}, k = L \end{cases}, \tag{6.4}$$

and

$$F_Z(k) = \begin{cases} 0, \ k < 0 \\ 1 - p^k, \ k = 0, 1, \cdots, L-1 \\ 1, \ k \geq L \end{cases}. \tag{6.5}$$

Besides, we can also show that

$$E[Z] = \sum_{k=1}^{L} \Pr(Z \geq k) = \sum_{k=1}^{L} p^{k-1} = \begin{cases} \frac{1-p^L}{1-p}, \ p \neq 1 \\ L, \ p = 1 \end{cases} \triangleq g_L(p). \tag{6.6}$$

With these results, we can prove the following results:

Proposition 6.1. *Let $X \sim BGeom(p, L)$ for $0 \leq p \leq 1$ and $Y \sim BGeom(q, L)$ for $0 \leq q \leq 1$ be independent, then*

1) $Z = \min(X, Y) \sim BGeom(pq, L)$ and $E[Z] = g_L(pq)$.

2) Let $Z = \max(X, Y)$, then

$$E[Z] = g_L(p) + g_L(q) - g_L(pq). \tag{6.7}$$

3) Let $Z \sim \mathrm{BGeom}(a, L-X)$ *for* $0 \le a \le 1$. *Then for* $0 \le p < 1$, *we have*

$$E[Z|\, X \le L-1] = \frac{1}{1-a}\left(1 - \frac{1-p}{1-p^{L-1}} a^{L-1} g_{L-1}\left(\frac{p}{a}\right)\right) \triangleq h_0(a, p; L). \tag{6.8}$$

4) Let $W \sim \mathrm{BGeom}(a, L-X)$ *for* $0 \le a \le 1$ *and* $T \sim \mathrm{BGeom}(b, L-Y)$ *for* $0 \le b \le 1$, *and define* $Z = \max(W, T)$. *Then for* $0 \le p, q < 1$, *we have*

$$\begin{aligned}
&E[Z|\max(X,Y) \le L-1] \\
&= \frac{1}{1-p^{L-1}}\left(g_{L-1}(a) - p^{L-1} g_{L-1}\left(\frac{a}{p}\right)\right) \\
&\quad + \frac{1}{1-q^{L-1}}\left(g_{L-1}(b) - q^{L-1} g_{L-1}\left(\frac{b}{q}\right)\right) \\
&\quad - \frac{1}{(1-p^{L-1})(1-q^{L-1})}\left(g_{L-1}(ab) + p^{L-1}q^{L-1} g_{L-1}\left(\frac{ab}{pq}\right)\right) \\
&\quad + \frac{1}{(1-p^{L-1})(1-q^{L-1})}\left(q^{L-1} g_{L-1}\left(\frac{ab}{q}\right) + p^{L-1} g_{L-1}\left(\frac{ab}{p}\right)\right) \\
&\triangleq h_1(a, b, p, q; L).
\end{aligned} \tag{6.9}$$

Proof. For property 1, we use (6.4) and (6.5) and can obtain

$$\begin{aligned}
f_Z(k) &= \Pr(X = k, Y > k) + \Pr(X > k, Y = k) + \Pr(X = Y = k) \\
&= (1-pq)\, p^{k-1} q^{k-1}
\end{aligned} \tag{6.10}$$

for $k = 1, 2, \cdots, L-1$ and $f_Z(L) = \Pr(X = Y = L) = p^{L-1}q^{L-1}$. Comparing with (6.4), we have $Z \sim \mathrm{BGeom}(pq, L)$.

For property 2, we have

$$\begin{aligned}
f_Z(k) &= \Pr(X = k, Y < k) + \Pr(X < k, Y = k) + \Pr(X = Y = k) \\
&= (1-p)\, p^{k-1} + (1-q)\, q^{k-1} - (1-pq)\, p^{k-1} q^{k-1}
\end{aligned} \tag{6.11}$$

for $k = 1, 2, \cdots, L-1$ and $f_Z(L) = p^{L-1} + q^{L-1} - p^{L-1}q^{L-1}$. By using (6.6), we can obtain (6.7) immediately.

For property 3, we need to first derive the conditional PMF of X given by

$$f_X\left(k\middle|\, X \leq L-1\right) = \frac{1-p}{1-p^{L-1}} p^{k-1} \tag{6.12}$$

for $k = 1, 2, \cdots, L-1$. Now we can calculate the conditional expectation of Z given X from (6.6) and obtain

$$\begin{aligned} E\left[Z\middle|\, X \leq L-1\right] &= E\left[\frac{1-a^{L-X}}{1-a}\middle|\, X \leq L-1\right] \\ &= \frac{1}{1-a}\left(1 - \frac{1-p}{1-p^{L-1}} \sum_{k=1}^{L-1} p^{k-1} a^{L-k}\right) \\ &= \frac{1}{1-a}\left(1 - \frac{1-p}{1-p^{L-1}} a^{L-1} g_{L-1}\left(\frac{p}{a}\right)\right). \end{aligned} \tag{6.13}$$

Finally for property 4, we have

$$\begin{aligned} f_W\left(k\middle|\, X \leq L-1\right) &= \Pr\left(W = k, X < L-k\middle|\, X \leq L-1\right) \\ &\quad + \Pr\left(W = k, X = L-k\middle|\, X \leq L-1\right) \\ &= \left(1-a\right) a^{k-1} \frac{1-p^{L-k-1}}{1-p^{L-1}} + a^{k-1} \frac{\left(1-p\right) p^{L-k-1}}{1-p^{L-1}} \\ &= \frac{a^{k-1}}{1-p^{L-1}} \left(1 - a + \left(a-p\right) p^{L-k-1}\right) \end{aligned} \tag{6.14}$$

for $k = 1, 2, \cdots, L-1$. Now we can compute the conditional CDF of W given by

$$\begin{aligned} F_W\left(k\middle|\, X \leq L-1\right) &= \frac{1-a}{1-p^{L-1}} \sum_{i=1}^{k} a^{i-1} + \frac{a-p}{1-p^{L-1}} p^{L-2} \sum_{i=1}^{k} a^{i-1} p^{-(i-1)} \\ &= \frac{1 - a^k - p^{L-1} + a^k p^{L-k-1}}{1-p^{L-1}} \end{aligned} \tag{6.15}$$

for $k = 0, 1, \cdots, L-1$. As W and T are independent, we have

$$\begin{aligned} F_Z\left(k\middle|\, \max\left(X, Y\right) \leq L-1\right) &= F_W\left(k\middle|\, X \leq L-1\right) F_T\left(k\middle|\, X \leq L-1\right) \\ &= 1 - \frac{1-p^{L-k-1}}{1-p^{L-1}} a^k - \frac{1-q^{L-k-1}}{1-q^{L-1}} b^k \\ &\quad + \frac{1 - p^{L-k-1} - q^{L-k-1} + \left(pq\right)^{L-k-1}}{\left(1-p^{L-1}\right)\left(1-q^{L-1}\right)} a^k b^k \end{aligned} \tag{6.16}$$

for $k = 0, 1, \cdots, L-1$. By using (6.6) and the relation

$$E\left[Z|\max\left(X,Y\right) \leq L-1\right] = \sum_{k=1}^{L-1} \Pr\left(Z \geq k|\max\left(X,Y\right) \leq L-1\right), \tag{6.17}$$

we can obtain (6.9) immediately. □

6.2 Single-Relay Case

In this section, we study several transmission strategies for the bidirectional relay channel shown in Fig. 1. We focus only on the single-relay case, i.e., $N = 1$. The multi-relay case will be discussed in the next section. For notational convenience, we use P_1, P_2 and P_R to represent the transmitted power of S_1, S_2 and R_1, respectively. The packet loss rate for the channel $i \rightarrow j$ is denoted by $q_{i,j}$ for $i, j \in \{1, 2, R\}$.

6.2.1 Direct Transmission

The simplest strategy is that the bidirectional communication goes only through the direct link between the two source nodes, while the relay node stays idle. Due to the half-duplex constraint, the two source nodes need to take turns to send a frame of K packets to the other end in a time-division multiplex manner. Each time after S_1 sends out a packet, three consequences may occur:

1. S_2 successfully decodes the packet and feeds back the ACK signal. Then S_1 will start to deliver the next packet.

2. S_2 fails to decode the packet, and the current transmission is not the Lth

attempt. Then S_1 will initiate AQR and retransmit the same packet once again.

3. S_2 fails to decode the packet, and the current transmission is the Lth attempt. Then the current packet is dropped, and S_1 will send out a new packet in the next time slot.

Note that for each packet of S_1, it will be either successfully delivered to S_2 (denoted by the event $I_1 = 1$) or dropped after expiration (denoted by the event $I_1 = 0$). It is also worth noting that if the current decoding fails, S_2 will drop the erroneous packet and attempt to decode the same packet based solely on the newly arrived signal in the future. Therefore, the total transmission times of each packet satisfy BGeom $(q_{1,2}, L)$ and is denoted by $T_1 = \min(X_1, L)$, where $X_1 \sim \text{Geom}(q_{1,2})$. The probability of successful delivery is given by

$$\Pr(I_1 = 1) = \Pr(X_1 \leq L) = 1 - q_{1,2}^L, \tag{6.18}$$

and the average transmission times are $E[T_1] = g_L(q_{1,2})$. Due to symmetry, the above metrics should be similar for any packet delivered along the reverse direction after properly modifying the subscripts. Therefore, the effective throughput of direct transmission is

$$\begin{aligned} \eta_{DT} &= \frac{K \times E[\mathbb{1}\{I_1 = 1\} + \mathbb{1}\{I_2 = 1\}]}{K \times E[T_1 + T_2]} = \frac{2 - q_{1,2}^L - q_{2,1}^L}{\frac{1-q_{1,2}^L}{1-q_{1,2}} + \frac{1-q_{2,1}^L}{1-q_{2,1}}} \\ &\overset{L \gg 1}{\approx} \frac{2(1-q_{1,2})(1-q_{2,1})}{2 - q_{1,2} - q_{2,1}}, \end{aligned} \tag{6.19}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function.

6.2.2 Pure Relaying

If the relay node is located at some place between the two source nodes, the relay link would be much better than the direct link. As a result, if the relay node happens to successfully decode the source packets, it can initiate the ARQ for the original data source to improve the transmission quality. In this subsection, we study the conventional pure relaying strategy [12], where the relay node forwards the individual packets in different time slots. The whole bidirectional communication is completed in three phases. In the first two phases, the two sources take turns to send out a frame of K packets. Then in the third phase, the relay node will forward all the successfully decoded packets in the buffer to the intended receivers.

We first consider the transmission phase of S_1. Each time after S_1 sends out a packet, the following consequences may occur:

1. S_2 successfully decodes the packet and feeds back the ACK signal. Then S_1 will start to deliver the next packet.

2. S_2 fails to decode the packet, but the relay R successfully decodes the packet. For E2E constraint, the extra requirement is that the current transmission is not the Lth attempt. Then R will store the packet in the local buffer and continue the unfinished ARQ for S_1 during the future relaying phase, and S_1 will start to deliver the next packet.

3. Neither S_2 nor R is able to decode the packet, and the current transmission is not the Lth attempt. Then S_1 will initiate AQR and retransmit the same packet once again.

4. Neither S_2 nor R is able to decode the packet, and the current transmission is the Lth attempt. For E2E constraint, this also includes the event that S_2 fails to decode the packet, but the relay R successfully decodes the packet upon the Lth attempt. Then the current packet is dropped, and S_1 will send out a new packet in the next time slot.

Note that in case 1), S_2 and R may decode the packet simultaneously. In that case, the transmission is successful and the relay node needs not to store that packet any more. Also for E2E constraint, the relay node will only store the packets that have been successfully decoded within the first $L-1$ attempts. This is because the E2E transmission session expires right after the Lth attempt, so the relay node should drop that packet even if the Lth decoding is successful. That is why we discriminate E2E constraint in case 2) and case 4). Eventually, each packet of S_1 will be either successfully delivered to S_2 (denoted by the event $I_1 = 1$), or stored at R but yet not decoded by S_2 (denoted by the event $I_1 = -1$), or dropped after expiration (denoted by the event $I_1 = 0$). As each packet transmission will terminate right after either S_2 or R is able to decode the packet, the transmission time is given by $T_1 = \min(T_{1,R}, T_{1,2})$, where $T_{1,2} = \min(X_{1,2}, L)$ and $T_{1,R} = \min(X_{1,R}, L)$, and $X_{1,2} \sim \text{Geom}(q_{1,2})$ and $X_{1,R} \sim \text{Geom}(q_{1,R})$ are independent. The probability that S_2 decodes the packet is given by

$$\Pr(I_1 = 1) = \sum_{n=1}^{L} \Pr(X_{1,2} = n, X_{1,R} \geq n) = \frac{1 - q_{1,2}}{1 - q_{1,R}q_{1,2}} \left(1 - q_{1,R}^L q_{1,2}^L\right). \quad (6.20)$$

The probability that the packet is stored at R but yet not decoded by S_2 is given

by

$$\Pr\left(I_1=-1\right)=\sum_{n=1}^{L}\Pr\left(X_{1,R}=n,X_{1,2}>n\right)=\frac{\left(1-q_{1,R}\right)q_{1,2}}{1-q_{1,R}q_{1,2}}\left(1-q_{1,R}^{L}q_{1,2}^{L}\right) \quad (6.21)$$

for per-hop constraint, and

$$\Pr\left(I_1=-1\right)=\sum_{n=1}^{L-1}\Pr\left(X_{1,R}=n,X_{1,2}>n\right)=\frac{\left(1-q_{1,R}\right)q_{1,2}}{1-q_{1,R}q_{1,2}}\left(1-q_{1,R}^{L-1}q_{1,2}^{L-1}\right) \quad (6.22)$$

for E2E constraint. By using property 1 of Proposition 6.1, we can also calculate the average transmission times of each packet as $E\left[T_1\right]=g_L\left(q_{1,R}q_{1,2}\right)$.

After the two source transmission phases, the relay R will forward the packets stored in the local buffer to the intended receiver. During this phase, both source nodes are receivers and the relay node will continue the unfinished ARQ. Let $\mathbb{D}_i$ be the set of packets of S_i for $i=1,2$ that are stored in the buffer, and the set size is denoted by $|\mathbb{D}_i|=D_i$. Besides, the relay also needs to manage a set $\mathbb{R}_i$ that records the maximum residual transmission times of each packet in $\mathbb{D}_i$. At the beginning of the relaying phase, the maximum residual transmission times are equal to L for per-hop constraint and $L-T_i$ for E2E constraint[1]. Note that the two sets $\mathbb{D}_i$ and $\mathbb{R}_i$ have the same size D_i that satisfies binomial distribution, i.e., $D_i\sim\mathrm{Bin}\left(Q_i,K\right)$ where $Q_i\triangleq\Pr\left(I_i=-1\right)$ given by (6.21) and (6.22). For pure relaying, the relay R simply delivers all the packets in the buffer one-by-one in different time slots. Due to symmetry, we only consider transmitting a packet from R to S_2. Because the $R\rightarrow S_2$ channel is a point-to-point channel, the transmission process mimics the direct transmission described in the last subsection, except that each packet

[1]The residual transmission times could be different for packets stored at the relay R. Here we omit the time index of packets for notational convenience.

may have different residual transmission times for E2E constraint. As a result, for each packet delivered from R to S_2 the total transmission times can be denoted by $T_{R,2} = \min(X_{R,2}, L)$ for per-hop constraint and $T_{R,2} = \min(X_{R,2}, L - T_1)$ with the condition that $T_1 \leq L-1$, where $X_{R,2} \sim \text{Geom}(q_{R,2})$. Let $\{I_{R,2} = 1\}$ and $\{I_{R,2} = 0\}$ represent the events of successful delivery and packet loss due to transmission expiration, respectively. Then for per-hop constraint, the probability of successful delivery is

$$\Pr(I_{R,2} = 1) = \Pr(X_{R,2} \leq L) = 1 - q_{R,2}^{L} \tag{6.23}$$

and we have $E[T_{R,2}] = g_L(q_{R,2})$. For E2E constraint, we have

$$\begin{aligned}\Pr(I_{R,2} = 1) &= \Pr(X_{R,2} \leq L - T_1 | T_1 \leq L - 1) \\ &= 1 - \frac{1 - q_{1,R}q_{1,2}}{1 - q_{1,R}^{L-1}q_{1,2}^{L-1}} q_{R,2}^{L-1} g_{L-1}\left(\frac{q_{1,R}q_{1,2}}{q_{R,2}}\right).\end{aligned} \tag{6.24}$$

By using property 1 and property 3 of Proposition 6.1, we have

$$E[T_{R,2}] = h_0(q_{R,2}, q_{1,R}q_{1,2}; L). \tag{6.25}$$

With the above results, we can write the total number of successfully delivered packets in three phases as

$$\begin{aligned}E[M] &= \sum_{i=1, j=\{1,2\}\backslash\{i\}}^{2} (KE[\mathbb{I}\{I_i = 1\}] + E[D_i] E[\mathbb{I}\{I_{R,j} = 1\}]) \\ &= K \sum_{i=1, j=\{1,2\}\backslash\{i\}}^{2} (\Pr(I_i = 1) + \Pr(I_i = -1) \Pr(I_{R,j} = 1)),\end{aligned} \tag{6.26}$$

and the total transmission times are given by

$$\begin{aligned}E[T] &= \sum_{i=1, j=\{1,2\}\backslash\{i\}}^{2} (KE[T_i] + E[D_i] E[T_{R,j}]) \\ &= K \sum_{i=1, j=\{1,2\}\backslash\{i\}}^{2} (E[T_i] + \Pr(I_i = -1) E[T_{R,j}]),\end{aligned} \tag{6.27}$$

After plugging (6.20)-(6.24) into the above two expressions, we obtain the closed-form throughput according to (6.3). When the maximum transmission times are sufficiently large (i.e., $L \gg 1$), the two types of constraints would lead to the same asymptotic throughput given by

$$\eta_{\text{Relay}} \overset{L\gg 1}{\approx} 2 \left[\sum_{i=1,j=\{1,2\}\backslash\{i\}}^{2} \frac{1 - q_{R,j} + (1 - q_{i,R})\, q_{i,j}}{(1 - q_{i,R} q_{i,j})\,(1 - q_{R,j})} \right]^{-1}. \tag{6.28}$$

6.2.3 Static Network Coding

Pure relaying is not bandwidth efficient as the relay node forwards all the source packets in orthogonal time slots. However, if both buffers are non-empty, the relay node can combine the two packets intended for different receivers to save the channel use. For example, suppose the relay node needs to deliver X_1 to S_2 and also deliver X_2 to S_1. Then instead of transmitting these two packets separately, the relay can perform network coding to combine these two packets, i.e., $X_R = X_1 \oplus X_2$, and then send out this single network-coded packet X_R. If S_1 is able to successfully decode this packet, it can also decode the message sent from S_2 by $X_2 = X_1 \oplus X_R$. Likewise, S_2 can also decode X_1 based on X_R through similar operation. In this way, both source nodes can get the desired message while the relay node only needs to broadcast a single packet.

For such network-coded relaying, the whole data exchange still occurs in three phases, i.e., two source transmission phases followed by one data relaying phase. The first two source transmission phases are exactly the same as what we studied in the last subsection. The only difference is how the relay node shall forward

the packets in the local buffer during the data relaying phase. Suppose the D_i packets of S_i stored in the buffer $\mathbb{D}_i$ are $\{X_i(1), X_i(2), \cdots, X_i(D_i)\}$ for $i = 1, 2$. Let $D_{NC} = \min(D_1, D_2)$ and $D_{REG} = |D_1 - D_2|$ be the number of network-coded packets and the number of regular packets, respectively. If any $\mathbb{D}_i$ is empty, the relay node only needs to forward data to a single terminal and all the packets are regular packets. Otherwise, without loss of generality we assume $D_1 \geq D_2 > 0$. Then the relay node will combine the first D_{NC} packets in $\mathbb{D}_1$ and $\mathbb{D}_2$ through $X_R(k) = X_1(k) \oplus X_2(k)$ for $k = 1, 2, \cdots, D_{NC}$. Afterwards, the relay node will forward D_{NC} network-coded packets $\{X_R(k)\}_{k=1}^{D_{NC}}$ intended for both source nodes, and D_{REG} regular packets $\{X_1(k)\}_{k=D_{NC}+1}^{D_2}$ intended for S_2 alone.

The transmission of regular packets is very much like pure relaying and has been studied in the last subsection. For the network-coded flow, it is very much like a combination of two unicast flows where each component flow is subject to its own maximum transmission constraint. Each time any source node is able to decode the desired component packet, or any component packet has reached the maximum transmission limit, that component packet will be dropped by the relay and the following information flow will only contain a single packet. The transmission of any network-coded flow will terminate until both component unicast flows are finished after successful delivery or expiration. For example, suppose $X_R = X_1 \oplus X_2$ where X_1 and X_2 can be transmitted up to 2 and 4 times. If neither source nodes can decode X_R in the first 2 attempts, then X_1 will be dropped and in the following time slots, the relay node will forward $X_R = X_2$ to S_1 alone up to 2 more times.

Next we analyze the effective throughput. The first thing to note is that the

packet loss rate remains the same for a network-coded packet and a regular packet. Consequently, the average number of successfully delivered packets is still given by (6.26). The total transmission times of all three phases are given by

$$\begin{aligned} E[T] &= KE[T_1] + KE[T_2] + E[T_R] \\ &= Kg_L(q_{1,2}q_{1,R}) + Kg_L(q_{2,1}q_{2,R}) + E[T_R], \end{aligned} \tag{6.29}$$

where $T_i \sim \text{BGeom}(q_{i,j}q_{i,R}, L)$ for $i = 1, 2$ and $j = \{1, 2\} \setminus \{i\}$ is the transmission times of a single packet sent from S_i, and

$$\begin{aligned} E[T_R] = {} & E[(D_1 - D_2)\mathbb{I}\{D_1 > D_2\}]E[T_{R,2}] + E[(D_2 - D_1)\mathbb{I}\{D_2 > D_1\}]E[T_{R,1}] \\ & + E[\min(D_1, D_2)]E[T_{R,NC}], \end{aligned} \tag{6.30}$$

is the total transmission times of relay R. Here $D_i \sim \text{Bin}(Q_i, K)$ with $Q_i \triangleq \Pr(I_i = -1)$ for $i = 1, 2$ is given by (6.21) and (6.22), and $E[T_{R,i}]$ for $i = 1, 2$ are the average transmission times of a regular packet sent from R to S_i and have been derived in the last subsection. As a result, we only need to compute the average transmission times of a network-coded packet, i.e., $E[T_{R,NC}]$. Since any network-coded flow will terminate until both component unicast flows are finished, we have $T_{R,NC} = \max(T_{R,1}, T_{R,2})$. For per-hop constraint, $T_{R,i} \sim \text{BGeom}(q_{R,i}, L)$ for $i = 1, 2$ are independent. By using property 2 of Proposition 6.1, we have

$$E[T_{R,NC}] = g_L(q_{R,1}) + g_L(q_{R,2}) - g_L(q_{R,1}q_{R,2}). \tag{6.31}$$

For E2E constraint, we have $T_{R,i} \sim \text{BGeom}(q_{R,i}, L - T_j)$ with the condition that $T_j \leq L - 1$ for $i = 1, 2$ and $j = \{1, 2\} \setminus \{i\}$. By using property 4 of Proposition 6.1,

we have

$$E\left[T_{R,NC}\right] = h_1\left(q_{R,1}, q_{R,2}, q_{1,R}q_{1,2}, q_{2,R}q_{2,1}; L\right). \tag{6.32}$$

6.2.4 Dynamic Network Coding

The network coding scheme studied in the last subsection is static in that the pairing pattern is fixed after scheduling. Specifically, once the two packets are determined to be combined by the scheduler, they cannot be paired with any other packets later. This scheme cannot leverage the network coding gain to the full. For example, if one unicast flow terminates earlier than the other one, the intended receiver of that unicast flow has to stay idle and wait until the other unicast flow terminates. So some channel use is wasted along the way. If the relay still has more packets intended for that receiver in the buffer, a more efficient way is to send these packets along with the remaining unicast flow and form a new network-coded packet. In this way, both relay-source channels are in use until the relay node has no more packets intended for any source node. We remark that a similar idea is also studied in [75], where different network-coded packets but not the raw source packets can be re-combined in some cases.

To realize the above idea, the pairing pattern must be determined dynamically depending on the decoding status. If the two buffers $\mathbb{D}_1$ and $\mathbb{D}_2$ are non-empty, the first packet in those two buffers are always combined to form a network-coded packet. Whenever any component unicast flow terminates due to successful decoding or transmission expiration, the related packet will be eliminated from the buffer.

Then the relay will pick up a new packet from that buffer and do network coding. This process continues until one buffer becomes empty. Since then the relay will forward the packets in the remaining buffer as regular packets successively. For example, suppose $\mathbb{D}_1 = \{X_1(1), X_1(2), X_1(3)\}$ and $\mathbb{D}_2 = \{X_2(1)\}$. Then the first network-coded packet is $X_R(1) = X_1(1) \oplus X_2(1)$. Suppose after some transmissions, $X_1(1)$ expires and is removed from $\mathbb{D}_1$ but $X_2(1)$ is still valid. Then the relay will form a new network-coded packet given by $X_R(2) = X_1(2) \oplus X_2(1)$. In contrast, for static network coding scheme, the relay node will keep sending $X_R(2) = X_2(1)$ after removing $X_1(1)$, during which period S_2 always remains idle and the channel $R \to S_2$ is not in use at all. For dynamic network coding, all the remaining packets are always combined whenever possible. In the above example, if $X_2(1)$ terminates earlier than $X_1(2)$, $\mathbb{D}_2$ becomes empty and $\mathbb{D}_1 = \{X_1(2), X_1(3)\}$. Then the relay will send $X_1(2)$ and $X_1(3)$ to S_2 successively as regular packets, which is similar to pure relaying.

Again, dynamic network coding will not affect the probability of successful delivery, and the total transmission times is still given by (6.29). The only difference is the total transmission times of the relay, i.e., $E[T_R]$. Note that given the buffer size D_i, the relay node only needs to deliver D_i packets $X_i(1), X_i(2), \cdots, X_i(D_i)$ to S_j for $i = 1, 2$ and $j = \{1, 2\} \setminus \{i\}$. These packets may be organized in the form of network-coded packets or regular packets depending on the decoding status. Nevertheless, the packet loss rate remains the same. Therefore, if we denote the total transmission times of all packets containing $X_i(k)$ by $T_{R,j}(k)$, we have $T_{R,j}(k) \sim \text{BGeom}(q_{R,j}, L)$ for per-hop constraint and $T_{R,j} \sim \text{BGeom}(q_{R,j}, L - T_i)$ with the condition that

$T_i \sim \text{BGeom}\left(q_{i,j}q_{i,R}, L\right) \leq L-1$ for E2E constraint. As the transmission flow to S_j will terminate when $\mathbb{D}_i$ becomes empty, its duration is equal to the summation of the individual packet transmission times, i.e., $\sum_{k=1}^{D_i} T_{R,j}\left(k\right)$. Finally, as the relay transmission will terminate when both $\mathbb{D}_1$ and $\mathbb{D}_2$ become empty, we have

$$T_R|_{D_1,D_2} = \max\left(\sum_{k=1}^{D_2} T_{R,1}\left(k\right), \sum_{k=1}^{D_1} T_{R,2}\left(k\right)\right). \tag{6.33}$$

Analytically, it is hard to compute $E\left[T_R\right]$ due to the maximum operation. In practice, we can use the following lower bound to get an estimate of $E\left[T_R\right]$, i.e.,

$$E\left[T_R|\, D_1, D_2\right] \geq \max\left(D_2 E\left[T_{R,1}\right], D_1 E\left[T_{R,2}\right]\right). \tag{6.34}$$

Here D_i and $T_{R,i}$ for $i = 1, 2$ have exactly the same distributions as in the static network coding case and have been given in the last subsection. Averaging the above expression over the distribution of D_i will lead to $E\left[T_R\right]$, and the theoretical throughput thus obtained is a tight upper bound as will be shown in simulations. In the special case when there is no maximum transmission constraint (i.e., $L \to \infty$), all the packets will be delivered successfully, and the exact relay transmission times can be obtained through the recursion given by

$$\begin{aligned}\left(1-q_{R,1}q_{R,2}\right) E\left[T_R|\, D_1, D_2\right] &= 1 + \left(1-q_{R,1}\right) q_{R,2} E\left[T_R|\, D_1, D_2-1\right] \\ &+ q_{R,1}\left(1-q_{R,2}\right) E\left[T_R|\, D_1-1, D_2\right] \\ &+ \left(1-q_{R,1}\right)\left(1-q_{R,2}\right) E\left[T_R|\, D_1-1, D_2-1\right]\end{aligned} \tag{6.35}$$

for $D_1, D_2 > 0$, and the boundary conditions are

$$E\left[T_R|\, D_1, D_2 = 0\right] = \frac{D_1}{1-P_{R,2}}, \tag{6.36}$$

$$E\left[T_R|\, D_1 = 0, D_2\right] = \frac{D_2}{1-P_{R,1}}. \tag{6.37}$$

6.2.5 Throughput Comparison

In this subsection, we compare the throughput of different schemes obtained above. To make the analysis tractable, we focus on the symmetric networks where the relay node is located at the middle between S_1 and S_2. Besides, we assume all nodes use the same power and transmit at the same rate. The path loss coefficient is supposed to be dependent on the distance only. Under these assumptions, we have $q_{1,R} = q_{2,R} = q_{R,1} = q_{R,2} \triangleq q_R$ and $q_{1,2} = q_{2,1} \triangleq q_S$. Besides, we assume that the maximum transmission constraint is sufficiently large, i.e., $L \gg 1$. From (6.19) and (6.28), we have

$$\eta_{\text{DT}} = 1 - q_S, \tag{6.38}$$

$$\eta_{\text{Relay}} = \frac{1 - q_R q_S}{1 + q_S} \overset{q_S=1}{=} \frac{1 - q_R}{2}. \tag{6.39}$$

For network coding schemes, the throughput also depends on the frame length K. When $K \gg 1$, we have

$$D_i \to K \frac{(1 - q_R)\, q_S}{1 - q_R q_S}. \tag{6.40}$$

After some simple algebra, we can show that

$$\eta_{\text{S-NC}} = \frac{2\,(1 - q_R q_S)\,(1 + q_R)}{2\,(1 + q_R) + (1 + 2q_R)\, q_S} \overset{q_S=1}{=} \frac{2\,(1 - q_R^2)}{3 + 4q_R}, \tag{6.41}$$

$$\eta_{\text{D-NC}} = \frac{2\,(1 - q_R q_S)}{2 + q_S} \overset{q_S=1}{=} \frac{2\,(1 - q_R)}{3}. \tag{6.42}$$

We first consider the special case when there is no direct link, i.e., $q_S = 1$. In this case, $\eta_{\text{DT}} \equiv 0$ because no information can be delivered through the direct link alone. Besides, we can show that $0 \le \eta_{\text{Relay}} \le \frac{1}{2}$ and $0 \le \eta_{\text{S-NC}}, \eta_{\text{D-NC}} \le \frac{2}{3}$, and the maximum throughput is achieved when $q_R = 0$. Next we consider the general case

with direct link, i.e., $0 \leq q_S < 1$. It is easy to see that $0 \leq \eta_{\text{DT}}, \eta_{\text{Relay}}, \eta_{\text{S-NC}}, \eta_{\text{D-NC}} \leq 1$, and the upper bound is achieved when $q_S = 0$. Note that when the direct link is in good quality (i.e., $q_S \ll 1$), using direct transmission alone is able to achieve the throughput bound. In that case, the relay nodes could stay idle to save the transmitted power and channel use. We remark again that such incremental relaying is the key difference between the current work and previous studies [31,74–77], where direct link is neglected and all the packet transmissions have to go through the relay link. Actually, incorporating the direct link is pretty important in that the throughput upper bound can increase from $\frac{2}{3}$ to 1.

Next we compare the relative throughput gain. By comparing direct transmission and pure relaying, we have

$$\frac{\eta_{\text{Relay}}}{\eta_{\text{DT}}} = \frac{1 - q_R q_S}{1 - q_S^2} > 1 \Leftrightarrow q_S > q_R. \tag{6.43}$$

Therefore, wireless relaying can improve the throughput if and only if the relay link has better quality than direct link. Next we study network coding gain, which is given by

$$1 \leq \frac{\eta_{\text{S-NC}}}{\eta_{\text{Relay}}} = 1 + \frac{q_S}{2 + 2q_R + q_S + 2q_R q_S} \leq \frac{4}{3}. \tag{6.44}$$

Clearly, static network coding is strictly better than pure relaying in terms of the achievable throughput. However, the relative throughput gain is bounded by 33.3%, which occurs when the direct link is always in outage and the relay link is in perfect condition. Finally, we study the dynamic network coding gain, which is given by

$$1 \leq \frac{\eta_{\text{D-NC}}}{\eta_{\text{S-NC}}} = 1 + \frac{q_R q_S}{2 + 2q_R + q_S + q_R q_S} \leq \frac{7}{6}. \tag{6.45}$$

It is observed that dynamic network coding can further improve the throughput. However, the relative gain is at most 16.7%. Note that the largest throughput gain is achieved when both direct link and relay link are in bad conditions. In most reasonable system settings, such dynamic network coding gain is not that significant, as will be seen in simulations later.

6.2.6 Power Allocation

In the last subsection, we compared the throughput by assuming equal power allocation. For practical wireless networks, the node distribution could be quite random, and the transmitted power should be properly allocated to address the near-far problem. However, the optimum transmitted power can be found only through exhaustive search, as the closed-form throughput is hard to manipulate. In the sequel, we seek to develop a near-optimum power allocation strategy with closed-form solution.

For the packet loss rate, we use the channel outage model given by (6.2). To make a step further, we intentionally neglect the maximum transmission constraint and the direct link, i.e., let $L \to \infty$ and $q_{1,2} = q_{2,1} = 1$. From (6.28), the throughput of pure relaying is then equal to

$$\begin{aligned}\eta_{\text{Relay}} &\approx 2\left(\exp\left(\frac{\alpha}{\lambda_{1,R}P_1}\right) + \exp\left(\frac{\alpha}{\lambda_{2,R}P_2}\right) + \exp\left(\frac{\alpha}{\lambda_{R,1}P_R}\right) + \exp\left(\frac{\alpha}{\lambda_{R,2}P_R}\right)\right)^{-1} \\ &\leq \frac{1}{2}\exp\left(-\frac{\alpha}{4}\left(\frac{1}{\lambda_{1,R}P_1} + \frac{1}{\lambda_{2,R}P_2} + \frac{\lambda_{R,1}+\lambda_{R,2}}{\lambda_{R,1}\lambda_{R,2}P_R}\right)\right), \end{aligned} \tag{6.46}$$

where $\alpha = 2^r - 1$ and we use the arithmetic-geometric mean inequality in the last step. To maximize the throughput, we use the upper bound instead. The power

allocation problem is formulated as

$$\begin{aligned} &\max \eta_{\mathrm{R}elay} \approx \min \left(\frac{1}{\lambda_{1,R} P_1} + \frac{1}{\lambda_{2,R} P_2} + \frac{\lambda_{R,1} + \lambda_{R,2}}{\lambda_{R,1} \lambda_{R,2} P_R} \right) \\ &\text{s.t. } P_1 + P_2 + P_R \leq 3P. \end{aligned} \tag{6.47}$$

By using the method of Lagrange multipliers, we can derive the optimum solution given by

$$\begin{cases} P_1 = \frac{3P\lambda_{1,R}^{-1/2}}{\lambda_{1,R}^{-1/2} + \lambda_{2,R}^{-1/2} + \left(\lambda_{R,1}^{-1} + \lambda_{R,2}^{-1}\right)^{1/2}} \\ P_2 = \frac{3P\lambda_{2,R}^{-1/2}}{\lambda_{1,R}^{-1/2} + \lambda_{2,R}^{-1/2} + \left(\lambda_{R,1}^{-1} + \lambda_{R,2}^{-1}\right)^{1/2}} \\ P_R = \frac{3P\left(\lambda_{R,1}^{-1} + \lambda_{R,2}^{-1}\right)^{1/2}}{\lambda_{1,R}^{-1/2} + \lambda_{2,R}^{-1/2} + \left(\lambda_{R,1}^{-1} + \lambda_{R,2}^{-1}\right)^{1/2}} \end{cases}. \tag{6.48}$$

For dynamic network coding, $L \to \infty$ and $q_{1,2} = q_{2,1} = 1$ imply that $D_1 = D_2 \equiv K$ in (6.33), which leads to

$$\begin{aligned} \eta_{\text{D-NC}} &\approx 2\left(\exp\left(\frac{\alpha}{\lambda_{1,R} P_1} \right) + \exp\left(\frac{\alpha}{\lambda_{2,R} P_2} \right) + \exp\left(\frac{\alpha}{\min\left(\lambda_{R,1}, \lambda_{R,2}\right) P_R} \right) \right)^{-1} \\ &\leq \frac{2}{3} \exp\left(-\frac{\alpha}{3} \left(\frac{1}{\lambda_{1,R} P_1} + \frac{1}{\lambda_{2,R} P_2} + \frac{1}{\min\left(\lambda_{R,1}, \lambda_{R,2}\right) P_R} \right) \right). \end{aligned} \tag{6.49}$$

Likewise, the power allocation problem can be formulated as

$$\begin{aligned} &\max \eta_{\text{D-NC}} \approx \min \left(\frac{1}{\lambda_{1,R} P_1} + \frac{1}{\lambda_{2,R} P_2} + \frac{1}{\min\left(\lambda_{R,1}, \lambda_{R,2}\right) P_R} \right) \\ &\text{s.t. } P_1 + P_2 + P_R \leq 3P. \end{aligned} \tag{6.50}$$

The optimizer is given by

$$\begin{cases} P_1 = \frac{3P\lambda_{1,R}^{-1/2}}{\lambda_{1,R}^{-1/2} + \lambda_{2,R}^{-1/2} + \max\left(\lambda_{R,1}^{-1/2}, \lambda_{R,2}^{-1/2}\right)} \\ P_2 = \frac{3P\lambda_{2,R}^{-1/2}}{\lambda_{1,R}^{-1/2} + \lambda_{2,R}^{-1/2} + \max\left(\lambda_{R,1}^{-1/2}, \lambda_{R,2}^{-1/2}\right)} \\ P_R = \frac{3P \max\left(\lambda_{R,1}^{-1/2}, \lambda_{R,2}^{-1/2}\right)}{\lambda_{1,R}^{-1/2} + \lambda_{2,R}^{-1/2} + \max\left(\lambda_{R,1}^{-1/2}, \lambda_{R,2}^{-1/2}\right)} \end{cases}. \tag{6.51}$$

6.3 Multi-Relay Case

In this section, we extend the above results to the multi-relay networks. The packet loss rates of $S_i \rightarrow R_k$ channel and $R_k \rightarrow S_i$ channel are respectively denoted by q_{i,R_k} and $q_{R_k,i}$ for $i = 1, 2$ and $k = 1, 2, \cdots, N$. The buffers of R_k are denoted by $\mathbb{D}_{1,R_k}$ and $\mathbb{D}_{2,R_k}$. The power of S_i is denoted by P_i and the power of R_k is denoted by P_{R_k}. We assume the transmitted power is fixed during packet transmission and as a result, the packet loss rate depends only on the distance between the transmitter and the receiver. Therefore, we can rank the quality of $R_k \rightarrow S_i$ channel according to the distance, and such ranking is assumed to be known to the whole network.

6.3.1 Successive Relaying

One simple multi-relay transmission strategy is successive relaying, where all the relays work in a time-division multiplexing manner. The whole bidirectional communication is still completed in three phases as in the single-relay case. During the source transmission phase, each source packet is either dropped due to exceeding the maximum transmission constraint, or decoded by the intended receiver, or decoded by some relay nodes. Note that if multiple relays happen to decode the same packet of S_i for $i = 1, 2$ simultaneously, we assume that only the relay with the best $R_k \rightarrow S_j$ channel quality for $j = \{1, 2\} \setminus \{i\}$ will store the packet and continue the unfinished ARQ later. If several relay nodes happen to decode the same packet and have the same channel quality, only one relay is randomly selected. Consequently, after the two source transmission phases, all relays may store some

mutually exclusive packets in the local buffers. Then in the third relaying phase, all relays take turns to send these packets to the intended receivers using either pure relaying, static network coding or dynamic network coding studied in the last subsection.

Next we study the throughput of such successive relaying. As the exact analysis is intractable, we only consider the special case when the maximum transmission constraint is sufficiently large, i.e., $L \gg 1$. Besides, we assume symmetric settings such that $q_{i,R_k} = q_{R_k,i} \triangleq q_R$ and $q_{1,2} = q_{2,1} \triangleq q_S$ for $i = 1, 2$ and $k = 1, 2, \cdots, N$. We first study the scenario that $N \gg K = 1$, in which case only one packet is exchanged between source nodes. Depending on the decoding status, either both packets are successfully delivered through direct link, or only one packet is delivered through direct link and the other one is decoded by the relay but not the intended receiver, or both source packets are decoded by some relay nodes. The probabilities of these three events are $(1 - q_S)^2$, $2q_S(1 - q_S)$ and q_S^2, respectively. As we assume $N \gg 1$, after each source transmission the packet will be either decoded by some relay node or successfully delivered to the intended receiver. As a result, the total transmission times of the two source transmission phases are exactly 2. For each regular packet delivered by the relay, the average transmission times is $\frac{1}{1-q_R}$. For each network-coded packet, the average transmission times is $\frac{1+2q_R}{1-q_R^2}$ from (6.31). Note that when $K = 1$, static network coding becomes equivalent to dynamic network coding, and network coding can only be performed when a single relay happens to decode both

source packets. After some simple algebra, we can obtain

$$\eta_{\text{Relay}} = \frac{1-q_R}{1-q_R+q_S}, \tag{6.52}$$

$$\eta_{\text{NC}} = \left(\frac{1-q_R+q_S}{1-q_R} - \frac{q_S^2}{2N\left(1-q_R^2\right)}\right)^{-1} \overset{N\gg 1}{\approx} \eta_{\text{Relay}}. \tag{6.53}$$

It is observed that network coding can barely boost throughput when the number of relays is much larger than the frame length. This is because the source packets will be decoded by different relays with large chance, in which case the network coding gain is pretty limited. Next we study the throughput when the frame length is sufficiently large, i.e., $K \gg N \gg 1$. On average, each relay node can decode $\frac{K}{N}q_S$ packets from both sources. So for dynamic network coding, the total transmission times of each relay node is approximately $\frac{K}{N}\frac{q_S}{1-q_R}$ from (6.34). Consequently, we have

$$\eta_{\text{D-NC}} = \frac{2\left(1-q_R\right)}{2-2q_R+q_S}. \tag{6.54}$$

The throughput of pure relaying is independent of K and is still given by (6.52). The relative throughput gain is

$$1 \leq \frac{\eta_{\text{D-NC}}}{\eta_{\text{Relay}}} = 1 + \frac{q_S}{2-2q_R+q_S} \leq 2 \tag{6.55}$$

for large K. It is observed that dynamic network coding can greatly improve throughput and the throughput gain is up to 100%.

6.3.2 Hybrid Network Coding

For successive relaying, each relay node applies DNC alone on the locally stored packets. However, the network coding gain is quite limited in the small frame length

case. In this section, we develop a hybrid network coding scheme to address this issue. Let us start with an example, where there are $N = 2$ relay nodes and there is $K = 1$ packet per frame. Suppose in a particular frame transmission, we have $\mathbb{D}_{1,R_1} = \{X_1\}$, $\mathbb{D}_{2,R_2} = \{X_2\}$ and $\mathbb{D}_{2,R_1} = \mathbb{D}_{1,R_2} = \phi$. That is, each relay node is able to decode only one source packet and for successive relaying, it is impossible to perform network coding at both relays. However, if the two relays send X_1 and X_2 simultaneously, the transmitted signal will be automatically combined in the air, which is a nature form of ANC [25]. The received signal at S_i is given by

$$y_i = h_{R_1,i}\sqrt{P_{R_1}}X_1 + h_{R_2,i}\sqrt{P_{R_2}}X_2 + n_i \tag{6.56}$$

for $i = 1, 2$. At S_1, it can fist subtract the self packet X_1 from the mixed signal and then make decoding of X_2. The packet loss rate is still given by $q_{R_2,1}$ after the self-interference is perfectly eliminated. Similar operations can be performed at S_2 to decode X_1. In this way, both packets may be delivered in a single channel use even though they are stored at different relays.

For the general case, each relay can first perform DNC to combine as many local packets as possible. Whenever one of the two buffers becomes empty, say $\mathbb{D}_1$, it can invite another relay node with non-empty $\mathbb{D}_1$ to jointly perform ANC. This process continues until it is impossible to pair any two packets within the relay array. Afterwards, all the relays will take turns to deliver the remaining regular packets. In some sense, the relay array behaves like a super relay node that has distributed buffers. It can adaptively choose to perform DNC or ANC depending on the location of the packets. However, the relay nodes do not need to know the

packets in other buffers, so the signaling overhead is very limited.

It is intractable to analyze the throughput of hybrid network coding for the general case. Instead, we still consider symmetric setting, i.e., $q_{i,R_k} = q_{R_k,i} \overset{\Delta}{=} q_R$ and $q_{1,2} = q_{2,1} \overset{\Delta}{=} q_S$ for $i = 1, 2$ and $k = 1, 2, \cdots, N$. Under this assumption, all the relay nodes have equal chance to decode a source packet, and the packet loss rate is identical for all source-relay links. As a result, we can regard the relay array as a single super relay node. The packet loss rate from any source to this super relay node is q_R^N, which is the probability that none of the N relays can decode the source packet. The packet loss rate from the relay array to any source is still given by q_R, as each packet is delivered only by one relay node, and the packet loss rate remains the same regardless of network coding scheme. Consequently, the throughput of network coding can be obtained from the single-relay result after replacing with $q_{1,R} = q_{2,R} = q_R^N$. When $L \gg 1$ and $N \gg 1$, we have

$$\eta_{\text{D-NC}} = \begin{cases} \dfrac{2\left(1-q_R\right)}{2-2q_R+q_S}, K \gg 1 & \text{(6.57a)} \\ \dfrac{2\left(1-q_R^2\right)}{2\left(1-q_R^2\right)+2q_S\left(1+q_R\right)-q_S^2}, K = 1 & \text{(6.57b)} \end{cases}$$

By comparing (6.54) and (6.57a), we can observe that hybrid network coding and successive network coding achieve the same asymptotic throughput for large K. For small K (i.e., $K = 1 \ll N$), the relative throughput gain of hybrid network coding is

$$1 \leq \frac{\eta_{\text{Hybrid}}}{\eta_{\text{Successive}}} = 1 + \frac{q_S^2}{2\left(1-q_R^2\right)+2q_S\left(1+q_R\right)-q_S^2} \leq \frac{4}{3}. \tag{6.58}$$

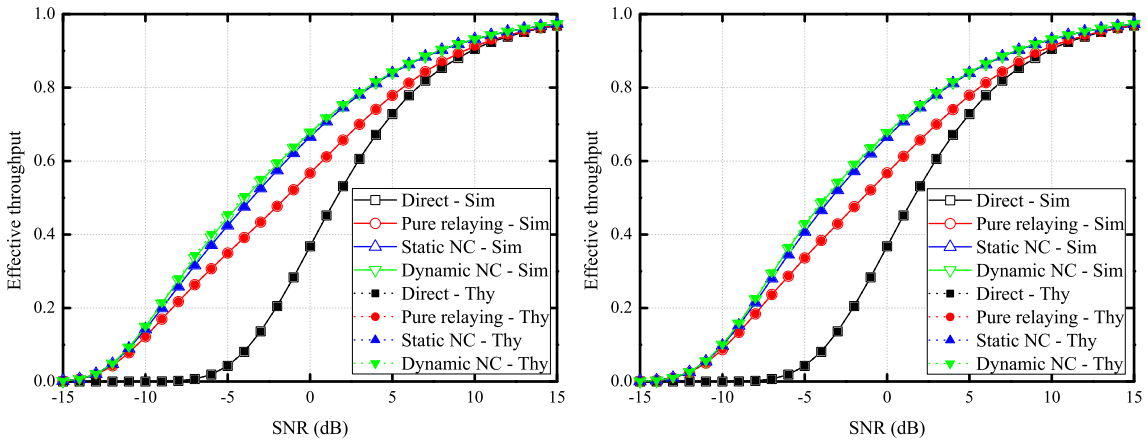
In this case, hybrid network coding can greatly improve the throughput, but the throughput gain is up to 33.3%.

6.4 Simulations

In this section, we present some simulation results to validate our study. Throughout simulations, we use the path loss model $\lambda = d^{-3}$, where λ is the path loss coefficient and d is the distance. The noise power is always normalized, and the average transmitted power of all nodes is referred to as SNR. The two source nodes are always located at $(0, 0)$ and $(1, 0)$, respectively.

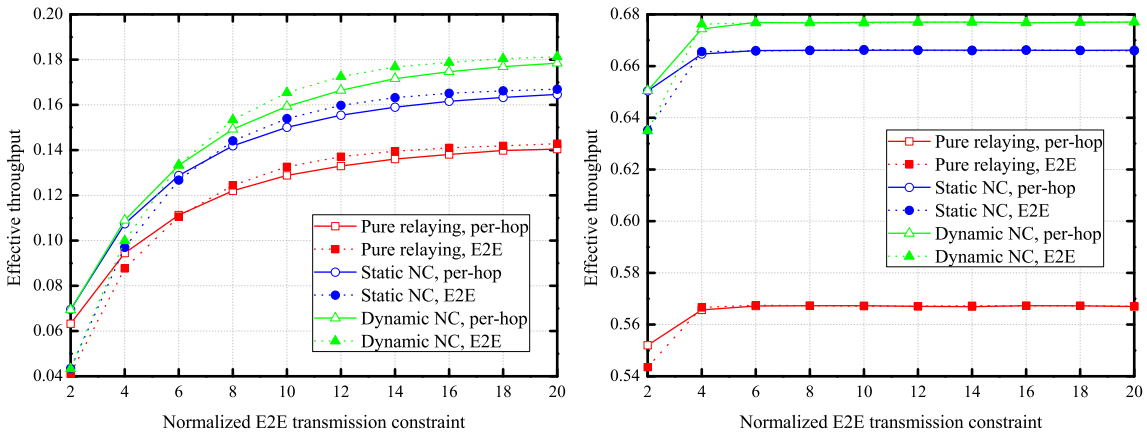
We first compare the throughput of the four transmission schemes in the single-relay networks in Figure 6.2. It is observed that the simulation results match perfectly with our theoretical results regardless of the type of transmission constraint. Compared with direct transmission, wireless relaying can greatly boost the throughput when the relay link is much better than the direct link. Network coding can further improve the throughput at moderate SNRs. For example, at 0dB the throughput goes from 0.57 to 0.68, and the relative gain is around 20%. Comparatively, dynamic network coding has the best performance under all situations, but the throughput gain against static network coding is very limited over the entire SNR range, which is consistent with the analytical results. It is also observed that the four schemes converge to the same performance at high SNRs, in which case most of the packets are delivered through the direct link and the relay link is active only occasionally.

By comparing Figure 6.2(a) and Figure 6.2(b), we also observe that per-hop constraint performs better than E2E constraint at low SNRs, while the throughput is almost identical at medium-to-high SNRs. This is because for per-hop constraint,



(a) Per-hop maximum transmission constraint. (b) E2E maximum transmission constraint.

Figure 6.2: Effective throughput versus SNR for $L = 4$ and $K = 10$. The relay node is located at $(0.5, 0)$.



(a) SNR $= -10$dB. (b) SNR $= 0$dB.

Figure 6.3: Effective throughput versus normalized transmission constraint for $K = 10$. The relay node is located at $(0.5, 0)$.

the same packet can be transmitted up to L times for each hop, whereas for E2E constraint the total transmission times are limited by L regardless of the sender. Therefore, the successful delivery probability is higher for per-hop constraint. To fairly compare these two types of constraints, we normalize the aggregate E2E trans-

mission times and compare the throughput with different maximum transmission constraint in Figure 6.3. For example, if $L = 4$ for E2E constraint, then the same packet can only be transmitted up to 2 times for the per-hop constraint. We observe that at low SNRs, the throughput increases quickly with the maximum transmission times because allowing more ARQs will improve the successful delivery probability. In contrast, at high SNRs the throughput almost remains the same when $L \geq 4$, as most packets can be delivered within 4 attempts. It is also observed that for all relaying schemes, E2E constraint performs better when the maximum transmission constraint is not that stringent. This is because for E2E constraint, the transmission chances are dynamically allocated between the two hops, whereas for per-hop constraint such split is fixed. So there is some throughput loss due to early give-up under per-hop constraint. In contrast, per-hop constraint leads to higher throughput for very stringent maximum transmission constraint. This is because when the first few attempts are failed, that packet should be dropped early to save the channel use for transmitting a new packet. Otherwise, even the relay node is able to decode the packet later, there are very limited ARQs left and the chance of successful delivery is still pretty low. As a result, early termination appears to be a better choice for small L.

Next we study the impact of power allocation in Figure 6.4. The relay node is located at $(D_{sr}, 0)$, and we plot the throughput with different relay locations. It is observed that our power allocation schemes (6.48) and (6.51) perform very close to the optimum ones that are obtained through exhaustive search. When the network topology is highly asymmetric, i.e., when the relay node is very close to

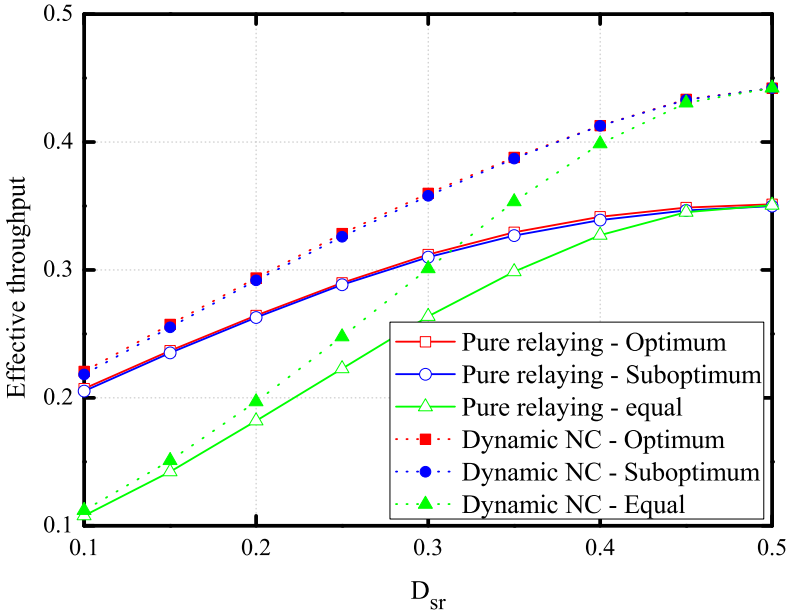


Figure 6.4: Effective throughput versus relay position with power allocation for SNR $= -5$dB, $K = 5$ and $L = \infty$.

one source, optimum power allocation can almost double the throughput against equal power allocation. This is because some source-relay link will become the system bottleneck, and that link limits the throughput of the whole system. Our power allocation schemes try to address this issue by allocating more power to the end terminal with larger path loss, such that the packet loss rate of that link got improved to some extent.

Finally we study the throughput of multi-relay network in Figure 6.5 and Figure 6.6, where all the relays are assumed to locate at $(0.5, 0)$. In Figure 6.5, we fix the number of relays and change the frame length. It is observed that successive relaying performs very close to pure relaying when the frame length is only 1, as the chance to perform network coding locally is pretty low. As the frame length

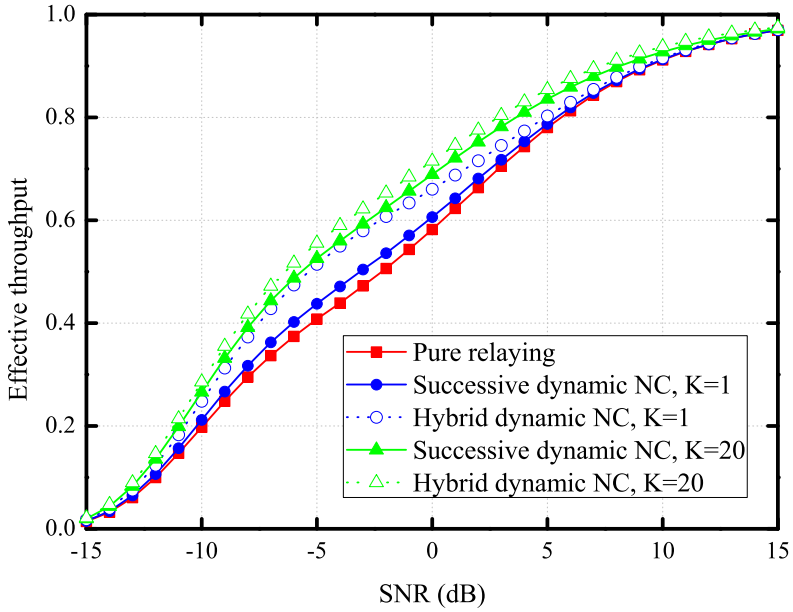


Figure 6.5: Effective throughput versus SNR with $N = 3$ relays for $L = \infty$. All relay nodes are located at $(0.5, 0)$.

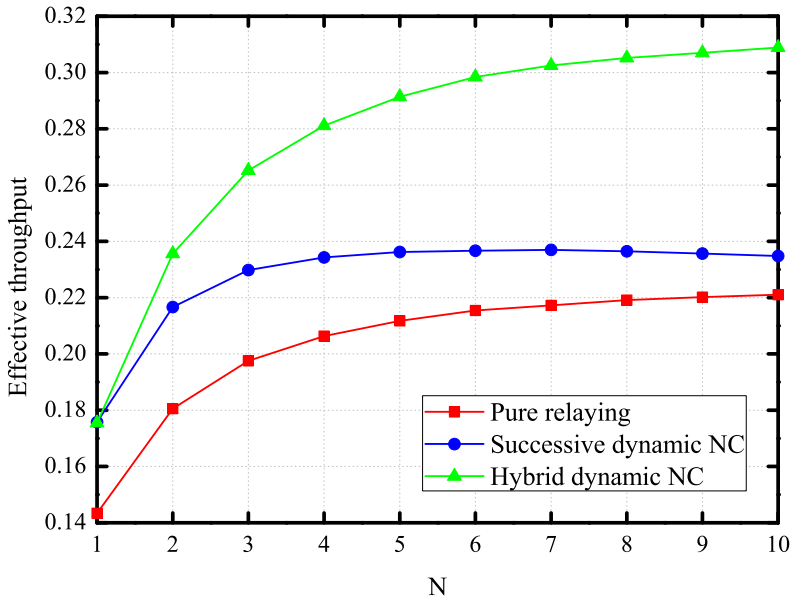


Figure 6.6: Effective throughput versus the number of relays for SNR $= -10$dB, $K = 3$ and $L = \infty$. All relay nodes are located at $(0.5, 0)$.

becomes much larger than the number of relays, most packets can be paired together to leverage the network coding gain. That is why the throughput is increasing with the frame length. In that case, successive network coding performs almost as well as hybrid network coding, which is consistent with our analysis. Then in Figure 6.6, we investigate the throughput with fixed frame length but different number of relays. It is observed that the gap between dynamic network coding and successive network coding becomes wider as N increases. Besides, the throughput of successive network coding converges to that of pure relaying when the number of relays exceeds the frame length. In sum, the throughput of successive network coding depends largely on the frame length and the number of relays, whereas hybrid network coding can overcome such shortcoming by smartly switching between DNC and ANC.

6.5 Conclusions

In this work, we studied the throughput of TWRC with network-coded ARQ and quantified the network coding gain. We showed that network coding can greatly improve the throughput, but the gain is well bounded. We also derived the near-optimum power allocation strategy and demonstrated that the end terminal of the bottleneck link should use more power to improve the throughput. Finally for the multi-relay network, we showed that successive relaying suffers significant throughput loss for small frame length, whereas hybrid network coding scheme can leverage the network coding gain to the full.

Chapter 7

Clustering Based Space-Time Network Coding

So far, we have focused on the wireless networks where there are dedicated relays. In the cases where there is no dedicated relay in the systems, user devices have to help each other to exploit the cooperative diversity gain. One such example is sensor network, where all the sensor nodes form clusters for data routing. The situation is very complicated, because users have to deliver both local data and relayed data to the intended receiver and at the same time, they also need to share their local data with others to achieve cooperative diversity gain. How to coordinate data sharing and data relaying in such networks has been an open design problem for a long time, and various user cooperation protocols [80–82] have been developed. One common drawback of those strategies is that they tend to pursue the largest diversity gain, while the bandwidth efficiency is relatively low. In all those methods, the data sharing phase and data relaying phase are separated via orthogonal channel use, therefore causing a huge loss of spectral efficiency.

So in this work, we aim to develop a new multi-user cooperation strategy that can achieve better tradeoff between diversity gain and bandwidth efficiency. We extend the conventional inter-user cooperation to the more general inter-cluster cooperation. To be specific, the whole network is divided into several small clusters, and different clusters help each other to relay the data. There is no intra-cluster co-

operation; however, transmit beamforming is used within each cluster to guarantee coherent combining of the same relayed data in the air. We no longer distinguish data relaying phase from the data sharing phase. Instead, each user would use linear network coding to combine the local data and the relayed data if the previous decoding is successful. Linear decorrelator is used at the receiver side to separate different source signals, and equal-gain combining (EGC) is performed to fully exploit cooperative diversity. We obtain both the exact SER and asymptotic SER of the M-ary phase-shift keying signal. It is shown that different tradeoffs between diversity gain and bandwidth efficiency can be achieved by adjusting the formation of clusters.

Notations: Boldface uppercase and lowercase letter represent matrix and column vector, respectively. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ stand for conjugate, transpose and conjugate transpose, respectively. We shall use abbreviation i.i.d. for independent and identically distributed, and denote $Z \sim \mathcal{CN}(\mu, \sigma^2)$ as a circularly symmetric complex Gaussian random variable. Finally, the probability of an event $\mathcal{A}$ and the PDF of a random variable Z are denoted by $\Pr(\mathcal{A})$ and $f(Z)$, respectively.

7.1 Transmission Strategy

Consider a wireless uplink channel where N source nodes send data to a single destination node d via M-ary PSK signals, as shown in Figure 7.1. All source nodes are divided into K clusters, each having Q nodes (i.e., $N = KQ$). The ith node in the nth cluster is denoted by i_n for $i = 1, 2, \cdots, Q$ and $n = 1, 2, \cdots, K$. Suppose

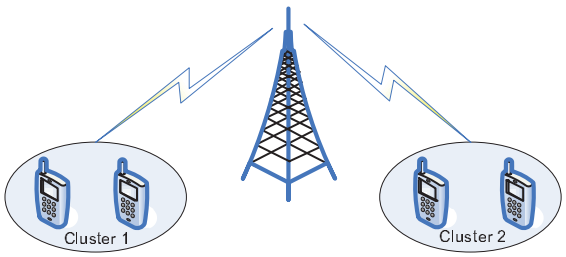


Figure 7.1: System model of the wireless uplink with user clustering

each node has a unique signature waveform $s_{i_n}(t)$, and the cross-correlation of those signature waverforms are defined as

$$\rho_{i_n,j_m} = \langle s_{i_n}(t), s_{j_m}(t) \rangle = \frac{1}{T_s} \int_0^{T_s} s_{i_n}(t) s^*_{j_m}(t) dt = \begin{cases} 1, \text{ if } i = j \text{ and } n = m \\ \rho, \text{ otherwise} \end{cases}, \quad (7.1)$$

where T_s is the symbol period. As [82], we assume those waveforms are known at each node. The cross-correlation coefficient $\rho \leq 1$ is a design parameter depending on the orthogonality of different waveforms. As will be clear later, ρ only determines the coding gain but is independent of the diversity gain.

Without loss of generality, the AWGN at any receiver is assumed to be i.i.d. $\mathcal{CN}(0, N_0)$, and the channel between any two nodes u and v is modeled as $h_{u,v} \sim \mathcal{CN}\left(0, \sigma^2_{u,v}\right)$. We assume all the channels experience slow fading, and each transmitter knows the phase of its own channel to the destination. When all the nodes operate in the same frequency band and the channels are reciprocal, this can be done by letting the destination broadcast a training sequence such that each node can perform channel estimation and thus acquire the channel phases.

As there are K clusters, each transmission phase is divided into K time slots. The nth time slot within the lth transmission phase is denoted by T^l_n. Each cluster

is allowed to transmit data only in its assigned time slot, and all the clusters take turns to transmit data in a TDMA manner. Without loss of generality, we assume that the nth time slot is assigned to the nth cluster. In each dedicated time slot, all nodes within a certain cluster would send data simultaneously[1], and all the other clusters and the destination node would listen and attempt to decode the desired data. The transmitted signal of each node is a linear combination of its local symbol and the most recent symbols of other clusters. Let the symbol transmitted by node i_n during time slot T_n^l be $x_{i_n}^l$. Then during time slot T_n^l, the transmitted signal of node i_n consists of the local symbol $x_{i_n}^l$ and the symbols sent by other clusters in the previous $K-1$ time slots $T_{n+1}^{l-1}, \cdots, T_{n-1}^l$, which are given by $x_{j_m}^l$ for $j = 1, 2, \cdots, Q$ and $m = 1, 2, \cdots, n-1$, and $x_{j_m}^{l-1}$ for $j = 1, 2, \cdots, Q$ and $m = n+1, n+2, \cdots, K$. Note that during the initial transmission phase (i.e., $l = 1$), the relayed symbols $x_{j_m}^{l-1}$ for $m = n+1, n+2, \cdots, K$ are equal to zero as there are no symbols received from those clusters yet.

With the above notations, the transmitted signal of node i_n during time slot T_n^l is given by

$$\begin{aligned} f_{i_n}\left(\mathbf{x}^{(n,l)}\right)(t) = & \frac{h_{i_n,d}^*}{|h_{i_n,d}|}\sqrt{\mu_{i_n,i_n}P}I_{i_n,i_n}^l x_{i_n}^l s_{i_n}(t) \\ & + \frac{h_{i_n,d}^*}{|h_{i_n,d}|}\sum_{j=1}^{Q}\sum_{m=1}^{n-1}\sqrt{\mu_{j_m,i_n}P}I_{j_m,i_n}^l x_{j_m}^l s_{j_m}(t) \\ & + \frac{h_{i_n,d}^*}{|h_{i_n,d}|}\sum_{j=1}^{Q}\sum_{m=n+1}^{K}\sqrt{\mu_{j_m,i_n}P}I_{j_m,i_n}^{l-1} x_{j_m}^{l-1} s_{j_m}(t), \end{aligned} \tag{7.2}$$

where the first term corresponds to the local symbol and the last two terms corre-

[1]We assume the transmitters are perfectly synchronized. The effect of synchronization errors is beyond the scope of this work.

spond to relayed symbols. P stands for the total transmitted power of each node, and μ_{j_m,i_n} is the portion of power allocated to relay node j_m's symbol. For power normalization, we have $\mu_{i_n,i_n} + \sum_{j=1}^{Q} \sum_{m \neq n} \mu_{j_m,i_n} = 1$. In this work, we assume the decoding errors could be perfectly detected, and the decoding state of symbol $x_{j_m}^l$ at node i_n is denoted by

$$I_{j_m,i_n}^l = \begin{cases} 1, \text{ if node } i_n \text{ decode } x_{j_m}^l \text{ correctly} \\ 0, \text{ otherwise} \end{cases} \tag{7.3}$$

for $m \neq n$. For notational convenience, we further define

$$I_{j_m,i_n}^l = \begin{cases} 1, \ j = i \\ 0, \text{ otherwise} \end{cases} \tag{7.4}$$

for $m = n$. If node i_n fails to decode a specific symbol $x_{j_m}^l$ of other clusters, that symbol would be dropped instantly and would not be relayed by node i_n in its next dedicated time slot. Another feature is that each node does not relay symbols of the nodes in the same cluster, because those nodes transmit in the same time slot and it is impossible to receive their signals due to half-duplex constraint. However, multiple nodes within the same cluster may happen to relay the same symbol for other clusters simultaneously. To guarantee coherent combining of those replicas in the air, transmit beamforming coefficient $\frac{h_{i_n,d}^*}{|h_{i_n,d}|}$ is used to counter the channel phase distortion. For notational convenience, we define $\mathbf{x}_k^l = \left(x_{1_k}^l, x_{2_k}^l, \cdots, x_{Q_k}^l\right)^T$ as the local symbol vector of all the nodes in the kth cluster during transmission phase T_k^l, and we also define the super symbol vector $\mathbf{x}^{(n,l)} = \left({\mathbf{x}_1^l}^T, \cdots, {\mathbf{x}_n^l}^T, {\mathbf{x}_{n+1}^{l-1}}^T, \cdots, {\mathbf{x}_K^{l-1}}^T\right)^T$.

7.2 Multiuser Detection

Because all the user signals are combined in the air, MUD should be used for separating the source signals at the receiver side. One widely used MUD scheme is linear decorrelator [96], which totally eliminates co-channel interference. In the sequel, we discuss the receiver operation at other source nodes and at the receiver, respectively.

7.2.1 Source Decoding

When a certain cluster is transmitting, all the remaining clusters would listen and attempt to decode the local symbols of that cluster. According to the transmitted signal model (7.2), the received signal during time slot T_m^l at the source node i_n for $n \neq m$ is given by

$$\begin{aligned} y_{i_n}^{T_m^l}(t) &= \sum_{j=1}^{Q} h_{j_m,i_n} f_{j_m}\left(\mathbf{x}^{(m,l)}\right)(t) + w_{i_n}^{T_m^l}(t) \\ &= \sum_{r=1}^{Q} \left\{ \sum_{k=1}^{m} b_{r_k,i_n}^{(m,l)} x_{r_k}^l s_{r_k}(t) + \sum_{k=m+1}^{K} b_{r_k,i_n}^{(m,l-1)} x_{r_k}^{l-1} s_{r_k}(t) \right\} + w_{i_n}^{T_m^l}(t), \end{aligned} \quad (7.5)$$

where

$$b_{r_k,i_n}^{(m,l)} = \sum_{j=1}^{Q} h_{j_m,i_n} \frac{h_{j_m,d}^*}{|h_{j_m,d}|} \sqrt{\mu_{r_k,j_m} P} I_{r_k,j_m}^l \quad (7.6)$$

is the equivalent channel coefficient from the mth cluster to node i_n for symbol $x_{r_k}^l$ during time slot T_m^l. Although the transmitted symbols are mixed in the air, a linear decorrelator can be used to decouple those symbols. To be specific, the received signal $y_{i_n}^{T_m^l}(t)$ is first fed into the matched filter bank $\left\{s_{u_q}(t)\right\}$ for $u = 1, 2, \cdots, Q$

and $q = 1, 2, \cdots, K$, the output of the u_qth branch is given by

$$\begin{aligned} y_{u_q,i_n}^{T_m^l} &= \left\langle y_{i_n}^{T_m^l}(t), s_{u_q}(t) \right\rangle \\ &= \sum_{r=1}^{Q} \left\{ \sum_{k=1}^{m} b_{r_k,i_n}^{(m,l)} x_{r_k}^{l} \rho_{r_k,u_q} + \sum_{k=m+1}^{K} b_{r_k,i_n}^{(m,l-1)} x_{r_k}^{l-1} \rho_{r_k,u_q} \right\} + w_{u_q,i_n}^{T_m^l} \end{aligned} \tag{7.7}$$

Those scalar outputs can be put together in a more compact matrix form as

$$\mathbf{y}_{i_n}^{T_m^l} = \left(y_{1_1,i_n}^{T_m^l}, \cdots, y_{Q_K,i_n}^{T_m^l} \right)^T = \mathbf{R}\mathbf{B}_{i_n}^{(m,l)} \mathbf{x}^{(m,l)} + \mathbf{w}_{i_n}^{T_m^l}, \tag{7.8}$$

where $\mathbf{R}$ is the $N \times N$ correlation matrix of the signature waveforms with 1 on the diagonal and all the off-diagonal elements being ρ,

$$\mathbf{B}_{i_n}^{(m,l)} = diag\left(b_{1_1,i_n}^{(m,l)}, \cdots, b_{Q_m,i_n}^{(m,l)}, b_{1_{m+1},i_n}^{(m,l-1)}, \cdots, b_{Q_K,i_n}^{(m,l-1)} \right)$$

is a diagonal matrix with the equivalent channel coefficients for the corresponding symbols on the main diagonal, and $\mathbf{w}_{i_n}^{T_m^l} \sim \mathcal{CN}\left(0, N_0\mathbf{R}\right)$ is the equivalent AWGN vector. Now the source symbols can be easily decoupled by pre-multiplying $\mathbf{y}_{i_n}^{T_m^l}$ by $\mathbf{R}^{-1}$, i.e.,

$$\tilde{\mathbf{y}}_{i_n}^{T_m^l} = \mathbf{R}^{-1}\mathbf{y}_{i_n}^{T_m^l} = \mathbf{B}_{i_n}^{(m,l)} \mathbf{x}^{(m,l)} + \tilde{\mathbf{w}}_{i_n}^{T_m^l}. \tag{7.9}$$

After the source signals are separated, the element-wise single symbol decoding can be performed on

$$\tilde{y}_{j_m,i_n}^{T_m^l} = b_{j_m,i_n}^{(m,l)} x_{j_m}^{l} + \tilde{w}_{j_m,i_n}^{T_m^l} = h_{j_m,i_n} \frac{h_{j_m,d}^*}{|h_{j_m,d}|} \sqrt{\mu_{j_m,j_m} P} x_{j_m}^{l} + \tilde{w}_{j_m,i_n}^{T_m^l} \tag{7.10}$$

to extract all the local symbols $\left\{ x_{j_m}^{l} \right\}_{j=1}^{Q}$ transmitted by the mth cluster during time slot T_m^l, where $\tilde{w}_{j_m,i_n}^{T_m^l} \sim \mathcal{CN}\left(0, N_0\rho_N\right)$ is the equivalent AWGN and

$$\rho_N = \frac{1 + (N-2)\,\rho}{1 + (N-2)\,\rho - (N-1)\,\rho^2} \tag{7.11}$$

is the noise enhancement factor due to decorrelation. The conditional SNR is given by

$$\gamma^{l}_{j_m,i_n|h_{j_m,i_n}} = \frac{\mu_{j_m,j_m}\gamma}{\rho_N}|h_{j_m,i_n}|^2, \tag{7.12}$$

where $\gamma = \frac{P}{N_0}$ is the reference system SNR.

7.2.2 Destination Decoding

When a certain cluster is transmitting, the destination node would listen and attempt to decode both the local symbols and the relayed symbols. After decorrelation, the vector output during time slot T^l_m is given by

$$\tilde{\mathbf{y}}^{T^l_m}_d = \mathbf{A}^{(m,l)}_d \mathbf{x}^{(m,l)} + \tilde{\mathbf{w}}^{T^l_m}_d, \tag{7.13}$$

where $\tilde{\mathbf{w}}^{T^l_m}_d \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{R}^{-1})$ is the equivalent AWGN, and

$$\mathbf{A}^{(m,l)}_d = diag\left(a^{(m,l)}_{1_1,d}, \cdots, a^{(m,l)}_{Q_m,d}, a^{(m,l-1)}_{1_{m+1},d}, \cdots, a^{(m,l-1)}_{Q_K,d}\right)$$

is a diagonal matrix with the equivalent channel coefficients

$$a^{(m,l)}_{i_n,d} = \sum_{j=1}^{Q} |h_{j_m,d}|\sqrt{\mu_{i_n,j_m}P}I^l_{i_n,j_m} \tag{7.14}$$

from the mth cluster to the destination on the main diagonal. It is easy to see that the signals from different nodes are combined coherently in the air due to transmit beamforming. Besides, as the destination can obtain a set of K replicas for any source symbol $x^l_{j_m}$ in the consecutive time slots $T^l_m, \cdots, T^{l+1}_{m-1}$, it can combine these replicas through EGC given by

$$\tilde{x}^l_{j_m,d} = \sum_{n=m}^{K} \tilde{y}^{T^l_n}_{j_m,d} + \sum_{n=1}^{m-1} \tilde{y}^{T^{l+1}_n}_{j_m,d} = \tilde{h}^l_{j_m,d}x^l_{j_m} + \tilde{w}^l_{j_m,d}, \tag{7.15}$$

where $\tilde{h}^l_{j_m,d} = \sum_{i_n \in {}_1\psi^l_{j_m} \cup \{j_m\}} |h_{i_n,d}| \sqrt{\mu_{j_m,i_n}}$ is the equivalent channel,

$$ {}_1\psi^l_{j_m} = \left\{ i_n : I^l_{j_m,i_n} = 1, n \neq m \right\}$$

is the set of source nodes that can decode $x^l_{j_m}$ correctly, and $\tilde{w}^l_{j_m,d} \sim \mathcal{CN}\left(0, N_0 \rho_N K\right)$ is the equivalent AWGN. As the equivalent channel $\tilde{h}^l_{j_m,d}$ is a function of both the decoding states and real channel coefficients, the conditional SNR is given by

$$\gamma^l_{j_m|\{h\},\left\{I^l_{j_m,i_n}\right\}} = \frac{\gamma}{\rho_N K} \left| \tilde{h}^l_{j_m,d} \right|^2. \tag{7.16}$$

It is worth noting that some soft symbols $\tilde{y}^{T^l_n}_{j_m,d}$ ($\tilde{y}^{T^{l+1}_n}_{j_m,d}$) may be pure noise. This may occur when no nodes in the nth cluster can decode $x^l_{j_m}$ correctly. So it seems better to exclude such soft symbols during EGC to suppress the noise power. However, this requires the destination to know the decoding states at all source nodes, which incurs tremendous amount of feedback overhead. So in this work, we assume the decoding states are only local information, and the destination would combine all the soft symbols no matter they contain the source information or not.

7.3 Performance Analysis

In this section, we study the error performance of the aforementioned transmission strategy. Both the exact SER and approximated SER at high SNRs are obtained.

7.3.1 Exact SER Analysis

Given the channel coefficient, the conditional SER for the M-ary PSK signal is [85]

$$\varphi\left(\gamma_{|\{h\}}\right) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi} \exp\left(-\frac{g_{psk}\gamma_{|\{h\}}}{\sin^2\theta}\right) d\theta, \tag{7.17}$$

where $g_{psk} = \sin^2\left(\frac{\pi}{M}\right)$ is a constant determined by the constellation size M. The unconditional SER can then be obtained after averaging the above expression over channel distribution.

However, the SER analysis for the proposed protocol is complicated as the SNR expression (7.16) also depends on the decoding states at other clusters. To facilitate the analysis, we define the decoding states vector for symbol $x_{j_m}^l$ as $\mathbf{I}_{j_m}^l = \left(\mathbf{I}_{j_m,1}^l, \cdots, \mathbf{I}_{j_m,m-1}^l, \mathbf{I}_{j_m,m+1}^l, \cdots, \mathbf{I}_{j_m,K}^l\right)$, where $\mathbf{I}_{j_m,n}^l = \left(I_{j_m,1_n}^l, I_{j_m,2_n}^l, \cdots, I_{j_m,Q_n}^l\right)$. Note that $\mathbf{I}_{j_m}^l$ is actually a random binary vector of length $Q(K-1)$, so it can also be represented by the corresponding base-ten number for notational convenience, i.e., $\left|\mathbf{I}_{j_m,n}^l\right|_2 = \left[\mathbf{I}_{j_m,1}^l, \cdots, \mathbf{I}_{j_m,m-1}^l, \mathbf{I}_{j_m,m+1}^l, \cdots, \mathbf{I}_{j_m,K}^l\right]_2$. Besides, as the decoding states at different nodes are independent, all the elements of the vector $\mathbf{I}_{j_m}^l$ are independent Bernoulli random variables with PDF

$$\Pr\left(I_{j_m,i_n}^l\right) = P_{j_m,i_n}^{1-I_{j_m,i_n}^l}\left(1-P_{j_m,i_n}\right)^{I_{j_m,i_n}^l}. \tag{7.18}$$

The PDF for the decoding states vector $\mathbf{I}_{j_m}^l$ is then given by

$$\Pr\left(\mathbf{I}_{j_m}^l\right) = \prod_{i=1}^{Q}\prod_{\substack{n=1\\n\neq m}}^{K}\Pr\left(I_{j_m,i_n}^l\right) = \prod_{i_n\in{}_0\psi_{j_m}^l} P_{j_m,i_n} \prod_{i_n\in{}_1\psi_{j_m}^l}\left(1-P_{j_m,i_n}\right), \tag{7.19}$$

where ${}_0\psi_{j_m}^l = \left\{i_n : I_{j_m,i_n}^l = 0, n\neq m\right\}$ is the set of source nodes that fail to decode

$x_{j_m}^l$ correctly, and

$$P_{j_m,i_n} = E_{h_{j_m,i_n}} \varphi \left(\frac{\mu_{j_m,j_m} \gamma}{\rho_N} |h_{j_m,i_n}|^2 \right) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \left(1 + \frac{\mu_{j_m,j_m} \sigma_{j_m,i_n}^2 g_{psk} \gamma}{\rho_N \sin^2 \theta} \right)^{-1} d\theta \tag{7.20}$$

is the decoding error for symbol $x_{j_m}^l$ at the source node i_n.

Now according to the law of total probability, the decoding error for symbol $x_{j_m}^l$ at the destination is

$$P_e \left(x_{j_m}^l \right) = \sum_{\left| \mathbf{I}_{j_m}^l \right|_2 = 0}^{2^{Q(K-1)}-1} P_e \left(x_{j_m}^l \left| \mathbf{I}_{j_m}^l \right. \right) \Pr \left(\mathbf{I}_{j_m}^l \right), \tag{7.21}$$

where $P_e \left(x_{j_m}^l \left| \mathbf{I}_{j_m}^l \right. \right)$ is the conditional SER given the decoding states. If $\left| {}_1\psi_{j_m}^l \right| = 0$, i.e., no source nodes are able to decode symbol $x_{j_m}^l$, then the decoding is based totally on the received signal from the direct link, and it is easy to show that

$$P_e \left(x_{j_m}^l \left| \left| \mathbf{I}_{j_m}^l \right|_2 = 0 \right. \right) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \left(1 + \frac{\mu_{j_m,j_m} \sigma_{j_m,d}^2 g_{psk} \gamma}{\rho_N K \sin^2 \theta} \right)^{-1} d\theta. \tag{7.22}$$

On the other hand, when $\left| {}_1\psi_{j_m}^l \right| > 0$ we actually have to evaluate the SER of PSK signal using EGC with $\left(\left| {}_1\psi_{j_m}^l \right| + 1 \right)$ branches. Unfortunately, no exact closed-form expression has been found in the past decades except for the special case with two branches. Alternatively, we apply the Gauss-Hermite quadrature approximation developed in [97], which is given by

$$P_e \left(x_{j_m}^l \left| \left| \mathbf{I}_{j_m}^l \right| > 0 \right. \right) \approx \frac{1}{2\pi^2} \int_0^{\frac{M-1}{M}\pi} \frac{1}{\sqrt{\eta_{j_m}^l (\theta)}} \sum_{k=1}^{N_p} H_{z_k} F_{j_m}^l \left(\frac{z_k}{\sqrt{\eta_{j_m}^l (\theta)}}, \theta \right) d\theta, \tag{7.23}$$

where the integrand is given by

$$F_{j_m}^l (\nu, \theta) = \mathrm{Re} \left\{ \left(X(\theta) + jY(\nu, \theta) \right) \prod_{i_n \in {}_1\psi_{j_m}^l \cup \{j_m\}} \left(A_{j_m,i_n}(\nu) + jB_{j_m,i_n}(\nu) \right) \right\}. \tag{7.24}$$

Here z_k are the zeros of the N_p-th order Hermite polynomial, and H_{z_k} are the weights tabulated in [87, 25.10]. It has been shown that $N_p = 20$ is enough to accurately characterize the SER greater than 10^{-5}. The other functions in (7.24) are defined respectively as

$$\eta_{j_m}^{l}(\theta) = \frac{\sin^2\theta}{2A_{psk}^2} + \frac{1}{4}\sum_{i_n \in {}_1\psi_{j_m}^{l} \cup \{j_m\}} \mu_{j_m,i_n}\sigma_{i_n,d}^2, \tag{7.25}$$

$$X(\theta) = \sqrt{\frac{\pi}{2}}\frac{\sin\theta}{A_{psk}}, \tag{7.26}$$

$$Y(\nu,\theta) = -\frac{\nu\sin^2\theta}{A_{psk}^2}{}_1F_1\left(\frac{1}{2};\frac{3}{2};\frac{\nu^2\sin^2\theta}{2A_{psk}^2}\right), \tag{7.27}$$

$$A_{j_m,i_n}(\nu) = {}_1F_1\left(-\frac{1}{2};\frac{1}{2};\frac{\nu^2\mu_{j_m,i_n}\sigma_{i_n,d}^2}{4}\right), \tag{7.28}$$

$$B_{j_m,i_n}(\nu) = \Gamma\left(\frac{3}{2}\right)\nu\sqrt{\mu_{j_m,i_n}\sigma_{i_n,d}^2}, \tag{7.29}$$

where ${}_1F_1(\cdot;\cdot;\cdot)$ is the Kummer confluent hypergeometric function, $\Gamma(\cdot)$ is the gamma function, and $A_{psk} = \sqrt{\frac{2\gamma g_{psk}}{K\rho_N}}$ is a constant. Finally, plugging (7.19), (7.22) and (7.23) back into (7.21) leads to the closed-form SER. As will be shown later, the above results match well with the simulations.

7.3.2 Asymptotic SER Analysis

To gain more insights into the benefits of the proposed cooperation strategy, we explore in this subsection the asymptotic SER in the high SNR regions, i.e., when $\gamma \gg 1$. It is easy to check that the source decoding error P_{j_m,i_n} in (7.20) would approach zero, so we can approximate (7.19) as

$$\Pr\left(\mathbf{I}_{j_m}^{l}\right) \approx \prod_{i_n \in {}_0\psi_{j_m}^{l}} P_{j_m,i_n}. \tag{7.30}$$

After plugging the above expression back into (7.21), we can obtain

$$P_e\left(x_{j_m}^l\right) \approx \sum_{\left|\mathbf{I}_{j_m}^l\right|_2=0}^{2^{Q(K-1)}-1} \left\{ E_{\{h\}}\varphi\left(\frac{\gamma}{\rho_N K} \left| \sum_{i_n \in {}_1\psi_{j_m}^l \cup \{j_m\}} |h_{i_n,d}| \sqrt{\mu_{j_m,i_n}} \right|^2 \right) \right.$$
$$\left. \times \prod_{i_n \in {}_0\psi_{j_m}^l} E_{\{h\}}\varphi\left(\frac{\mu_{j_m,j_m}\gamma}{\rho_N} |h_{j_m,i_n}|^2 \right) \right\}. \tag{7.31}$$

A direct observation is that all terms within the summation actually have the similar form of $E_{\tilde{h}}\varphi\left(c\gamma\tilde{h}^2\right)$, where c is a constant and $\tilde{h} = \sum_{i=1}^{L} |h_i|$ is the equivalent channel with $h_i \sim \mathcal{CN}\left(0, \sigma_i^2\right)$ being independent random variables. It has been proved in [98] that at high SNRs, this metric depends only on the behavior of the distribution of $\tilde{h}$ around the origin. Using Taylor series expansion, we can obtain

$$f_{\tilde{h}}\left(x\right) \approx \frac{2^L}{\Gamma\left(2L\right) \prod_{l=1}^{L} \sigma_l^2} x^{2L-1} + o\left(x^{2L}\right). \tag{7.32}$$

With the above result, it is easy to show that

$$E_{\tilde{h}}\varphi\left(c\gamma\tilde{h}^2\right) \approx \frac{\Gamma\left(L\right) 2^{L-1}}{\Gamma\left(2L\right) \left(g_{psk}c\gamma\right)^L \prod_{l=1}^{L} \sigma_l^2} G(M,L), \tag{7.33}$$

where

$$G(M,L) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \sin^{2L}\theta d\theta \tag{7.34}$$

is a constant depending on the constellation size M and the number of combining branches L. Finally, plugging (7.33) back into (7.31) and doing some manipulation, we can obtain the asymptotic SER given by

$$P_e\left(x_{j_m}^l\right) \approx \left(\frac{\rho_N}{g_{psk}\gamma}\right)^{N-Q+1}$$
$$\times \sum_{\left|\mathbf{I}_{j_m}^l\right|_2=0}^{2^{Q(K-1)}-1} \frac{K^{\left|{}_1\psi_{j_m}^l\right|+1} \Gamma\left(\left|{}_1\psi_{j_m}^l\right|+1\right) 2^{\left|{}_1\psi_{j_m}^l\right|} G\left(M, \left|{}_1\psi_{j_m}^l\right|+1\right) \left(G(M,1)\right)^{\left|{}_0\psi_{j_m}^l\right|}}{\Gamma\left(2\left(\left|{}_1\psi_{j_m}^l\right|+1\right)\right) \prod_{i_n \in {}_0\psi_{j_m}^l} \mu_{j_m,j_m}\sigma_{j_m,i_n}^2 \prod_{i_n \in {}_1\psi_{j_m}^l \cup \{j_m\}} \mu_{j_m,i_n}\sigma_{i_n,d}^2}. \tag{7.35}$$

7.3.3 Discussions

According to the definition (1.1), the diversity gain of the proposed method is equal to $N - Q + 1$. If M-ary PSK modulation is used throughout the network, the symbol rate is equal to $R_s = \log_2 M$. As all the clusters transmit in a TDMA way, each node has one transmission chance every $K = N/Q$ time slots. Consequently, the bandwidth efficiency is equal to $R_d = \frac{R_s}{K} = \frac{Q}{N} R_s$. We observe that the diversity gain is an increasing function of the cluster size Q, whereas the bandwidth efficiency is a decreasing function of the cluster size Q. Therefore, there is a tradeoff between diversity gain and bandwidth efficiency, and different tradeoffs can be achieved by changing the formation of clusters. In the special case where there is a single node in each cluster (i.e., $Q = 1$ and $K = N$), it achieves the highest diversity gain of N at a large loss of bandwidth efficiency. However, as will be seen in simulations, high diversity gain does not always promise better throughput, especially when the required bandwidth efficiency is high and we are forced to use higher-order modulations to compensate for the rate loss. In that case, a better solution is to sacrifice some diversity gain by forming larger clusters and as a result, much smaller constellations could be used to achieve the same bandwidth efficiency. From (7.35), it is also observed that the diversity gain is independent of the cross correlation factor ρ. However, as ρ_N is an increasing function of ρ, there would be some loss in coding gain if the orthogonality of the signature waveforms degrades.

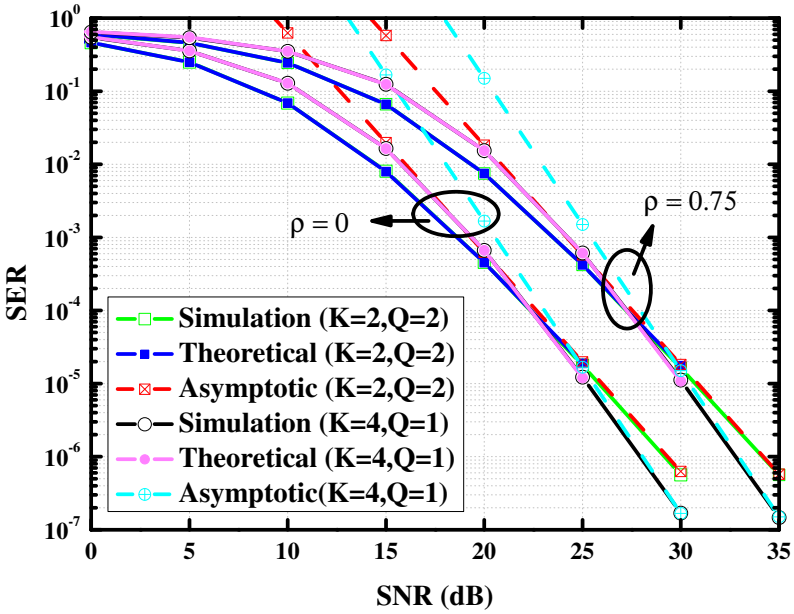


Figure 7.2: SER performances with QPSK modulations.

7.4 Simulations

In this section, we present some simulation results to study the performances of the proposed scheme. All channels are independent, and the channel gain is modeled as $\sigma^2 = d^{-3}$, where d is the distance between the associated two nodes. In all cases, the transmitted power at each node is equally allocated to transmit the local symbol and relayed symbols.

The SER performances with and without clustering are given in Figure 7.2. The four nodes are symmetrically located on a unit circle and the destination is at the center. For the case $K = 2$, the two nodes on the same diameter are clustered together, and QPSK signals are used by all nodes. Clearly, simulation results match well with the theoretical analysis (7.21) when SER is greater than 10^{-5}. When SER is low, the Gauss-Hermit quadrature is not a good approximation, but the

asymptotic analysis (7.35) is tight since SNR is high in such cases. It is also observed that the curve corresponding to $K = 4$ clusters has steeper slope than that having $K = 2$ clusters in the high SNR regions. This is consistent with our analysis, as the former has a diversity gain of 4 while the latter only has a diversity gain of 3. Somewhat surprising, at low-to-modest SNRs node clustering shows better SER performance even though its diversity gain is lower. This is because in the case of 2 clusters, each node only needs to transmit 1 local symbol and 2 relayed symbols and the total power is equally split into 3 portions, whereas in the other case with 4 clusters, each node has to transmit 1 local symbol and 3 relayed symbols and the total power is equally split into 4 portions. When the channels are not in good conditions, local symbol generally requires more transmitted power to guarantee successful decoding at the neighboring clusters and trigger inter-cluster cooperation. That is why having clusters shows better SER performance in the low-to-modest SNR regions. Lastly, it is also observed that when the cross correlation factor ρ becomes larger, the proposed scheme would suffer some loss in coding gain, though the diversity order is still the same. This is because the noise enhancement factor ρ_N in (7.11) is actually an increasing function of ρ. So the equivalent noise power at the decorrelator output will increase accordingly.

To gain more insights into the advantage of the proposed protocol, we compare with the conventional TDMA and STNC protocol [82] after carefully normalizing the data rate and power. The four source nodes are now randomly generated on a square and the destination is always at the center. We show the results in Figure 7.3 and Figure 7.4 where the square size is 2x2 and 4x4, respectively. In the case

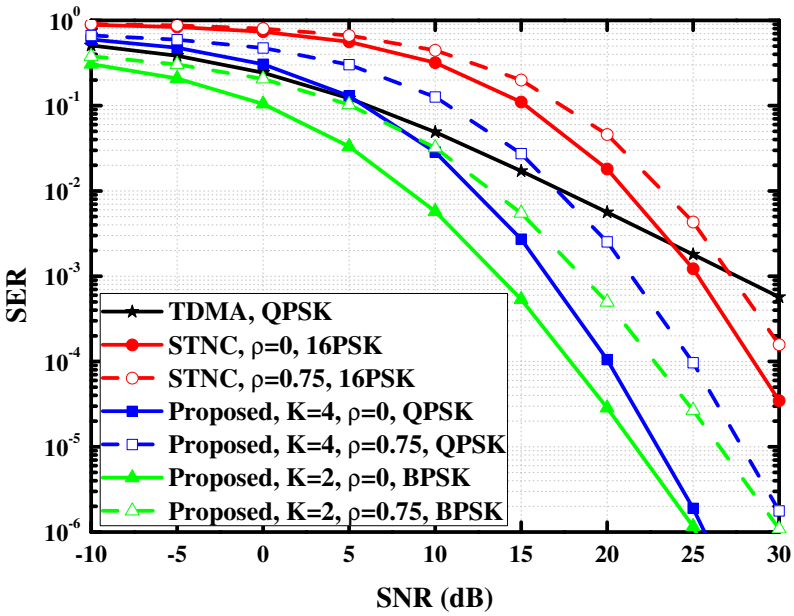


Figure 7.3: SER comparison in a 2x2 network.

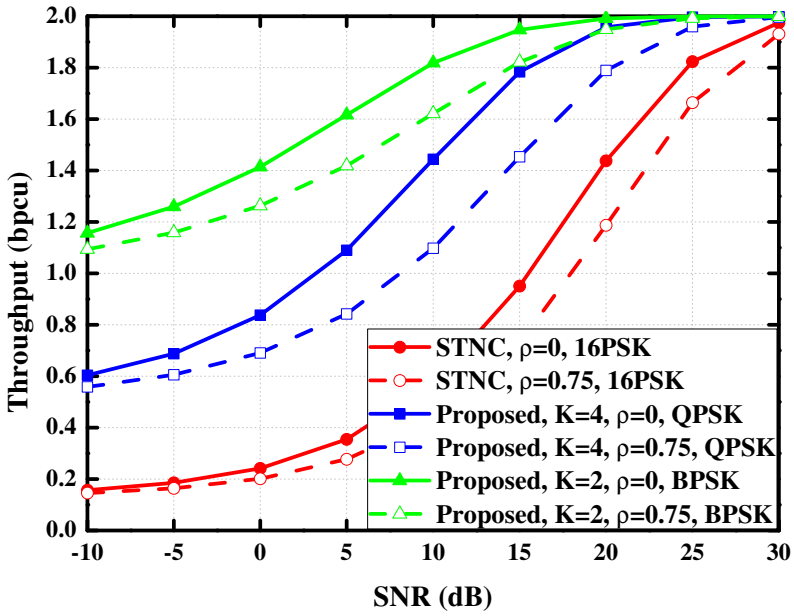


Figure 7.4: Throughput comparison in a 4x4 network.

of clustering (i.e., $K = 2$), the two source nodes having the largest distance are always grouped together to improve inter-cluster connectivity. It is observed that the proposed scheme has a huge performance gain. Compared with STNC, which also achieves a diversity gain of 4 as the proposed scheme with $K = 4$ clusters, the coding gain is about 8dB in the high SNR regions in Figure 7.3. This is because the STNC scheme is not bandwidth efficient due to the separation of data sharing phase and data relaying phase, whereas in our scheme those two phases have been combined through smart design of the space-time network coding. It is also observed that both STNC and the proposed scheme with $K = 4$ perform worse than TDMA in the low SNR regions in Figure 7.3. This is because for the nodes located far away from each other, the cooperation is not effective due to higher decoding error. So part of the transmitted power reserved to relay symbols is actually wasted in most cases. Another interesting observation is that the proposed scheme with $K = 2$ actually performs best in all cases, though the diversity gain is only 3. The reason is that the nodes having large distance are always grouped in the same cluster, so the inter-cluster communications are really reliable. Figure 7.4 shows the throughput in a 4x4 network, which is defined as the number of bits per channel use (bpcu) that can be successfully delivered to the destination. Simulation results confirm again the huge throughput gain of our scheme due to more efficient use of channel resources. Comparatively, the throughput gain is more eminent in the low-to-modest SNR regions, in which case the user cooperation is not fully effective and to improve the bandwidth efficiency is more important to achieve a better overall performance.

7.5 Conclusions

We proposed a novel clustering based space-time network coding protocol to achieve cooperative diversity gain for wireless uplink. Both the exact and asymptotic SER expressions were derived and it was shown that there is a basic tradeoff between diversity gain and bandwidth efficiency. Depending on the channel conditions, sacrificing some diversity gain could result in a large improvement on the bandwidth efficiency and thus lead to much better performance. Future work may concern the asynchronization problem within the cluster. One may also develop the clustering algorithm and address the rate and power allocation issues.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

In this thesis, we have studied the cooperative communication systems with wireless network coding. For uncoded systems, we explored the diversity gain and drew the important conclusion that network-coded cooperation cannot achieve the same diversity performance as the conventional diversity schemes like multiple-antenna diversity. The diversity loss is caused by error propagation, co-channel interference or the use of non-coherent transmissions, which are the unique features of wireless communication and have been neglected more or less in the existing study of wireline network coding. However, the diversity loss is not severe and most of the time only occurs at modest SNRs, and full dominant diversity gain is still achievable at extremely high SNRs. In sum, network-coded cooperation is still a good substitute for the conventional diversity schemes, especially in a dense network where there are abundant dedicated relays.

For the networks without dedicated relays, user devices have to help each other for data relaying. But due to half-duplex constraint, user devices cannot transmit and receive at the same time. As a result, there is conflict between diversity gain and bandwidth efficiency. To be specific, having more cooperating users relay data would lead to higher diversity gain, but the chance to send the own data also got

reduced. Our study showed that highest diversity gain does not always lead to best performance because of the tremendous loss in bandwidth efficiency; instead, diversity gain and bandwidth efficiency should be compromised for different system settings. The transmission strategy we developed is able to achieve good tradeoff by simply changing the formation of clusters, and it is very suitable for the CDMA cellular uplink.

For a large wireless network, the relaying channels are precious resources and may not meet the needs of all users. Our study demonstrated that for coded systems, network dynamics could be exploited to use cooperative diversity more flexibly. When the channels are in good conditions, most of the time wireless relaying is not necessary at all, and it could be a backup transmission strategy just in case the direct link is in outage. To further enhance the relay sharing efficiency, we developed a network-coded ARQ strategy in which the relays only retransmit the mixed source messages when necessary. When there is error detection mechanism, DNC is quite robust to error propagation and it is thus widely accepted as a better scheme than ANC. However, our study revealed that using DNC alone is not able to fully leverage the network coding gain when there are multiple relays. This is because each relay may only decode a small portion of the source messages, so the chance to mix different source messages locally through DNC is pretty low. In this case, ANC is a good complement to DNC because it allows mixing messages in the air directly. Therefore, a hybrid network coding scheme is expected to achieve better performance for wireless applications.

8.2 Future Work

8.2.1 User Scheduling

In this work, we have shown that network coding is not always superior to the conventional orthogonal relaying. This is especially true for ANC, where different user signals may become co-channel interference to each other because of random channel distortion. Consequently, network coding gain depends largely on the channel conditions of different users. Intuitively, proper user scheduling may help to solve this problem by pairing users together only when they are a good match. Besides user pairing, how to pair the users is yet another design issue. In the current 4G cellular systems and the Wifi networks, orthogonal frequency-division multiplexing (OFDM) technique is widely used. One distinct feature of OFDM is that the whole bandwidth is divided into small sub-carriers, and each sub-carrier is regarded as a separate channel. So the user pairing patterns could be quite different across the sub-carriers. In classical OFDM systems where each sub-carrier accommodates only one user, the total transmitted power could be allocated in a water-filling way across all sub-carriers to maximize the capacity [2]. When network coding is used, each sub-carrier can potentially serve multiple users, how to smartly allocate the transmitted power remains a pretty interesting topic. In sum, the user scheduling algorithm needs to address when to pair, who to pair, and how to pair.

8.2.2 Energy Saving

So far, we focus only on the scenarios where the devices have limited power but unlimited energy, and we have studied the corresponding diversity performance. In practice, the dual problem is equally important. That is, given that a certain quality-of-service (QoS) goal has to be achieved, how much energy could be saved by applying network coding. In that case, power limit is no longer a big concern, and the top priority should be placed on QoS requirements like data rate target. Network coding reduces the channel use for data relying, so the relays could potentially use less power to achieve the same QoS goal. At the receiver side, base-band processing also consumes energy. As each network-coded message inherently carries information for multiple source messages, the receiver could decode much fewer packets and thus save energy too. From a network view, the overall benefit is much more substantial. This is especially true for sensor networks, where the sensor nodes have very limited energy and the whole network would break down when certain key sensor nodes run out of battery. Consequently, how to use network coding to save energy and extend the network lifetime is a very important practical issue and deserves more exploration.

Bibliography

[1] J. Proakis, *Digital Communications*, 4th ed., New York: McGraw-Hill, 2001.

[2] D. N. C. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, 2005.

[3] K. J. Ray Liu, A. K. Sadek, W. Su, and A. Kwasinski, *Cooperative Communications and Networking*, Cambridge University Press, 2008.

[4] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecomn.*, vol. 10, pp. 585-595, Nov. 1999.

[5] L. Z. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.

[6] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998.

[7] S. M. Alamouti, "A simple transmit diversity technique for wireless communiucations," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[8] G. J. Foschini and M. Gans, "On the limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311-335, Mar. 1998.

[9] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Info. Theory*, vol. 25, no. 9, pp. 572-584, Sep. 1979.

[10] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927-1938, Nov. 2003.

[11] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939-1948, Nov. 2003.

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

[13] W. F. Su, A. K. Sadek, and K. J. Ray Liu, "Cooperative communication protocols in wireless networks: Performance analysis and optimum power allocation," *Wireless Personal Commun.*, vol. 44, no. 2, pp. 181–217, Jan. 2008.

[14] K. G. Seddik, A. K. Sadek, W. F. Su, and K. J. Ray Liu, "Outage analysis and optimal power allocation for multinode relay networks," *IEEE Signal Process. Lett.*, vol. 14, no. 6, pp. 377-380, Jun. 2007.

[15] A. Ribeiro, X. D. Cai, and G. B. Giannakis, "Symbol error probabilities for general cooperative links," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1264-1273, May 2005.

[16] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. Ray Liu, "Cooperative communications with relay selection: When to cooperate and whom to cooperate with?" *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2814-2827, Jul. 2008.

[17] K. G. Seddik, A. K. Sadek, A. S. Ibrahim, and K. J. Ray Liu, "Design criteria and performance analysis for distributed space-time coding," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2280-2292, Jul. 2008.

[18] X. J. Tang and Y. B. Hua, "Optimal design ofnon-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398-1407, Apr. 2007.

[19] W. Guan and H. W. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517-519, Jul. 2008.

[20] W. Guan, H. W. Luo, and W. Chen, "Linear relaying scheme for MIMO relay system with QoS requirements," *IEEE Signal Process. Lett.*, vol. 15, pp. 697-700, 2008.

[21] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379-389, Feb. 2007.

[22] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204-1216, Jul. 2000.

[23] T. Ho and D. S. Lun, "Network coding: An introduction," Cambridge University Press, 2008.

[24] S. Y. R. Li, R. W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371-381, Feb. 2003.

[25] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proc. ACM Conference of the Special Interest Group on Data Communication*, Aug. 2007.

[26] M. C. Valenti, D. Torrieri, and T. Ferrett, "Noncoherent physical-layer network coding with FSK modulation: Relay receiver design issues," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2595-2604, Sep. 2011.

[27] S. Zhang, S. C. Liew, and P. P. Lam, "Physical-layer network coding," in *Proc. ACM International Conference on Mobile Computing and Networking*, Sep. 2006.

[28] R. H. Y. Louie, Y. H. Li, and B. Vucetic, "Practical physical layer network coding for two-way relay channels: Performance analysis and comparison," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 764-777, Feb. 2010.

[29] T. K. Akino, P. Popovski and V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *IEEE J. Sel. Areas Commun.*, Vol. 27, No. 5, pp. 773-787 , Jun. 2009.

[30] P. Larsson, N. Johansson, and K.-E. Sunell, "Coded bi-directional relaying," in *Proc. IEEE Vehicular Technology Conference*, vol. 2, pp. 851-855, May 2006.

[31] P. Popovski and H. Yomo,"The anti-packets can increase the achievable throughput of a wireless multi-hop network," in *Proc. IEEE International Conference on Communication*, pp. 3885-3890, Jun. 2006.

[32] P. Popovski and H. Yomo, "Physical network coding in two-way wireless relay channels", in *Proc. IEEE International Conference on Communication*, pp. 707-712, Jun. 2007.

[33] E. C. Y. Peh, Y.-C. Liang, and Y. L. Guan, "Power control for physical-layer network coding in fading environments," in *Proc. IEEE Personal, Indoor and Mobile Radio Communication*, pp. 1-5, Sep. 2008.

[34] T. Cui, F. F. Gao, and C. Tellambura, "Differential modulation for two-way wireless communications: A perspective of differential network coding at the physical layer," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 2977-2987, Oct. 2009.

[35] W. Guan and K. J. Ray Liu, "Performance analysis of two-way relaying with non-coherent differential modulation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 2004-2014, Jun. 2011.

[36] W. Guan and K. J. Ray Liu, "Two-way denoise-and-forward relaying with non-coherent differential modulation," in *Proc. IEEE Global Telecommunications Conference*, Dec. 2011.

[37] T. Himsoon, W. P. Siriwongpairat, W. F. Su, and K. J. Ray Liu, "Differential modulation with threshold-based decision combining for cooperative communications," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3905-3923, Jul. 2007.

[38] F. A. Onat, A. Adinoyi, Y. Fan, H. Yanikomeroglu, J. Thompson, and I. Marsland, "Threshold selection for SNR-based selective digital relaying in cooperative wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4226-4237, Nov. 2008.

[39] F. A. Onat, Y. J. Fan, H. Yanikomeroglu, and J. S. Thompson, "Asymptotic BER analysis of threshold digital relaying schemes in cooperative wireless systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4938-4947, Dec. 2008.

[40] T. Wang, G. B. Giannakis, and R. Q. Wang, "Smart regenerative relays for link-adaptive cooperative communications," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1950-1960, Nov. 2008.

[41] T. Wang, A. Cano, G. B. Giannakis, and J. N. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," *IEEE Trans. Commun.*, vol. 55, no. 7, pp. 1427-1438, Jul. 2007.

[42] M. D. Selvaraj, R. K. Mallik, and R, Goel, "Optimum receiver performance with binary phase-shift keying for decode-and-forward relaying," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1948-1953, May 2011.

[43] Y. Chen, S. Kishore, and J. Li, "Wireless diversity through network coding," in *Proc. IEEE Wireless Communications and Networking Conference*, vol. 3, pp. 1681-1686, Apr. 2006.

[44] Z. Han, X. Zhang, and H. V. Poor, "Cooperative transmission protocols with high spectral efficiency and high diversity order using multiuser detection and network coding," *IEEE Trans. Wireless Commun.*, vol. 8, No. 5, pp. 2352-2361, May 2009.

[45] M. Xiao and M. Skoglund, "Multiple-user cooperative communications based on linear network coding," *IEEE Trans. Commun.*, vol. 58, No. 12, pp. 3345-3351, Dec. 2010.

[46] G. Al-Habian, A. Ghrayeb, M. Hasna, and A. Abu-Dayya, "Threshold-based relaying in coded cooperative networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 123-135, Jan. 2011.

[47] A. Nasri, R. Schober, and M. Uysal, "Error rate performance of network-coded cooperative diversity systems," in *Proc. IEEE Global Telecommunications Conference*, pp. 1-6, Dec. 2010.

[48] M. Iezzi, M. D. Renzo, and F. Graziosi, "Closed-form error probability of network-coded cooperative wireless networks with channel-aware detectors," in *Proc. IEEE Global Telecommunications Conference*, Dec. 2011.

[49] W. Guan and K. J. Ray Liu, "Mitigating error propagation for wireless network coding," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3632-3643, Oct. 2012.

[50] W. Guan and K. J. Ray Liu, "Anti error propagation methods for wireless uplink using network coding," in *Proc. IEEE Global Telecommunications Conference*, Dec. 2012.

[51] D. Q. Chen, K. Azarian, and J. N. Laneman, "A case for amplify-forward relaying in the block-fading multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3728-3733, Aug. 2008.

[52] M. Badr and J. C. Belfiore, "Distributed space time codes for the amplify-and-forward multiple-access relay channel," in *Proc. IEEE International Symposium on Information Theory*, pp. 2543-2547, Jul. 2008.

[53] S. A. Jafar, K. S. Gomadam, and C. C. Huang, "Duality and rate optimization for multiple access and broadcast channels with amplify-and-forward relays," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3350-3370, Oct. 2007.

[54] Z. G. Ding, T. Ratnarajah, and K. K. Leung, "On the study of network coded AF transmission protocol for wireless multiple access channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 118–123, Jan. 2009.

[55] K. S. Gomadam and S. A. Jafar, "The effect of noise correlation in amplify-and-forward relay networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 731-745, Feb. 2009.

[56] S. Berger and A. Wittneben, "Cooperative distributed multiuser MMSE relaying in wireless ad-hoc networks," in *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 1072-1076, Oct. 2005.

[57] E. Koyuncu and H. Jafarkhani, "Distributed beamforming in wireless multiuser relay-interference networks with quantized feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4538-4576, Jul. 2012.

[58] W. Guan and K. J. Ray Liu, "Diversity analysis of analog network coding with multi-user interferences," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 668-679, Feb. 2013.

[59] W. Guan and K. J. Ray Liu, "Error performances of multiple access system using analog network coding," in *Proc. IEEE International Conference on Communication*, Jun. 2012.

[60] D. Chen and J. N. Laneman, "Cooperative diversity for wireless fading channels without channel state information," in *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 1307-1312, Nov. 2004.

[61] R. Annavajjala, P. C. Cosman, and L. B. Milstein, "On the performance of optimum noncoherent amplify-and-forward reception for cooperative diversity," in *Proc. IEEE Military Communications Conference*, pp. 3280-3288, Oct. 2005.

[62] Y. Zhu, P. Y. Kam, and Y. Xin, "Non-coherent detection for amplify-and-forward relay systems in a Rayleigh fading environment," in *Proc. IEEE Global Telecommunications Conference*, pp. 1658-1662, Nov. 2007.

[63] G. Farhadi and N. C. Beaulieu, "A low complexity receiver for noncoherent amplify-and-forward cooperative systems," *IEEE Trans. Wireless Commun.*, vol. 58, no. 9, pp. 2499-2504, Sep. 2010.

[64] M. R. Souryal, "Non-coherent amplify-and-forward generalized likelihood ratio test receiver," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2320-2327, Jul. 2010.

[65] J. H. Yuan, Y. H. Li, and L. Chu, "Differential modulation and relay selection with detect-and-forward cooperative relaying," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 261-268, Jan. 2010.

[66] L. Y. Song, G. Hong, B. L. Jiao, and M. Debbah, "Joint relay selection and analog network coding using differential modulation in two-way relay channels," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2932-2939, Jul. 2010.

[67] I. Krikidis, Z. G. Ding, and C. D. Charalambous, "Noncoherent energy detection with orthogonal signaling for an uncoded two-way relay channel," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 404-409, Jan. 2012.

[68] J. Tian, Q. Zhang, and F. Q. Yu, "Non-coherent detection for two-way AF cooperative communications in fast Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2753-2762, Oct. 2011.

[69] W. Guan and K. J. Ray Liu, "Diversity analysis of non-coherent wireless network coding," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1800-1811, Apr. 2013.

[70] W. Guan and K. J. Ray Liu, "On analysis of wireless uplink using analog network coding with non-coherent modulations," in *Proc. IEEE Global Telecommunications Conference*, Dec. 2012.

[71] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, "Wireless broadcasting using network coding," *IEEE Trans. Veh. Technol.*, Vol. 58, pp. 914-925, Feb. 2009.

[72] P. Fan, C. Zhi, C. Wei, and K. B. Letaief, "Reliable relay assisted wireless multicast using network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 749-762, Jun. 2009.

[73] P. Larsson., B. Smida, T. Koike-Akino, and V. Tarokh, "Analysis of network coded HARQ for multiple unicast flows," in *Proc. IEEE International Conference on Communication*, May 2010.

[74] P. Popovski and H. Yomo, "Wireless network coding by amplify-and-forward for bi-directional traffic flows", *IEEE Commun. Lett.*, Vol. 11, No. 1, pp. 16-19, Jan. 2007.

[75] Q. T. Vien, L. N. Tran, and H. X. Nguyen, "Efficient ARQ retransmission schemes for two-way relay networks," *Journal of Communication Software and Systems*, vol. 7, no.1, pp. 9-15, Mar. 2011.

[76] Z. Y. Chen, Q. S. Gong, C. Zhang, and G. Wei, "ARQ protocols for two-way wireless relay systems: Design and performance analysis," *International Journal of Distributed Sensor Networks*, vol. 2012, 2012.

[77] F. Iannello and O. Simeone, "Throughput analysis of type-I HARQ strategies in two-way relay channels," in *Proc. IEEE Conference on Information Sciences and Systems*, pp. 539-544, Mar. 2009.

[78] W. Guan and K. J. Ray Liu, "Network-coded ARQ for two-way relay channels," *IEEE Trans. Commun.*, to be submitted.

[79] W. Guan and K. J. Ray Liu, "On analysis of two-way relaying with network-coded ARQ," to appear in *Proc. IEEE Global Telecommunications Conference*, Dec. 2013.

[80] A. K. Sadek, W. Su, and K. J. Ray Liu, "Multinode cooperative communications in wireless networks," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 341-355, Jan. 2007.

[81] H. Q. Lai, A. S. Ibrahim, and K. J. Ray Liu, "Wireless network cocast: Location-aware cooperative communications using network coding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3844-3854, Jul. 2009.

[82] H. Q. Lai and K. J. Ray Liu, "Space-time network coding," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1706-1718, Apr. 2011.

[83] W. Guan and K. J. Ray Liu, "Clustering based space-time network coding," in *Proc. IEEE Global Telecommunications Conference*, Dec. 2012.

[84] D. Chen and J. N. Laneman, "Modulation and demodulation for cooperative diversity in wireless systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1785-1794, Jul. 2006.

[85] M. K. Simon and M. S. Alouini, "A unified approach to the performance analysis of digital communications over generalized fading channels," *Proc. IEEE*, vol. 86, no. 9, pp. 1860-1877, Sep. 1998.

[86] K. H. Biyari and W. C. Lindsey, "Statistical distributions of hermitian quadratic forms in complex gaussian variables," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 1076-1082, May 1993.

[87] M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover, 1972.

[88] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York: Academic, 2007.

[89] M. O. Hasna and M. S. Alouini, "A performance study of dual-hop transmissions with fixed gain relays," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1963-1968, Nov. 2004.

[90] Y. D. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3524-3536, Dec. 2006.

[91] J. Salo, H. M. El-Sallabi, and P. Vainikainen, "Impact of double-rayleigh fading on system performance," in *Proc. International Symposium on Wireless Pervasive Computing*, Jan. 2006.

[92] M. Kobayashi and X. Mestre, "Impact of CSI on distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2580-2591, May 2009.

[93] X. Giraud, E. Boutillon, and J. C. Belfiore, "Algebraic tools to build modulation schemes for fading channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 938-952, May 1997.

[94] W. F. Su, Z. Safar, and K. J. Ray Liu, "Full-rate full-diversity space-frequency codes with optimum coding advantage," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 229-249, Jan. 2005.

[95] G. K. Karagiannidis and A. S. Lioumpas, "An improved approximation for the Gaussian Q-function," *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 644-646, Aug. 2007.

[96] S. Verdu, *Multiuser Detection*. Cambridge University Press, 1998.

[97] M. S. Alouini and M. K. Simon, "Performance analysis of coherent equal gain combiing over Nakami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 1449-1462, Dec. 2001.

[98] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389-1398, Aug. 2003.