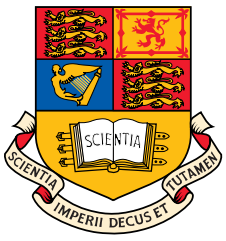
COOPERATIVE STRATEGIES DESIGN BASED ON THE DIVERSITY AND MULTIPLEXING TRADEOFF

HAISHI NING

SUPERVISORS
DR. CONG LING AND PROF. KIN K. LEUNG

A Thesis submitted in fulfilment of requirements for the degree
of Doctor of Philosophy of Imperial College London



Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London
January 2012

To my parents

Abstract

This thesis focuses on designing wireless cooperative communication strategies that are either optimal or near-optimal in terms of the tradeoff between diversity and multiplexing gains. Starting from classical cooperative broadcast, multiple-access and relay channels with unit degree of freedom, to more general cooperative interference channels with higher degrees of freedom, properties of different network topologies are studied and their unique characteristics together with several advanced interference management techniques are exploited to design cooperative transmission strategies in order to enhance data rate, reliability or both at the same time. Moreover, various algorithms are proposed to solve practical implementation issues and performance is analyzed through both theoretical verifications and simulations.

Acknowledgment

I want to thank my supervisors Dr. Cong Ling and Prof. Kin K. Leung for their generous help and guidance. I still remember the difficult time in my first year when I was lost and anxious about the research direction, Cong never stopped encouraging me and approving my work, even many of them were only naive and vague ideas. I wouldn't have realized the importance of the diversity and multiplexing tradeoff in cooperative strategy design without Cong's suggestions and discussions. Kin's explanation of research as "re" and "search" still vividly resides in my mind and is now my first step to tackle any technical problem. I can't give enough thanks to them for allowing me to participate their research projects, from which I have learned broader knowledge and expanded my network. When I decide to further my career in the financial industry, Cong was totally thoughtful and supportive. Not to mention their financial support, without that I simply could neither start nor finish my PhD.

Although Cong is an expert in lattice, information and coding theory and Kin is an outstanding figure in networking and computer science, I didn't become an expert in any of their fields. However, through their supervision, I have been able to use information theoretical approaches to seek insights and solve practical networking problems. Without their help, the ideas in this thesis may never come out.

Finally, I would like thank my parents and my family, for their continuous support about everything I do. They make me who I am now, they make me enjoy my life and my work and they make me happy.

Publications

1. H. Ning, C. Ling, and K. Leung, ”Near-optimal relaying strategy for cooperative broadcast channels,” in *IEEE International Conference on Communications, 2009*, Jun. 2009.

2. H. Ning, C. Ling, and K. Leung, ”Active physical-layer network coding for cooperative two-way relay channels,” in *6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2009*, Jun. 2009.

3. H. Ning, C. Ling, and K. Leung, ”Relay-aided interference alignment: Feasibility conditions and algorithm,” in *IEEE International Symposium on Information Theory, 2010*, Jun. 2010.

4. H. Ning, C. Ling, and K. Leung, ”Interference alignment with diversity,” in *Sensor Signal Processing for Defence, 2010*, Sep. 2010.

5. H. Ning, C. Ling, and K. Leung, ”Generalized sequential slotted amplify and forward strategy in cooperative communications,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1968-1974, Apr. 2011.

6. H. Ning, C. Ling, and K. Leung, ”Feasibility condition for interference alignment with diversity,” *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2902-2912, May 2011.

Contents

List of Figures

Abbreviations

AF: Amplify and forward

BER: Bit error rate

BPCU: Bits per channel use

CBC: Cooperative broadcast channels

CBRC: Cooperative broadcast relay channels

CER: Current effective relay

CIC: Cooperative interference channels

CMA: Cooperative multiple access channels

CMAR: Cooperative multiple access relay channels

CMR: Cooperative multiple relay channels

CSIR: Channel side information at receiver

DF: Decode and forward

DDF: Dynamic decode and forward

DMT: Diversity and multiplexing tradeoff

DNC: Digital network coding

DoF: Degree of freedom

FH: Fischer-Huber loading algorithm

GSSAF: Genelized sequential slotted amplify and forward

IC: Interference cancellation

SSAF: Sequential slotted amplify and forward

MAC Multiple access channels

MIMO Multiple-input multiple-output

MISO: Multiple-input single-output

ML: Maximum likelihood

MMSE: Minimum mean squared error

MRC: Maximal ratio combining

NAF: Nonorthogonal amplify and forward

NER: Next effective relay

PAM: Pulse amplitude modulation

PDF: Probability density function

QoS: Quality of service

RAXN: Relay-aided X networks

SINR: Signal-to-interference-plus-noise ratio

SIMO: Single-input Multiple-output

SNR: Signal-to-noise ratio

TWXN: Two-way exchange network

WNC: Wireless network coding

XOR: Exclusive-or

GBN: Generalized butterfly network

Mathematical Symbols

ρ :	Signal to noise ratio (SNR)
E :	Average short term energy or expectation
σ^2 :	Noise Variance
b :	Exponential order of $f(\rho)$
$\doteq$:	Order equal to
$\dot{\leqslant}$:	Order less or equal to
$\dot{\geqslant}$:	Order larger or equal to
$\{C\}$:	A family of codes
R :	Data rate or relay node
P_E :	Maximum likelihood error probability
r :	Multiplexing gain
d :	Diversity gain or destination node
L :	Codeword length
M :	Number of transmit antennas
N :	Number of receive antennas
K :	Number of user-pairs in CIC or number of relays
d^* :	Optimal DMT curve
$d^*_{\max}$:	Maximum diversity gain on the optimal DMT curve
$r^*_{\max}$:	Maximum multiplexing gain on the optimal DMT curve
s :	Source node or transmitted symbol
t :	Destination node, used where to avoid confusion with diversity gain d

C : Capacity

$[\cdot]^+$: The larger of $[\cdot]$ and 0

x : Transmit signal

y : Receive signal

I : Time slot

n : Additional white Gaussian noise

$\mathcal{Q}$: Node set

$\mathbf{1}_1(\cdot)$: Dirac delta function

$\tilde{N}$: Number of equivalent relays

α, β, γ and λ : Normalization factor

v, u and w: Exponential order of variables

P_O : Outage probability

$\mathbb{R}$: Field of real numbers

$\mathcal{O}$: Events set

$\mathcal{O}^c$: Complementary events set

$\log$: Logarithm

$\det$: Determinant

$\mathbf{v}$: Precoding filter

$\mathbf{u}$: Receiving filter

$\mathbf{H}$: Channel matrix

$\mathbf{Q}$: Covariance matrix

I : Total leakage interference power

$\nu_{\min}$: Eigenvector corresponding to the smallest eigenvalue

Var : Variance

p : Probability density function

$\mathrm{Re}\{\cdot\}$: The real part of

S : Source node

D : Destination node

θ and ϕ : Phases used in modulation and demodulation

$|\cdot|$: Number of elements in a set

$\setminus$: Exclude a element from a set

$\bigcup$: Union of sets

$\bigcap$: Intersection of sets

Chapter 1

Introduction

1.1 Overview

In the past decade, the successful deployment of many powerful point-to-point wireless communication technologies has significantly changed our life. As a result, wireless communications nowadays are network based with multiple terminals almost everywhere. Thus, network-level transmission strategy design becomes inevitable and plays an important role to meet consumer expectations on continuously growing faster and more reliable communications.

Cooperative communication, one of the network communication technologies, has become more and more important in modern wireless networks. Cooperation technologies enable physically disconnected users to transmit, relay and receive signal in a coordinated way such that desired performance targets can be met economically. Typical applications include but not limited to increasing data rate and reducing error probability, etc.

Since cooperative communication is inevitably network based with multiple terminals, cooperative strategy is needed to decide which terminal at what time or frequency to transmit, relay or receive what signal. Due to the wireless broadcast nature, cooperative strategy design plays an essential role in managing undesired interference and maximizing desired communication robustness and efficiency.

1.2 Cooperative communications: promises and challenges

The pioneer work by Sendonaris *et al.* [1, 2] shows that user cooperation can be beneficial in terms of higher data rate and lower outage probability for a cellular system. These two improvements imply reduced power for a mobile user in order to achieve a certain data rate with a certain outage probability requirement. Moreover, the reduced power requirement may also be translated into increased battery life or coverage area for a communication terminal. The benefits of user cooperation come from the fact that the user with the better channel can help other users achieve some acceptable level of performance while sacrificing only a small fraction of its own data rate.

However, how to enable the distributed terminals to cooperate is not a simple task. Due to the wireless broadcast nature, all terminals in a network share the single wireless medium. This means when one terminal transmits, all others are possible to receive the signal at the same time. Moreover, due to different cooperation processes, there is a tradeoff between the gain one can achieve for the data rate and the gain one can achieve for the outage probability. The communication terminals used in [1, 2] are assumed to be full-duplex which means they can transmit and receive signal at the same time. In practice, most communication terminals are half-duplex because of the significant difference between the transmit power and the receive power. This constraint incurs a further tradeoff: in order to help other users achieve better communication performance, one has to be silent for some time in order to receive other's signal, while the silent period can be potentially used to transmit its own signal. Thus, if not properly designed, cooperation may even decrease the system performance due to detrimental side effects such as undesired interference, poor relaying channels, long silent period, time-division transmission requirement and overwhelming overhead, etc.

1.3 Diversity and multiplexing tradeoff: a fundamental performance measurement

Because there is always a tradeoff between a transmission strategy's robustness and efficiency, a more fundamental performance measurement of a communication systen is their tradeoff rather than any of them alone because of its inherent fair comparison property. This tradeoff is often referred to as the diversity and multiplexing tradeoff (DMT) in communication and information theory. In this thesis, we use this measurement to analyze and design strategies for different cooperative networks wherever possible. Since one can always design a coding scheme to achieve a particular point on a strategy's tradeoff curve, such a designing metric makes our work fundamentally important to cooperative strategy deign no matter what future requirements are in practical applications.

Let the transmit signal-to-noise ratio (SNR) of a physical link be defined as $\rho = \frac{E}{\sigma^2}$, where E is the average short term energy at the transmitter and σ^2 is the noise variance at the receiver. We say b is the exponential order of $f(\rho)$ if $\lim_{\rho\to\infty} \frac{\log(f(\rho))}{\log(\rho)} = b$ and denote $f(\rho)$ as $f(\rho)\dot{=}\rho^b$. $\dot{\leqslant}$ and $\dot{\geqslant}$ are similarly defined. Consider a coding scheme as a family of codes $\{C(\rho)\}$ with data rate $R(\rho)$ bits per channel use (BPCU) and average maximum-likelihood (ML) error probability $P_E(\rho)$. The multiplexing gain[1] r and the diversity gain[2] d are defined as

$$r = \lim_{\rho\to\infty} \frac{R(\rho)}{\log(\rho)}, \tag{1.1}$$

and

$$d = -\lim_{\rho\to\infty} \frac{\log(P_E(\rho))}{\log(\rho)}. \tag{1.2}$$

The diversity and multiplexing gains are two of the most important asymptotic measures of any communication system, which characterize the system outage probability and data rate respectively. The DMT was initially studied in the context of point-to-point

[1] In this thesis, we use multiplexing gain and degree of freedom (DoF) interchangeably.
[2] Diversity gain is referred to as spatial diversity gain if not specified otherwise.

MIMO communications. It is well known that multiple antennas can be used to increase either the diversity gain or the multiplexing gain. In [3], it is shown that for a MIMO system, both types of gains can be obtained simultaneously, but there is a fundamental tradeoff between them. Thus, the DMT is a more fundamental measure of the system performance than either of the diversity gain or the multiplexing gain alone. A well known analogy is that a codeword with longer redundancy may have more powerful error correcting ability. However, the closeness to the Shannon limit is a more fundamental measure than either the rate or the error correcting ability alone.

MIMO communication [4,5] is highly related to our research because the DMT of a fully connected MIMO or multiple-input single-output (MISO) system naturally serves as an upper bound for the DMT of a corresponding distributed cooperative system. Thus, for a cooperative network with multiple source-destination pairs, we are interested in finding a transmission strategy which can approach the DMT of a corresponding MIMO or MISO system wherever possible. Even in cases where the DMT can not be explicitly derived, we still aim to use the DMT principle to design cooperative strategies in order to trade one measure for the improvement of the other measure.

In [3], it has been shown that if $L \geqslant M + N - 1$, where L is the codeword length and M and N are the number of transmit and receive antennas respectively, then the optimal tradeoff curve d^* is given by the piecewise-linear function connecting the points $(k, d^*(k))$, $k = 0, 1, ..., \min\{M, N\}$, where

$$d^*(k) = (M - k)(N - k). \tag{1.3}$$

In particular, $d^*_{\max} = MN$ and $r^*_{\max} = \min\{M, N\}$. As an illustrative example, the DMT of a 5×5 MIMO channel is shown in Fig. 1.1. As special cases, the DMTs of MISO and single-input multiple-output (SIMO) systems are lines connecting the points $(0, mn)$ and $(1, 0)$. Fig 1.2 shows the DMT of a 5×1 MISO channel.

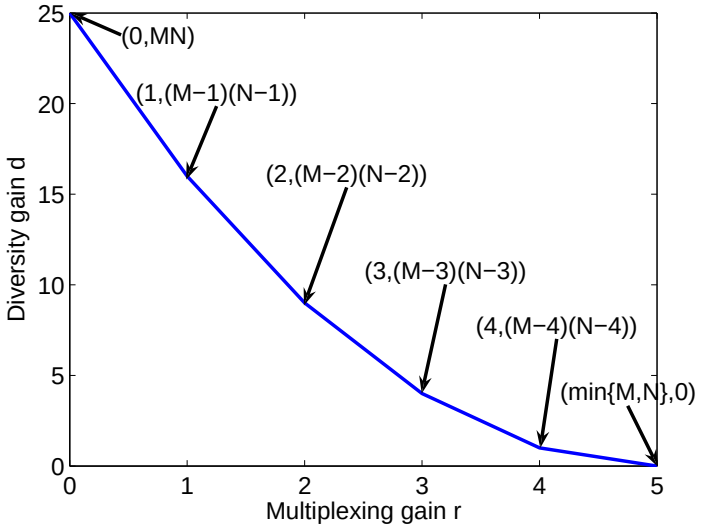


Figure 1.1: Diversity and multiplexing tradeoff of a 5×5 MIMO channel.

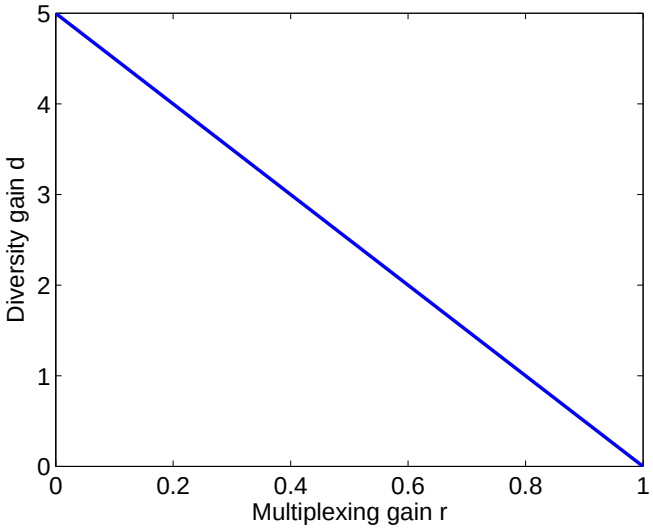


Figure 1.2: Diversity and multiplexing tradeoff of a 5×1 MISO channel.

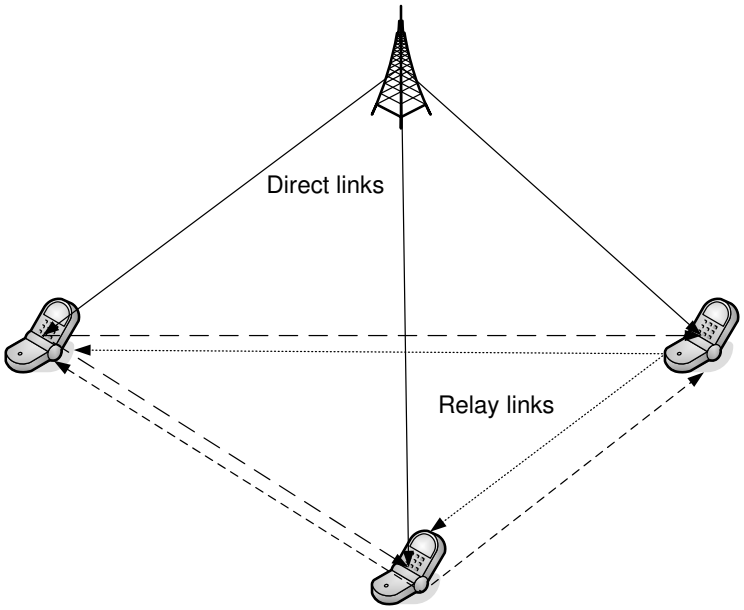


Figure 1.3: Cooperative broadcast channels.

1.4 Cooperative networks with unit degree of freedom

In this section, we introduce several representative cooperative networks with unit DoF and their associated strategies. In particular, we classify primitive cooperative networks with unit DoF into three categories including cooperative broadcast channels (CBC), cooperative multiple access channels (CMA) and cooperative multiple relay channels (CMR). Other unit multiplexing gain cooperative networks can be regarded as combinations of these three primitive topologies.

Popular cooperative strategies include fixed relaying schemes such as decode and forward (DF) and amplify and forward (AF), and selection relaying schemes based on channel estimations [6]. In the class of DF strategies, there are fixed DF [6] scheme and dynamic DF (DDF) scheme [7]. In the class of AF strategies, there are orthogonal AF scheme [6], nonorthogonal AF (NAF) scheme [7] and sequential slotted AF (SSAF) scheme [8, 9].

1.4.1 Cooperative broadcast channels

As illustrated in Fig. 1.3, CBC is important because it is practically similar to a single cell downlink phase, which is the main area most researchers and companies are working to match the ongoing higher data rate and reliability requirements.

The development of cooperative strategies with optimal DMT for CBC was pioneered by Azarian *et al.* [7]. In order to achieve the full multiplexing gain, the authors in [7] employ a modified version of the DDF scheme for CBC, which was referred to as CBC-DDF. In CBC-DDF strategy, any node will start helping others only after it has successfully decoded the entire message. Although full multiplexing gain can be achieved, this strategy fails to exploit the spatial diversity gain in the high multiplexing gain regime because of the high unsuccessful decoding probability for each node in this region.

As the definition of the DMT suggests, ensuring large multiplexing gain or diversity gain alone is often not optimal, while better tradeoff characteristic is more attractive. In this thesis, we propose a generalized sequential slotted amplify and forward (GSSAF) strategy for CBC, which is based on the class of SSAF strategies. Our strategy allows each destination to act in turn as a relay and forward its previously received signal to other destinations sequentially. While GSSAF is not a full multiplexing gain strategy for CBC, the loss is negligible when the number of destinations is large. Moreover, GSSAF allows each destination to be protected by the maximum number of extra paths in order to achieve the near-optimal diversity gain in the high multiplexing gain regime. A DMT lower bound for GSSAF is derived which suggests that our proposed strategy approaches the MISO DMT upper bound and is therefore asymptotically optimal.

1.4.2 Cooperative multiple access channels

As shown in Fig. 1.4, the CMA can be used to increase system capacity and robustness in a single cell uplink phase. Simple relaying strategies like AF and DF are ready to be applied to the CMA for applications with low quality of service (QoS) and complexity requirements, although they may not be preferable in terms of the fundamental performance measures such as the DMT [6].

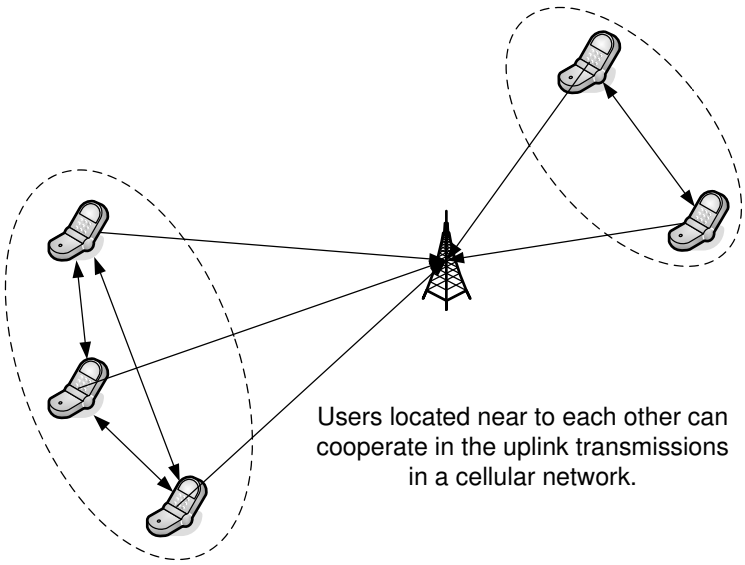


Figure 1.4: Cooperative multiple access channels.

In [7], a modified version of NAF scheme for the CMA is used in order to achieve the MISO DMT upper bound. In particular, the cooperative multiple access problem is considered across several cooperation frames. Each of the source nodes transmits only once in its uniquely assigned symbol transmission interval per cooperation frame. During each symbol transmission interval, a source transmits a linear combination of its own symbol and the signal it observed during its most recent symbol reception interval. This assignment naturally enables the transmission of a new independent symbol as well as a relayed symbol in every symbol interval, and thus eliminate the half-duplex and orthogonal constraints. It is shown this CMA-NAF strategy can help a CMA to achieve the optimal DMT as a corresponding fully connected MISO system.

1.4.3 Cooperative multiple relay channels

The CMR as illustrated in Fig. 1.5 is one of the most studied channel models in classical information theory [10–14]. However, the exact capacity of the relay channel is still an open problem [14]. Recent work has focused on deriving its DoF or multiplexing gain, which becomes increasingly accurate estimation of the exact capacity when the SNR is sufficiently large.

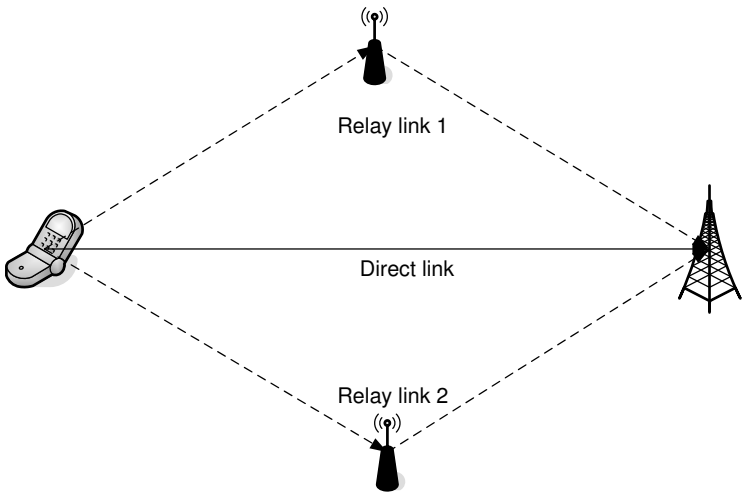


Figure 1.5: Cooperative multiple relay channels.

In [6], the authors there evaluate the performance of fixed relaying strategies including AF and DF schemes, selection relaying strategies and incremental relaying strategies for the CMR. It is shown that AF has better DMT characteristic than DF, while incremental AF scheme can dramatically improve the spectral efficiency over both fixed and selection relaying schemes.

Also as stated in [6], a key area of further research is ”exploring cooperative diversity protocols in the high spectral efficiency regime”. Two useful observations can be made here. Firstly, from the *Data Processing Theorem*, we know soft value contains more information than its hard decision. Thus, we should always prefer the class of AF schemes than the class of DF schemes. Secondly, in order to achieve high spectral efficiency, i.e., high multiplexing gain, we should allow the sources to transmit as often as possible.

In [7], the authors there point out that the suboptimality of the relaying strategies reported in [6] stems from the use of orthogonal signal space in the time domain. Thus, several relaying strategies with the use of nonorthogonal signal space are proposed including NAF and DDF schemes. In particular, for the CMR, DDF scheme is shown to outperform either DF or NAF scheme.

Since the remaining significant loss of diversity in the high multiplexing gain regime

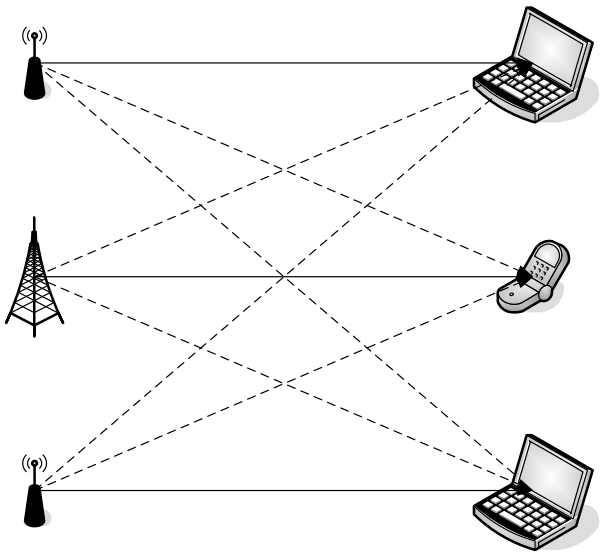


Figure 1.6: Cooperative interference channels.

for DDF scheme comes from the half-duplex constraint, the authors in [8] proposed SSAF scheme which makes use of soft information (i.e., it is in the class of AF schemes), eliminates the orthogonal constraint (i.e., it allows nonorthogonal simultaneous transmissions) and loosens the half-duplex constraint (i.e., it is a sequential relaying strategy and thus the effect of half-duplex constraint can be ignored asymptotically). It is shown that SSAF scheme can achieve the optimal MISO DMT upper bound for the CMR asymptotically when the number of relays is large enough.

1.5 Cooperative networks with higher degree of freedom

Cooperative networks with higher DoF are mostly studied in the context of cooperative interference channels (CIC) with more than two source-destination pairs as shown in Fig. 1.7. The exact capacity region of the general interference network has been an open problem to information theorists for decades. Even for the two-user case, capacity region is only known for special cases such as those with strong and very strong interference [15, 16]. The best known result for the general two-user Gaussian interference network can determine the capacity region within 0.5 bit for real cases or 1 bit for complex cases [17, 18] by using a modified version of the Han-Kobayashi scheme [19].

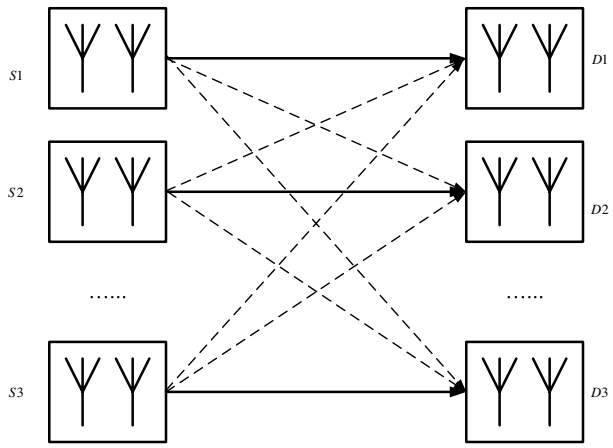


Figure 1.7: General K-user cooperative interference channels.

For the general K-user CIC as shown in Fig. 1.7, where $K > 2$, neither the capacity region nor the DMT upper bound is known. Most research focused on DoF region [20, 21], which characterizes the capacity scaling behavior with respect to the SNR. In [21], Cadambe and Jafar showed that the sum capacity of the general $(1 \times 1)^K$ CIC [3]can be approximated as

$$C(\text{SNR}) = \frac{K}{2}\log(\text{SNR}) + o(\log(\text{SNR})), \tag{1.4}$$

where $f(x) = o(g(x))$ denotes $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$. The DoF characterization $\frac{K}{2}$, which is also known as the multiplexing gain, becomes increasingly accurate as $o(\log(\text{SNR}))$ tends to be negligible compared to $\frac{K}{2}\log(\text{SNR})$ in the high SNR regime.

1.5.1 Interference management in cooperative interference channels

The most distinctive feature that differentiates CIC from its unit DoF counterparts is there exist desired and undesired signals. While in unit multiplexing gain networks where every signal transmission can be potentially unharmful, it is less so in CIC where undesired signal may play the role of interference that significantly decreases the performance. Thus, in order to benefit from the cooperation processes in networks containing CIC, interfer-

[3] $(M \times N)^K$ is used to denote a K-user interference network, where each transmitter has M antennas, each receiver has N antennas, and each user wants to achieve d DoF per channel use.

ence management technologies are necessary to control the detrimental side effects.

Conventionally, interference is often managed in an orthogonal or near-orthogonal approach, where either time-division multiplexing access (TDMA), frequency-division multiplexing access (FDMA) or code-division multiplexing access (CDMA) with minimum total squared correlation is used to separate transmissions into nonintersecting domains. However, this cake-cutting perspective has been proved to be suboptimal to achieve the multiplexing gain upper bound of CIC [21].

In the next two sections, we introduce two recently proposed interference management technologies called network coding and interference alignment, both of which improve the performance of CIC by using smart interference control rather than complete interference avoidance. They will be used in later chapters to design transmission strategies for cooperative networks containing CIC.

1.5.2 Network coding

Network coding was initially proposed in [22] to achieve the capacity of a single-session multicast network by permitting intermediate nodes to encode received data rather than just to do traditional routing operations. For a single-session multicast network, it was shown in [23] that linear codes are sufficient to achieve the multicast capacity. A polynomial time algorithm for network code construction was proposed in [24]. Later, a distributed random linear code construction approach was proposed in [25], which was also shown to be asymptotic valid given a sufficiently large field size. For a multiple-session network, it was shown in [26, 27] that linear network coding may be insufficient to achieve the capacity. Moreover, finding a network coding solution for a network with multiple sessions was shown to be a NP-hard problem [28, 29].

Although optimal network coding solutions for multiple-session networks are generally unknown, simple network coding solutions are able to offer tremendous throughput improvement for wireless cooperative networks, which was famously demonstrated by [30–33]. The main motivation behind using network coding in cooperative communications is the wireless broadcast nature, which means when a node transmits its signal,

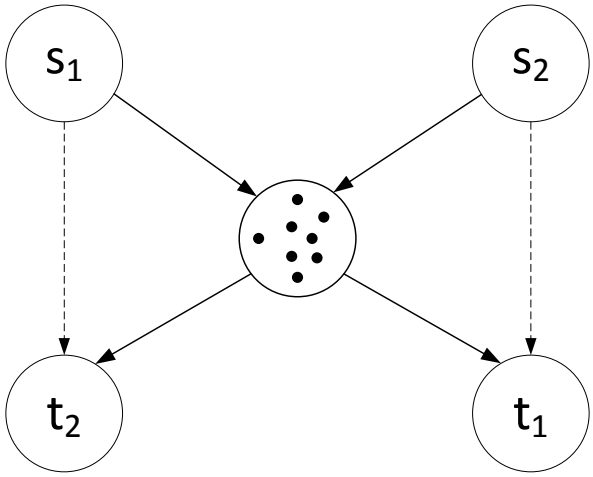


Figure 1.8: Relay-aided X network with a cluster of N relays.

every other node can potentially overheard this information. Conventionally, the overheard signal is treated as noise or interference and thus completely ignored. However, as shown in [31], smartly controlled interference can be used to greatly improve the total network throughput. While interference is harmful in a conventional perspective, if a node has previously transmitted or overheard this interference, its detrimental effects can be completely removed and thus increase the chances of conveying more information in a single transmission.

To illustrate the use of network coding in cooperative communications, let us consider a relay-aided X network (RAXN) as shown in Fig. 1.8. This is a cooperative network with combining features of CBC, CMA, CMR and CIC. There are conventional as well as network coding based cooperative strategies for RAXN, while those network coding based strategies have been shown to be powerful to improve the system performance.

Traditional multihop routing

Traditional multihop routing strategy transmits information over multiple hops along paths from the sources to the destinations. It uses only point-to-point coding, treating all interference as noise and the information is fully decoded at each intermediate relay. Much of current protocol development activity is based on the idea of multihop routing. From

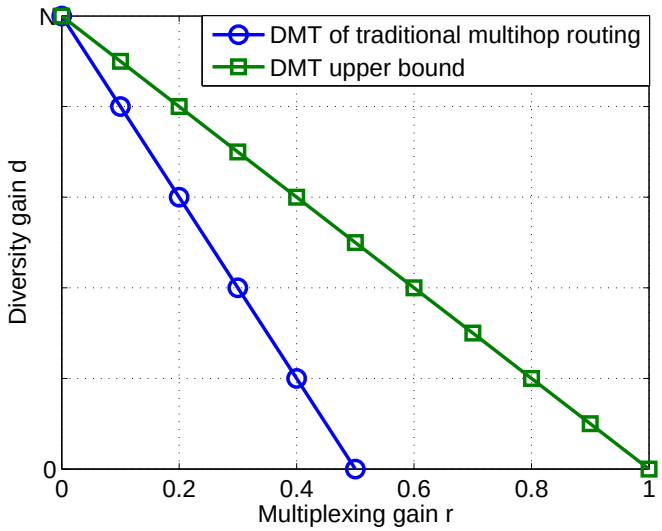


Figure 1.9: DMT of traditional multihop routing.

the transport capacity point of view, several network information theorists have justified the order optimality of multihop routing in the relatively high attenuation scenario [34–36]. This order optimality of the transport capacity characterizes the achievable throughput in the error-free case. In practice, the slope of the bit error rate (BER) is also important because we want to set up the communication with some acceptable QoS, and thus the DMT characteristic is also an important measurement.

Using traditional multihop routing strategy, the idea of interference avoidance is often used to achieve an acceptable QoS. Thus, we need four time slots to complete the communication task, i.e., each source uses one time slot to transmit its information to the relays. The relays fully decode each source's information and then forward each of them to its intended destination using one time slot respectively. This multihop routing strategy is indeed a realization of the DF strategy as shown in [6]. For clarity, we show DMT for this strategy, which is the same as that of DF strategy in Fig. 1.9.

Digital network coding

Consider the transmissions from the sources to the relays as the multiple-access phase and the transmissions from the relays to the destinations as the broadcast phase.

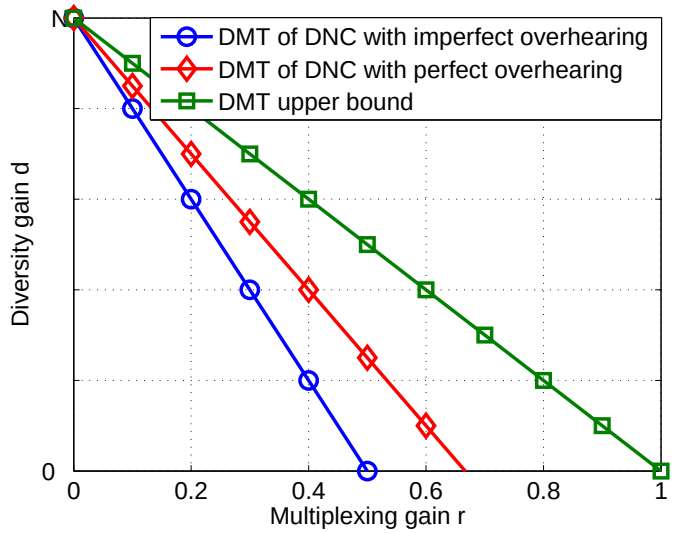


Figure 1.10: DMT of digital network coding.

The multiple-access phase of digital network coding (DNC) strategy is just like that of traditional multihop routing strategy, where each source transmits its information to the relays sequentially. Moreover, due to the wireless broadcast nature, each source's unintended destination can also receive the signal emitted by the undesired source, i.e., t_1 can receive signal from s_2 and t_2 can receive signal from s_1.

Firstly, we assume each destination can perfectly decode this overheard information and stores it in its memory stack. Then, the broadcast phase of DNC strategy involves the exclusive-or (XOR) between the two sources' information at the relays. Again, due to the wireless broadcast nature, when the relays broadcast the network coded information, both destinations can receive the signal. After decoding the XOR of the two sources' information, each destination XORs it again with its previously stored overheard information, in order to extract its desired information. Thus, it only needs three time slots to complete the communication task using DNC strategy with perfect overhearing. The DMT characteristic of DNC in this case is similar to that of DF strategy with improved efficiency (multiplexing gain).

Secondly, we consider the situation with imperfect overhearing, i.e., when one or both of the destinations cannot perfectly decode the overheard information. In this case, the destinations simply discard the imperfectly overheard information. Thus, since

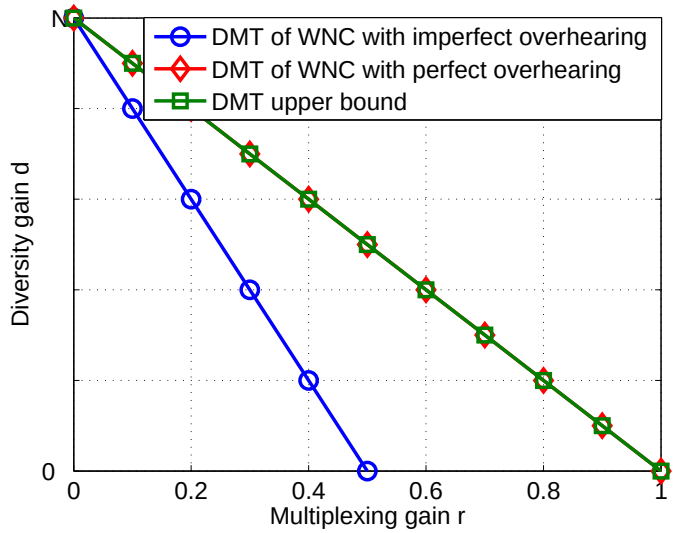


Figure 1.11: DMT of wireless network coding.

there is not enough overheard information to help in the decoding process, the relays cannot broadcast network coded packets to the destinations. Instead, with imperfect overhearing, DNC strategy falls back to traditional multihop routing strategy and still uses four time slots to complete the communication task. The complete DMT characteristic of DNC strategy is shown in Fig. 1.10.

Wireless network coding

In [32, 33], it was shown that the network coding operations at the relays in DNC can also be done in the air using electromagnetic waves. The broadcast phase of wireless network coding (WNC) strategy is just like that of DNC, i.e., in the case with perfect overhearing, the relays broadcast the network coded information to both destinations in one time slot due to the wireless broadcast nature; in the case with imperfect overhearing, WNC also falls back to traditional multihop routing strategy and uses two time slots in the second phase.

If the overheard information can be perfectly decoded by both destinations, then two sources can transmit their information simultaneously to the relays in the multiple-access phase of WNC. Instead of decoding each source's information separately, the

relays decode the XOR of two sources' information directly using the superpositioned signal. Thus, WNC saves one time slot compared to DNC in the multiple-access phase and only needs two time slots to complete the entire communication task with perfect overhearing. However, if the overheard information is imperfect, then the relays need to transmit each source's information separately to the corresponding destination. Thus, WNC still needs two time slots for each source to transmit its information to the relays in the multiple-access phase with imperfect overhearing, and thus four time slots to complete all. In summary, with perfect overhearing, WNC has the same DMT characteristic as that of the AF strategy in [37], and otherwise, WNC essentially falls back to multihop routing. The DMT characteristic of WNC strategy is illustrated in Fig. 1.11.

Network coding opportunities in practice

The tremendous throughput improvement offered by network coding among multiple sessions was most famously demonstrated by [30, 31]. However, it was shown in [27] that linear network coding may not be sufficient for multiple sessions network information flow problems. Moreover, finding a network coding solution for a network with multiple sessions was shown to be a NP-hard problem [28].

It has been proved that previously well known approaches for code construction for a single session multicast network is optimal for a multiple sessions network, if and only if we can find a edge-disjoint subgraph for each session which can at the same time support the multicast capacity for each session. In [38], random coding was applied to multiple sessions after transforming the network topology to construct pollution-free subgraphs.

In practice, such qualified edge-disjoint subgraphs may not exist most of the time because of the overlapping of edges belong to multiple sessions. The benefits of the overlapping among multiple sessions were studied in [39]. The authors there proposed two metrics: overlap ratio and overlap width, to measure the benefits that a system can achieve by coding across multiple sessions. The main idea there is to divide multiple sessions into different groups based on those two metrics and construct a linear network

coding solution for each group. Simulation results in [39] showed that such a scheme can achieve about 30% higher throughput than coding within the same session.

Given the inherent difficulty of multiple sessions network information flow problems, the authors in [40] considered network coding for only two unicast sessions and gave conditions under which a linear network coding solution exists. Further, the equivalent conditions under which there exists a linear network coding solution for two multicast sessions were proved in [41]. Since the butterfly network is well known to admit network coding gain, the authors in [42,43] focused on decomposing the whole network into many butterfly based structures and used linear programming to find a network coding solution.

To the best of our knowledge, there are no generalized conditions for the existence of a linear network coding solution for a multiple unicasts (more than 2) information flow problem. This problem is important because the prevalent multiple pair communication problems are ubiquitous. Moreover, before we actually apply any network coding solution to a network with arbitrary topology, we need these conditions to check whether there is any network coding gain, how much is the network coding gain and how to construct the network codes.

In Appendix A, we try to characterize the generalized conditions which guarantee to provide network coding gain for a multiple unicasts information flow problem. In particular, we want to show the generalized conditions are equivalent to the existence of the generalized butterfly networks. By doing this, we convert the problem of finding a network coding solution to another equivalent qualified paths finding problem, where the qualified paths should form a generalized butterfly network.

1.5.3 Interference alignment

The scheme used in [21] to achieve the DoF upper bound of CIC is interference alignment, which controls the interference contamination such that all interference signals are aligned into a certain signal subspace and leaves the remaining signal subspace interference-free for desired signal. Equation (1.4) implies that on average, each user can almost achieve half the rate as if there were no interference at all, no matter how

many of them share the resource. Thus, in the high SNR regime, the sum capacity scales linearly with the number of users.

Prosperous research works follow to construct interference alignment solutions using various techniques [44–51] and to apply similar ideas to several different applications [52, 53]. However, although being theoretical powerful, interference alignment may not be feasible for certain network configurations. In [54], the feasibility conditions for interference alignment were analyzed. The interference alignment problem was viewed as a multivariate polynomial system, and a $(M \times N)^K$ interference network is feasible to achieve interference alignment without symbol extension only if

$$M + N \geqslant (K + 1)d, \tag{1.5}$$

because only under this condition, the number of variables exceeds the number of equations so that a solution may exist.

For single-antenna interference networks, practical scheme only exists for special cases that can achieve the DoF upper bound exactly with finite complexity. Symbol extension is widely regarded as necessary to asymptotically approach the DoF upper bound [21] for more general cases. However, this is only true for time varying or frequency selective fading channels. For deterministic networks with constant channel coefficients, simple symbol extension will generate a scaled identity matrix, which can not be used to separate the desired and interference signals into different signal subspaces.

1.6 Research objectives

Even for the relatively simple primitive networks with unit DoF, cooperative strategy design is far from perfect. Some previously proposed strategies were designed to increase the network throughput and some others were designed to reduce the network failure probability. A lot of these strategies were even designed to achieve one performance measure at the price of the other. Although optimal DoF bound for CIC is known, most achieving strategies require infinite complexity and unrealistic channel conditions.

We aim to develop methodologies and algorithms about how to use the DMT principle to design cooperative strategies. In cases where the DMT optimal bounds can be explicitly derived, we provide cooperative strategy designing rules to achieve the optimal bounds. In other cases, we show how to use the DMT principle to design cooperative strategies to meet certain performance targets.

Moreover, we aim to simplify future cooperative strategy designing processes. We show that due to the DMT, one can always design a cooperative strategy to maximize one measure and then gradually trade a portion of it for the improvement of the other measure whatever the particular requirement is.

1.7 Thesis structure

The structure of this thesis closely follows our procedure to step-by-step achieve the research objectives, i.e., to generalize and simplify cooperative strategy designing rules based on the DMT principle.

Chapter 2 focuses on cooperative networks with unit DoF including three primitive topologies CBC, CMA and CMR. These networks have only unit DoF because of the cut-set bound constraint and thus every signal transmission can be made potentially useful. We propose a GSSAF strategy that provides DMT optimality for every of these three networks. Moreover, we also prove GSSAF is asymptotically optimal for all cooperative networks with unit DoF, even for compound networks consisting combinations of these three primitive topologies such as cooperative broadcast relay channels (CBRC) and cooperative multiple access relay channels (CMAR), etc. This is our first step to generalize cooperative strategy design tasks using the DMT measure.

Chapter 3 continues to study cooperative strategies design for RAXN. This is a special cooperative network topology with unit DoF that has combining features of three primitive topologies. Due to its special properties, we propose a WNC based partial interference cancelation strategy that can use arbitrary level of overhearing information to improve the network throughput and robustness. The research in this chapter completes the strategy design problem for cooperative networks with unit DoF. Appendix I

also presents a distributed scheme to find out whether there is any network coding gain for a given network with arbitrary topology

Chapter 4 moves to network topologies with more than unit DoF, typically represented by CIC with more than two source-destination pairs. The main problem for such networks is the competence of shared resource between different users, which leads to the suboptimality of achievable DoF. To meet our objectives to generalize and simplify cooperative strategy designing rules, we analyze the special tradeoff between diversity and DoF in CIC and provide insights that although DMT in CIC is infeasible to be explicitly derived, the general DMT principle still applies.

Chapter 5 proposes a noncoherent interference alignment strategy based on our insights observed in Chapter 5. Our proposed strategy trades signal power for intermediate diversity gain towards ultimate multiplexing improvement. Our strategy is generalized and simplified in the sense that it is universal DMT optimal for general single-antenna CIC with minimum scheduling efforts and only two symbol extensions.

Chapter 6 conclude this thesis.

Chapter 2

Generalized Sequential Slotted Amplify and Forward Strategy in Cooperative Communications

2.1 Introduction

From last chapter, we have seen that cooperative communication can significantly improve the efficiency and robustness of wireless communication systems. Many relaying strategies for cooperative networks with various topologies have been proposed with the trend of allowing the source nodes to transmit as much as possible and using nonorthogonal signal subspace as much as possible [7]. These cooperative strategies are often compared using the diversity and DMT [3], which is a fundamental measure that characterizes throughput and error performance simultaneously.

For CMA with multiple cooperative source nodes, a single destination node but without any dedicated relay node, NAF strategy, which protects half of sources' signal by allowing a source node to relay its previously received signal simultaneously with another source node transmitting a new independent message, has been shown to be optimal [7]. For CMR with a single source node, a single destination node and multiple dedicated relay nodes, SSAF strategy which permits each relay node to forward its previously received signal in a specific assigned time slot sequentially, has been shown to be asymptotically

optimal [8]. The best known relaying strategy for CBC with a single source node, multiple cooperative destination nodes but without any dedicated relay node is CBC-DDF, which is based on the DDF strategy that allows a destination node to help others only after it has successfully decoded the desired information. However, CBC-DDF has been shown to be suboptimal in the high multiplexing gain regime [7].

The optimality of a strategy for unit multiplexing gain cooperative networks is defined as the ratio between the DMT upper bound for any strategy and the achievable DMT lower bound for this specific strategy that tends to 1 as the total number of nodes in the network increases. In a wireless cooperative network, one natural DMT upper bound for a destination node is the MISO DMT upper bound, which is obtained by viewing all other nodes in the network as being physically connected through a genie with perfect knowledge about the source information. For CMA with N source nodes, CMR with $N-1$ relay nodes and CBC with N destination nodes, the MISO DMT upper bound is [3]

$$d(r) = N(1-r)^{+}. \tag{2.1}$$

The suboptimality of the CBC-DDF strategy motivated us to develop a better transmission strategy for CBC which can approach the DMT upper bound. More importantly, in practical wireless networks, nearly every transmission is a broadcast process, nearly every reception is a multiple access process and a large number of wireless nodes need to function as relays. Thus, we are particularly interested in finding a transmission strategy that is universally optimal for CMR, CBC and CMA.

In this chapter, we propose a generalized SSAF (GSSAF) strategy for cooperative networks. We firstly analyze the achievable DMT of GSSAF in each cooperation sub frame. Then, together with simulation results, GSSAF is shown to be able to achieve the DMT upper bounds for CMR, CBC and CMA. Finally, GSSAF is shown to be a unified asymptotic optimal strategy for wireless cooperative networks with unit multiplexing gains.

2.1.1 Notations

In this chapter, we use $s_m, 1 \leqslant m \leqslant M$ to denote the M source nodes, $t_n, 1 \leqslant n \leqslant N$ to denote the N destination nodes and $e_k, 1 \leqslant k \leqslant K$ to denote the K relay nodes. We always omit the subscript of s_m, t_n or e_k when there is only one source node, one destination node or one relay node if no confusion can be raised. $(M \times K \times N)$ is used to denote a wireless cooperative network with M sources, K relays and N destinations.

Every node is constrained by average energy E. All source nodes transmit independent information at the same rate R. We use $x_{s_m,l}$ and $x_{e_k,l}$ to represent the transmit signal from the mth source node and the kth relay node at the lth time slot, and we use $y_{e_k,l}$ and $y_{t_n,l}$ to represent the receive signal at the kth relay node and the nth destination node at the lth time slot. h_{s_m,e_k} is used to denote the channel gain between the mth source node and the kth relay node and h_{s_m,t_n} and h_{e_k,t_n} are similarly defined. We assume the existing physical links are all quasi-static flat Rayleigh-fading.

2.2 Generalized SSAF strategy for cooperative networks

Wireless cooperative networks using SSAF strategies have several attractive properties. The relays have low processing complexity because they only need to scale and retransmit their previously received signal. Decoding is only needed at the destinations and not until the end of one cooperation frame. Moreover, the scheduling complexity is also low because every node operates sequentially and equally.

2.2.1 GSSAF description

While original SSAF strategy was designed only for $(1 \times K \times 1)$, our proposed GSSAF works for general $(M \times K \times N)$ which treats every node other than the source or its single destination as a relay in each sub frame level. Specifically, our proposed GSSAF strategy for a $(M \times K \times N)$ cooperative network is defined as follows:

- One cooperation frame consists M sub frames.
- The m-th sub frame consists $K_m + N_m + 1 - \mathbf{1}_1(N_m)$ time slots, where $1 \leqslant m \leqslant M$,

$K_m + N_m = M + K + N - 1$ and K_m and N_m are the numbers of equivalent dedicated relays and destinations of s_m respectively.

  - s_m keeps transmitting a new message in every time slot.
  - From the second time slot, every node except for s_m and its intended single destination is selected in one time slot to forward its received signal in the previous time slot sequentially.

- Each source operates in a time-sharing fashion across different sub frames.
- Each destination starts decoding its desired messages after the whole cooperation frame.

2.2.2 Optimality of GSSAF

The main result about GSSAF strategy is the summarized in the following theorem.

Theorem 1. *The achievable DMT lower bound of GSSAF in the m-th sub frame is*

$$\begin{aligned} d(r) \quad > \quad & [(\tilde{N}-3) + 3\mathbf{1}_1(N_m) - (\tilde{N}+1)r]^+ \\ & + [1 - \frac{\tilde{N}+1}{\tilde{N}+\mathbf{1}_1(N_m)}r]^+, \end{aligned} \tag{2.2}$$

where $\tilde{N} = K_m + N_m - \mathbf{1}_1(N_m)$.

Proof. When $N_m = 1$, GSSAF in the m-th sub frame specializes to the original SSAF strategy in [8] for CMR with K_m relays and $K_m + 1$ slots, whose DMT was proven to be

$$d(r) = (1-r)^+ + [K_m - (K_m+1)r]. \tag{2.3}$$

For cases when $N_m > 1$, several techniques can be used to simplify the analysis. Firstly, we note that the highest probability of error occurs when s_m only transmits common messages and all destinations want to decode every transmitted message from the source [7]. Secondly, the diversity gain of a destination that does not transmit in the second or the last slot is not better than the diversity gains of those who do transmit in those two slots (because one more slot transmitted signal is protected by an extra path).

Thus, we only investigate the diversity gain of a destination $q_l \in \mathcal{Q}_1$ when s_m only transmits common messages, for $1 < l < K_m + N_m$, to give a DMT lower bound. Because we only consider the m-th sub frame, we ignore the subscript m in all analytical expressions and use $\tilde{N}$ to denote $K_m + N_m$ for simplicity of expressions.

Since there is no difference in processing different symbols, we assume each transmission contains only one symbol. The received signal vector for the destination node q_l is

$$\mathbf{y}_{q_l} = \mathbf{H}_{q_l,(\tilde{N}+1)\times(\tilde{N}+1)} \cdot \mathbf{x}_{q_l,(\tilde{N}+1)\times 1} + \mathbf{n}_{q_l,(\tilde{N}+1)\times 1}. \tag{2.4}$$

Assume the channel gain between s and q_l is h_{s,q_l}, for $1 \leqslant l \leqslant \tilde{N}$, and the channel gain between q_l and q_{l+1} is $h_{q_l,q_{l+1}}$, for $1 \leqslant l \leqslant \tilde{N}-1$. We will later propose a relay pre-ordering algorithm that aims to choose the channel gain between consecutive used relays, i.e., $|h_{q_l,q_{l+1}}|$ as small as possible. Under independent Rayleigh-fading channel realizations, when the number of destination nodes is large, there is a high probability that a bad CER to NER link exists with $|h_{q_l,q_{l+1}}|$ being very small, for $1 \leqslant l \leqslant \tilde{N}-1$. Thus, the signal from CER is small interference at NER, which can be viewed as small noise enhancement at NER. From the analysis for the accumulated noise later, we know this small noise enhancement does not affect the DMT analysis. Thus, for analytical simplicity, we assume $|h_{q_l,q_{l+1}}| = 0$. Under reciprocal channel realizations, we also have $|h_{q_{l+1},q_l}| = |h_{q_l,q_{l+1}}| = 0$. Finally, due to the half-duplex constraint, we know that $|h_{q_l,q_l}| = 0$, for $1 \leqslant l \leqslant \tilde{N}$.

From the GSSAF strategy description earlier, we know that $y_{q_l,l+1} = 0$. Thus, the effective received signal vector changes to

$$\tilde{\mathbf{y}}_{q_l} = \tilde{\mathbf{H}}_{q_l,\tilde{N}\times\tilde{N}} \cdot \tilde{\mathbf{x}}_{q_l,\tilde{N}\times 1} + \tilde{\mathbf{n}}_{q_l,\tilde{N}\times 1} \tag{2.5}$$

where

$$\tilde{\mathbf{H}}_{q_l,\tilde{N}\times\tilde{N}} = \begin{pmatrix} \mathbf{A}_{q_l,l\times l} & \mathbf{0}_{q_l,l\times(\tilde{N}-l)} \\ \mathbf{0}_{q_l,(\tilde{N}-l)\times l} & \mathbf{B}_{q_l,(\tilde{N}-l)\times(\tilde{N}-l)} \end{pmatrix}_{\tilde{N}\times\tilde{N}}, \tag{2.6}$$

$$\mathbf{A}_{q_l} = \begin{pmatrix} h_{s,q_l} & 0 & \dots & 0 \\ h_{q_1,q_l}\beta_{q_1}h_{s,q_1} & h_{s,q_l} & \dots & 0 \\ & \dots\dots & & \\ 0 & \dots & 0 & h_{s,q_l} \end{pmatrix}, \tag{2.7}$$

and

$$\mathbf{B}_{q_l} = \begin{pmatrix} h_{s,q_l} & 0 & \dots & 0 \\ & \dots\dots & & \\ 0 & \dots & h_{q_{\tilde{N}},q_l}\beta_{q_{\tilde{N}}}h_{s,q_{\tilde{N}}} & h_{s,q_l} \end{pmatrix}. \tag{2.8}$$

β_{q_l} denotes the normalization factor at destination node q_l (the selected lth relay node) so that its forwarded signal satisfies its average energy constraint E.

Thus, we have:

1. For time slots $k = 1, l$ and $(l+2)$

$$y_{q_l,k} = h_{s,q_l} \cdot x_{s,k} + n_{q_l,k}. \tag{2.9}$$

2. For time slots $2 \leqslant k \leqslant (\tilde{N}+1)$, $k \neq l, l+1$ and $(l+2)$

$$\begin{aligned} y_{q_l,k} &= \breve{\mathbf{H}}_{q_l,k} \cdot \breve{\mathbf{x}}_{q_l,k} + n_{q_l,k} \\ &= \begin{pmatrix} h_{q_{k-1},q_l}\beta_{q_{k-1}}h_{s,q_{k-1}} & h_{s,q_l} \end{pmatrix} \begin{pmatrix} x_{s,k-1} \\ x_{s,k} \end{pmatrix} \\ &\quad + n_{q_l,k}. \end{aligned} \tag{2.10}$$

Note that, in the second case above, the term $n_{q_l,k}$ is actually the accumulated noise from both relay and destination nodes, which can be represented as

$$n_{q_l,k} = \hat{n}_{q_l,k} + h_{q_{k-1},q_l}\beta_{q_{k-1}}n_{q_{k-1},k-1} \tag{2.11}$$

where the normalization factor $\beta_{q_{k-1}}$ at the relay node q_{k-1} should satisfy the energy constraint

$$|\beta_{q_{k-1}}|^2 \leqslant \frac{E}{E|h_{s,q_{k-1}}|^2 + \sigma^2} = \frac{\rho}{\rho|h_{s,q_{k-1}}|^2 + 1}. \tag{2.12}$$

Let h be complex standard normal distributed and v denote the exponential order of $\frac{1}{|h|^2}$. The probability density function (pdf) of v can be written as [7]

$$p_v \doteq \begin{cases} \rho^{-\infty} = 0, & \text{for } v < 0 \\ \rho^{-v}, & \text{for } v \geqslant 0. \end{cases} \tag{2.13}$$

Thus, for $\tilde{N}$ independent identically distributed (i.i.d) variables $\{v_j\}_{j=1}^{\tilde{N}}$, the probability that $(v_1, \ldots, v_{\tilde{N}})$ belongs to a set O is

$$P_O \doteq \rho^{-d_O}, \quad \text{for } d_O = \inf_{(v_1,\ldots,v_{\tilde{N}}) \in O^+} \sum_{j=1}^{N} v_j \tag{2.14}$$

given that $\mathbb{R}^{\tilde{N}+}$ denotes the set of nonnegative $\tilde{N}$-tuples and $O^+ = O \bigcap \mathbf{R}^{\tilde{N}+}$ is not empty. So, the exponential order of P_O depends only on O^+ and is dominated by the realization with the largest exponential order.

From [55], it can be shown that under the consideration of outage events belonging to set O^+, proper selection of the normalization factor $\beta_{q_{k-1}}$ will make its exponential order vanish in all the DMT analytical expressions and the noise enhancement problem does not affect the DMT result. Thus, the DMT of our proposed strategy depends only on the channel matrix and not on the variance of the accumulated noise. So, for analytical simplicity, we assume the accumulated noise equals to the noise at each destination node which does not affect the DMT analysis.

In order to get a DMT lower bound, we want to first upper-bound the probability of error of the ML decoder. Using Bayes' theorem, we can write

$$\begin{aligned} P_E(\rho) &= P_O(R) P_{E|O} + P_{E,O^c} \\ &\leqslant P_O(R) + P_{E,O^c} \end{aligned} \tag{2.15}$$

where the outage events set O and its complement set O^c are chosen such that $P_O(R)$ dominates P_{E,O^c}, i.e.,

$$P_{E,O^c} \dot{\leqslant} P_O(R). \tag{2.16}$$

Thus, we have

$$P_E(\rho) \dot{\leqslant} P_O(R). \tag{2.17}$$

From (2.14) and the simplified channels described by (2.9) and (2.10), we let

$$P_O(R) \doteq \rho^{-d_O(r)}, \tag{2.18}$$

for

$$d_O(r) = \inf_{(\mathbf{v},\mathbf{u})\in O^+} [\sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} (v_{k-1}+u_{k-1}) + v_l]. \tag{2.19}$$

Thus, $d_O(r)$ provides a lower bound on the diversity gain achieved by our proposed CBC-SSAF strategy for destination node q_l.

From the definition of the outage probability [4], for the $\tilde{N}+1$ messages from the source node s to the destination node q_l, we know that

$$\begin{aligned} P_O(R) &= P[I(\tilde{\mathbf{x}}_{q_l}; \tilde{\mathbf{y}_{q_l}}) < R] \\ &\leqslant P[\sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} I(\check{\mathbf{x}}_{q_l,k}; y_{q_l,k}) + \\ &\qquad \sum_{k=1,l,l+2} I(x_{s,k}; y_{q_l,k}) < (\tilde{N}+1) r \log\rho]. \end{aligned} \tag{2.20}$$

For time slots $k = 1, l$ and $(l+2)$

$$\begin{aligned} &\lim_{\rho\to\infty} \frac{I(x_{s,k}; y_{q_l,k})}{\log\rho} \\ = &\lim_{\rho\to\infty} \frac{\log(1+\rho|h_{s,q_l}|^2)}{\log\rho} = (1-v_l)^+. \end{aligned} \tag{2.21}$$

For time slots $2 \leqslant k \leqslant \tilde{N}+1$, $k \neq l, l+1, l+2$

$$\begin{aligned} &\lim_{\rho\to\infty} \frac{I(\check{\mathbf{x}}_{q_l,k}; y_{q_l,k})}{\log\rho} = \lim_{\rho\to\infty} \frac{\log(1+\rho\check{\mathbf{H}}_{q_l,k}\check{\mathbf{H}}^{\dagger}_{q_l,k})}{\log\rho} \\ = &\lim_{\rho\to\infty} \frac{\log(1+\rho|h_{q_{k-1},q_l}\beta_{q_{k-1}}h_{s,q_{k-1}}|^2+\rho|h_{s,q_l}|^2)}{\log\rho} \\ = &(\max\{1-v_{k-1}-u_{k-1}, 1-v_l\})^+, \end{aligned} \tag{2.22}$$

where $[\cdot]^{\dagger}$ is used to denote the matrix conjugated transposition operation. Thus, from (2.20), (2.21) and (2.22), the outage events set O^{+} should be defined as

$$\begin{aligned} O^{+} = \{(\mathbf{v},\mathbf{u}) \in \mathbb{R}^{(2\tilde{N}-5)+} | 3(1-v_l)^{+} \\ + \sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} (\max\{1-v_{k-1}-u_{k-1}, 1-v_l\})^{+} \\ < (\tilde{N}+1)r\}. \end{aligned} \tag{2.23}$$

From (2.23), we can easily see that, in order to let the outage events happen, the following constraints must be simultaneously satisfied (necessary but not sufficient conditions):

1. For v_l, we have

$$v_l > (1 - \frac{\tilde{N}+1}{\tilde{N}} r)^{+}. \tag{2.24}$$

2. For $\sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1}(v_{k-1}+u_{k-1})$, we have

$$\sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} (v_{k-1}+u_{k-1}) > [(\tilde{N}-3)-(\tilde{N}+1)r]^{+}. \tag{2.25}$$

From (2.16) and (2.17), we see that (2.24) and (2.25) can be used to lower-bound the diversity gain $d(r)$ if and only if the outage events described by (2.23) dominate the probability of error of the ML decoder. To see this, we first observe that the channel described by (2.5) can be seen as a coherent linear Gaussian channel as

$$\tilde{\mathbf{y}}_{q_l,\tilde{N}\times 1} = \tilde{\mathbf{H}}_{q_l,\tilde{N}\times\tilde{N}} \cdot \tilde{\mathbf{x}}_{q_l,\tilde{N}\times 1} + \tilde{\mathbf{n}}_{q_l,\tilde{N}\times 1}. \tag{2.26}$$

Thus, the average pairwise error probability (PEP) of the ML decoder at high SNR can be upper-bounded by [3, Eqn. 7]

$$\begin{aligned} P_{PE} &\leqslant \det(\mathbf{I}_{\tilde{N}} + \frac{1}{2}\rho\tilde{\mathbf{H}}_{q_l}\tilde{\mathbf{H}}_{q_l}^{\dagger})^{-1} \\ &\doteq \det(\mathbf{I}_{\tilde{N}} + \rho\tilde{\mathbf{H}}_{q_l}\tilde{\mathbf{H}}_{q_l}^{\dagger})^{-1} \end{aligned} \tag{2.27}$$

where we set $\sum_{\tilde{\mathbf{n}}_{q_l}} = \sigma^2\mathbf{I}_{\tilde{N}}$ because such manipulation does not affect the DMT. From the

formula for multiple-input multiple-output channel capacity [4, 5], we know that

$$\begin{aligned} & \sum_{k=1,l,l+2} I(x_{s,k}; y_{q_l,k}) + \sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} I(\check{\mathbf{x}}_{q_l,k}; y_{q_l,k}) \\ \leqslant \quad & \log[\det(\mathbf{I}_{\tilde{N}} + \rho\tilde{\mathbf{H}}_{q_l}\tilde{\mathbf{H}}_{q_l}^{\dagger})]. \end{aligned} \tag{2.28}$$

Thus,

$$\begin{aligned} & \rho^{\sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1}(\max\{1-v_{k-1}-u_{k-1},1-v_l\})^+ + 3(1-v_l)^+} \\ & \dot{\leqslant} \det(\mathbf{I}_{\tilde{N}} + \rho\tilde{\mathbf{H}}_{q_l}\tilde{\mathbf{H}}_{q_l}^{\dagger}). \end{aligned} \tag{2.29}$$

Let the $\tilde{N}+1$ messages from the source node s to the destination node q_l form a codeword of length $\tilde{N}+1$. The data rate is $R = (\tilde{N}+1)r\log\rho$ BPCU and we have a total of $\rho^{(\tilde{N}+1)r}$ codewords. Thus, we can bound $P_E(\rho)$ as

$$\begin{aligned} P_E(\rho) \quad \dot{\leqslant} \quad & \rho^{-\sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1}(\max\{1-v_{k-1}-u_{k-1},1-v_l\})^+} \\ & \cdot\rho^{-3(1-v_l)^+ + (\tilde{N}+1)r}. \end{aligned} \tag{2.30}$$

Therefore, P_{E,O^c} can be written as [7]

$$P_{E,O^c} \dot{\leqslant} \int_{O^{c+}} \rho^{-d_e(r,\mathbf{v},\mathbf{u})}\mathbf{d_v}\mathbf{d_u} \tag{2.31}$$

where

$$\begin{aligned} d_e(r,\mathbf{v},\mathbf{u}) = & -(\tilde{N}+1)r + 3(1-v_l)^+ \\ & + \sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} (\max\{1-v_{k-1}-u_{k-1}, 1-v_l\})^+ \\ & + [\sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} (v_{k-1}+u_{k-1}) + v_l]. \end{aligned} \tag{2.32}$$

Because P_{E,O^c} is dominated by the smallest term of $d_e(r,\mathbf{v},\mathbf{u})$ over O^{c+}, we can write

$$P_{E,O^c} \dot{\leqslant} \rho^{-d_e(r)}, \quad \text{for } d_e(r) = \inf_{(\mathbf{v},\mathbf{u})\in O^{c+}} d_e(r,\mathbf{v},\mathbf{u}). \tag{2.33}$$

Comparing (2.18) and (2.33), we see that for (2.16) to be met, O^+ should be defined as

$$\begin{aligned} O^+ = \{(\mathbf{v},\mathbf{u}) \in \mathbb{R}^{(2\tilde{N}-5)+} | 3(1-v_l)^+ \\ + \sum_{k=2,k\neq l,l+1,l+2}^{\tilde{N}+1} (\max\{1-v_{k-1}-u_{k-1}, 1-v_l\})^+ \\ \leqslant (\tilde{N}+1)r\} \end{aligned} \tag{2.34}$$

which contains the outage events set (2.23). So, we conclude that the outage events described by (2.23) also satisfy (2.16) and therefore dominate the probability of error of the ML decoder. Thus, we can use (2.24) and (2.25) to lower-bound $d_O(r)$ (which further provides a lower bound for $d(r)$) as

$$d_O(r) > [(\tilde{N}-3)-(\tilde{N}+1)r]^+ + (1-\frac{\tilde{N}+1}{\tilde{N}}r)^+. \tag{2.35}$$

□

2.2.3 Practical implementation

In each sub frame, every node other than the source or its intended single destination can be used as relays. We dynamically order the nodes to act as relays before the start of each sub frame, such that the next effective relay (NER) is chosen as the worst destination for the current effective relay (CER). The purpose of such relay pre-ordering is to separate consecutively used relays as much as possible by making the channel gain between them as small as possible.

A simple method to implement the relay pre-ordering operations can be done using Algorithm 1. Before the start of the m-th sub frame, let other nodes except for the source or its intended single destination sequentially broadcast short *probe frames*. These operations consume $K_m + N_m - \mathbf{1}_1(N_m)$ *probe frame* time slots. After reception of these *probe frames*, the source can choose the first relay based on step 3 in Algorithm 1. Because of the wireless broadcast nature, every other node can at the same time estimate its local $K_m + N_m - 1 - \mathbf{1}_1(N_m)$ relay-to-relay channel gains. If we assume the wireless reciprocity property holds, a node can use this information to choose its NER

Algorithm 1 Relay pre-ordering algorithm for GSSAF.

Input: Before the start of the m-th sub frame, set $i = 1$ and $s = s_m$. Moreover, define a node set $\mathcal{Q}_1 = \emptyset$ and a node set $\mathcal{Q}_2 = \{q_1, q_2, \ldots, q_{K_m+N_m-\mathbf{1}_1(N_m)}\}$, which is a entry-wise mapping of the node set $\{s_1, \ldots, s_{m-1}, s_{m+1}, \ldots, s_M, e_1, \ldots, e_{K_l}, t_1, \ldots, t_{N_m-\mathbf{1}_1(N_m)}\}$

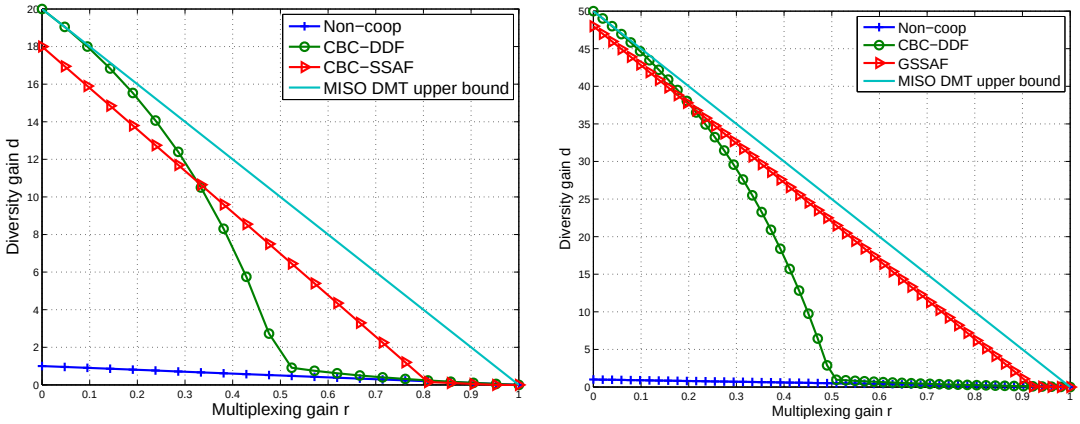
Output: Ordered node set $\mathcal{Q}_1$

1: **while** $i \leqslant K_m + N_m - \mathbf{1}_1(N_m)$ **do**
2: **if** $i = 1$ **then**
3: Choose $q_j \in \mathcal{Q}_2$ such that $h_{s,q_j} \geqslant h_{s,q_n}, \forall q_n \in \mathcal{Q}_2, n \neq j$
4: **else**
5: Choose $q_j \in \mathcal{Q}_2$ such that $h_{q_{i-1},q_j} \leqslant h_{q_{i-1},q_n}, \forall q_n \in \mathcal{Q}_2, n \neq j$
6: **end if**
7: Swap the indexes for q_j and q_i
8: Add q_i to $\mathcal{Q}_1$
9: Delete q_i from $\mathcal{Q}_2$
10: $i = i + 1$
11: **end while**

based on step 5 in Algorithm 1. Then, each node sends back a short *feedback frame* to the source with only its chosen NER's unique ID embedded in it. These operations consume $K_m + N_m - \mathbf{1}_1(N_m)$ *feedback frame* time slots. After decoding these *feedback frames*, the source can construct a linked list locally and the relay pre-ordering operations can be completed. In each *data frame* time slot, the source only needs to embed a unique ID in its signature. Each node extracts this ID and if it matches its own, it acknowledges it should function as a relay in the next *data frame* time slot. The extra payloads of these relay pre-ordering operations are $K_m + N_m - \mathbf{1}_1(N_m)$ *probe frame* time slots and $K_m + N_m - \mathbf{1}_1(N_m)$ *feedback frame* time slots, which can be well assumed to be much shorter than the $K_m + N_m + 1 - \mathbf{1}_1(N_m)$ *data frame* time slots. Moreover, considering the underlying block fading assumption, the cost of the scheduling algorithm is negligible.

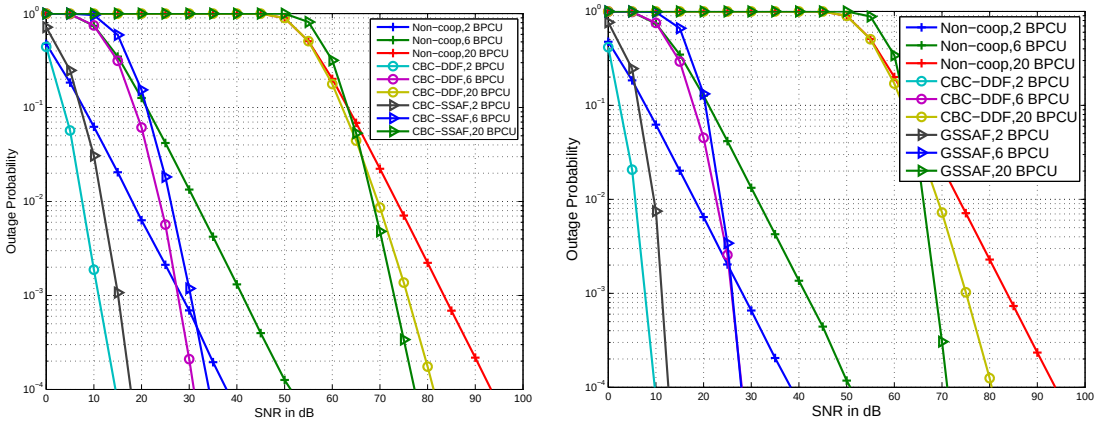
2.3 Conclusion

The proposed GSSAF strategy has no specific requirements on M, K or N, and is ready to be used as a generalized relaying rule in cooperative networks with different topologies and connectivity statuses. The key implication is that in each sub frame, it is asymptotically optimal to treat every other node as a relay by using a sequential AF relaying rule.



(a) 20 users. (b) 50 users.

Figure 2.1: DMTs of GSSAF strategy for CBC.



(a) 5 users. (b) 10 users.

Figure 2.2: Outage behaviors of GSSAF strategy for CBC.

Remark 1. *GSSAF is asymptotically optimal for CMR, CBC and CMA.*

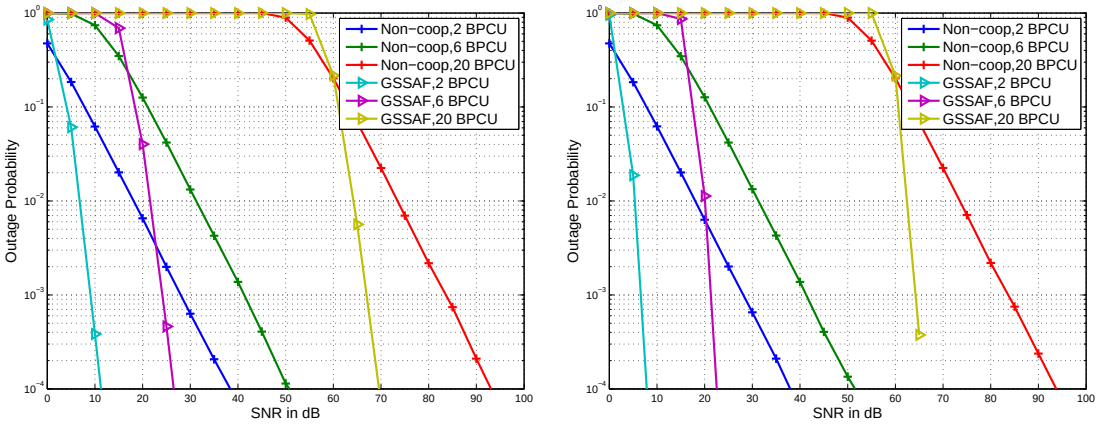
It is trivial to verify its optimality for CMR since the original SSAF is a special case of GSSAF.

For CBC, It is easy to argue its asymptotic optimality by setting $M = 1$ and $K = 0$ in (2.2) for a $(M \times K \times N)$ network. Compared to currently best known relaying strategy for cooperative broadcast channels CBC-DDF, GSSAF offers tremendous DMT improvement in the high multiplexing gain regime. This is the region where the CBC-DDF strategy is incapable to beat direct transmission strategy. As shown in Fig. 2.1, the DMT analytical results state that the GSSAF strategy approaches the DMT upper bound as N increases and is thus asymptotically optimal. Moreover, from the outage behaviors in Fig. 2.2, it is easy to see the diversity order of GSSAF is almost dominantly better than those of non-cooperative and CBC-DDF strategies in the high spectral efficiency regime.

Similarly, we can verify its optimality for CMA by setting $K = 0$ and $N = 1$ and considering M sub frames at the destination jointly [9, 55]. For CMA, GSSAF can exactly achieve the MISO DMT upper bound with any number of source nodes and finite cooperation frame length. Fig. 2.3 shows the outage behaviors of GSSAF for CMAs with various numbers of cooperative users. It can be seen that the diversity order of GSSAF is much higher than that of the non-cooperative strategy. Moreover, the slops of the outage curves remains the same even in the high spectral efficiency regime, which demonstrates the exact optimality of GSSAF for CMA.

Remark 2. *GSSAF can be used as a unified relaying strategy for general wireless cooperative networks, and it is asymptotically optimal for those networks with unit multiplexing gains.*

Due to the cut-set bound, wireless cooperative networks with unit multiplexing gains considered in this chapter are constrained to those with either a single source and/or a single destination (an exception is a (2×2) interference network with only private messages). Conventionally, even for the simplest cases such as CMR, CBC and CMA, a node needs to use different strategies under different circumstances, not to mention compound networks such as cooperative broadcast relay channels (CBRC) or coopera-



(a) 5 users. (b) 10 users.

Figure 2.3: Outage behaviors of GSSAF strategy for CMA.

tive multiple access relay channels (CMAR). For wireless cooperative networks with unit multiplexing gains, the MISO DMT upper bound still holds. Moreover, from (2.2), it is easy to see that GSSAF is able to asymptotically approach this upper bound. It follows that a node in cooperative networks with unit multiplexing gains can use GSSAF as a unifying rule no matter its specific role as a source , a relay or a destination, to achieve the DMT optimality.

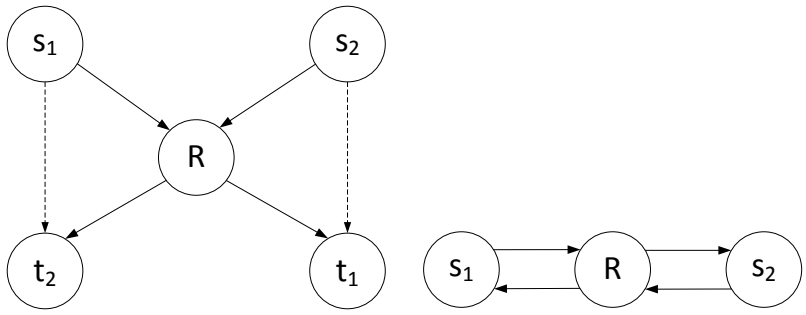
Chapter 3

Diversity and Multiplexing Tradeoff of Wireless Network Coding for Relay-Aided X Networks

3.1 Introduction

After generalization of transmission strategies for primitive cooperative networks with unit multiplexing gains, we come to consider a special network topology called the relay-aided X network (RAXN). This channel model is ubiquitous in practice because the source and destination nodes do not have to be the true communication end-users. It can happen as long as two traditional routing paths intersect at some point and share one or more intermediate relay nodes. Because of the shared use of the resources, there is higher throughput and reliability requirements at the shared relay nodes. This motivates us to develop new transmission strategies to meet the ongoing higher and higher QoS requirements. Moreover, this network has combining features of those primitive unit multiplexing topologies as well as properties of CIC. Thus, it requires interference management technologies to control the information flow in order to maximize the system performance.

In this chapter, we propose a WNC based partial interference cancelation strategy for RAXN. The main concept of WNC can be demonstrated using the two-way exchange network (TWXN) as shown in Fig. 3.1(a), where s_1 and s_2 want to exchange information



(a) TWXN. (b) RAXN.

Figure 3.1: Illustrative channel models for the use of WNC in RAXN.

through the help of a relay R. Without WNC, the conventional hop-by-hop transmission strategy needs 4 transmissions including $s_1 \rightarrow R$, $R \rightarrow s_2$, $s_2 \rightarrow R$ and $R \rightarrow s_1$. With WNC, only 2 transmissions are needed including $(s_1, s_2) \rightarrow R$ and $R \rightarrow (s_1, s_2)$.

Interference cancelation (IC) is often used in the decoding process of WNC to retrieve the desired signal [31–33, 56], which involves the use of priori known information to reconstruct the interference signal. The priori known information comes from either a nodes's previously transmitted information or its overheard information. While it is reasonable to assume a node's previously transmitted information to be perfect (without fading or noise), it is less so to assume the overheard information to be lossless.

As pointed out in [33], for a simple RAXN as shown in Fig. 3.1(b), WNC increases the network throughput on the expenses of higher BER because of the imperfect overhearing. Since one can always trade a strategy's achievable diversity gain for its achievable multiplexing gain [3], it is not immediately clear that which one of WNC or the conventional hop-by-hop transmission strategy is fundamentally better. This question can be equivalently interpreted as: For RAXN with imperfect overhearing, if we allow the conventional hop-by-hop transmission strategy to have higher BER as that of WNC, can it transmit as fast as WNC? Or, if we force WNC to use more redundancy to reduce its BER as that of the conventional hop-by-hop transmission strategy, can its throughput still be higher? This is a fundamental problem needs to be investigated before the prevalent use of WNC.

Due to the problems introduced, WNC for RAXN is not a straightforward exten-

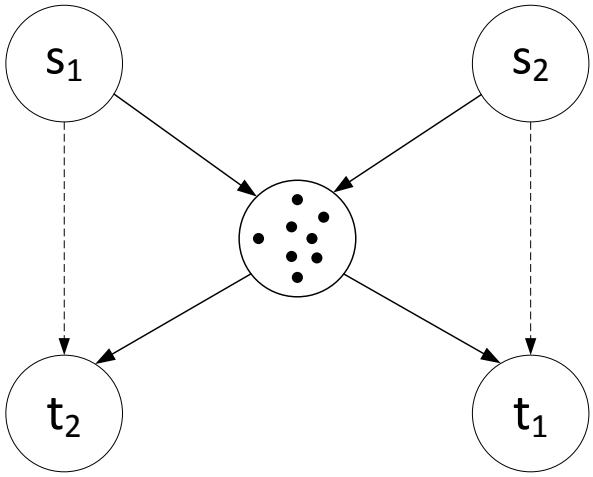


Figure 3.2: Relay-aided X network with a cluster of N relays.

sion of that for TWXN, because of the wireless imperfect overhearing. Thus, the main objective of this chapter is to study how WNC can be used in strategy design to improve the performance of RAXN with imperfect overhearing and what is its ultimate fundamental performance, in terms of the DMT. While WNC is the underlying technology used to improve the multiplexing gain, we also consider the use of clustered relays to increase the diversity gain as shown in Fig. 3.2. We propose a WNC based partial IC strategy for RAXN, which could smartly use the imperfect overhearing to cancel part of the interference. The DMT of the strategy is analyzed which proves the fundamental superiority of WNC over the conventional hop-by-hop transmission strategy even with imperfect overhearing.

3.1.1 Notations and preliminaries

Throughout this chapter, we use $\mathcal{S} = \{s_1, s_2\}$ to denote the two sources, $\mathcal{T} = \{tx_1, t_2\}$ to denote the two destinations and $\mathcal{R} = \{R_1, R_2, \ldots, R_N\}$ to denote the N relays. We use x_{s_1} and x_{s_2} to denote the signal transmitted from the two sources and x_{R_n} to denote the signal transmitted from the nth relay, for $1 \leqslant n \leqslant N$. Similarly, y_{R_n} is used to represent the received signal at the nth relay and y_{t_1} and y_{t_2} are used to denote the received signal at the two destinations respectively.

Every node is constrained by average energy E. All sources transmit independent information at the same rate R. h_{s_m,R_n}, h_{s_m,t_k} and h_{R_n,t_k} are used to denote the channel gain between the mth source and the nth relay, the channel gain between the mth source and the kth destination and the channel gain between the nth relay and the kth destination, where $1 \leqslant m \neq k \leqslant 2$ and $1 \leqslant n \leqslant N$. We assume the existing physical links are all quasi-static flat Rayleigh-fading, which means the channel gains are constant during each frame but change independently between different frames.

We characterize the channels between the sources and their unintended destinations using the amount of information the destinations overheard from their undesired sources in the first time slot. In order to evaluate the performance of our proposed WNC based partial IC strategy when the overheard information is imperfect, we assume that in the first time slot, each destination can only decode part of its undesired source's information correctly, i.e., t_1 can decode R_{t_1} amount of information from s_2 correctly, and t_2 can decode R_{t_2} amount of information from s_1 correctly, where $0 \leqslant R_{t_1}, R_{t_2} \leqslant R$.

Let $d_1 = 1 - Ar_1$ and $d_2 = 1 - Br_2$ be two linear functions that denote the DMTs of two independent messages, where A and B are two constants. The overall DMT is obtained by adding the multiplexing gains up subject to equal diversity gains, and can be written as

$$d = 1 - \frac{AB}{A+B} r. \tag{3.1}$$

3.2 Problems of naive interference cancelation strategies with imperfect overhearing

Firstly consider the simple RAXN as shown in Fig. 3.1(b), where s_1 and s_2 want to send independent information to t_1 and t_2 respectively through the help of a relay R. With WNC, the signalling is as follows:

1. In the first slot, s_1 and s_2 transmits x_{s_1} and x_{s_2} simultaneously. The received signal

at other nodes are

$$\begin{aligned} y_{t_1,1} &= h_{s_2,t_1}x_{s_2} + n_{t_1,1}, \\ y_{t_2,1} &= h_{s_1,t_2}x_{s_1} + n_{t_2,1}, \\ y_{R,1} &= h_{s_1,R}x_{s_1} + h_{s_2,R}x_{s_2} + n_{R,1}. \end{aligned} \tag{3.2}$$

2. In the second slot, R amplifies and forwards (AF) its previously received signal to both t_1 and t_2, whose received signal are

$$\begin{aligned} y_{t_1,2} &= h_{R,t_1}\beta_R h_{s_1,R}x_{s_1} + h_{R,t_1}\beta_R h_{s_2,R}x_{s_2} \\ &\quad + h_{R,t_1}\beta_R n_{R,1} + n_{t_1,2}, \\ y_{t_2,2} &= h_{R,t_2}\beta_R h_{s_1,R}x_{s_1} + h_{R,t_2}\beta_R h_{s_2,R}x_{s_2} \\ &\quad + h_{R,t_2}\beta_R n_{R,1} + n_{t_2,2}. \end{aligned} \tag{3.3}$$

In order to retrieve the desired signal at each destination, there are two possible approaches to do IC using the overheard signal. Take t_1 for instance:

1. Without decoding the overheard information, IC can be done by

$$\begin{aligned} \tilde{y}_{t_1} &= y_{t_1,2} - \frac{h_{R,t_1}\beta_R h_{s_2,R}}{h_{s_2,t_1}} y_{t_1,1} \\ &= h_{Rt_1}\beta_R h_{s_1,R}x_{s_1} + \tilde{n}_{t_1}, \end{aligned} \tag{3.4}$$

   where $\tilde{n}_{t_1} = \frac{h_{R,t_1}\beta_R h_{s_2,R}}{h_{s_2,t_1}} n_{t_1,1} + n_{t_1,2}$. It is easy to verify the accumulative noise variance is infinitely large because $\frac{1}{x}$ is not integrable for an exponential distributed variable x. This fact clearly prohibits the use of such an IC method.

2. With decoding the overheard information, IC can be done by

$$\begin{aligned} \tilde{y}_{t_1} &= y_{t_1,2} - h_{R,t_1}\beta_R h_{s_2,R}\hat{x}_{s_2} \\ &= h_{R,t_1}\beta_R h_{s_1,R}x_{s_1} \\ &\quad + h_{R,t_1}\beta_R h_{s_2,R}(x_{s_2} - \hat{x}_{s_2}) + n_{t_1,2}, \end{aligned} \tag{3.5}$$

where $\hat{x}$ represents the estimated overheard information. Ideally, we would like the overhearing links to be strong enough such that $x_{s_2} = \hat{x}_{s_2}$. However, due to wireless fading and noise corruption, the decoded overheard information is highly possible to be subject to certain errors and it is not reasonable to assume more resource (e.g., coding or retransmission requests) to be used to ensure the overhearing reliability. Thus, such an IC method would lead to error propagation, which will result in incorrect decoding of the desired information.

From the analysis, when the overhearing is imperfect, in order to use WNC to do IC, we must denoise and at the same time avoid error propagation. In the next section, we propose a WNC based partial IC strategy which solve these two problems simultaneously.

3.3 WNC based on partial interference cancelation strategy for RAXN with imperfect overhearing

Consider the RAXN as shown in Fig. 3.2, our propose WNC based partial IC strategy is defined as follows:

1. There are two time slots in each transmission frame.
2. In the first time slot, s_1 and s_2 broadcast their independent information x_{s_1} and x_{s_2} simultaneously to the N shared relays and their unintended destinations respectively.
3. The destinations decode the overheard signal and store the decoded information (may contain errors) in their own memory stacks.
4. In the second time slot, the N shared relays normalize their received signal in the first time slot and broadcast the normalized signal simultaneously to the two destinations.
5. Each destination exploits physical layer hints to divide its decoded overheard information in the first time slot into clean and faulty parts, and use the clean part to reconstruct and cancel part of the interference from the received signal in the second time slot to retrieve their desired information.

The main difference between this WNC based partial IC strategy and the conventional WNC strategy has two folds, which include its "activeness" and its novel approach to use the imperfect overhearing to improve the exchange performance. The activeness means the relays actively and intentionally mix the two sources' signals no matter whether the overheard information is perfect or not. Conventional WNC falls back to the hop-by-by transmission strategy when the overhearing is imperfect (because in this case interference can not be canceled to retrieve the desired signal), even when there are only a few incorrectly overheard symbols. We will show that even when the overhearing is imperfect, our proposed strategy can still help to improve the overall system performance by smartly using the available but imperfect overhearing to cancel part of the interference.

3.3.1 Practical considerations

Channel side information

We assume channel side information is only available at the receivers (CSIR), which is practicable by inserting a negligibly short training sequence into the message sequences. Moreover, we let the relays broadcast their estimated CSIR and normalization factors by embedding them into the training sequence with negligible overhead compared to the original message length.

Synchronization

We do not consider the synchronization issue and assume all the relays are fully synchronized. The synchronization issue arising from the simultaneous relaying operations can be overcome by a distributed relay selection algorithm which chooses only the best relay to forward its received signal from the sources in the first time slot. Moreover, from (3.23) and (3.24) stated later, the performance of simultaneous relaying is dominated by the best two-hop link between the sources and the corresponding destinations. Thus, a distributed relay selection algorithm does not entail a cost on the achievable DMT. However, we only consider simultaneous relaying in this chapter for mathematical simplicity. The same asymptotic DMT performance can be achieved with a suitable distributed relay

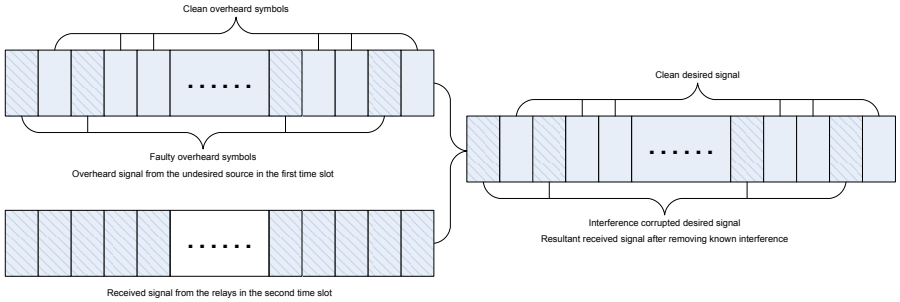


Figure 3.3: Decoding at the destinations of WNC based partial IC strategy.

selection algorithm such as that in [57].

3.3.2 Decoding at the destinations

For the source-to-relay and relay-to-destination links that suffer deep fading, it is feasible to use some physical layer error correction codes or even higher layer protocols to ensure they are error-free because they are the designed transmissions. However, for the overheard links, it is infeasible in practice to spend extra resource to ensure their reliability. Thus, when the destinations try to decode their overheard information in the first time slot, the decoded packets may be imperfect with symbol errors. We divide the decoded overheard information at the destinations in the first time slot into two parts: one part with high probability to be correct (clean overheard symbols); the other part with high probability to be incorrect (faulty overheard symbols). Possible methods to mark the decoded symbols as clean or faulty can use the output of many well known soft decoders or the confidence values calculated from the physical layer signal as shown in [58, 59]. The comparison between different marking methods and their associated error propagation effects are out of the scope of this thesis. The purpose for such division is that we want to use the clean overheard symbols to remove part of the interference in the received signal in the second time slot.

When each destination receives the signal from the relays in the second time slot, it uses the clean overheard symbols together with the channel side information and the normalization factors to reconstruct part of the interference. Then, it subtracts such reconstructed interference from its received signal in the second time slot at the corre-

sponding positions. This results in dividing the desired signal at the destinations into two parts: one part without interference and the other part with unknown interference. For the second part with unknown interference signal, we will use traditional decoding method for multiple access channels (MAC) to extract the desired information. Fig. 3.3 illustrates the decoding process at the destinations.

3.4 DMT analysis

The main result about the WNC based partial IC strategy can be summarized in the following theorem.

Theorem 2. *The achievable DMT lower bound for the proposed WNC based partial IC strategy is*

$$d(r) > N[1 - \frac{2R(2R - R_{t_1} - R_{t_2})}{2R^2 - R_{t_1}^2 - R_{t_2}^2} r], 0 < r < 1. \tag{3.6}$$

3.4.1 Signaling and DMT analysis for the part of the signal with known interference

Since there is no difference in processing different symbols, we assume the first part of the received signal at the nth relay R_n is a combination of two super-symbols. In the first time slot, the received signal vector at the N relays is

$$\begin{aligned} \mathbf{y}_{\mathcal{R},N\times 1} &= \mathbf{H}_{\mathcal{S},\mathcal{R},N\times 2} \cdot \mathbf{x}_{\mathcal{S},2\times 1} + \mathbf{n}_{\mathcal{R},N\times 1} \\ &= \begin{pmatrix} h_{s_1,R_1} & h_{s_2,R_1} \\ h_{s_1,R_2} & h_{s_2,R_2} \\ & \cdots \\ h_{s_1,R_N} & h_{s_2,R_N} \end{pmatrix} \cdot \begin{pmatrix} x_{s_1} \\ x_{s_2} \end{pmatrix} + \begin{pmatrix} n_{R_1} \\ n_{R_2} \\ \cdots \\ n_{R_N} \end{pmatrix}. \end{aligned} \tag{3.7}$$

In the second time slot, the signal transmitted from the nth relay is

$$x_{R_n} = \beta_{R_n} \cdot y_{R_n}, \text{ for } n = 1, 2, \ldots, N, \tag{3.8}$$

where β_{R_n} is the normalization factor at the relay R_n which is used to ensure the relay to satisfy its average energy constraint E. Each relay chooses its energy normalization factor β_{R_n} based on its own received signal energy. We use an equal power allocation scheme with the average energy constraint E for each relay in the second time slot of one transmission frame. A more advanced power allocation scheme will enhance the performance in terms of the throughput and outage probability only in the low SNR regime. However, such improvement becomes trivial in the high SNR regime and a simple equal power allocation scheme is sufficient to achieve the same DMT as that of the optimal power allocation scheme.

The received signal at the destinations in the second time slot can be written as

$$y_{t_1} = \mathbf{H}_{\mathcal{R},t_1,1\times N} \cdot \mathbf{x}_{\mathcal{R},N\times 1} + n_{t_1} \tag{3.9}$$

where $\mathbf{H}_{\mathcal{R},t_1,1\times N} = [h_{R_1,t_1}, h_{R_2,t_1}, ..., h_{R_N,t_1}]$, $\mathbf{x}_{\mathcal{R},N\times 1} = [x_{R_1}, x_{R_2}, ...x_{R_N}]^T$ and $[\cdot]^T$ denotes the matrix transposition. From (3.7) and (3.8), we know $\mathbf{x}_{\mathcal{R},N\times 1}$ can also be written as

$$\begin{aligned} \mathbf{x}_{\mathcal{R},N\times 1} &= \beta_{N\times N} \cdot \mathbf{y}_{\mathcal{R},N\times 1} \\ &= \begin{pmatrix} \beta_{R_1} & 0 & \dots & 0 \\ 0 & \beta_{R_2} & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \beta_{R_N} \end{pmatrix} \cdot \begin{pmatrix} y_{R_1} \\ y_{R_2} \\ \dots \\ y_{R_N} \end{pmatrix}. \end{aligned} \tag{3.10}$$

Substituting (3.7) and (3.10) into (3.9), we can get

$$\begin{aligned} y_{t_1} &= \mathbf{H}_{\mathcal{R},t_1,1\times N} \cdot \mathbf{x}_{\mathcal{R},N\times 1} + n_{t_1} \\ &= \mathbf{H}_{\mathcal{R},t_1,1\times N} \cdot \beta_{N\times N} \cdot \mathbf{y}_{\mathcal{R},N\times 1} + n_{t_1} \\ &= \mathbf{H}_{\mathcal{R},t_1} \cdot \beta \cdot (\mathbf{H}_{\mathcal{S},\mathcal{R}} \cdot \mathbf{x}_{\mathcal{S}} + \mathbf{n}_{\mathcal{R}}) + n_{t_1} \\ &= [h_{R_1,t_1}, h_{R_2,t_1}, ..., h_{R_N,t_1}] \\ &\quad \cdot \begin{pmatrix} \beta_{R_1} & 0 & \dots & 0 \\ 0 & \beta_{R_2} & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \beta_{R_N} \end{pmatrix} \cdot \begin{pmatrix} h_{s_1,R_1} & h_{s_2,R_1} \\ h_{s_1,R_2} & h_{s_2,R_2} \\ & \dots \\ h_{s_1,R_N} & h_{s_2,R_N} \end{pmatrix} \\ &\quad \cdot \begin{pmatrix} x_{s_1} \\ x_{s_2} \end{pmatrix} + \tilde{n}_{t_1}. \end{aligned} \tag{3.11}$$

Thus,

$$\begin{aligned} y_{t_1} &= \sum_{n=1}^{N} h_{s_1,R_n} \beta_{R_n} h_{R_n,t_1} x_{s_1} + \sum_{n=1}^{N} h_{s_2,R_n} \beta_{R_n} h_{R_n,t_1} x_{s_2} \\ &\quad + (\sum_{n=1}^{N} h_{R_n,t_1} \beta_{R_n} n_{R_n} + n_{t_1}). \end{aligned} \tag{3.12}$$

With CSIR and normalization factors received from the relays and estimated by the destinations themselves, each destination can remove its known interference from its received signal in the second time slot. For destination t_1, $\sum_{n=1}^{N} h_{s_2,R_n} \beta_{R_n} h_{R_n,t_1} x_{s_2}$ is the known signal and thus can be removed from (3.12). Thus, we can write

$$y_{t_1} = \sum_{n=1}^{N} h_{s_1,R_n} \beta_{R_n} h_{R_n,t_1} x_{s_1} + (\sum_{n=1}^{N} h_{R_n,t_1} \beta_{R_n} n_{R_n} + n_{t_1}). \tag{3.13}$$

The accumulated noise at the first destination t_1 from both relay and destinations can be written as

$$\tilde{n}_{t_1} = \sum_{n=1}^{N} h_{R_n,t_1} \beta_{R_n} n_{R_n} + n_{t_1} \tag{3.14}$$

where the normalization factor β_{R_n} is chosen to satisfy energy constraint

$$\begin{aligned} |\beta_{R_n}|^2 &\leqslant \frac{E}{E|h_{s_1,R_n}|^2 + E|h_{s_2,R_n}|^2 + \sigma^2} \\ &= \frac{\rho}{\rho|h_{s_1,R_n}|^2 + \rho|h_{s_2,R_n}|^2 + 1}. \end{aligned} \tag{3.15}$$

where σ^2 is the noise variance.

Let w_n denote the exponential order of $|\beta_{R_n}|^2$ and $v_{i,n}$ and $u_{n,j}$ denote the exponential orders of $\frac{1}{|h_{s_i,R_n}|^2}$ and $\frac{1}{|h_{R_n,t_j}|^2}$ respectively, for $i,j = 1,2$ and $1 \leqslant n \leqslant N$. Thus, from (3.15), we can easily see that

$$w_n \leqslant \min(v_{1,n}, v_{2,n}, 1). \tag{3.16}$$

For (3.16) to be met, we choose w_n as

$$w_n = (v_{1,n}, v_{2,n})^- \tag{3.17}$$

where we use $(x)^-$ to mean $\min\{x, 0\}$ and $(x)^+$ to mean $\max\{x, 0\}$. This choice for w_n will ensure β_{R_n} to satisfy the energy constraint (3.15). This, under the consideration of outage events belonging to set O^+ as stated in (2.14) will make w_n, i.e., the exponential order of β_{R_n} vanish in all the DMT analytical expressions.

Let $w_{\tilde{n}_{t_1}}$ and $w_{n_{t_1}}$ denote the exponential orders of the variances of $\tilde{n}_{t_1}$ and n_{t_1}. From (3.14) and (3.17), we know

$$w_{\tilde{n}_{t_1}} = \max_{n=1,2,\ldots,N} \{(-u_{n,1})^+\} + w_{n_{t_1}} = w_{n_{t_1}}. \tag{3.18}$$

Thus, the DMT of the strategy depends only on the channel matrix and not on the variance of the accumulated noise. So, for analytical simplicity, we assume the accumulated noise equals to the noise at each destination which does not affect the DMT analysis. Thus, we can rewrite (3.13) as

$$y_{t_1} = \sum_{n=1}^{N} h_{s_1,R_n} \beta_{R_n} h_{R_n,t_1} x_{s_1} + n_{t_1}. \tag{3.19}$$

and y_{t_2} can be similarly written as

$$y_{t_2} = \sum_{n=1}^{N} h_{s_2,R_n}\beta_{R_n}h_{R_n,t_2}x_{s_2} + n_{t_2}. \tag{3.20}$$

Observing (3.19) and (3.20), we immediately notice that they are very similar to multiple-input single-output (MISO) channels and thus should have similar DMT characteristics. We obtain a lower bound of the DMT by firstly approximating the exponential order of the error probability of ML decoder by that of the outage probability. From the definition of the outage probability, we know that

$$\begin{aligned} P_{O_1} &= P[I(x_{s_1}; y_{t_1}|x_{s_2}) < R_{t_1}] \\ &= P[\log(1+\rho\sum_{n=1}^{N}|h_{s_1,R_n}|^2|\beta_{R_n}|^2|h_{R_n,t_1}|^2) \\ &\quad < r_{t_1}\log\rho] \end{aligned} \tag{3.21}$$

and

$$\begin{aligned} P_{O_2} &= P[I(x_{s_2}; y_{t_2}|x_{s_1}) < R_{t_2}] \\ &= P[\log(1+\rho\sum_{n=1}^{N}|h_{s_2,R_n}|^2|\beta_{R_n}|^2|h_{R_n,t_2}|^2) \\ &\quad < r_{t_2}\log\rho]. \end{aligned} \tag{3.22}$$

In the high SNR regime, the exponential order of β_{R_n} vanishes and thus we have

$$\begin{aligned} & \lim_{\rho\to\infty}\frac{I(x_{s_1}; y_{t_1}|x_{s_2})}{\log\rho} \\ = & \lim_{\rho\to\infty}\frac{\log(1+\rho\sum_{n=1}^{N}|h_{s_1,R_n}|^2|\beta_{R_n}|^2|h_{R_n,t_1}|^2)}{\log\rho} \\ = & \max_{n=1,2,\cdots,N}\{1-v_{1,n}-u_{n,1}\}^+ \end{aligned} \tag{3.23}$$

and

$$\begin{aligned} & \lim_{\rho\to\infty} \frac{I(x_{s_2}; y_{t_2}|x_{s_1})}{\log\rho} \\ = & \lim_{\rho\to\infty} \frac{\log(1+\rho\sum_{n=1}^{N}|h_{s_2,R_n}|^2|\beta_{R_n}|^2|h_{R_n,t_2}|^2)}{\log\rho} \\ = & \max_{n=1,2,\ldots,N}\{1-v_{2,n}-u_{n,2}\}^+. \end{aligned} \tag{3.24}$$

Thus, from (3.21), (3.22), (3.23) and (3.24), the outage events sets O^+ should be defined as

$$O_1^+ = \{(\mathbf{v},\mathbf{u}) \in \mathbb{R}^{2N+} | \max_{n=1,2,\ldots,N}\{1-v_{1,n}-u_{n,1}\}^+ < r_{t_1}\} \tag{3.25}$$

and

$$O_2^+ = \{(\mathbf{v},\mathbf{u}) \in \mathbb{R}^{2N+} | \max_{n=1,2,\ldots,N}\{1-v_{2,n}-u_{n,2}\}^+ < r_{t_2}\}. \tag{3.26}$$

From (3.25) and (3.26), we can easily see that, in order for the outage events to happen, the following constraints should be satisfied:

1. $v_{1,n} + u_{n,1} > 1 - r_{t_1}, \forall n = 1, 2, \ldots, N.$
2. $v_{2,n} + u_{n,2} > 1 - r_{t_2}, \forall n = 1, 2, \ldots, N.$

From (2.14), we know the outage probability should be dominated by the probability of the outage event with the largest exponential order, i.e., the outage event with the smallest $d_O(r)$. Thus, we can write

$$P_{O_1} \doteq \rho^{-d_{O_1}(r_{t_1})}, \tag{3.27}$$

for $d_{O_1}(r_{t_1}) = \inf_{(\mathbf{v},\mathbf{u})\in O^+}[\sum_{n=1}^{N}(v_{1,n}+u_{n,1})]$. And

$$P_{O_2} \doteq \rho^{-d_{O_2}(r_{t_2})},$$

for $d_{O_2}(r_{t_2}) = \inf_{(\mathbf{v},\mathbf{u})\in O^+}[\sum_{n=1}^{N}(v_{2,n}+u_{n,2})]$.

Thus, we can lower-bound $d_{O_1}(r_{t_1})$ and $d_{O_2}(r_{t_2})$ as

$$d_{O_1}(r_{t_1}) \geqslant \inf_{(\mathbf{v},\mathbf{u})\in O^+} [\sum_{n=1}^{N}(v_{1,n}+u_{n,1})] > N(1-r_{t_1}) \tag{3.28}$$

and

$$d_{O_2}(r_{t_2}) \geqslant \inf_{(\mathbf{v},\mathbf{u})\in O^+} [\sum_{n=1}^{N}(v_{2,n}+u_{n,2})] > N(1-r_{t_2}). \tag{3.29}$$

As $d_{O_1}(r_{t_1})$ and $d_{O_2}(r_{t_2})$ also serve as lower bounds for $d_1(r_{t_1})$ and $d_2(r_{t_2})$, we can further write

$$d_1(r_{t_1}) > N(1-r_{t_1}) \tag{3.30}$$

and

$$d_2(r_{t_2}) > N(1-r_{t_2}). \tag{3.31}$$

Now, we show the DMTs of $N(1-r_{t_1})$ and $N(1-r_{t_2})$ are actually also upper bounds for the parts of the received signal with known interference at the two destinations with R_{t_1} and R_{t_2} amount of overheard information. Assume in the first time slot, the two sources can transmit their independent R_{t_1} amount of x_{s_1} and R_{t_2} amount of x_{s_2} reliably to the N relays. In practice, this may not be done due to the wireless fading environment and noise corruption. However, this assumption is sufficient to give a DMT upper bound for this part of the received signal. In the second time slot, let the N relays fully cooperate through a genie. Because the two destinations cannot cooperate, the best achievable performance is obtained by viewing the transmissions from the relays to the destinations in the second time slot as two MISO channels.

Without any cooperation between the destinations, the DMTs of the two MISO channels are $N(1-r_{t_1})$ and $N(1-r_{t_2})$ respectively. Considering two independent MISO channels together does not increase the diversity gain. This is because the two MISO channels in the second time slot are statistically independent without any cooperation

and thus provide no further diversity gain by considering them jointly. Thus, we can get upper bounds as

$$d_1(r_{t_1}) < N(1 - r_{t_1}) \tag{3.32}$$

and

$$d_2(r_{t_2}) < N(1 - r_{t_2}). \tag{3.33}$$

Combining (3.30), (3.31), (3.32) and (3.33), we know the DMTs of the strategy for the interference known part of the signal at the two destinations are

$$d_1(r_{t_1}) = N(1 - r_{t_1}) \tag{3.34}$$

and

$$d_2(r_{t_2}) = N(1 - r_{t_2}). \tag{3.35}$$

In order to get the relationship between the diversity gain d and the multiplexing gain r, we need to map points in (3.34) from a coordinate system with r_{t_1} as x-axis to a coordinate system with r as x-axis. Because $r_{t_1} = \frac{R_{t_1}}{R} r$, points $(0, N)$ and $(1, 0)$ are mapped to points $(0, N)$ and $(\frac{R_{t_1}}{R}, 0)$. Thus, in the new coordinate system, the DMT (3.34) changes to

$$d_1(r_{t_1}) = N(1 - \frac{R}{R_{t_1}} r), \tag{3.36}$$

and similarly we have

$$d_2(r_{t_2}) = N(1 - \frac{R}{R_{t_2}} r). \tag{3.37}$$

Finally, from (3.1) and taking the consumption of two time slots in one cooperation frame into consideration, we can get the final overall DMT for the interference known part of the

received signal at both destinations as

$$d_I(r_{t_1}, r_{t_2}) = N[1 - (\frac{2R}{R_{t_1} + R_{t_2}})r], \tag{3.38}$$

where $d_I(r_{t_1}, r_{t_2})$ is used to denote the diversity gain of the interference known part of the received signal at both destinations.

3.4.2 Signaling and DMT analysis for the part of the signal with unknown interference

For this part of the received signal, as we have no information correctly overheard, we cannot remove the interference term from (3.12). Thus, we have no choice but to use traditional decoding method for MAC.

We first note that from the received signal at both destinations, we can at least extract $I(x_{s_1}; y_{t_1}) + I(x_{s_2}; y_{t_2})$ amount of desired information by treating interference as noise. Moreover, in the high SNR regime, for a 2-user interference channel, the total achievable multiplexing gain is 1. Because the multiplexing gain for MAC $\{x_{s_1}, x_{s_2}\} \to y_{t_1}$ is also 1, thus we have $\lim_{\rho\to\infty} \frac{I(x_{s_1}, x_{s_2}; y_{t_1})}{\log \rho} = \lim_{\rho\to\infty} \frac{I(x_{s_1}; y_{t_1}) + I(x_{s_2}; y_{t_2})}{\log \rho}$ and it is sufficient to consider the interference unknown part of the received signal at both destinations as an N-to-1 MISO channel with capacity $I(x_{s_1}, x_{s_2}; y_{t_1}) = I(x_{s_2}; y_{t_1}) + I(x_{s_1}; y_{t_1}|x_{s_2})$ to give a DMT lower bound. In the high SNR regime, the exponential order of β_{R_n} vanishes and thus we have

$$\begin{aligned}
& \lim_{\rho\to\infty} \frac{I(x_{s_2}; y_{t_1})}{\log \rho} \\
= & \lim_{\rho\to\infty} \frac{\log(1 + \frac{\rho \sum_{n=1}^{N} |h_{s_2,R_n}|^2 |\beta_{R_n}|^2 |h_{R_n,t_1}|^2}{\rho \sum_{n=1}^{N} |h_{s_1,R_n}|^2 |\beta_{R_n}|^2 |h_{R_n,t_1}|^2 + 1})}{\log \rho} \\
= & [\max_{n=1,2,\ldots,N} \{1 - v_{2,n} - u_{n,1}\} \\
& - (\max_{n=1,2,\ldots,N} \{1 - v_{1,n} - u_{n,1}\})^+]^+.
\end{aligned} \tag{3.39}$$

From the definition of the outage probability, we know that

$$\begin{aligned} P_O &\doteq P[I(x_{s_1}; y_{t_1}) + I(x_{s_2}; y_{t_2}|x_{s_1}) \\ &\quad < R - R_{t_1} + R - R_{t_2}] \\ &= P[\log(1 + \frac{\rho \sum_{n=1}^{N} |h_{s_1,R_n}|^2 |\beta_{R_n}|^2 |h_{R_n,t_1}|^2}{\rho \sum_{n=1}^{N} |h_{s_2,R_n}|^2 |\beta_{R_n}|^2 |h_{R_n,t_1}|^2 + 1}) \\ &\quad + \log(1 + \rho \sum_{n=1}^{N} |h_{s_2,R_n}|^2 |\beta_{R_n}|^2 |h_{R_n,t_2}|^2) \\ &\quad < (r_{t_1}^c + r_{t_2}^c) \log \rho], \end{aligned} \tag{3.40}$$

where $r_{t_1}^c$ denotes the multiplexing gain of x_{s_1} at destination t_1 with $R - R_{t_1}$ amount of information and unknown interference signal and $r_{t_2}^c$ is similarly defined. From (3.24), (3.39) and (3.40), the outage events set O^+ should be defined as

$$\begin{aligned} O^+ &= \{(\mathbf{v}, \mathbf{u}) \in \mathbb{R}^{3N+} | [\max_{n=1,2,\dots,N} \{1 - v_{2,n} - u_{n,1}\} \\ &\quad - (\max_{n=1,2,\dots,N} \{1 - v_{1,n} - u_{n,1}\})^+]^+ \\ &\quad + [\max_{n=1,2,\dots,N} \{1 - v_{1,n} - u_{n,1}\}]^+ < r_{t_1}^c + r_{t_2}^c\}. \end{aligned} \tag{3.41}$$

Thus, in order for the outage events to happen, the following constraints must be satisfied:

$$v_{2,n} + u_{n,1} > 1 - r_{t_1}^c - r_{t_2}^c, \forall n = 1, 2, \dots, N. \tag{3.42}$$

From (2.14) and (3.41), we can lower-bound the DMT of the interference unknown part of the received signal at both destinations $d_{II}(r_{t_1}^c, r_{t_2}^c)$ as

$$\begin{aligned} d_{II}(r_{t_1}^c, r_{t_2}^c) &\geqslant d_O(r_{t_1}^c, r_{t_2}^c) \\ &\geqslant \inf_{(\mathbf{v},\mathbf{u}) \in O^+} [\sum_{n=1}^{N} (v_{1,n} + u_{n,1} + v_{2,n})] \\ &> N(1 - r_{t_1}^c - r_{t_2}^c). \end{aligned} \tag{3.43}$$

Because $r_{t_1}^c = \frac{R - R_{t_1}}{R} r$ and $r_{t_2}^c = \frac{R - R_{t_2}}{R} r$, using the same technique from (3.34) to (3.36),

we can further write

$$d_{II}(r_{t_1}^c, r_{t_2}^c) > N[1 - (\frac{R}{R - R_{t_1}} + \frac{R}{R - R_{t_2}})r]. \tag{3.44}$$

3.4.3 Overall result

From (3.38), (3.44) and (3.1), we know the overall achievable DMT for the proposed strategy is lower-bounded as

$$d(r) > N[1 - \frac{2R(2R - R_{t_1} - R_{t_2})}{2R^2 - R_{t_1}^2 - R_{t_2}^2}r]. \tag{3.45}$$

As shown in Fig. 3.4, the more information overheard by the destinations, the better DMT the proposed WNC based partial IC strategy can achieve. In the case of perfect overhearing, it can achieve the DMT upper bound, which is obtained by viewing the channel model as two two-hop fully cooperative MISO channels; in the case of imperfect overhearing, conventional WNC strategy falls back to hop-by-hop transmission strategy, while the proposed strategy can still improve the overall system performance by using the partial correctly overheard information to cancel part of the interference. This indicates that, although introducing bi-directional interference to both destinations due to the imperfect overhearing, WNC in general improves the overall system throughput and robustness regardless of the quality of the overhearing links, through an appropriate approach to utilize the imperfect overhearing information.

Theorem 1 also answers the fundamental question that motivated the research in this chapter: if WNC uses more redundancy to achieve the same BER as that of the conventional hop-by-hop transmission strategy, which one has higher throughput? Or, if the conventional hop-by-hop transmission strategy increases its rate to that of WNC, which one has lower BER? From Fig. 3.4, it can be easily seen that, setting either performance measurement (the diversity and multiplexing gains that represents the error and throughput performance respectively) being the same, the other performance measurement of WNC is always strictly better than that of the conventional hop-by-hop transmission strategy. This proves the fundamental superiority of WNC even with imperfect overhearing.

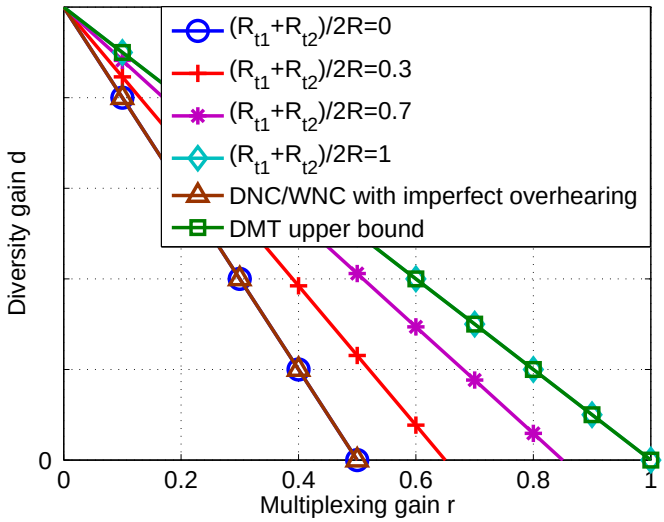


Figure 3.4: DMT of WNC based partial IC strategy with imperfect overhearing.

3.5 Conclusion

The throughput improvement of WNC often comes with the assumption of perfect overhearing or worse error performance. Thus, it is not immediately clear whether or not WNC is fundamentally better than the conventional hop-by-hop transmission strategy for a RAXN with imperfect overhearing. In this chapter, we studied how to use WNC to improve the performance of RAXN with imperfect overhearing and what is its ultimate fundamental performance, from the perspective of the DMT. Through analyzing the DMT lower bound of the proposed achievable strategy, our result gave a strong justification of the superiority of WNC over the conventional hop-by-hop transmission strategy even with imperfect overhearing.

It is worth mentioning that the proposed IC strategy in this chapter is only one of the achievable schemes. We use this partial IC strategy to show the fundamental advantage of WNC over the conventional scheme. It does not exclude other WNC based schemes to exploit the imperfect overhearing to improve the systems performance. In fact, the joint decoding schemes such as that in [60] is a good example of other means to use the overheard information, while its fundamental DMT performance itself is an interesting topic to study.

Chapter 4

Interference Alignment with Diversity

4.1 Introduction

The exact capacity of a general interference network is a long-standing open problem to information theorists, which leads to the difficulty to explicitly derive its optimal DMT. However, the DoF upper bound is known to be achievable by using the interference alignment technology. In this chapter, we analyze the feasibility conditions for CIC that interference alignment and diversity gain can be achieved simultaneously. The results tell us the general DMT principle still applies to CIC. In the next chapter, we will use this insight to design a noncoherent interference alignment strategy for CIC to achieve the DoF upper bound exactly with finite complexity.

Interference alignment is a powerful tool in controlling the interference contamination in a way such that all interference seen by a receiver falls into a certain signal subspace and leaves the remaining subspace interference-free [61]. In [21], Cadambe and Jafar showed that the sum capacity of the K-user interference channel can be approximated as

$$C(\mathrm{SNR}) = \frac{K}{2}\log(\mathrm{SNR}) + o(\log(\mathrm{SNR})). \tag{4.1}$$

Thus, in the high SNR regime, the capacity scales linearly with the number of users.

Equivalently, this means each user can almost achieve a constant rate no matter how many of them share the wireless medium. However, the achievable scheme in [21] requires infinite symbol extensions in a time-varying or frequency-selective environment to asymptotically achieve the DoF upper bound. Moreover, it also assumes global channel knowledge is available at every node which may generate overwhelming overhead in practice.

Almost all currently known interference alignment strategies are designed to achieve the multiplexing gain upper bound promised by interference alignment, while they hardly consider the symbol error rate performance, which can be characterized by the diversity gain. However, these algorithms are not distinguishable in terms of the multiplexing gain in the high SNR regime. Thus, it is important to study their difference from another perspective, which is why we study the diversity gains different interference alignment strategies can provide.

In this thesis, we categorize different interference alignment strategies into diversity interference alignment or zero-forcing interference alignment by whether they use only the interference channel matrices in order to completely null the interference or they use both desired and interference channel matrices in order to maximize or minimize a well chosen utility function. Among our considerations, the zero-forcing interference alignment strategies include the closed-form eigenvector solution for three-user MIMO interference channels in [21] and the iterative zero-forcing algorithm in [44]. Correspondingly, the diversity interference alignment strategies include the minimum mean squared error (MMSE) algorithm in [62] and the iterative max-SINR algorithm in [44]. In particular, we use the zero-forcing and max-SINR algorithms in [44] as examples to evaluate the performance of these two-type strategies in terms of the diversity gain through simulations.

4.2 Problem formulation

Consider the K-user $(M \times N)$ CIC as shown in Fig. 4.1. There are K interfering users sharing the same wireless medium. Each transmitter is equipped with M antennas, each

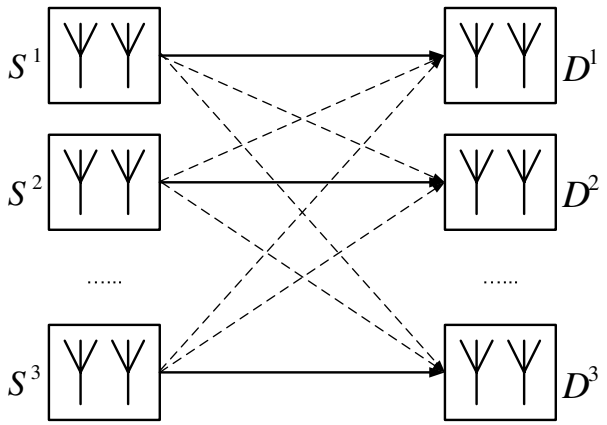


Figure 4.1: A K-user $(M \times N)$ interference network.

receiver is equipped with N antennas and each user has 1 DoF or stream of information to send. The channel between each pair of transmit and receive antennas is complex Gaussian distributed with zero mean and unit variance, and thus its envelop is Rayleigh distributed and its power is exponential distributed.

In this chapter, we use $\mathbf{H}^{[ji]}$ to denote the channel matrix between the i-th transmitter S_i and the j-th receiver D_j, $\mathbf{v}^{[i]}$ to denote the precoding filter at S_i and $\mathbf{u}^{[j]}$ to denote the receiving filter at D_j, for $1 \leqslant i, j \leqslant K, i \neq j$. $s^{[i]}$ and $\hat{s}^{[i]}$ are used to denote the symbol transmitted by S_i with power constraint P and the symbol estimated by D_i after the receiving filter respectively, for $1 \leqslant i \leqslant K$. The noise at D_j is denoted as $\mathbf{n}^{[j]}$ and assumed to be complex Gaussian distributed with zero mean and covariance matrix $\sum_{\mathbf{n}^{[j]}} = \mathbf{I}_{N \times N}$, for $1 \leqslant j \leqslant K$. Thus, the received signal at D_j can be written as

$$\mathbf{y}^{[j]} = \mathbf{H}^{[jj]}\mathbf{v}^{[j]}s^{[j]} + \sum_{i=1, i\neq j}^{K} \mathbf{H}^{[ji]}\mathbf{v}^{[i]}s^{[i]} + \mathbf{n}^{[j]}. \tag{4.2}$$

For the self-completeness of this thesis, we briefly describe the iterative zero-forcing and max-SINR algorithms here, while more details can be found in [44].

4.2.1 Zero-forcing algorithm

The total leakage interference at D_j due to all undesired transmitters is

$$I^{[j]} = P\mathbf{u}^{[j]\dagger}\mathbf{Q}^{[j]}\mathbf{u}^{[j]}, \tag{4.3}$$

where $\mathbf{Q}^{[j]} = \sum_{i=1,i\neq j}^{K} \mathbf{H}^{[ji]}\mathbf{v}^{[i]}\mathbf{v}^{[i]\dagger}\mathbf{H}^{[ji]\dagger}$ is the covariance matrix of the interference signal at D_j. Thus, in order to minimize the total leakage interference, $\mathbf{u}^{[j]}$ is chosen as

$$\mathbf{u}^{[j]} = \nu_{\min}[\mathbf{Q}^{[j]}], \tag{4.4}$$

where $\nu_{\min}[\mathbf{Q}^{[j]}]$ is the eigenvector corresponding to the smallest eigenvalue of $\mathbf{Q}^{[j]}$. Due to the wireless reciprocal property, the signal space along which a receiver sees the least leakage interference is also the signal space along which it causes the least leakage interference in the reciprocal network. Thus, exchanging the roles of $\mathbf{u}^{[i]}$ and $\mathbf{v}^{[i]}$, i.e., setting $\overleftarrow{\mathbf{v}}^{[i]} = \mathbf{u}^{[i]}$ and $\overleftarrow{\mathbf{u}}^{[i]} = \mathbf{v}^{[i]}$ does not increase the total leakage interference, where $\overleftarrow{\mathbf{v}}^{[i]}$ is the precoding filter in the reciprocal network and $\overleftarrow{\mathbf{u}}^{[i]}$, $\overleftarrow{\mathbf{Q}}^{[i]}$ and $\overleftarrow{\mathbf{I}}^{[i]}$ are similarly defined.

In the reciprocal network, the total leakage interference can be written as

$$\overleftarrow{I}^{[j]} = P\overleftarrow{\mathbf{u}}^{[j]\dagger}\overleftarrow{\mathbf{Q}}^{[j]}\overleftarrow{\mathbf{u}}^{[j]}, \tag{4.5}$$

where $\overleftarrow{\mathbf{Q}}^{[j]} = \sum_{i=1,i\neq j}^{K} \overleftarrow{\mathbf{H}}^{[ji]}\overleftarrow{\mathbf{v}}^{[i]}\overleftarrow{\mathbf{v}}^{[i]\dagger}\overleftarrow{\mathbf{H}}^{[ji]\dagger}$. Similarly, in order to minimize the total leakage interference $\overleftarrow{I}^{[j]}$, $\overleftarrow{\mathbf{u}}^{[j]}$ is chosen as

$$\overleftarrow{\mathbf{u}}^{[j]} = \nu_{\min}[\overleftarrow{\mathbf{Q}}^{[j]}], \tag{4.6}$$

where $\nu_{\min}[\overleftarrow{\mathbf{Q}}^{[j]}]$ is the eigenvector corresponding to the smallest eigenvalue of $\overleftarrow{\mathbf{Q}}^{[j]}$. Again, setting $\mathbf{v}^{[i]}$ as the newly updated $\overleftarrow{\mathbf{u}}^{[i]}$ does not increase the total leakage interference because of the wireless reciprocity property.

Thus, the algorithm alternates between the original and reciprocal networks and because each step iteration only reduces the utility function, convergence of the algorithm

is guaranteed.

4.2.2 Max-SINR algorithm

The zero-forcing algorithm aims at minimizing the total leakage interference by iteratively updating the precoding and receiving filters. However, it is designed only to create some interference-free signal subspace to achieve interference alignment, while it does not take any desired signal power into consideration.

The suboptimality of the zero-forcing algorithm becomes non ignorable especially when our ultimate goal can be expressed as some utility function such as the throughput

$$R = d\log(1 + \frac{P_S}{P_I + P_N}), \tag{4.7}$$

where R is the rate, d is the DoF and P_S, P_I and P_N are signal, interference and noise power respectively. Although the zero-forcing algorithm indeed minimizes P_I, the rate R, which is directly related to SINR $= \frac{P_S}{P_I+P_N}$ is unnecessarily to be maximized at the same time. Recall (4.2), the SINR at D_j can be written as

$$\text{SINR}^{[j]} = P\frac{\mathbf{u}^{[j]\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}\mathbf{v}^{[j]\dagger}\mathbf{H}^{[jj]\dagger}\mathbf{u}^{[j]}}{\mathbf{u}^{[j]\dagger}(\mathbf{I} + P\sum_{i=1,i\neq j}^{K}\mathbf{H}^{[ji]}\mathbf{v}^{[i]}\mathbf{v}^{[i]\dagger}\mathbf{H}^{[ji]\dagger})\mathbf{u}^{[j]}}. \tag{4.8}$$

Thus, in order to maximize $\text{SINR}^{[j]}$, $\mathbf{u}^{[j]}$ is chosen as

$$\mathbf{u}^{[j]} = \frac{(\mathbf{I} + P\sum_{i=1,i\neq j}^{K}\mathbf{H}^{[ji]}\mathbf{v}^{[i]}\mathbf{v}^{[i]\dagger}\mathbf{H}^{[ji]\dagger})^{-1}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}}{||(\mathbf{I} + P\sum_{i=1,i\neq j}^{K}\mathbf{H}^{[ji]}\mathbf{v}^{[i]}\mathbf{v}^{[i]\dagger}\mathbf{H}^{[ji]\dagger})^{-1}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}||}. \tag{4.9}$$

The iterative process of the max-SINR algorithm is the same as that of the zero-forcing algorithm, except the process of updating the precoding filter by the receiving filter directly aims at the ultimate utility function. Hence, it performs better than the zero-forcing algorithm, especially in terms of the symbol error rate performance. Moreover, we will later show the maximizing process is crucial to achieve potential diversity benefit.

4.3 Feasibility condition for diversity interference alignment

In order to determine whether diversity interference alignment can be achieved, we view a diversity interference alignment scheme as a system of multivariate polynomial equations. Because each user sends only 1 DoF, each channel coefficient appears only once at the receiver side. Thus, the considered multivariate polynomial system is always generic. Moreover, it is known that a generic multivariate polynomial system has non-trivial solution(s) if and only if the number of equations does not exceed the number of variables [63]. So, in order to determine the feasibility condition for diversity interference alignment, our task becomes calculating and comparing the number of equations and the number of variables of the multivariate polynomial system corresponding to a diversity interference alignment scheme.

For interference alignment to be achieved, the precoding and receiving filters need to satisfy the following conditions:

$$\begin{aligned} \mathbf{u}^{[j]\dagger}\mathbf{H}^{[ji]}\mathbf{v}^{[i]} &= 0, \\ \mathrm{rank}(\mathbf{u}^{[i]\dagger}\mathbf{H}^{[ii]}\mathbf{v}^{[i]}) &= 1, \end{aligned} \tag{4.10}$$

where $1 \leqslant i, j \leqslant K, i \neq j$. It is easy to see (4.10) corresponds to $N_{e,\text{alignment}}$ equations which can be expressed as

$$N_{e,\text{alignment}} = K(K-1). \tag{4.11}$$

However, these conditions do not consider the desired channel matrices at all and thus can not guarantee any diversity benefit.

4.3.1 One-sided diversity

Firstly, let us consider only one-sided diversity, e.g., receiver-side diversity. In order to achieve receive diversity gain, it is well known that maximal ratio combining (MRC) filters are the optimal receiving filters to maximize the output SNR for single-stream information

without interference, which need to be designed as

$$\mathbf{u}^{[j]} = \frac{\mathbf{H}^{[jj]}\mathbf{v}^{[j]}}{||\mathbf{H}^{[jj]}\mathbf{v}^{[j]}||}, \tag{4.12}$$

or equivalently

$$\mathbf{u}^{[j]\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]} = ||\mathbf{H}^{[jj]}\mathbf{v}^{[j]}||, \tag{4.13}$$

for $1 \leqslant j \leqslant K$. Obviously, these MRC filters are completely determined by $\mathbf{H}^{[jj]}\mathbf{v}^{[j]}$ and independent of $\mathbf{H}^{[jk]}\mathbf{v}^{[k]}$, for $1 \leqslant j, k \leqslant K, j \neq k$. However, due to the channel randomness, conditions (4.13) are too harsh and such MRC filters will fail to satisfy (4.10) almost surely. Thus, MRC filters are not good choices to achieve diversity interference alignment.

Now, let us consider the following diversity conditions:

$$\begin{aligned} \mathbf{u}^{[j]} &= \frac{\mathbf{B}^{[j]}\mathbf{B}^{[j]\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}}{||\mathbf{B}^{[j]}\mathbf{B}^{[j]\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}||}, \\ \text{subject to SINR}^{[j]} &= \frac{P\mathbf{u}^{[j]\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}\mathbf{v}^{[j]\dagger}\mathbf{H}^{[jj]\dagger}\mathbf{u}^{[j]}}{\mathbf{u}^{[j]\dagger}(\mathbf{I} + P\sum_{i=1,i\neq j}^{K}\mathbf{H}^{[ji]}\mathbf{v}^{[i]}\mathbf{v}^{[i]\dagger}\mathbf{H}^{[ji]\dagger})\mathbf{u}^{[j]}} \text{ is maximized,} \end{aligned} \tag{4.14}$$

where $\mathbf{B}^{[j]}$ is a matrix with orthonormal columns which are independent of $\mathbf{H}^{[jj]}\mathbf{v}^{[j]}$, for $1 \leqslant j \leqslant K$.

To see why such a diversity condition gives a good comprise between multiplexing and diversity gains, let us consider a receiving filter $\mathbf{u}^{[j]}$ which maximizes $\text{SINR}^{[j]}$ such that

$$\mathbf{u}^{[j]} = arg\max_{\mathbf{u}} \frac{P\mathbf{u}^{\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}\mathbf{v}^{[j]\dagger}\mathbf{H}^{[jj]\dagger}\mathbf{u}}{\mathbf{u}^{\dagger}(\mathbf{I} + P\sum_{i=1,i\neq j}^{K}\mathbf{H}^{[ji]}\mathbf{v}^{[i]}\mathbf{v}^{[i]\dagger}\mathbf{H}^{[ji]\dagger})\mathbf{u}} \tag{4.15}$$

In the high SNR regime, where the diversity order is meaningful, such a receiving filter naturally leads to (4.10) to be satisfied because any $\mathbf{u}^{[j]}$ such that $\mathbf{u}^{[j]\dagger}\mathbf{H}^{[ji]}\mathbf{v}^{[i]} \neq 0$ has $\text{SINR}^{[j]}$ being upper-bounded by a constant, and thus must not be the optimum one whose corresponding $\text{SINR}^{[j]}$ grows linearly with P when $\mathbf{u}^{[j]\dagger}\mathbf{H}^{[ji]}\mathbf{v}^{[i]} = 0$.

With a matrix with orthonormal columns $\mathbf{B}^{[j]}$ acting as an agent rotation matrix, (4.15) can be equivalent written as (4.14). Comparing with (4.12), it is clear that (4.14) is

a more relaxed condition. This is because giving a receiving filter $\mathbf{u}^{[j]}$ that satisfies (4.10), although we can not guarantee (4.12) to be satisfied, we can however ensure (4.14) to be satisfied with a proper chosen matrix $\mathbf{B}^{[j]}$. The reason why we introduce the agent matrix $\mathbf{B}^{[j]}$ is because (4.15) is not analytical. In order to analyze the output SINR, we want to use $\mathbf{B}^{[j]}$ to produce an easier analytical formula.

It is worth to mention again that $\mathbf{B}^{[j]}$ is used as an agent rotation matrix to link the diversity and multiplexing requirements. The causal relationship is obtaining $\mathbf{u}^{[j]}$ first and then getting the implied $\mathbf{B}^{[j]}$, but not the other way round (which is to explicitly obtain $\mathbf{B}^{[j]}$ and then get $\mathbf{u}^{[j]}$). The reason why we are doing this is because the algorithms are numerical iterative methods which lead to intractable analytical complexity, and we want to use a single agent matrix $\mathbf{B}^{[j]}$ to simplify the analysis as shown in the later sections. Because of this, explicit calculation of $\mathbf{B}^{[j]}$ is not the focus. As long as the final SINR is expressed in some analytical form, we should be satisfied.

The maximizing condition here is crucial. For a MIMO interference network, there may be many combinations of precoding and receiving filters such that the alignment conditions (4.10) can be satisfied. However, among all the qualified combinations, only a subset of them such that the output $\text{SINR}^{[j]}$ at each receiver can be maximized. Denote this maximum SINR at each receiver as $(\lambda^{[j]})^2$, for $1 \leqslant j \leqslant K$, and the defined conditions (4.14) can be equivalently written as

$$\mathbf{u}^{[j]\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]} = |\lambda^{[j]}|. \tag{4.16}$$

for $1 \leqslant j \leqslant K$. These conditions will further impose $N_{e,\text{diversity}}$ equations which can be expressed as

$$N_{e,\text{diversity}} = K. \tag{4.17}$$

Combining (4.11) and (4.17), we know the multivariate polynomial system for the receive diversity interference alignment scheme satisfying (4.10) and (4.16) has N_e equations

which can be written as

$$N_e = N_{e,\text{alignment}} + N_{e,\text{diversity}} = K^2. \tag{4.18}$$

Now, we count the number of variables of this multivariate polynomial system. For a given set of channel matrices, $\lambda^{[j]}$ can be completely described by $\mathbf{H}^{[jk]}$ and $\mathbf{v}^{[k]}$, for $1 \leqslant j, k \leqslant K$, and can not be counted as an independent variable. Thus, all the variables of this multivariate polynomial system come from the precoding and receiving filters.

With M transmit antennas, the signal space of the precoding filters can be written as

$$\text{span}(\mathbf{v}^{[i]}) = \text{span}\left(\begin{pmatrix} v_1^{[i]} \\ v_2^{[i]} \\ v_3^{[i]} \\ \cdots \\ v_M^{[i]} \end{pmatrix} \right) = \text{span}\left(\begin{pmatrix} 1 \\ \frac{v_2^{[i]}}{v_1^{[i]}} \\ \frac{v_3^{[i]}}{v_1^{[i]}} \\ \cdots \\ \frac{v_M^{[i]}}{v_1^{[i]}} \end{pmatrix} \right), \tag{4.19}$$

for $1 \leqslant i \leqslant K$. So, after removing the effect of the superfluous variable, each precoding filter can provide $(M-1)$ variables. Similarly, with N receive antennas, each receiving filter can provide $(N-1)$ variables. Thus, the total number of variables N_v introduced by all precoding and receiving filters is

$$N_v = K(M-1+N-1) = K(M+N-2). \tag{4.20}$$

Comparing (4.18) and (4.20), we know the feasibility condition for the single-stream diversity interference alignment scheme satisfying (4.10) and (4.16) is

$$\begin{aligned} & N_v \geqslant N_e \\ \Rightarrow \quad & K(M+N-2) \geqslant K^2 \\ \Rightarrow \quad & M+N \geqslant K+2. \end{aligned} \tag{4.21}$$

This is indeed a necessary and sufficient condition for a single-stream interference

alignment scheme with diversity gain. The sufficiency is because when $M+N \geqslant K+2$, the multivariate polynomial system consisting (4.10) and (4.16) does have a solution and we will show in next section, this solution enjoys diversity gain. The necessity is because when $M+N \leqslant K+1$, $N_{e,\text{alignment}} \geqslant N_v$ and any more structure (corresponding more conditions to be satisfied) between $\mathbf{u}^{[j]}$ and $\mathbf{v}^{[j]}$ for potential diversity gain will make the corresponding multivariate polynomial system unsolvable (because the number of equations will exceed the number of variables).

4.3.2 Two-sided diversity

Secondly, let us consider both transmitter-side and receiver-side diversity. Again, in order to realize receive diversity gain, the optimal receiving filters need to be

$$\mathbf{u}^{[j]} = \frac{\mathbf{H}^{[jj]}\mathbf{v}^{[j]}}{||\mathbf{H}^{[jj]}\mathbf{v}^{[j]}||} = \alpha \mathbf{H}^{[jj]}\mathbf{v}^{[j]}, \tag{4.22}$$

where $1 \leqslant j \leqslant K$ and α is a constant in $\mathcal{R}$ that does not affect the diversity gain. Assume $\alpha = 1$, with (4.10) being satisfied, the output SNR is

$$\text{SNR}^{[j]} = \mathbf{v}^{[j]\dagger}\mathbf{H}^{[jj]\dagger}\mathbf{H}^{[jj]}\mathbf{v}^{[j]}. \tag{4.23}$$

In order to achieve maximum diversity gain with both transmit and receive diversity, the output SNR needs to be maximized and thus the precoding filters needs to satisfy

$$\mathbf{v}^{[j]} = \nu_{\max}(\mathbf{H}^{[jj]\dagger}\mathbf{H}^{[jj]}), \tag{4.24}$$

i.e., $\mathbf{v}^{[j]}$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{H}^{[jj]\dagger}\mathbf{H}^{[jj]}$. Thus, for maximum diversity gain to be achieved, the precoding and receiving filters are deterministic and only related to the desired channel matrices $\mathbf{H}^{[jj]}$, for $1 \leqslant j \leqslant K$. Thus, all previously considered variables in precoding and receiving filters become superfluous and $N_v = 0$. However, there are still $N_{e,\text{alignment}}$ equations need to be satisfied in order to realize interference alignment conditions. Because of the channel randomness, such $\mathbf{v}^{[j]}$ and $\mathbf{u}^{[j]}$ will fail to satisfy (4.10) with probability 1. So, a maximum diversity interference

alignment scheme is almost surely infeasible to be achieved.

Remark 3. *When $M+N \geqslant K+2$, single-stream diversity interference alignment satisfying* (4.10) *and* (4.16) *are feasible to be achieved simultaneously. Thus, one-sided diversity serves as a lower bound for the achievable diversity gain of a feasible single-stream diversity interference alignment system.*

Remark 4. *Maximal ratio transmission and single-stream interference alignment are not feasible to be achieved simultaneously. Two-sided diversity is naturally an upper bound for the achievable diversity gain of any interference alignment scheme.*

Remark 5. *Because all the variables come from the precoding and receiving filters, for a diversity interference alignment scheme to be feasible, the total number of equations of the corresponding multivariate polynomial system is constrained by the number of transmit and receive antennas. Thus, there is an obvious tradeoff between the level of interference alignment and the level of the diversity gain a strategy can achieve. The higher level of interference alignment is achieved, the more alignment conditions need to be satisfied, and thus the less freedom left to satisfy diversity conditions, and vise-versa.*

4.4 Diversity gains of different interference alignment strategies

We classify different interference alignment strategies into two categories called diversity interference alignment and zero-forcing interference alignment. Diversity interference alignment aims to simultaneously create some interference-free signal subspace at the receivers to satisfy (4.10) and choose appropriate $\mathbf{B}^{[j]}$ to maximize the output SINR to satisfy (4.16), while zero-forcing interference alignment aims solely to satisfy the interference alignment conditions in (4.10). Although these two types of solutions are not distinguishable in terms of the multiplexing gain, we show in this section their difference lies in the fact that they have different diversity gains. Typical diversity interference alignment solutions include max-SINR algorithm in [44] and MMSE algorithm in [62], while typical zero-forcing interference alignment solutions include zero-forcing algorithm in [44] and closed-form solution for three-user MIMO channels in [21]. In this thesis, we use a 3-

user (2×3) system to show diversity interference alignment schemes can achieve higher diversity order than zero-forcing interference alignment schemes. Similar results can be shown for other system settings using the same proof procedure by replacing the probability distribution for two-dimensional precoding filters with the probability distribution for higher-dimensional precoding filters.

Without loss of generality, we only focus on the first user. When $N = K = 3 \geqslant M = 2$, from (4.21), we know interference alignment with diversity gain is possible with receive diversity being preferred and it is feasible for a diversity interference alignment strategy to satisfy the following conditions simultaneously:

$$\begin{aligned} \mathbf{u}^{[1]\dagger}\mathbf{H}^{[1i]}\mathbf{v}^{[i]} &= 0, \\ \mathrm{rank}(\mathbf{u}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}) &= 1, \\ \mathbf{u}^{[1]} &= \frac{\mathbf{B}^{[1]}\mathbf{B}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}}{||\mathbf{B}^{[1]}\mathbf{B}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}||}. \end{aligned} \tag{4.25}$$

for $2 \leqslant i \leqslant 3$. Thus, the estimate of the transmitted symbol after the receiving filter can be written as

$$\begin{aligned} \hat{s}^{[1]} &= \mathbf{u}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}s^{[1]} + \mathbf{u}^{[1]\dagger}\sum_{i=2}^{3}\mathbf{H}^{[1i]}\mathbf{v}^{[i]}s^{[i]} + \tilde{n}^{[1]} \\ &= ||\mathbf{B}^{[1]}\mathbf{B}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}||s^{[1]} + \tilde{n}^{[1]}, \end{aligned} \tag{4.26}$$

where $\tilde{n}^{[1]} = \mathbf{u}^{[1]\dagger}\mathbf{n}^{[1]}$,

$$E(\tilde{n}^{[1]}) = E(\mathbf{u}^{[1]\dagger}\mathbf{n}^{[1]}) = E(\mathbf{u}^{[1]\dagger})E(\mathbf{n}^{[1]}) = 0, \tag{4.27}$$

and

$$\mathrm{Var}(\tilde{n}^{[1]}) = \mathbf{u}^{[1]\dagger}\mathbf{I}\mathbf{u}^{[1]} = 1. \tag{4.28}$$

From (4.26) and (4.28), we also know the output SINR at the first receiver is

$$\mathrm{SINR}^{[1]} = P||\mathbf{B}^{[1]}\mathbf{B}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}||^2 \tag{4.29}$$

Because $\mathbf{B}^{[1]}$ is an orthonormal matrix, we know

$$\text{SINR}^{[1]} = P|\mathbf{B}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}|^2. \tag{4.30}$$

Let

$$\begin{aligned}\tilde{\mathbf{h}}^{[1]} &= \mathbf{H}^{[11]}\mathbf{v}^{[1]} \\ &= \begin{pmatrix} h_{1,1}^{[11]} & h_{1,2}^{[11]} \\ h_{2,1}^{[11]} & h_{2,2}^{[11]} \\ h_{3,1}^{[11]} & h_{3,2}^{[11]} \end{pmatrix} \begin{pmatrix} v_1^{[1]} \\ v_2^{[1]} \end{pmatrix} = \begin{pmatrix} \tilde{h}_1^{[1]} \\ \tilde{h}_2^{[1]} \\ \tilde{h}_3^{[1]} \end{pmatrix},\end{aligned} \tag{4.31}$$

where $\tilde{h}_j^{[1]} = \sum_{i=1}^{2} h_{j,i}^{[11]} v_i^{[1]}$, for $1 \leqslant j \leqslant 3$.

Denote the probability distribution of each term in $\tilde{h}_j^{[1]}$ as μ-distributed for convenience. Now, we are ready to show approximating $\tilde{h}_j^{[1]}$ as a complex Gaussian distributed random variable does not change the diversity order. Because the sum of Gaussian random variables is also a Gaussian random variable, the proof is simplified to show approximating a μ-distributed random variable as a complex Gaussian distributed random variable does not change the diversity order. Taking $h_{1,1}^{[11]} v_1^{[1]}$ for instance, because the signal direction is unbiased and one does not have a desired signal space due to the channel randomness, the precoding filter $\mathbf{v}^{[1]}$ is isotropic, i.e., $p_\theta(\omega) = \frac{1}{2\pi}$, for $0 \leqslant \omega \leqslant 2\pi$. Moreover, from the geometric relation between $v_1^{[1]}$ and θ as shown in Fig. 4.2, we have

$$|v_1^{[1]}|^2 = |\cos\theta|^2, \quad \text{for} \quad 0 \leqslant \theta \leqslant 2\pi. \tag{4.32}$$

Note the inverse function of (4.32) is not monotonic and can be written as

$$\theta = \pm \arccos(\pm\sqrt{|v_1^{[1]}|^2}), \quad \text{for} \quad 0 \leqslant |v_1^{[1]}|^2 \leqslant 1. \tag{4.33}$$

Thus, the probability density function (PDF) of $|v_1^{[1]}|^2$ is

$$p_{|v_1^{[1]}|^2}(m) = \beta|\frac{d\theta}{dm}|p_\theta, \tag{4.34}$$

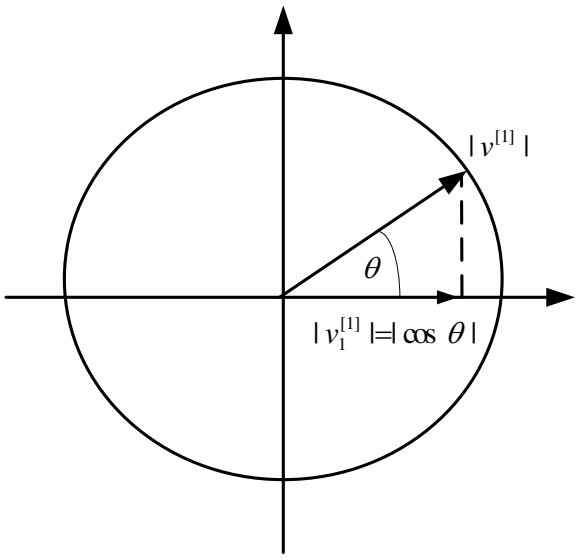


Figure 4.2: Geometric illustration of precoding filter $\mathbf{v}^{[1]}$.

where $\beta = 4$ is the number of solutions for (4.33). So,

$$\begin{aligned} p_{|v_1^{[1]}|^2}(m) &= 4\left|-\frac{1}{2\sqrt{m}\sqrt{1-m}}\right|\frac{1}{2\pi} \\ &= \frac{1}{\pi\sqrt{m}\sqrt{1-m}}, \quad \text{for} \quad 0 \leqslant m \leqslant 1. \end{aligned} \tag{4.35}$$

When $|v_1^{[1]}|^2 = 0$, we know $v_1^{[1]} = 0$ and $|h_{1,1}^{[11]} v_1^{[1]}|^2 = 0$. Since we are only interested in the near-zero behavior of $|h_{1,1}^{[11]} v_1^{[1]}|^2$, we exclude the event $|v_1^{[1]}|^2 = 0$ and only consider

$$p_{|v_1^{[1]}|^2}(m) = \frac{1}{\pi\sqrt{m}\sqrt{1-m}}, \quad \text{for} \quad \tau \leqslant m \leqslant 1, \tag{4.36}$$

where τ is an arbitrary small positive real number. This does affect the diversity order of $|h_{1,1}^{[11]} v_1^{[1]}|^2$ because whether considering the event $|v_1^{[1]}|^2 = 0$ only affects its zero-point PDF but not the near-zero points set. Moreover, the area under the PDF of a single point is essentially arbitrary small and this single-point event will eventually only happen with arbitrary small probability.

Since $h^{[11]}$ is standard complex Gaussian distributed, its instantaneous power is

exponential distributed whose PDF can be written as

$$p_{|h_{1,1}^{[11]}|^2}(n) = \frac{1}{2}e^{-\frac{n}{2}}, 0 \leqslant n. \tag{4.37}$$

Thus, the PDF of the instantaneous power of $h_{1,1}^{[11]}v_1^{[1]}$ is

$$\begin{aligned} p_{|h_{1,1}^{[11]}v_1^{[1]}|^2}(x) &= \int_{\mathcal{A}:|h_{1,1}^{[11]}|^2|v_1^{[1]}|^2=x} p_{(|h_{1,1}^{[11]}|^2,|v_1^{[1]}|^2)}(x)d\mathcal{A} \\ &= \int_{\tau}^{1} p_{|h_{1,1}^{[11]}|^2}(\frac{x}{m})p_{|v_1^{[1]}|^2}(m)dm \\ &= \int_{\tau}^{1} \frac{1}{2}e^{-\frac{x}{2m}}\frac{1}{\pi\sqrt{m}\sqrt{1-m}}dm. \end{aligned} \tag{4.38}$$

Because $d\arcsin(\sqrt{m}) = \frac{1}{2\sqrt{m}\sqrt{1-m}}dm$, (4.38) can be further written as

$$\begin{aligned} p_{|h_{1,1}^{[11]}v_1^{[1]}|^2}(x) &= \int_{\tau}^{1} \frac{1}{\pi}e^{-\frac{x}{2m}}d\arcsin(\sqrt{m}) \\ &= [\frac{1}{\pi}e^{-\frac{x}{2m}}\arcsin(\sqrt{m})]_{\tau}^{1} - \int_{\tau}^{1}\frac{1}{\pi}\arcsin(\sqrt{m})de^{-\frac{x}{2m}} \\ &= [\frac{1}{\pi}e^{-\frac{x}{2m}}\arcsin(\sqrt{m})]_{\tau}^{1} - \int_{\tau}^{1}\frac{1}{\pi}\arcsin(\sqrt{m})e^{-\frac{x}{2m}}\frac{x}{2m^2}dm. \end{aligned} \tag{4.39}$$

When $x \to 0^+$, the first term of (4.39) can be written as

$$\begin{aligned} \lim_{x\to 0^+}[\frac{1}{\pi}e^{-\frac{x}{2m}}\arcsin(\sqrt{m})]_{\tau}^{1} &= \lim_{x\to 0^+}\frac{1}{\pi}[e^{-\frac{x}{2}}\frac{\pi}{2} - e^{-\frac{x}{2\tau}}\arcsin(\sqrt{\tau})] \\ &= \frac{1}{2} - \frac{1}{\pi}\arcsin(\sqrt{\tau}) \\ &\doteq \frac{1}{2}x^0e^{-\frac{x}{2}} + o(x^0), \end{aligned} \tag{4.40}$$

where we define a function $f(x)$ of x as $o(x)$ if $\lim_{x\to 0^+}\frac{f(x)}{x} = 0$, i.e., $f(x)$ is of lower order than x as $x \to 0^+$. Similarly, when $x \to 0^+$, the second term of (4.39) can be expressed as

$$\lim_{x\to 0^+}\int_{\tau}^{1}\frac{1}{\pi}\arcsin(\sqrt{m})e^{-\frac{x}{2m}}\frac{x}{2m^2}dm = 0 \doteq o(x^0). \tag{4.41}$$

Thus, the near-zero behavior of PDF for $|h_{1,1}^{[11]}v_1^{[1]}|^2$ is

$$\lim_{x\to 0^+} p_{|h_{1,1}^{[11]}v_1^{[1]}|^2}(x) = \frac{1}{2}x^0 e^{-\frac{x}{2}} + o(x^0). \tag{4.42}$$

From [64], we know the diversity order of a channel fading coefficient depends only on the near-zero behavior of the PDF of its instantaneous power. Thus, the diversity order of the μ-distributed random variable $h_{1,1}^{[11]}v_1^{[1]}$ is $1 = 0 + 1$, which is the diversity order of a complex Gaussian random variable. So, approximating the μ-distributed random variables as complex Gaussian distributed random variables does not change the diversity order. Consequently, each term in $\tilde{\mathbf{h}}^{[1]}$ can be well approximated as complex Gaussian distributed without changing the diversity order. Thus, $||\tilde{\mathbf{h}}^{[1]}||^2 = \sum_{j=1}^{3}|\tilde{h}_j^{[1]}|^2$ is χ^2-distributed with $6 = 2 \times 3$ degrees of freedom whose PDF can be written in the form

$$p_{||\tilde{\mathbf{h}}^{[1]}||^2}(x) = x^2 e^{-\frac{x}{2}}. \tag{4.43}$$

From (4.30), we know

$$\begin{aligned} \mathrm{SINR}^{[1]} &= P|\mathbf{B}^{[1]\dagger}\mathbf{H}^{[11]}\mathbf{v}^{[1]}|^2 \\ &= P||\mathbf{B}^{[1]}||^2||\tilde{\mathbf{h}}^{[1]}||^2|\cos\phi^{[1]}|^2 \\ &= P||\tilde{\mathbf{h}}^{[1]}||^2|\cos\phi^{[1]}|^2, \end{aligned} \tag{4.44}$$

where $\phi^{[1]}$ is the angle between $\mathbf{B}^{[1]}$ and $\tilde{\mathbf{h}}^{[1]}$.

Now, we come to the point where the maximizing conditions make significant difference. Firstly, let us make clear that for feasible diversity interference alignment systems, i.e., where $M + N \geqslant K + 2$, because $N_v > N_{e,\text{alignment}}$, thus without considering the diversity conditions (4.14), there may be more than one solution satisfying (4.10). This means if we fix the desired channel matrices and only allow the interference channel matrices to change randomly, each instant set of interference channel matrices may output several qualified combinations of $\mathbf{v}^{[1]}$ and $\mathbf{u}^{[1]}$, with each combination having a unique $\phi^{[1]}$ and corresponding $|\cos\phi^{[1]}|$.

Think the PDF of $\phi^{[1]}$ at a specific angle $0 \leqslant \eta \leqslant \pi$ as the sum of that at many

discrete instants of channel matrices. Thus, we have

$$p_{\phi^{[1]}}(\eta) = \sum_{\mathrm{H}^{jk}} p_{\phi^{[1]}|\mathrm{H}^{jk}}(\eta) p_{\mathrm{H}^{jk}} \tag{4.45}$$

for $1 \leqslant j, k \leqslant K$. Although we do not know the exact PDF of each term in (4.45), the channel randomness means the number of different channel instants is arbitrary large. Thus, we can well approximate the PDF of $\phi^{[1]}$ as Gaussian distributed within a meaningful interval $[a, b]$, for $0 \leqslant a \leqslant b \leqslant \pi$.

For zero-forcing interference alignment solutions, there is no preference among different combinations of $\mathbf{v}^{[1]}$ and $\mathbf{u}^{[1]}$ that satisfying (4.10). Thus, all solutions to (4.10) are qualified to be the precoding and receiving filters. Hence, due to the channel randomness, $\mathbf{v}^{[j]}$ and $\mathbf{u}^{[j]}$ are independent of $\mathbf{H}^{[ji]}$ and $E(\mathbf{u}^{[j]\dagger}\mathbf{H}^{[ji]}\mathbf{v}^{[j]}) = 0$ which means $E(\phi^{[1]}) = \frac{\pi}{2}$. Thus, for zero-forcing interference alignment algorithm, we can approximate the PDF of $\phi^{[1]}$ as

$$p_{\phi^{[1]}}(\eta) = e^{-\frac{(\eta-\frac{\pi}{2})^2}{2}}, \quad \text{for} \quad 0 < a \leqslant \eta \leqslant b < \pi. \tag{4.46}$$

From (4.46), it is easy to see zero-forcing interference alignment solutions are not good in terms of diversity gain. This is because the most probable event is $\phi^{[1]} = \frac{\pi}{2}$, which corresponds to $\mathrm{SINR}^{[1]} = 0$. Moreover, from (4.35), we know the PDF of $|\cos\phi^{[1]}|^2$ can be written as

$$p_{|\cos\phi^{[1]}|^2}(m) = \frac{1}{\pi\sqrt{m}\sqrt{1-m}} e^{-\frac{(\arccos\sqrt{m}-\frac{\pi}{2})^2}{2}}, \quad \text{for} \quad \epsilon \leqslant m \leqslant 1. \tag{4.47}$$

where $\epsilon = \min[\cos(a), \cos(b)]$. Thus, the PDF of $||\tilde{\mathbf{h}}^{[1]}||^2|\cos\phi^{[1]}|^2$ is

$$\begin{aligned} p_{||\tilde{\mathrm{h}}^{[1]}||^2|\cos\phi^{[1]}|^2}(x) &= \int_{\mathcal{A}:||\tilde{\mathrm{h}}^{[1]}||^2|\cos\phi^{[1]}|^2=x} p_{(||\tilde{\mathrm{h}}^{[1]}||^2,|\cos\phi^{[1]}|^2)}(x)d\mathcal{A} \\ &= \int_{\epsilon}^{1} p_{||\tilde{\mathrm{h}}^{[1]}||^2}(\frac{x}{m}) p_{|\cos\phi^{[1]}|^2}(m)dm \\ &= \int_{\epsilon}^{1} \frac{x^2}{m^2} e^{-\frac{x}{2m}} \frac{1}{\pi\sqrt{m}\sqrt{1-m}} e^{-\frac{(\arccos\sqrt{m}-\frac{\pi}{2})^2}{2}} dm. \end{aligned} \tag{4.48}$$

Using similar techniques as in (4.39)(4.40)(4.41), we can show $0 <$

$\lim_{x\to 0^+} p_{||\tilde{\mathbf{h}}^{[1]}||^2|\cos\phi^{[1]}|^2}(x)$, which means

$$\lim_{x\to 0^+} p_{||\tilde{\mathbf{h}}^{[1]}||^2|\cos\phi^{[1]}|^2}(x) \dot{\geqslant} \gamma x^0 + o(x^0), \tag{4.49}$$

where γ is a constant that does not affect the diversity order. This proves the zero-forcing interference alignment solutions do not provide any diversity gain.

For diversity interference alignment solutions, one does have preference among all the possible solutions satisfying (4.10). In particular, a diversity interference alignment solution chooses only a subset of the many possibilities, such that the output SINR is locally maximized and then eliminates the other bad combinations which generate low output SINR. In other words, when there are multiple qualified outputs, the combinations with $\phi^{[1]}$ close to $\frac{\pi}{2}$ are always eliminated. This can be easily seen the diversity conditions (4.14) or (4.15), where any bad combination will be eliminated because it does not give the locally maximum SINR. Thus, $E(\mathbf{u}^{[j]\dagger}\mathbf{H}^{[ji]}\mathbf{v}^{[i]}) > 0$ and $E(\phi^{[1]}) < \frac{\pi}{2}$. With the fact that the PDF of $\phi^{[1]}$ of diversity interference alignment solutions can also be approximated as Gaussian distributed, we can write

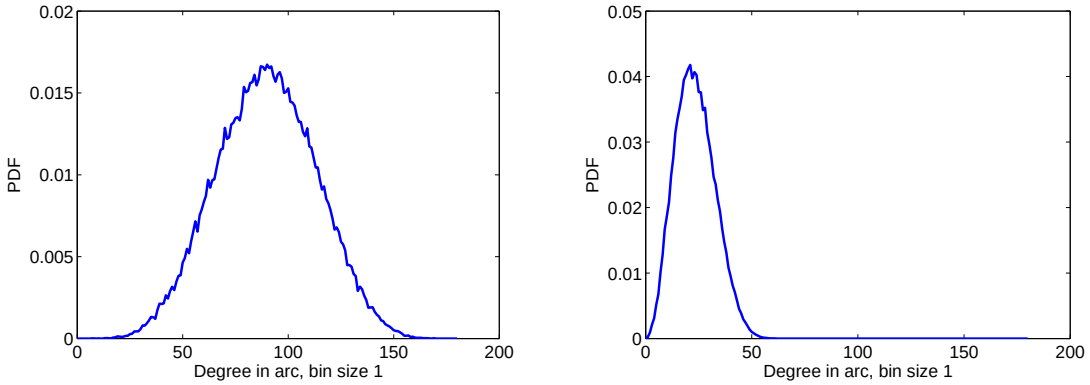
$$p_{\phi^{[1]}}(\eta) = e^{-\frac{(\eta-\psi)^2}{2}}, \quad \text{for} \quad 0 < a \leqslant \eta \leqslant b < \frac{\pi}{2}. \tag{4.50}$$

where $0 \leqslant \psi \leqslant \frac{\pi}{2}$. Let $\epsilon = \min[\cos a, \cos b]$, where ϵ is a constant only related to the channel fading characteristics. From (4.44), we know

$$\text{SINR}^{[1]} \geqslant ||\tilde{\mathbf{h}}^{[1]}||^2\epsilon. \tag{4.51}$$

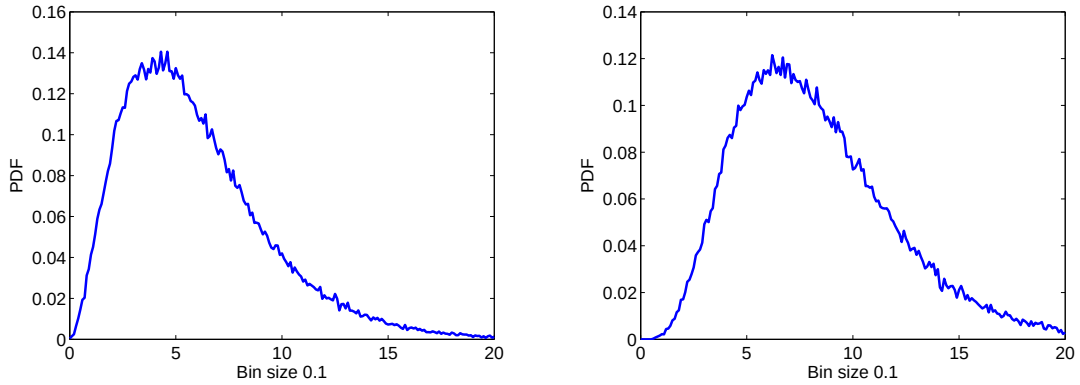
Because the constant ϵ only affects the SNR shift but not the diversity gain, we can conclude with diversity interference alignment solutions, the output PDF of the output SINR is at least χ^2-distributed with 6 degrees of freedom. From the results in [64], it is easy to see that for such an instantaneous SINR PDF, the largest diversity order at a fixed rate is 3.

In order to support the previous theoretical analysis, we present Fig. 4.3 to show the PDF of the angle $\phi^{[j]}$ between $\mathbf{B}^{[j]}$ and $\mathbf{H}^{[ji]}\mathbf{v}^{[i]}$. Moreover, Fig. 4.4 and Fig. 4.5



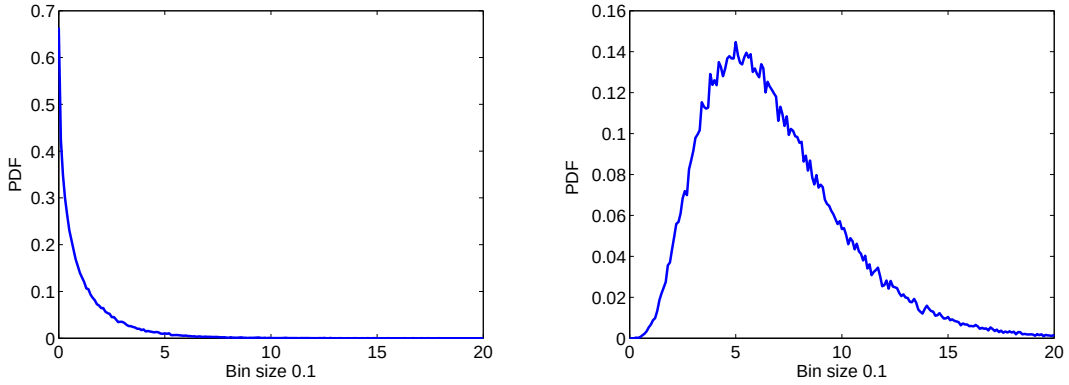
(a) Zero-forcing algorithm (b) Max-SINR algorithm

Figure 4.3: PDF of the angle $\phi^{[j]}$ between $\mathbf{B}^{[j]}$ and $\mathbf{H}^{[ij]}\mathbf{v}^{[j]}$.



(a) Zero-forcing algorithm (b) Max-SINR algorithm

Figure 4.4: PDF of the received signal power $|\mathbf{H}^{[ij]}\mathbf{v}^{[j]}|^2$ before applying the receiving filters.



(a) Zero-forcing algorithm (b) Max-SINR algorithm

Figure 4.5: PDF of the received signal power $|\mathbf{u}^{[j]\dagger}\mathbf{H}^{[ij]}\mathbf{v}^{[j]}|^2$ after the receiving filters.

illustrate the PDFs of the received signal power before and after applying the receiving filters respectively. All these results are obtained by practical simulations.

Note that this is for the case of $N = K = 3 \geqslant M = 2$ with receive diversity gain being preferred. When $M = K = 3 \geqslant N = 2$, transmit diversity gain is preferred and the achievable diversity order is at least 3. Thus, the achievable diversity order of diversity interference alignment solutions satisfying (4.10) and (4.16) is at least 3, for 3-user (2×3) or 3-user (3×2) systems.

To conclude this part, it is worth mentioning why an infeasible diversity interference alignment system, i.e., where $M+K \leqslant K+2$, does not promise to give any diversity gain. Taking a 3-user (2×2) system for instance. For such an interference channel, since $N_v = N_{e,\text{alignment}}$, one is already struggling to generate the combination of precoding and receiving filters to satisfy the interference alignment conditions (4.10) solely. Thus, after finding such $\mathbf{v}^{[j]}$ and $\mathbf{u}^{[j]}$ to satisfy the diversity conditions (4.10), one does not have any extra freedom to choose appropriate $\mathbf{B}^{[j]}$ to rotate the signals in order to satisfy (4.16). In other words, for a 3-user (2×2) system, the angle ϕ is determined by the interference channel matrices but not the desired channel matrices. This means given a set of interference channel matrices, the relation between $\mathbf{v}^{[j]}$ and $\mathbf{u}^{[j]}$ are fixed no matter what the desired channel matrices are, which further implies we lose the randomness (diversity) of the desired channel matrices.

4.5 Simulations

In this section, we use the max-SINR and zero-forcing algorithms proposed in [44] as examples of diversity and zero-forcing interference alignment strategies to evaluate their symbol error rate performance. In order to verify the analytical results, we use fixed rate uncoded BPSK modulation for both zero-forcing and max-SINR algorithms. For clearer illustration, the BER performance of two base schemes, i.e., 1×1 direct transmission with diversity gain 1 and 2×1 Alamouti scheme with diversity gain 2, are also presented.

Example 1. *The first example we use is to illustrate the diversity gains of infeasible diversity interference alignment systems, i.e., systems where interference alignment and*

diversity gains can not be achieved simultaneously.

For a 3*-user* (2×2) *system, we have* $M = 2$*,* $N = 2$*,* $K = 3$ *and* $M + N = K + 1$*. Thus, from* (4.21)*, we know such a system does not satisfy diversity interference alignment conditions and is only able to achieve interference alignment solely without any diversity benefit. Simulation results in Fig. 4.6 verify our conjecture and shows that although max-SINR algorithm performs better than zero-forcing algorithm in every SNR value, they exhibit the same diversity order of* 1*.*

Example 2. *The second example we use is to illustrate the diversity benefit that feasible diversity interference alignment systems can offer.*

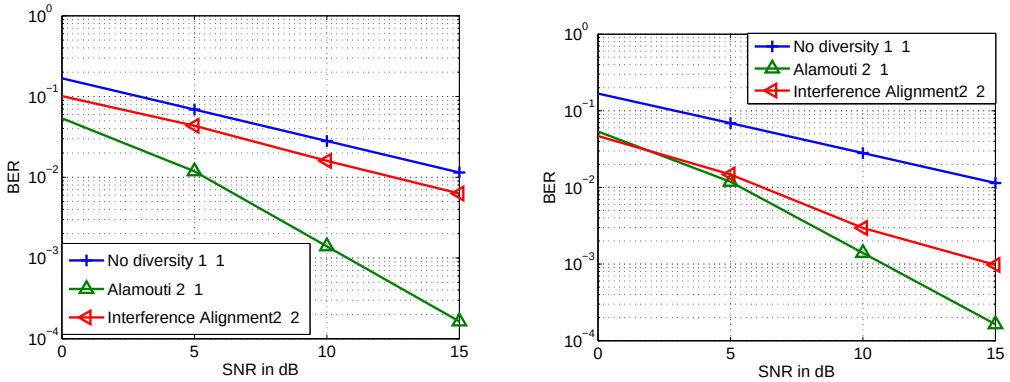
For a 3*-user* (2×3) *system, we have* $M = 2$*,* $N = 2$*,* $K = 3$ *and* $M + N = K + 2$*. Thus, from our analysis, diversity interference alignment is feasible. Similarly, a* 3*-user* (2×3) *system with* $M = 2$*,* $N = 2$*,* $K = 3$ *and* $M + N = K + 2$ *is also feasible to achieve diversity interference alignment. So, for such systems, interference alignment and diversity gains can be obtained simultaneously as can be clearly seen in Fig. 4.7.*

Example 3. *The third example we use is to illustrate the difference between the diversity gain offered by a diversity interference alignment strategy and that of a zero-forcing interference alignment strategy.*

For feasible diversity interference alignment system settings, e.g., 3*-user* (2×3) *and* 3*-user* (2×3) *interference channels, the multiplexing gains provided by diversity interference alignment and zero-forcing interference alignment strategies are the same. We have shown that for such systems, a diversity interference alignment strategy can provide higher achievable diversity gain than a zero-forcing interference alignment scheme. Simulation results in Fig. 4.7 and Fig. 4.8 demonstrate such superior of diversity interference alignment strategies in terms of the symbol error rate performance or diversity gain.*

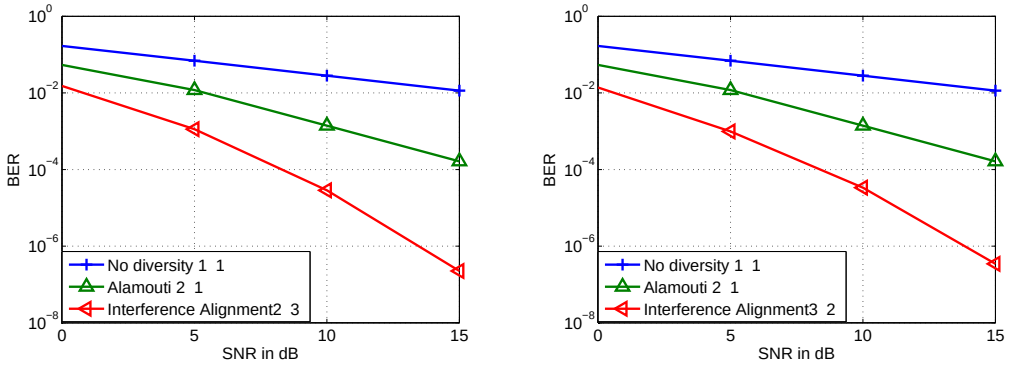
4.6 Conclusion and remarks

In this chapter, we analyzed the feasibility conditions for diversity interference alignment and the diversity gains that different interference alignment strategies can achieve. While almost all previous work about interference alignment were aiming at either showing higher multiplexing gain can be achieved by interference alignment or developing algo-



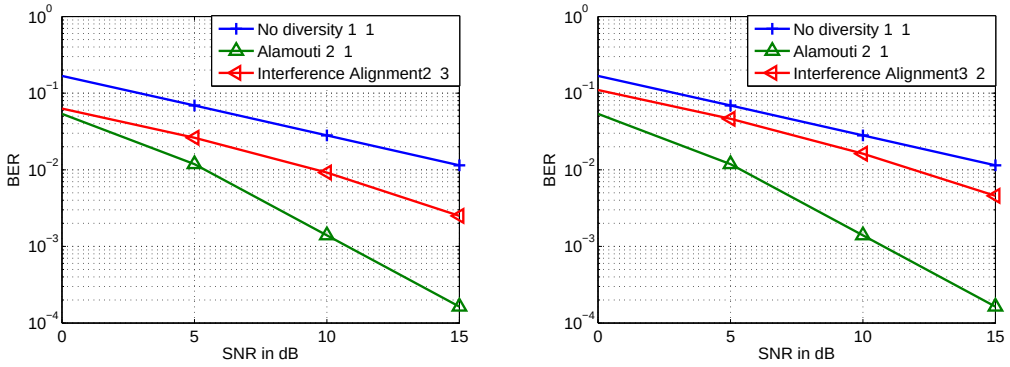
(a) Zero-forcing algorithm (b) Max-SINR algorithm

Figure 4.6: BER performance of different interference alignment strategies for a 3-user 2×2 interference channel.



(a) 3-user 2×3 interference channel. (b) 3-user 3×2 interference channel.

Figure 4.7: Feasible diversity interference alignment systems with diversity interference alignment strategies.



(a) Zero-forcing interference alignment solution. (b) Zero-forcing interference alignment solution.

Figure 4.8: Feasible diversity interference alignment systems with zero-forcing interference alignment strategies.

rithms to achieve practical alignment. We have proved that interference alignment can at the same time provide diversity gain under certain conditions, which was not observed before. Thus, with both multiplexing and diversity gains, diversity interference alignment truly becomes the preferable transmission strategy for CIC in terms of a wide range of designing criteria. Moreover, although the analysis in this chapter is derived for interference alignment strategies without any symbol extension, the results are applicable to a wide range of interference networks due to recent development in [65] which shows exact interference alignment is always feasible for small systems with only 2 symbol extensions.

To conclude this chapter, two important remarks can be made:

1. For multi-user interference channels, an optimal precoding and receiving filters designing method needs to take both desired and interference channel matrices into consideration. Besides conventional interference alignment conditions, the precoding or receiving filters should preserve certain structures for the equivalent combined channels in order to simultaneously achieve interference alignment and realize the potential diversity gain. This structure will impose extra conditions to be satisfied and thus the tradeoff between the level of interference alignment and the level of achievable diversity gain is obvious.

2. For a 3-user (3×2) interference channel, [66] conjectured that the maximum achievable diversity gain is only 1 when interference alignment is achieved. That result was based on the assumption that separation of interference alignment precoding filters design and space-time codes design is optimal. Although we don't claim the precoding filters and space-time codes should always be designed jointly to achieve maximum diversity gain, we use counter examples together with simulation verifications to show the diversity gain is unnecessarily restricted by applying the precoding and receiving filters, i.e., can be higher than 1 for a 3-user (3×2) interference channel. The achievable scheme we use is the max-SINR algorithm which is one of the diversity interference alignment strategies.

The results in this chapter give insights on our research work in the next chapter. Briefly speaking, because of the tradeoff we have observed between the achievable

diversity gain and interference alignment, we can design a strategy to increase the network diversity for CIC first and then modify it in order to trade the diversity benefit for the multiplexing gain improvement to achieve the DoF upper bound promised by interference alignment.

Chapter 5

Interference Alignment with Phase Randomization

5.1 Introduction

Although interference alignment mainly focused on improving the multiplexing gain of CIC, the other important asset a system possess is the diversity gain. In conventional point-to-point MIMO channels, it has been proved that there is a fundamental tradeoff between the achievable multiplexing and diversity gains of a communication system [3]. Similarly, in network level transmission strategy designs, one can also purposely tradeoff one asset for the other in order to maximize the desired network performance [9].

In Chapter 2, we have shown that a $(M \times N)^K$ CIC is feasible to achieve interference alignment without symbol extension only if

$$M + N \geqslant (K + 1)d. \tag{5.1}$$

It is easy to see that the reason for the infeasible interference alignment system is because there is not enough diversity, i.e., M and N are too small[1]. Similar problem exists for the interference alignment schemes with naive symbol extension [21]. In those schemes, simple symbol extension generates sparse channel matrix with only diagonal or block diagonal elements. As a result, the scheme in [21] is only able to asymptotically

[1]The maximum single-user diversity gain increases as $M + N$ increases because $MN \leqslant \frac{(M+N)^2}{4}$.

Figure 5.1: An indirect approach to achieve interference alignment via varying signal power to provide more diversity.

achieve the DoF upper bound with infinitely large symbol extension, i.e., when diversity is high enough. The diversity insufficiency problem becomes even more catastrophic for deterministic interference networks with constant channel coefficients. In such deterministic cases, symbol extension does not even asymptotically achieve the DoF upper bound because the extension itself does not increase diversity.

Inspired by the research about DMT and our insights from the last chapter, we adopt an indirect approach to obtain the DoF benefit offered by interference alignment as shown in Fig. 5.1. Firstly, we trade signal power for diversity improvement by using noncoherent transmissions with random phase offsets at both transmitters and receivers, which is done by distinctly scaled signal between each transmitter-receiver pair. While diversity is not our ultimate goal, we then further trade the increased system diversity for multiplexing improvement in order to achieve the DoF upper bound promised by interference alignment.

5.1.1 System model

We focus on $(1 \times 1)^K$ interference networks with each transmitter or receiver having only 1 antenna. $\mathbf{H}^{[ji]}$ is used to denote the channel matrix between the i-th transmitter S_i and the j-th receiver D_j after symbol extension, for $1 \leqslant i, j \leqslant K$. The diagonal elements in $\mathbf{H}^{[ji]}$ are independent real Gaussian distributed scalars for Gaussian interference networks or real constant scalars for deterministic interference networks. In this chapter, we mainly consider real Gaussian channel coefficients if not particularly specified, while the real deterministic cases will be discussed separately.

$\mathbf{v}^{[i]}$ is used to denote the precoding matrix at S_i and $\mathbf{u}^{[j]}$ is used to denote the receiving matrix at D_j, for $1 \leqslant i, j \leqslant K$. $s^{[i]}$ and $\hat{s}^{[i]}$ are used to denote the baseband

precoded symbols to be transmitted by S_i with short-term power constraint P and the symbols estimated by D_i, for $1 \leqslant i \leqslant K$. The additive noise at D_j is denoted as $\mathbf{n}^{[j]}$ and assumed to be Gaussian distributed with zero mean and covariance matrix $\sum_{\mathbf{n}^{[j]}} = \mathbf{I}_{N\times N}$, for $1 \leqslant j \leqslant K$. $\mathbf{x}^{[i]}$ and $\mathbf{y}^{[j]}$ are used to denote the bandpass transmitted signal at S_i and the bandpass received signal at D_j respectively, for $1 \leqslant i, j \leqslant K$.

We assume each transmitter or receiver uses advanced coding or decoding techniques in order to approach the Shannon limit. Moreover, pulse amplitude modulation (PAM) with sine wave being the carrier signal is adopted as the modulation scheme. For single-antenna interference networks, bandpass representation of the transmitted signal at S_i can be written as

$$\begin{aligned} x^{[i]} &= \mathrm{Re}\{s^{[i]} e^{j2\pi f_c t + \theta^{[i]}}\} \\ &= s^{[i]} \cos(2\pi f_c t + \theta^{[i]}), \end{aligned} \tag{5.2}$$

where f_c is the carrier frequency and $\theta^{[i]}$ is the random phase offset brought in by S_i. Similarly, the received signal (ignoring noise) from S_i to D_j can be written as

$$y^{[ji]} = h^{[ji]} s^{[i]} \cos(2\pi f_c t + \theta^{[i]}). \tag{5.3}$$

At D_j, a demodulator f with random phase offset $\varphi^{[j]}$ processes the received signal from S_i as

$$\begin{aligned} & f_{\varphi}^{[j]}(y^{[ji]}) \\ = \; & 2\int_0^T y^{[ji]} \cos(2\pi f_c t + \varphi^{[j]}) dt \\ = \; & 2\int_0^T h^{[ji]} s^{[i]} \cos(2\pi f_c t + \theta^{[i]}) \cos(2\pi f_c t + \varphi^{[j]}) dt \\ = \; & \cos(\varphi^{[j]} - \theta^{[i]}) h^{[ji]} s^{[i]}, \end{aligned} \tag{5.4}$$

where T is a whole symbol interval.

Conventionally, a coherent demodulator should track the phase change and reproduce the carrier signal such that $\theta^{[i]}$ and $\varphi^{[j]}$ are as close as possible. We will later show

that it is these intentionally chosen noncoherent random phase offsets that provides us more diversity. Roughly speaking, if the transmitter and receiver phase offsets between two antennas happen to be close to each other, the channel gain between them is large; On the other hand, if they happen to be far away, the channel gain between them is small.

5.2 The diversity insufficiency problem

From (5.1), we know that interference alignment is not feasible without symbol extension for some network configurations because the diversity is not large enough to be traded for multiplexing improvement to achieve the DoF upper bound. Thus, it is widely conjectured that symbol extension, either in time or frequency domain, is needed to bridge the information theoretically powerful interference alignment to practical applications. In particular, symbol extension is mostly desired in the following two common scenarios:

1. In cases that the feasibility condition (1.5) for a MIMO interference network is not satisfied, it is natural to consider symbol extension (either in time or frequency domain) to increase M and/or N.

2. For single-antenna interference networks, one may want to use symbol extension to increase M and N (and thus the resultant equivalent channel matrices are MIMO) so that MIMO interference alignment schemes can be applied to achieve the DoF upper bound exactly.

5.2.1 Problems of naive symbol extension

Although conceptually simple, there are certain limitations so that naive symbol extension themselves are not able to resolve all the problems:

1. Naive symbol extension increase $M + N$ and d proportionally, and thus an original infeasible system remains infeasible.

2. More importantly, the equivalent MIMO channel matrices after naive symbol extension possess a diagonal or block diagonal structure such that conventional MIMO

interference alignment schemes are not feasible to produce proper interference alignment precoding and receiving matrices.

The first problem is straightforward from (5.1). To better illustrate the second problem, let us consider a $(1 \times 1)^3$ interference network. With 2 symbol extension in the time domain, the equivalent channel matrix between S_i and D_j becomes

$$\mathbf{H}^{[ji]} = \begin{bmatrix} h_1^{[ji]} & 0 \\ 0 & h_2^{[ji]} \end{bmatrix}, \text{ for } 1 \leqslant i, j \leqslant 3, \tag{5.5}$$

where $h_\tau^{[ji]}$ is used to denote the channel coefficient value between S_i and D_j at time instant τ, for $1 \leqslant \tau \leqslant 2$. Since the received signal at each receiver is a 2×1 vector, the interference alignment conditions can be written as

$$\mathbf{H}^{[12]}\mathbf{v}^{[2]} = \alpha\mathbf{H}^{[13]}\mathbf{v}^{[3]}, \tag{5.6}$$

$$\mathbf{H}^{[21]}\mathbf{v}^{[1]} = \beta\mathbf{H}^{[23]}\mathbf{v}^{[3]}, \tag{5.7}$$

$$\mathbf{H}^{[31]}\mathbf{v}^{[1]} = \gamma\mathbf{H}^{[32]}\mathbf{v}^{[2]}, \tag{5.8}$$

where α, β and γ are three scalers. From (5.6)-(5.8), it is easy to see:

$$\mathbf{v}^{[2]} = (\mathbf{H}^{[12]})^{-1}\alpha\mathbf{H}^{[13]}\mathbf{v}^{[3]}, \tag{5.9}$$

$$\mathbf{v}^{[3]} = (\beta\mathbf{H}^{[23]})^{-1}\mathbf{H}^{[21]}\mathbf{v}^{[1]}, \tag{5.10}$$

$$\mathbf{v}^{[1]} = (\mathbf{H}^{[31]})^{-1}\gamma\mathbf{H}^{[32]}\mathbf{v}^{[2]}. \tag{5.11}$$

Thus, in order to align all interference at each receiver into the same signal subspace, the precoding matrices need to be designed as

$$\mathbf{v}^{[1]} = \mathbf{E}\mathbf{v}^{[1]}, \tag{5.12}$$

$$\mathbf{v}^{[2]} = \mathbf{F}\mathbf{v}^{[1]}, \tag{5.13}$$

$$\mathbf{v}^{[3]} = \mathbf{G}\mathbf{v}^{[1]}, \tag{5.14}$$

where

$$\mathbf{E} = (\mathbf{H}^{[31]})^{-1}\gamma\mathbf{H}^{[32]}(\mathbf{H}^{[12]})^{-1}\alpha\mathbf{H}^{[13]}(\beta\mathbf{H}^{[23]})^{-1}\mathbf{H}^{[21]}, \tag{5.15}$$

$$\mathbf{F} = (\gamma\mathbf{H}^{[32]})^{-1}\mathbf{H}^{[31]}, \tag{5.16}$$

$$\mathbf{G} = (\beta\mathbf{H}^{[23]})^{-1}\mathbf{H}^{[21]}. \tag{5.17}$$

Therefore, $\mathbf{v}^{[1]}$ must be a linearly scaled version of an eigenvector of $\mathbf{E}$. Since $\mathbf{E}$ is a diagonal matrix, its eigenvectors are $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$. Whichever eigenvector $\mathbf{v}^{[1]}$ is related to, one of its entries is 0, i.e., S_1 is silent in one of the two time instants. So do S_2 and S_3 because of the diagonal structure of $\mathbf{F}$ and $\mathbf{G}$.

Thus, with diagonal equivalent channel matrices through naive 2 symbol extension, every transmitter transmits in one time instant and keeps silent in the other time instant. Consequently, every receiver receives superpositioned desired and interference signal in one time instant and receive nothing but noise in the other time instant. Although all interference is aligned into the same signal subspace, the desired signal is in the same signal subspace and inseparable from the interference.

5.2.2 The insufficiency of coherent demodulation

From the last section, we see that the sparse diagonal channel matrix is infeasible to achieve interference alignment because of the lack of diversity. One technique to resolve the diagonality problem is to add some scaled versions of the received signal across several symbol extension together, in order to artificially generate the non-diagonal terms for the equivalent MIMO channel matrices. However, such operations will scale the desired and all interference signal equally so that although individual channel matrix is not diagonal, $\mathbf{E}$, $\mathbf{F}$ and $\mathbf{G}$ in (5.15)-(5.17) are still diagonal.

Let us consider again the $(1 \times 1)^3$ interference network with 2 symbol extension in the time domain. As a first step to resolve the diagonality problem, at each receiver, we add a scaled version of the received signal in the first time instant to the received signal in the second time instant. Thus, the equivalent channel matrix between S_i and D_j

becomes

$$\mathbf{H}^{[ji]} = \begin{bmatrix} h_1^{[ji]} & 0 \\ \lambda_j^i h_1^{[ji]} & h_2^{[ji]} \end{bmatrix}, \text{ for } 1 \leqslant i, j \leqslant 3, \tag{5.18}$$

where λ_j^i is the scaling factor at D_j for received signal from S_i. Because the desired and interference signal are superpositioned to each other and inseparable at this stage, we must have $\lambda_j^1 = \lambda_j^2 = \lambda_j^3$.

At D_j, we have

$$\begin{aligned} & (\mathbf{H}^{[jp]})^{-1}\mathbf{H}^{[jq]} \\ = & \begin{bmatrix} h_1^{[jp]} & 0 \\ \lambda_j^p h_1^{[jp]} & h_2^{[jp]} \end{bmatrix}^{-1} \begin{bmatrix} h_1^{[jq]} & 0 \\ \lambda_j^q h_1^{[jq]} & h_2^{[jq]} \end{bmatrix} \\ = & \begin{bmatrix} \frac{1}{h_1^{[jp]}} & 0 \\ -\frac{\lambda_j^p}{h_2^{[jp]}} & \frac{1}{h_2^{[jp]}} \end{bmatrix} \begin{bmatrix} h_1^{[jq]} & 0 \\ \lambda_j^q h_1^{[jq]} & h_2^{[jq]} \end{bmatrix} \\ = & \begin{bmatrix} \frac{h_1^{[jq]}}{h_1^{[jp]}} & 0 \\ -\frac{\lambda_j^p h_1^{[jq]}}{h_2^{[jp]}} + \frac{\lambda_j^q h_1^{[jq]}}{h_2^{[jp]}} & \frac{h_2^{[jq]}}{h_2^{[jp]}} \end{bmatrix} \\ = & \begin{bmatrix} \frac{h_1^{[jq]}}{h_1^{[jp]}} & 0 \\ 0 & \frac{h_2^{[jq]}}{h_2^{[jp]}} \end{bmatrix}, \text{ for } 1 \leqslant j, p, q \leqslant 3, j \neq p, q. \end{aligned} \tag{5.19}$$

Thus, from (5.15) and (5.19), it is easy to see $\mathbf{E}$ still possesses a diagonal structure and $\mathbf{v}^{[1]}$ must have a zero entry. Similarly to the argument in the last section, since $\mathbf{F}$ and $\mathbf{G}$ are both diagonal matrices as $\mathbf{E}$, $\mathbf{v}^{[2]}$ and $\mathbf{v}^{[3]}$ both have zero entries in the same position as in $\mathbf{v}^{[1]}$. Therefore, although 2 symbol extension is used, only 1 time instant is used to transmit information by each transmitter and the received desired and interference signal at each receiver is still inseparable.

From (5.19), we know the reason that simple artificial superposition technique does not work properly is $-\frac{\lambda_j^p h_1^{[jq]}}{h_2^{[jp]}}$ and $\frac{\lambda_j^q h_1^{[jq]}}{h_2^{[jp]}}$ cancels each other because $\lambda_j^p = \lambda_j^q$. In other words, the main problem is such operations scale all interference components equally.

Thus, it is obvious that finding a function which can provide unequal scaling to

different components in superpositioned signal is very important. If we restrict ourselves to the baseband representation of the received signal (ignoring noise), at time instant τ, we have

$$y_\tau^{[j]} = y_\tau^{[j1]} + y_\tau^{[j2]} + y_\tau^{[j3]}, \tag{5.20}$$

where $y_\tau^{[j]}$ is the overall received signal at D_j and $y_\tau^{[ji]}$ is the individual received signal at D_j from S_i, for $1 \leqslant i, j \leqslant 3$. What we need is to find a function f such that

$$f(y_\tau^{[j]}) = \lambda_j^1 y_\tau^{[j1]} + \lambda_j^2 y_\tau^{[j2]} + \lambda_j^3 y_\tau^{[j3]} \tag{5.21}$$

and

$$\lambda_j^1 \neq \lambda_j^2 \neq \lambda_j^3, \text{ for } 1 \leqslant j \leqslant 3. \tag{5.22}$$

(5.21) implies that such a function f should only produce linear combinations of different signal components, with unequal combining coefficients. Unfortunately, to our best knowledge, such a function f does not exist under baseband signal representation after coherent demodulation, where the phase offsets at all transmitters and receivers are equal, i.e, $\theta^{[i]} = \varphi^{[j]}$, $1 \leqslant i, j \leqslant 3$.

5.3 Joint noncoherent demodulation and interference alignment

We have seen that baseband signal representation after coherent demodulation does not have enough freedom to be manipulated to meet our needs. However, noncoherent demodulation with random phase offset at each transmitter or receiver provides us extra diversity (opportunity of unequal scaling) between each transmitter-receiver pair.

Coming back to the previous $(1 \times 1)^3$ interference network with 2 symbol extension in the time domain. If $s_1^{[i]}$ and $s_2^{[i]}$ are the baseband precoded signal to be transmitted by S_i across two time instants, the following transmitting strategy is used:

1. In the first time instant, S_i modulates $s_1^{[i]}$ with random phase offset $\theta_1^{[i]}$.

2. In the second time instant, S_i modulates $s_2^{[i]}$ with random phase offset $\theta_2^{[i]}$.

Correspondingly, if $y_1^{[j]}$ and $y_2^{[j]}$ are the bandpass received signal across two time instants, D_j creates 2 artificial signalling branches and the following receiving strategy is used:

1. In the first signalling branch, D_j firstly demodulates $y_1^{[j]}$ with random phase offset $\varphi_{1,1}^{[j]}$ and $y_2^{[j]}$ with random phase offset $\varphi_{1,2}^{[j]}$. Then, it adds the two signal together to generate baseband received signal $\tilde{y}_1^{[j]}$.

2. In the second signalling branch, D_j firstly demodulates $y_1^{[j]}$ with random phase offset $\varphi_{2,1}^{[j]}$ and $y_2^{[j]}$ with random phase offset $\varphi_{2,2}^{[j]}$. Then, it adds the two signal together to generate $\tilde{y}_2^{[j]}$.

Thus, at D_j and the first time instant, if we apply a function $f_{\varphi_{1,1}^{[j]}}$ and $f_{\varphi_{1,2}^{[j]}}$ to the overall bandpass received signal $y_1^{[j]}$ and $y_2^{[j]}$ respectively, we have

$$\begin{aligned} & \tilde{y}_1^{[j]} \\ = \; & f_{\varphi_{1,1}^{[j]}}(y_1^{[j]}) + f_{\varphi_{1,2}^{[j]}}(y_2^{[j]}) \\ = \; & f_{\varphi_{1,1}^{[j]}}(y_1^{[j1]} + y_1^{[j2]} + y_1^{[j3]}) + f_{\varphi_{1,2}^{[j]}}(y_2^{[j1]} + y_2^{[j2]} + y_2^{[j3]}) \\ = \; & \sum_{i=1}^{3}[\cos(\varphi_{1,1}^{[j]} - \theta_1^{[i]})h_1^{[ji]}s_1^{[i]} + \cos(\varphi_{1,2}^{[j]} - \theta_2^{[i]})h_2^{[ji]}s_2^{[i]}]. \end{aligned} \tag{5.23}$$

Similarly, at D_j and the second time instant, if we apply functions $f_{\varphi_{2,1}^{[j]}}$ and $f_{\varphi_{2,2}^{[j]}}$ to the overall received signal $y_1^{[j]}$ and $y_2^{[j]}$, we have

$$\tilde{y}_2^{[j]} = \sum_{i=1}^{3}[\cos(\varphi_{2,1}^{[j]} - \theta_1^{[i]})h_1^{[ji]}s_1^{[i]} + \cos(\varphi_{2,2}^{[j]} - \theta_2^{[i]})h_2^{[ji]}s_2^{[i]}]. \tag{5.24}$$

Therefore, the equivalent channel matrix between S_i and D_j across these 2 symbol extension should be

$$\mathbf{H}^{[ji]} = \begin{bmatrix} \cos(\varphi_{1,1}^{[j]} - \theta_1^{[i]})h_1^{[ji]} & \cos(\varphi_{1,2}^{[j]} - \theta_2^{[i]})h_2^{[ji]} \\ \cos(\varphi_{2,1}^{[j]} - \theta_1^{[i]})h_1^{[ji]} & \cos(\varphi_{2,2}^{[j]} - \theta_2^{[i]})h_2^{[ji]} \end{bmatrix}, \tag{5.25}$$

for $1 \leqslant i, j \leqslant 3$.

It is easy to see that with joint noncoherent demodulation and interference alignment, the scaling factor at D_j for the received signal from S_i becomes

$$\lambda_j^i = \frac{\cos(\varphi_{2,1}^{[j]} - \theta_1^{[i]})}{\cos(\varphi_{1,1}^{[j]} - \theta_1^{[i]})}. \tag{5.26}$$

Because each transmitter or receiver has a unique random phase offset, (5.26) implies that with joint noncoherent demodulation and interference alignment, $\lambda_j^1 \neq \lambda_j^2 \neq \lambda_j^3$ and the diagonality problem is now fully resolved.

Besides the conventional channel diversity, the unequal scaling of different signal components in superpositioned signal here uses the extra phase diversity provided by each distinct transmitter-receiver pair. It is also worth mentioning that the proposed joint noncoherent demodulation and interference alignment scheme also works for deterministic interference networks, where symbol extension themselves do not provide extra diversity. This is because with random phase offsets for each symbol extension at all transmitters and receivers, we literately improved the system diversity by the distinct phase difference of each transmitter-receiver pair at each symbol extension use.

5.4 Generalized noncoherent interference alignment for $(1 \times 1)^K$

Up to this point, we have used the $(1 \times 1)^3$ interference network to illustrate why to use symbol extension, the diagonality problem of naive symbol extension because of diversity insufficiency and how our proposed joint noncoherent demodulation and interference alignment scheme resolves the problem by jointly considering bandpass modulation/decomulation and interference alignment. This section generalizes the scheme to general $(1 \times 1)^K$ interference networks.

5.4.1 How many artificial signalling branches are needed

Let us restrict ourselves to use 2 symbol extension only (in time or frequency domain). Since each user has only 1 transmit or receive antenna, total achievable DoF upper bound is $\frac{K}{2}$, i.e., on average, each user wants to achieve $\frac{1}{2}$ DoF in one channel use or 1 DoF across 2 symbol extension, i.e., $d = 1$.

Also, with 2 symbol extensions, the equivalent channel matrices are MIMO and in the form of $N \times 2$, where N is the total number of artificial signalling branches we need to create at each receiver. From (1.5), we know that interference alignment is feasible only if

$$\begin{aligned} & N + 2 \geqslant (K+1)d \\ \Leftrightarrow \quad & N \geqslant K - 1. \end{aligned} \tag{5.27}$$

5.4.2 Generalized scheme description

- Before the start of transmission, S_i passes the original symbol through its unique interference alignment precoding matrix to generate the precoded signal to be transmitted across 2 symbol extension.
- In the first channel use, S_i modulates the first component of its precoded signal with random phase offset $\theta_1^{[i]}$.
- In the second channel use, S_i modulates the second component of its precoded signal with random phase offset $\theta_2^{[i]}$.
- D_j creates $K-1$ artificial signalling branches. The output of the k-th branch, for $1 \leqslant k \leqslant K-1$, is the addition of the demodulated signal of the received signal in first channel use with random phase offset $\varphi_{k,1}^{[j]}$ and the demodulated signal of the received signal in second channel use with random phase offset $\varphi_{k,2}^{[j]}$.
- After the demodulation process, D_j passes all the demodulated signal through its unique interference alignment receiving matrix to remove all interference from its undesired transmitters.

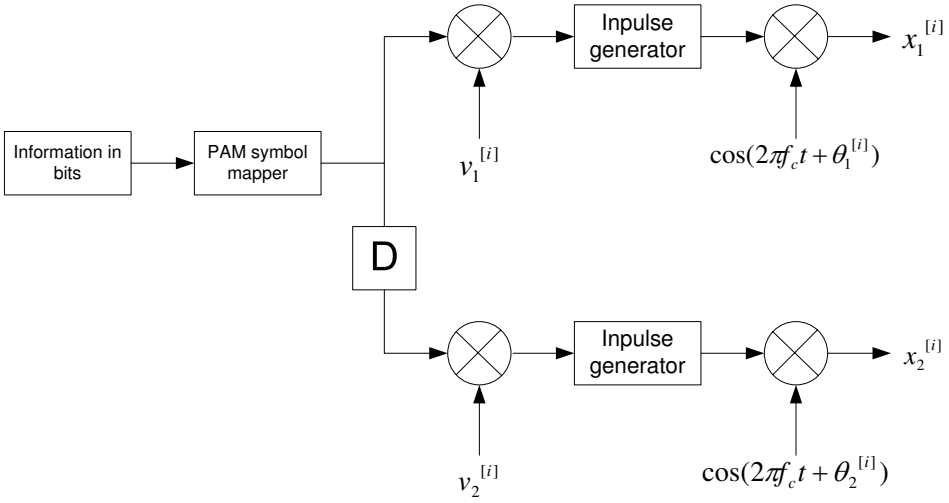


Figure 5.2: Transmitter side model of the noncoherent interference alignment scheme, where "D" is the delay component, v are precoding coefficients and θ are transmitter random phase offsets.

Correspondingly, the block diagram for the generalized scheme can be illustrated in Fig. 5.2 and Fig. 5.3.

5.4.3 DoF optimality

From the last subsection, it is easy see that the equivalent channel matrix between S_i and D_j of a $(1 \times 1)^K$ interference network using the noncoherent interference alignment scheme becomes

$$\mathbf{H}^{[ji]} = \begin{bmatrix} \cos(\varphi_{1,1}^{[j]} - \theta_1^{[i]})h_1^{[ji]} & \cos(\varphi_{1,2}^{[j]} - \theta_2^{[i]})h_2^{[ji]} \\ \cos(\varphi_{2,1}^{[j]} - \theta_1^{[i]})h_1^{[ji]} & \cos(\varphi_{2,2}^{[j]} - \theta_2^{[i]})h_2^{[ji]} \\ \cdots & \\ \cos(\varphi_{K-1,1}^{[j]} - \theta_1^{[i]})h_1^{[ji]} & \cos(\varphi_{K-1,2}^{[j]} - \theta_2^{[i]})h_2^{[ji]} \end{bmatrix}, \tag{5.28}$$

for $1 \leqslant i, j \leqslant K$. Now, we need to show with such equivalent channel matrices, a total of K DoF can be achieved, such that on average, each user can achieve $\frac{1}{2}$ DoF per channel use. We assume the random phase offset at each transmitter or receiver is a rational multiple of π and the channel coefficients are Gaussian rational numbers. A phase offset of a rational multiple of π can be obtained by a finite precision sampling of the carrier

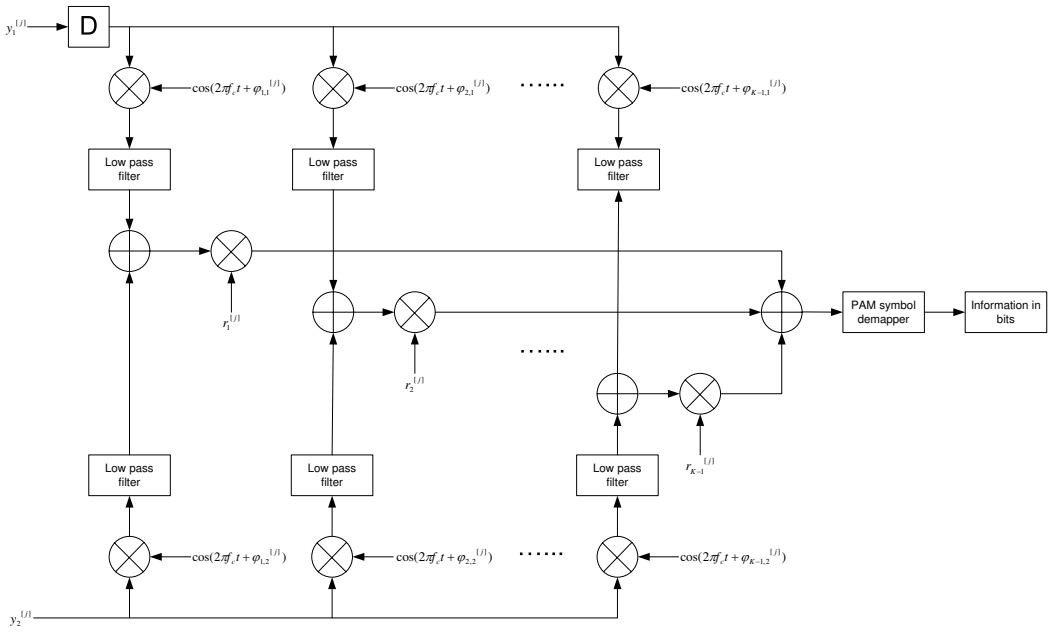


Figure 5.3: Receiver side model of the noncoherent interference alignment scheme, where r are receiving coefficients and φ are receiver random phase offsets.

wave. Similarly, a finite precision sampling of the received signal before the receiving matrix will result an equivalent quantized Gaussian rational distributed channel.

Lemma 1. $\cos(q\pi)$ *is a root of a monic polynomial with integer coefficients for any rational* q*.*

Proof. We only give a brief proof here for the self-containess of this thesis, while interested readers can refer to [67] for details.

Define a linear fractional transformation and its similarity parameter as

$$f(x) = \frac{ax + b}{cx + d}, \text{ and } \sigma = \frac{(a + d)^2}{ad - bc}. \tag{5.29}$$

where a, b, c, d are complex constants. The iterations of two linear fractional transformations in the complex plane are geometrically similar if and only if they have the same similarity parameter. Now, consider the following function

$$g(x) = \frac{(1 + m)x + (1 - m)}{(1 - m)x + (1 + m)}, \tag{5.30}$$

where m is a complex constant. For an random initial value of x_0, the n-th iteration of $g(x)$

can be written as

$$g^n(x) = \frac{(1+m^n)x_0 + (1-m^n)}{(1-m^n)x_0 + (1+m^n)}. \tag{5.31}$$

If we require $g^n(x)$ to be cyclic with period p, then we must have $m^p = 1$ for an integer p. This means m must be of the form

$$m = e^{j2\pi\frac{q}{p}}, \text{ for } q = 1, 2, \ldots, p-1. \tag{5.32}$$

Thus, the similarity parameter of $g(x)$ is

$$\begin{aligned} \sigma_{g(x)} &= \frac{4(1+m)^2}{(1+m)^2 - (1-m)^2} \\ &= m + \frac{1}{m} + 2 \\ &= 2\cos(\frac{2q\pi}{p}) + 2 \\ &= 4\cos(\frac{q\pi}{p})^2, \text{ for } q = 1, 2, \ldots, p-1. \end{aligned} \tag{5.33}$$

Now, consider another linear fractional transformation with one of the similarity parameters in (5.33) as

$$h(x) = 1 - \frac{1}{\sigma_{g(x)}x}. \tag{5.34}$$

Therefore, we know $h^p(x) = x$ because of their geometric similarity. Let P_p denote a polynomial of $\sigma_{g(x)}$ with period p, then $h^p(x) = x$ implies (after some algebra)

$$\frac{-[\sigma_{g(x)}x^2 - \sigma_{g(x)}x + 1)P_p}{\sigma_{g(x)}[P_p x - P_{p-1}]} = 0, \text{ for odd } p \tag{5.35}$$

and

$$\frac{-[\sigma_{g(x)}x^2 - \sigma_{g(x)}x + 1)P_p}{[\sigma_{g(x)}P_p x - P_{p-1}]} = 0, \text{ for even } p. \tag{5.36}$$

Thus, the polynomials P_p satisfy

$$P_{p+1} = \begin{cases} P_p - P_{p-1}, & \text{for } p \text{ is odd,} \\ \sigma_{g(x)} P_p - P_{p-1}, & \text{for } p \text{ is even.} \end{cases} \tag{5.37}$$

It is easy to verify that $\sigma_{g(x)}$ with period p are roots to monic polynomials P_p with integer coefficients and this completes the proof. □

Lemma 2. *Gauss's Lemma: Any root of a monic polynomial with integer coefficients must either be an integer or irrational number.*

Proof. Again, only brief proof is provided for the completeness of this thesis. It is trivial to verify that polynomials of degree 1 with integer coefficients only have integer roots. Let us define the degree of a real number x as the degree of the minimal monic polynomial with integer coefficients having x as a root. For a degree k monic polynomial $f(x)$ with integer coefficients and constant C, define another polynomial $g(x)$ as

$$g(x) = \frac{f(x) - C}{x}. \tag{5.38}$$

It is easy to see $g(x)$ is a degree $k-1$ monic polynomial with integer coefficients.

Assume $f(x)$ has a rational non-integer root r. Our task now becomes to deriving a contradiction under this assumption. Let r be a non-integer root of $f(x)$ and it must be of degree k. Thus, we have $f(r) = 0$ and $g(r) = -\frac{C}{r}$, where $-\frac{C}{r}$ must be a non-integer.

Define N as the smallest integer such that $Ng(r)$ is an integer. Thus, the product of N and any polynomial in r of degree $k-1$ is an integer and we have $N' = N[g(r) - \lfloor g(r) \rfloor]$ is an integer. Moreover, we have

$$\begin{aligned} N'g(r) &= N[g(r) - \lfloor g(r) \rfloor]g(r) \\ &= Ng(r)^2 - \lfloor g(r) \rfloor Ng(r) \end{aligned} \tag{5.39}$$

also being an integer. This is because $g(r)^2$ can be expressed as a polynomial of degree $k-1$ by reducing every higher power of r by substituting from the expression for r^k given by $f(r) = 0$. However, this contradicts the fact that N is the smallest integer such that $Ng(r)$ is an integer because $0 < [g(r) - \lfloor g(r) \rfloor] < 1$ and $N' < N$. Thus, $f(x)$ does not have a non-integer rational root and this completes the proof. □

Lemma 3. *The product of a rational number and an irrational number is irrational with probability* 1.

Lemma 4. *If all elements of fully connected channel matrices are irrational algebraic numbers, then the total achievable DoF is* $\frac{K}{2}$.

Proof. Please refer to Theorem 1 in [18]. □

Theorem 3. *The noncoherent interference alignment scheme is DoF optimal such that each user can achieve* $\frac{1}{2}$ *DoF per channel use.*

Proof. The proof for the theorem is a combination of the previously introduced lemmas. Firstly, from Lemma 1 and Lemma 2, the terms $\cos(\varphi^{[j]} - \theta^{[i]})$ in (5.28) are irrational numbers. Then, from Lemma 3, we know the products $\cos(\varphi^{[j]} - \theta^{[i]})h^{[ji]}$ are also irrational with probability 1. Finally, from Lemma 4, the equivalent channel matrices offered by noncoherent interference alignment are able to achieve the DoF upper bound $\frac{K}{2}$. □

5.4.4 Extensions to real deterministic interference networks

It is easy to see that the proposed noncoherent interference alignment scheme does not distinguish between Gaussian or deterministic interference networks. This is because either of them lacks of sufficient diversity to be traded for multiplexing improvement. The extra diversity we were manipulating comes from the distinct phase difference of each transmitter-receiver pair which does not depend on the underlying physical channels. Thus, noncoherent interference alignment works for Gaussian as well as deterministic interference networks.

5.5 Simulation results

In this section, we investigate the performance of our proposed noncoherent interference alignment scheme. The modulation scheme we employ is PAM, interference alignment precoding and receiving matrices are derived from closed-form solution for three-user MIMO interference channel in [21] and iterative zero-forcing and max-SINR solutions in [44]. The capacity upper bound we use is from (1.4) with the trivial $o(\log(\text{SNR}))$ term

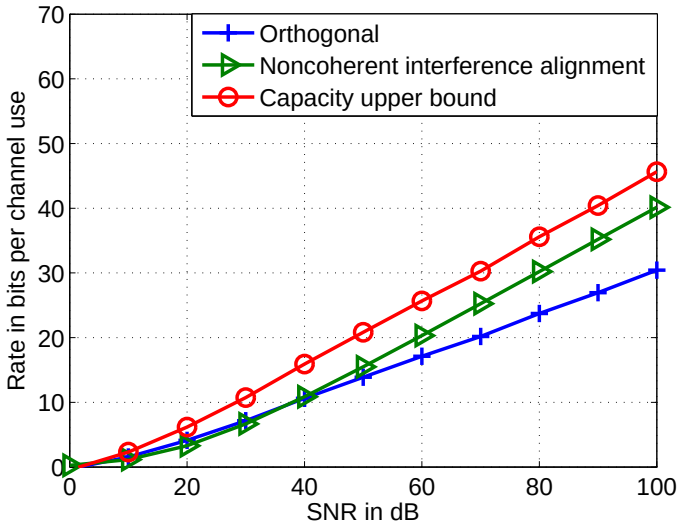


Figure 5.4: Rate performance of noncoherent interference alignment for $(1 \times 1)^3$.

being ignored. Such approximation becomes increasingly accurate in the medium-to-high SNR regime.

From the simulation results in Fig. 5.4 and Fig. 5.5, we can see our proposed noncoherent interference alignment scheme is able to achieve the DoF upper bound, i.e., its achievable throughput increases as the SNR with slope $\frac{K}{2}$. However, we obverse that the achievable throughput is better than that of the orthogonal transmission scheme only in the high SNR regime. This is because the noncoherent interference alignment scheme (actually almost every other interference alignment scheme) is only DoF optimal but not capacity optimal. Thus, there is a constant gap (in the medium-to-high SNR regime where interference rather than noise is the dominating factor that affects the throughput) between the achievable throughput and the capacity upper bound. The reasons for the gap is explained in detail as follows.

Firstly, for any interference alignment scheme, one has to sacrifice some signal subspaces in order to align and remove all interference signals from undesired transmitters. This, however, will at the same time remove the desired signal in those signal subspaces and the overall signal energy is almost definitely reduced. Fortunately, this only results in a fixed SNR offset and does not affect the achievable DoF. Actually, one important contribution of our work is to propose the first known scheme to tradeoff fixed

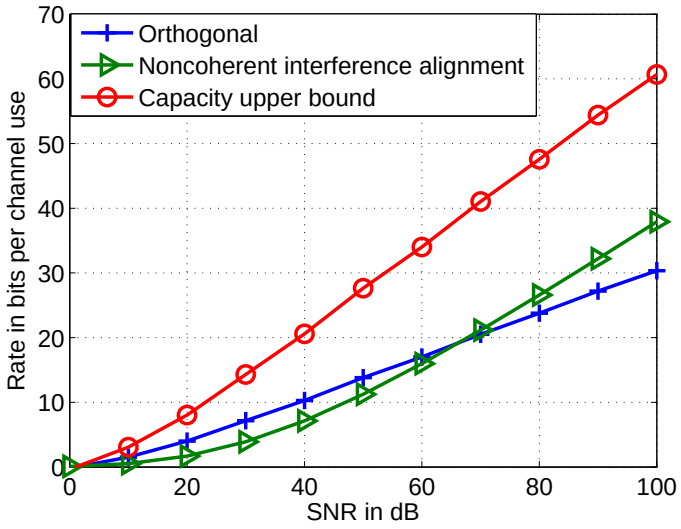


Figure 5.5: Rate performance of noncoherent interference alignment for $(1 \times 1)^4$.

power (or SNR offset) for DoF improvement.

Secondly, the use of random phase offsets in noncoherent interference alignment results in energy loss of desired signal in the demodulation process. From (5.4), we know desired signal energy is always reduced to some extent in order to create the "unequal scaling". In the simulation results presented above, the phase offset between each transmitter-receiver pair is drawn from a continuous uniform distribution set $[0^o, 360^o)$. It is easy to verify that such operations halve the average received signal power at each receiver compared to the case when coherent detection is used.

Finally, in the process of creating artificial signalling branches in order to meet the equivalent MIMO channel feasibility condition, we raised the noise level. From the noncoherent interference alignment scheme description, it is easy to see that the output of the k-th, for $1 \leqslant k \leqslant K-1$, signalling branch is the addition of the demodulated signal of the received signal across two channel uses. The addition operation includes the addition of desired and interference signals and also the addition of noise signals across two channel uses. While all interference signals can be removed, nothing can be done about the increased random noise.

Next, we consider the BER performance of our proposed noncoherent interference alignment scheme. As shown in Fig. 5.6, due to the reasons explained in the last several

paragraphs, and the fact that our proposed noncoherent interference alignment scheme trades the diversity gain for the multiplexing improvement, the BER performance is not good in its original form. In order to recover the diversity benefit in the finite rate case, we employ the Fischer-Huber (FH) loading algorithm [68]. In particular, we formulate an optimization problem to maximize the minimum distance of different PAM symbols so that the BER is minimized, subject to the total rate constraint. The rate and power allocation is applied to the equivalent parallel channels after and before and interference alignment precoding and receiving matrices respectively. For our real (PAM) case, the rate and power for each channel under the FH algorithm is allocated as

$$R_i = \frac{R_T}{d_i} + \frac{1}{2d_i} \log\left(\frac{|h_i|^{2d_i}}{\prod_{l\in I} |h_l|^2}\right), \tag{5.40}$$

and

$$P_i = \frac{\frac{P_T 2^{2R_{Q_i}}}{|h_i|^2}}{\sum_{l\in I} \frac{2^{2R_{Q_l}}}{|h_l|^2}}, \tag{5.41}$$

where d_i, R_{Q_i} and P_i are the DoF, rate and power respectively from S_i to D_i and h_i and Q_i are the channel gain and the precoding matrix of the i-th parallel channel, for $1 \leqslant i \leqslant K$. All the above variables are considered only in I which corresponds to the set of channels actually in use after the loading algorithms. Simulation results in Fig. 5.6 and Fig. 5.7 show that with rate and power loading algorithms, the proposed interference alignment scheme can achieve the DoF upper bound in the high SNR regime, while at the same time maintain acceptable BER performance in the low SNR regime.

5.6 Conclusion

In this chapter, we proposed a practical and universal interference alignment strategy for general $(1 \times 1)^K$ CIC, which trades signal power for intermediate diversity improvement towards the ultimate multiplexing requirement. The problems of naive symbol extension, which is a conventional diversity increasing technique to do interference alignment, was analyzed in details. It was identified that lack of diversity is the main problem such that

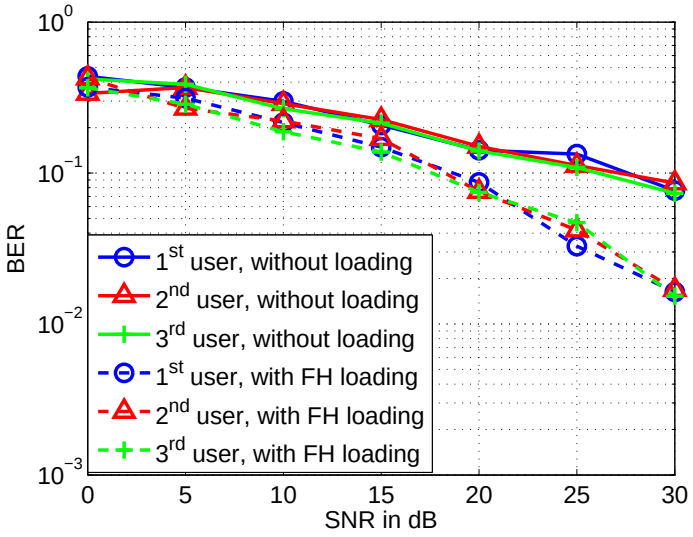


Figure 5.6: Error performance of noncoherent interference alignment for $(1 \times 1)^3$.

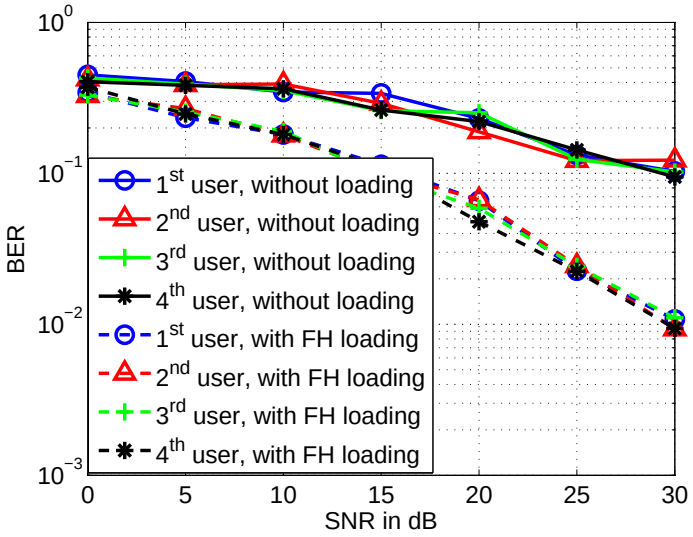


Figure 5.7: Error performance of noncoherent interference alignment for $(1 \times 1)^4$.

there is not enough freedom to be manipulated to meet desired alignment conditions for some network configurations. An noncoherent interference alignment scheme with joint noncoherent bandpass modulaiton/demodulation and interference alignment was then proposed to resolve the diagonality problem of naive symbol extension and the simple superposition technique. This scheme was then generalized to $(1 \times 1)^K$ interference networks, either Gaussian or deterministic, and was proven to be DoF optimal. Simulation results verified its correctness and showed significant DoF improvement in the high SNR regime.

As a conclusion to this chapter, we want to emphasize that although noncoherent energy loss of desired signal may cost us SNR offset, energy increase of interference signal will cause damaging error floor and decrease the slope of the achievable rate curve. Therefore, noncoherent interference alignment is preferable in the wide sense. An interesting future work would be analysis of the optimal region and distribution of the random phase offsets such that the total SNR loss is minimized. The challenge is to derive the tradeoff between the noncoherent loss and signal subspace loss (less noncoherent loss will result more signal subspace loss due to the increased closeness of desired and interference signal vectors).

Chapter 6

Conclusion and Future Works

6.1 Conclusion

This thesis focuses on cooperative strategy designing problems using the DMT principle. We start from primitive cooperative networks with unit DoF including CBC, CMA and CMR, to complex cooperative networks with higher DoF such as RAXN and CIC. We provide methodologies, algorithms and discussions about how to use the DMT principle to design cooperative strategies for different network topologies.

We start in Chapter 2 with classical broadcast, multiple-access and relay channels and proposes a GSSAF strategy that is asymptotic optimal for wireless cooperative networks with unit DoF in terms of the DMT. This unified strategy is easy to implement with ignorable overhead while it's gain over conventional non-cooperative and conventional cooperative strategies is significant. More importantly, GSSAF does not distinguish between different network topologies as long as they all have unit DoF constrained by the cut-set bound. Thus, it follows that a node in cooperative networks with unit DoF can use GSSAF as a unifying rule no matter it's specific role as a source, a relay or a destination to achieve DMT optimality.

Chapter 3 deals with RAXN which is a special network topology with unit DoF and combing features of classical unit DoF networks introduced in Chapter 2. WNC has been shown to be a promising technology to increase the exchange throughput for RAXN, while it's gain often comes with unrealistic assumptions such as perfect overhearing and

undesired side effects such as worse error performance. It is not immediately clear that whether WNC is still fundamentally better than conventional hop-by-hop transmission strategy for RAXN, when these assumptions are relaxed and fair comparisons are made. Through analyzing the DMT of our proposed WNC based partial interference cancelation strategy, we provide a positive proof to justify the fundamental superiority of WNC even with imperfect overhearing.

Chapter 4 focuses on using the DMT measure to analyze the tradeoff between different strategies for cooperative network topologies with higher DoF, typically represented by CIC with more than two source-destination pairs. We identify that although interference alignment is mainly a technique to achieve multiplexing gain upper bound for CIC, improving the spatial diversity is a promising method to tackle its implementation difficulties and strict communication channel assumptions because of the inherent DMT of every communication system. A feasibility condition for interference alignment with diversity gain is proposed and proven to give insights about the fundamental cooperative strategy designing criteria. Our analysis and simulations reveal that when designing interference alignment algorithms, we should take consideration of both desired channel matrices as well as interference channel matrices. Besides, carefully designed convergence measures such as maximizing SINR or minimizing the minimizing the MSE help to achieve the DoF upper bound for CIC with desired diversity gains.

Chapter 5 proposes a noncoherent interference alignment strategy based on the insights observed in Chapter 4. Although direct spatial diversity enhancement is not always possible for an interference channel, intentional random phase offsets at both transmitter and receiver sides will make the channel between them in different time slots being “almost” or pseudo-random, and thus enhance the spatial diversity of the overall system across different dimensions. Our proposed strategy trades signal power for intermediate diversity gain towards ultimate multiplexing improvement. Our strategy is generalized and simplified in the sense that it is universal DMT optimal for general signal-antenna CIC with minimum scheduling efforts and only two symbol extensions.

6.2 Future works

Our research focuses on cooperative strategies design based on the DMT principle and several open problems could have great impact for the DMT principle to play an even more important role in designing future cooperative strategies.

Firstly, our proposed WNC based partial interference cancelation algorithm is only one of the achieving strategy for RAXN for analysis purpose, while we think a joint detection and decoding technique based on soft values can potentially give better results. This is obviously an interesting and practical direction to follow.

Secondly, our designing framework solved the DMT upper bounds and designed cooperative strategies to achieve these bounds for most primitive network topologies. However for CIC, the explicit achievable DMT upper bound is still not available . In order to evaluate the “absolute” performance of a strategy, we need to compare our achievable DMT with the upper bound and check whether our strategy has successfully closed the gap. Thus, deriving the DMT for CIC is important and its success also gives insights on the achievable schemes designing issues.

Thirdly, due to the lack of feasibility conditions for non-generic linear systems, we mainly studied CIC with single antennas at each transmitter and receiver. In future ad-hoc configurations, it is likely that each distributed terminal adopts multiple antennas for better performance. Thus, solving the feasibility conditions for interference alignment with diversity in more general multiple-antenna interference networks is an important yet challenging task. Also, our research as well as most others assume the full channel information is known at both transmitter and receiver sides. This is a reasonable and practical assumption for small systems, but not so much for large systems where the interference alignment technology is most useful for, especially when multiple-antenna terminals are deployed. As a consequence, how to relax the channel side information assumption is an equally important question. Several researchers have proposed to use blind interference alignment with pre-defined switching patterns which need more thought about how to develop an practical integrated solution to solve these problems simultaneously.

Fourthly, our noncoherent interference alignment scheme is designed for real

Gaussian or deterministic interference networks. For complex Gaussian interference networks, we still only send the real symbols. However, in order to achieve the DoF upper bound, we need to recover all the desired transmitted real symbols at each receiver. Using phase randomization with oversampling, the equivalent MIMO channel between each transmitter and receiver pair becomes rank deficient but with extended degree of freedom. A natural extension is to design an interference alignment scheme based on this equivalent channel which aligns all interference signals into the extendted signal subspace and leave the other one interference-free for desired signal.

Finally, for our designing framework to be used more extensively in the future, we eventually need an easy to follow coding scheme to achieve the DMT promised by the designed strategy. Thus, an important work to follow is to further develop a general pattern about codes design corresponding to strategy design to approach the promised performance.

Appendix A

Distributed coding opportunity searching algorithm

Firstly, we restrict the coding operations to be linear for practically easy implementation purpose. As we said, this is often sub-optimal yet a tractable simplification. Secondly, our coding upon routing algorithm does not awake current sleeping nodes, i.e., if an intermediate node is made inactive by traditional routing algorithms, they remain inactive in our coding upon routing algorithm. In other words, we only use the active nodes, i.e., the nodes which need to transmit some information no matter whether we use network coding or not. This means our algorithm is built upon the network layer in the sense that we use the existing established routes and aim to further use network coding to save transmissions or to improve the transmission efficiency. The idea of using existing routes makes our algorithm sub-optimal, because sometimes longer or more congested routes can actually provide more coding opportunities or higher network coding gains. However, such cross-layer approaches may need to consider from physical layer to transport layer jointly and can make analysis quite complicated. To our best knowledge, no practical scheme exists for general communication networks with arbitrary topology and multiple unicast sessions. Thus, we turn to a sub-optimal solution to build our coding upon routing algorithm between the network layer and the transport layer. The advantages of such sub-optimal solutions are similar to that of the OSI model, i.e., we can separately design either better routing algorithms or better coding opportunity detection algorithms to improve

the network performance. Other advantages include easy implementation by inserting a coding layer between the network layer and the transport layer. This may be even more important in practice because the network performance can be improved by software modifications without replacing the expensive network core with network coding enabling devices.

In summary, our goal is to find linear network coding opportunity, gain and solution upon routing. Thus, routing is the underlying technology and we are aiming to explore possible linear coding opportunities along the existing routes to save bandwidth and improve efficiency without major change to current network core.

A.1 Network coding for multiple unicast sessions

A.1.1 A coding gain upper bound

Assume there are K unicast sessions, denoted as $f_1, f_2, \ldots, f_K$. Each unicast session has a unique source node S_k and a unique destination node D_k, for $k = 1, 2, \ldots, K$.

Each unicast session has its individual maxflow mincut bound C_k. This means we can find a group of C_k edge disjoint paths from S_k to D_k. There may be several groups of such paths, and the group with the minimum number of transmissions needs N_k transmissions. Assume the unicast sessions are ordered so that $N_1 \leqslant N_2 \leqslant \ldots \leqslant N_K$.

Denote the set of intermediate nodes along the group of paths with minimum number of transmissions as $\mathcal{R}_k$. Clearly,

$$N_k = C_K + |\mathcal{R}_k|. \tag{A.1}$$

With only routing, it is obvious that the overall minimum number of transmissions

needed for all the unicast sessions is

$$\begin{aligned} & N_1 + N_2 + ... + N_K \\ = \; & (C_1 + |\mathcal{R}_1|) + (C_2 + |\mathcal{R}_2|) + ... + (C_K + |\mathcal{R}_K|) \\ = \; & \sum_{i=1}^{K} C_i + \sum_{i=1}^{K} |\mathcal{R}_i|. \end{aligned} \tag{A.2}$$

With network coding, we give the minimum possible number of transmissions, which corresponds to a coding gain upper bound, as follows:

1. For unicast session f_1 only, we need at least N_1 transmissions.

2. (a) For unicast session f_2 given unicast session f_1 is completed, we need at least C_2 transmissions to inject the messages for this unicast session into the network.

   (b) Assume the C_1 messages emitted by S_1 and C_2 messages emitted by S_2 can be overheard by D_2 and D_1 respectively through a genie without any extra transmission. Then, the nodes in the intersection of $\mathcal{R}_1 \bigcap \mathcal{R}_2$ can serve both unicast sessions by coding their messages together.

   (c) Unicast session f_2 needs to use the nodes in $\mathcal{R}_2$. Now that the nodes in $\mathcal{R}_1 \bigcap \mathcal{R}_2$ has already been used, it only needs $|\mathcal{R}_2 \backslash \mathcal{R}_1|$ more transmissions.

   (d) Thus, for unicast session f_2, we need at least $C_2 + |\mathcal{R}_2 \backslash \mathcal{R}_1|$ transmissions.

3. Similarly, for unicast session f_k, the number of transmissions we need is at least

$$C_k + |\mathcal{R}_k \backslash \mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_{k-1}|, \tag{A.3}$$

   where $k = 1, 2, ..., K$ and $\mathcal{R}_0$ is a empty set.

So, with network coding, the minimum possible number of transmissions needed

is

$$\begin{aligned}
& N_1 + (C_2 + |\mathcal{R}_2 \backslash \mathcal{R}_1|) + (C_3 + |\mathcal{R}_3 \backslash \mathcal{R}_1, \mathcal{R}_2|) + ... + (C_K + |\mathcal{R}_K \backslash \mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_{K-1}|) \\
= \; & (C_1 + |\mathcal{R}_1|) + (C_2 + |\mathcal{R}_2 \backslash \mathcal{R}_1|) + (C_3 + |\mathcal{R}_3 \backslash \mathcal{R}_1, \mathcal{R}_2|) + ... + (C_K + |\mathcal{R}_K \backslash \mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_{K-1}|) \\
= \; & \sum_{i=1}^{K} C_i + (|\mathcal{R}_1| + |\mathcal{R}_2 \backslash \mathcal{R}_1| + |\mathcal{R}_3 \backslash \mathcal{R}_1, \mathcal{R}_2| + ... + |\mathcal{R}_K \backslash \mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_{K-1}|) \\
= \; & \sum_{i=1}^{K} C_i + |\bigcup_{i=1}^{K} \mathcal{R}_i|. \qquad \text{(A.4)}
\end{aligned}$$

Thus, we can upper bound the coding gain, which is defined as the ratio of the number of transmissions needed by routing and the number of transmissions needed by coding as

$$\begin{aligned}
\eta &= \frac{\sum_{i=1}^{K} C_i + \sum_{i=1}^{K} |\mathcal{R}_i|}{\sum_{i=1}^{K} C_i + |\bigcup_{i=1}^{K} \mathcal{R}_i|} \\
&\leqslant \frac{\sum_{i=1}^{K} C_i + K|\bigcup_{i=1}^{K} \mathcal{R}_i|}{\sum_{i=1}^{K} C_i + |\bigcup_{i=1}^{K} \mathcal{R}_i|} \\
&\leqslant \frac{K|\bigcup_{i=1}^{K} \mathcal{R}_i|}{|\bigcup_{i=1}^{K} \mathcal{R}_i|} \\
&= K. \qquad \text{(A.5)}
\end{aligned}$$

A.1.2 An achievable example

From the derivation of the coding gain upper bound, we observe in order to achieve this upper bound, the following conditions need to be satisfied:

1. Each source's messages must be overheard by its unintended destinations.
2. All the unicast sessions should use the same intermediate nodes.
3. The number of intermediate nodes should be much larger than the sum of their individual maxflow mincut bounds.

Based on these observations, we now construct an example which can achieve this coding gain upper bound asymptotically. As shown in Fig. A.1, K unicast sessions share the same path $(R_1, R_2, ..., R_N)$. With only routing, the minimum number of transmissions needed is $K(N+1)$. With simple network coding, i.e., mix the messages at R_1

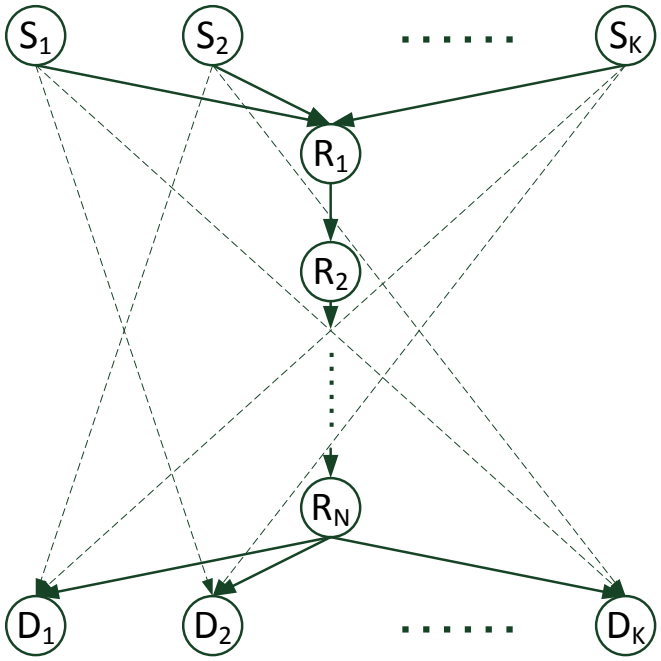


Figure A.1: An example to asymptotically achieve the coding gain upper bound.

and broadcast the mixed message at R_N, we only need $K+N$ transmissions. Thus, when the number of intermediate nodes is large, the coding gain is

$$\eta = \lim_{N\to\infty} \frac{K(N+1)}{K+N} = \lim_{N\to\infty} \frac{KN+K}{K+N} = \lim_{N\to\infty} \frac{KN}{N} = K. \tag{A.6}$$

However, the achievable coding gain for an arbitrary network with multiple unicast sessions is related to its topology and connectivity.

A.2 Generalized butterfly network

A.2.1 Definition

Definition 4. *A level-K generalized butterfly network (level-K GBN) is defined as a network satisfying the following conditions:*

1. *There are K unicast sessions with K distinct source nodes and K distinct corresponding destination nodes.*

2. *The paths from the source nodes to their corresponding destination nodes decided by a chosen existing routing algorithm for these K unicast sessions have intersections at some common nodes.*

3. *There exist $(K-1)$ opportunistic paths from any subset of $(K-1)$ source nodes to each destination node, which do not travel through the intersections.*

We model the level-K GBN as a directed acyclic graph, $GBN_K = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the node set and $\mathcal{E}$ is the edge set. An edge e can be represented by an ordered node pair (x, y), where $x, y \in \mathcal{V}$. y is called the head of the edge and x is called the tail of the edge. The messages can only be transmitted from x to y. The incoming edge set and the outgoing edge set of a node v are defined as

$$\begin{aligned} \mathcal{E}_{in}(v) &= \{(x, y) \mid (x, y) \in \mathcal{E}, y = v\}, \\ \mathcal{E}_{out}(v) &= \{(x, y) \mid (x, y) \in \mathcal{E}, x = v\}. \end{aligned} \tag{A.7}$$

$\mathcal{S}$ is used to denote the node set containing all source nodes and $\mathcal{D}$ is used to denote the node set containing all destination nodes. $\mathcal{B}$ represents the node set containing the intersecting nodes of all unicast sessions of the GBN_K. $\mathcal{R}$ represents the node set containing all the nodes involved in existing paths to accommodate the traffic demand for all unicast sessions. $\mathcal{W}$ represents the node set containing all the nodes involved in the opportunistic paths for all the destination nodes. $g = (K-1) \mid \mathcal{B} \mid$ and $p = \mid \mathcal{W} \setminus \mathcal{R} \mid$ are called the *network coding saving* and *network coding penalty* respectively. The saving of transmissions (SoT) by using network coding compared to traditional routing is the network coding saving minus the network coding penalty and can be written as

$$SoT = (g - p)^+ \tag{A.8}$$

where we use $(x)^+$ to mean $\max\{x, 0\}$. For a level-K GBN, network coding can save transmissions if $SoT > 0$; Otherwise, network coding is not better than traditional routing.

Using graph-theoretic characterization, a level-K GBN is a directed acyclic graph $GBN_K = (\mathcal{V}, \mathcal{E})$ containing K unicast sessions $(s_1 \to d_1)$, $(s_2 \to d_2)$,....., $(s_K \to d_K)$,

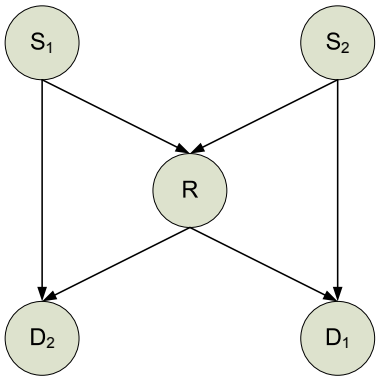


Figure A.2: Traditional butterfly network.

where $s_k \in \mathcal{S}$ and $d_k \in \mathcal{D}$ for $1 \leqslant k \leqslant K$. The paths decided by a chosen existing routing algorithm for all unicast sessions travel through $b_i \in \mathcal{B}$ for $i = 1, ..., \mid \mathcal{B} \mid$. For each destination, there are $(K-1)$ opportunistic pahts from any subset of the $(K-1)$ source nodes to it with edges chosen from $\mathcal{E}_o \in \mathcal{E} \setminus \mathcal{E}_{out}(v)$, where $v \in \mathcal{B}$.

A.2.2 Supporting examples

Example 4. *Traditional butterfly network as shown in Fig. A.2 is a level-2 GBN with the following definitions:*

1. $GNB_2 = (\mathcal{V}, \mathcal{E})$ *with two unicast sessions* $S_1 \to D_1$ *and* $S_2 \to D_2$.
2. $\mathcal{S} = \{S_1, S_2\}$, $\mathcal{D} = \{D_1, D_2\}$ *and* $\mathcal{R} = \{S_1, S_2, D_1, D_2\}$.
3. $\mathcal{B} = \{R\}$ *and* $\mathcal{W} = \{S_1, S_2, D_1, D_2\}$.
4. $g = \mid \mathcal{B} \mid = 1$ *and* $p = \mid \mathcal{W} \setminus \mathcal{R} \mid = 0$.
5. $SoT = (g-p)^+ = 1$.

Example 5. *A two-way exchange network as shown in Fig. A.3 contains a level-2 GBN with the following definitions:*

1. $GNB_2 = (\mathcal{V}, \mathcal{E})$ *with three unicast sessions* $A \to B$ *and* $B \to A$.
2. $\mathcal{S} = \{A, B\}$, $\mathcal{D} = \{A, B\}$ *and* $\mathcal{R} = \{A, R, B\}$.
3. $\mathcal{B} = \{R\}$ *and* $\mathcal{W} = \{A, B\}$.

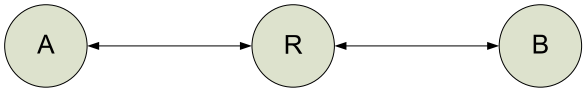


Figure A.3: Two-way exchange network.

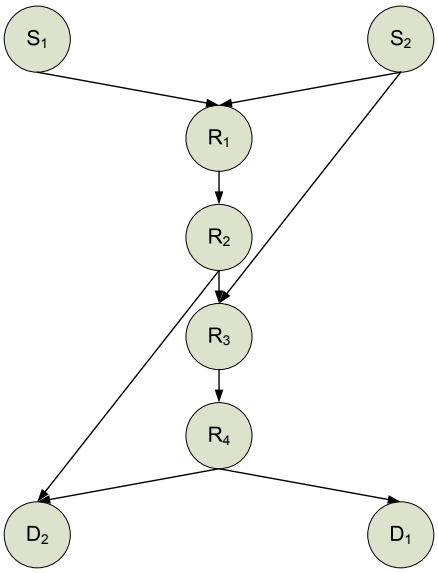


Figure A.4: The grail network.

4. $g = \mid \mathcal{B} \mid = 1$ *and* $p = \mid \mathcal{W} \setminus \mathcal{R} \mid = 0$.

5. $SoT = (g - p)^+ = 1$.

Example 6. *The grail network as shown in Fig. A.4 is a level-2 GBN with the following definitions:*

1. $GNB_2 = (\mathcal{V}, \mathcal{E})$ *with two unicast sessions* $S_1 \rightarrow R_3$ *and* $S_2 \rightarrow D_2$.

2. $\mathcal{S} = \{S_1, S_2\}$, $\mathcal{D} = \{R_3, D_2\}$ *and* $\mathcal{R} = \{S_1, S_2, R_1, R_2, R_3, R_4, D_1, D_2\}$.

3. $\mathcal{B} = \{R_1, R_2\}$ *and* $\mathcal{W} = \{S_2, R_3, R_4, D_2\}$.

4. $g = \mid \mathcal{B} \mid = 2$ *and* $p = \mid \mathcal{W} \setminus \mathcal{R} \mid = 0$.

5. $SoT = (g - p)^+ = 2$.

Example 7. *The star network as shown in Fig. A.5 is a level-3 GBN with the following definitions:*

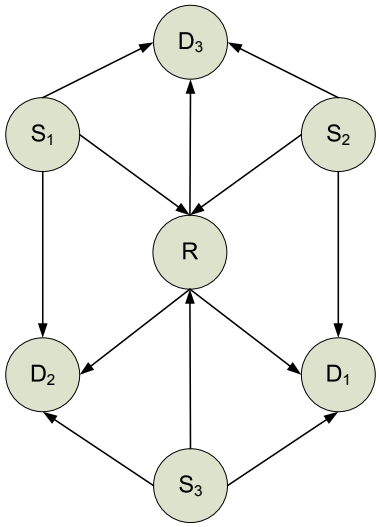


Figure A.5: Three-user star network.

1. $GNB_3 = (\mathcal{V}, \mathcal{E})$ *with three unicast sessions* $S_1 \to D_1$, $S_2 \to D_2$ *and* $S_3 \to D_3$.
2. $\mathcal{S} = \{S_1, S_2, S_3\}$, $\mathcal{D} = \{D_1, D_2, D_3\}$ *and* $\mathcal{R} = \{S_1, S_2, S_3, R, D_1, D_2, D_3\}$.
3. $\mathcal{B} = \{R\}$ *and* $\mathcal{W} = \{S_1, S_2, S_3, R_1, R_2, R_3\}$.
4. $g = 2 \mid \mathcal{B} \mid = 2$ *and* $p = \mid \mathcal{W} \setminus \mathcal{R} \mid = 0$.
5. $SoT = (g - p)^+ = 2$.

Example 8. *A network model with three source-destination pairs is shown in Fig. A.6, which contains a level-2 GBN with the following definitions:*

1. $GNB_2 = (\mathcal{V}, \mathcal{E})$ *with three unicast sessions* $S_1 \to D_1$ *and* $S_2 \to D_2$.
2. $\mathcal{S} = \{S_1, S_2\}$, $\mathcal{D} = \{D_1, D_2\}$ *and* $\mathcal{R} = \{S_1, S_2, R_1, D_1, D_2\}$.
3. $\mathcal{B} = \{R_1\}$ *and* $\mathcal{W} = \{S_1, S_2, R_2, D_1, D_2\}$.
4. $g = \mid \mathcal{B} \mid = 1$ *and* $p = \mid \mathcal{W} \setminus \mathcal{R} \mid = 1$.
5. $SoT = (g - p)^+ = 0$.

Example 9. *A wireless access network with two mobile terminals is shown in Fig. A.7, which contains a level-2 GBN with the following definitions:*

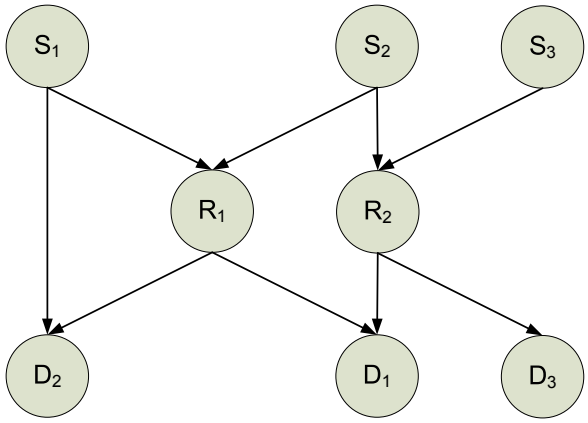


Figure A.6: A network model with three source-destination pairs.

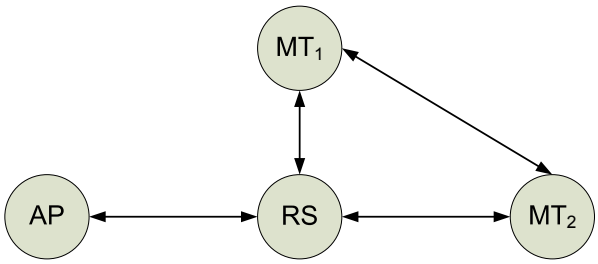


Figure A.7: A wireless access network with two mobile terminals.

1. $GNB_2 = (\mathcal{V}, \mathcal{E})$ *with three unicast sessions* $AP \rightarrow MT_2$ *and* $MT_1 \rightarrow AP$.
2. $\mathcal{S} = \{AP, MT1\}$, $\mathcal{D} = \{MT_2, AP\}$ *and* $\mathcal{R} = \{AP, RS, MT_1, MT_2\}$.
3. $\mathcal{B} = \{RS\}$ *and* $\mathcal{W} = \{AP, MT_1, MT_2\}$.
4. $g =\mid \mathcal{B} \mid = 1$ *and* $p =\mid \mathcal{W} \setminus \mathcal{R} \mid = 0$.
5. $SoT = (g - p)^+ = 1$.

Example 10. *A P2P network with server's download capacity* C_0*, each peer's upload capacity* C_i *and each peer's download capacity assumed to be sufficiently large as shown in Fig. A.8 has been proved to have no network coding gain [69]. The basic idea there is to first determine the throughput of traditional routing by solving a linear program, which finds explicit formulas in two cases based on the values of* C_0 *and* $\sum_{i=1}^{n} C_i$*. Then for each case, network coding is shown to be unable to improve the throughput. We argue the same conclusion using our approach.*

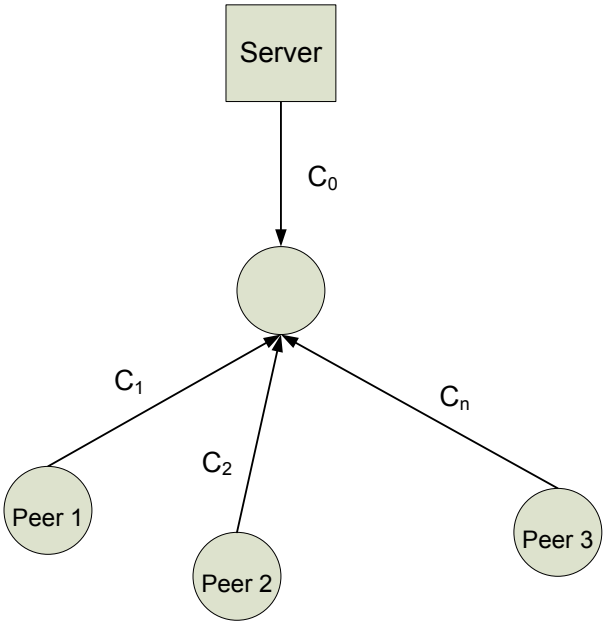


Figure A.8: A P2P network.

We should note that our previous examples focus on directed wireless networks. For wired networks without the wireless broadcast advantage, the network coding gain should be defined as $g = (K-1)(|\mathcal{B}|-1)$*, while the network coding penalty remains unchanged. Thus, no matter how many generalized butterfly networks we can find, the number of intersecting nodes is not larger than 1. This means for such a P2P network, the network coding gain* $g = (K-1)(|\mathcal{B}|-1) = 0$*. Because we will later show generalized butterfly networks are the necessary element to admit network coding gain, we can conclude such a P2P network does not have any coding advantage.*

Example 11. *The single relay network and single cell cellular network models as shown in Fig. A.9 and Fig. A.10 do not contain any generalized butterfly network, and thus do not have any network coding gain.*

A.2.3 Necessary condition for network coding gain

Lemma 5. *K unicast sessions can be coded together to admit network coding gain if these K unicast sessions can form a level-K GBN. In other words, level-K GBN is the necessary element that K unicast sessions can be coded together to provide network coding gain.*

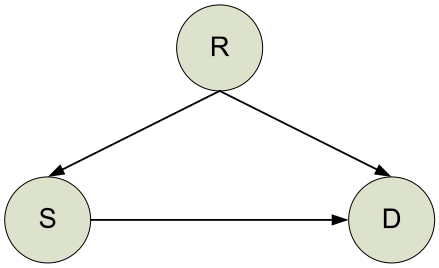


Figure A.9: The single relay network.

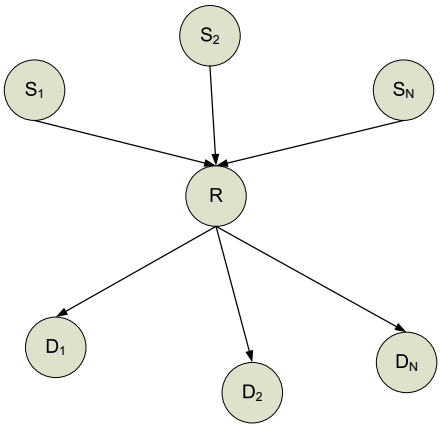


Figure A.10: The single cell cellular network.

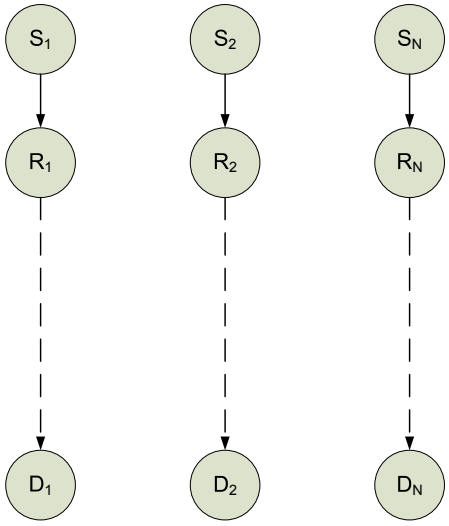


Figure A.11: Edge disjoint multiple unicast flows.

Proof. The proof is equivalent to show if we can apply coding to these K unicast sessions, there must exist a level-K GBN. According to Definition 4, this is equivalent to show the following conditions must be satisfied:

1. There are K unicast sessions with K distinct source nodes and K distinct corresponding destination nodes.

2. The paths from the source nodes to their corresponding destination nodes decided by a chosen existing routing algorithm for these K unicast sessions have intersections at some common nodes.

3. There exist $(K-1)$ opportunistic paths from any subset of $(K-1)$ source nodes to each destination node, which do not travel through the intersections.

First, we note that if the existing paths for the multiple unicast sessions are disjoint from each other as shown in Fig. A.11, there is no network coding gain. This is obvious because if there is no intersections among the existing paths, independent routing is sufficient to achieve minimum number of transmissions, and network coding does not provide gain. Thus, for networks admitting network coding gain, there must be at least two unicast flows intersecting with each other.

Pick any K unicast flows, which can be coded together to admit network coding gain. The intersections mean these K unicast flows travel through the same set of the

nodes $\mathcal{B} = \{b_1, b_2, ..., b_{|\mathcal{B}|}\} \subset \mathcal{R}$. However, the network coding operation mixes the K unicast sessions' messages, i.e., the information conveyed by edges $e \in \mathcal{E}_{out}(v \in \mathcal{B})$ is a function of the K unicast sessions' messages. The function is a linear function if we only consider linear network coding from practical easy implementation and low complexity perspectives.

In order for each destination node to decode its desired information, it must have another $K - 1$ opportunistic paths rather than the existing path to construct a system of equations (to remove the interference). Such opportunistic paths should be independent of the main routes (and thus be independent of the intersections) in order to be able to help the destination nodes to construct non-degraded systems of equations (to remove the interference).

From the argument of last two paragraphs, a network with multiple unicast sessions can be coded together to admit network coding gain if it satisfies all the conditions in Definition 4. Thus, it can now be concluded if K unicast sessions can be coded together, there must exist a level-K GBN. As a consequence, if there is no level-K GBN, not all the K unicast sessions can be coded together. □

Lemma 6. *The paths of two unicast sessions decided by traditional routing algorithms can only intersect once before reaching their destination nodes.*

Proof. We prove this theorem by using contradiction. Assume two paths for two unicast sessions decided by traditional routing algorithms intersect more than once before reaching their destination nodes. Thus, after the first intersection, the two paths need to apart for the second intersection. However, traditional routing algorithms will choose the best possible path from the end of the first intersection to the start of the second intersection. This means between the end of the first intersection and the start of the second intersection, the two paths will travel through the same intermediate nodes. Thus, the apartness is not possible, which contradicts the necessary condition for multiple intersections. So, we conclude that two unicast sessions' paths decided by traditional routing algorithms can intersect at most once before reaching their destination nodes. □

Lemma 7. *A generalized butterfly network can provide network coding gain if the network coding saving is larger than the network coding penalty.*

Proof. Consider a generalized butterfly network satisfying all the conditions in Definition 4. As shown in Lemma 5, network coding gain comes from the congestion in the intersections (containing nodes in the node set $\mathcal{B}$) of the existing paths (containing nodes in the node set $\mathcal{R}$), which can be potentially resolved by coding several unicast sessions' messages together. The interference introduced by the network coding operation can be removed by the opportunistic paths (containing nodes in the node set $\mathcal{W}$) of the generalized butterfly network. However, the nodes in the opportunistic paths which are not used by the existing routes (containing nodes in the node set $\mathcal{W} \setminus \mathcal{R}$) must be used by paths for other unicast sessions decided by traditional routing algorithms. Otherwise, they are inactive and contradict the generalized butterfly network conditions. As a result, in order to use network coding, these nodes in $\mathcal{W} \setminus \mathcal{R}$ must transmit some extra messages which are not needed in traditional routing algorithms. This is the penalty paid by using network coding.

Thus, the saving of transmissions (SoT) by using network coding compared to traditional routing, which is a characterization of the overall network coding gain, is the network coding saving minus the network coding penalty and can be written as $SoT = (g - p)^+$. □

A.3 Four-way handshaking coding opportunity detection algorithm

We have shown in Lemma 6 that possible network coding gain comes from the generalized butterfly networks, which are formed by intersections of multiple unicast sessions' paths. Thus, it is obvious to see finding network coding gain and network coding solution is equivalent to checking whether there are generalized butterfly networks with network coding gain around the intersections of multiple unicast sessions.

Lemma 7 tells us that multiple unicast sessions' paths may contain several generalized butterfly networks. However, they must have the same set of intersecting nodes $\mathcal{B}$. Since the network coding gain comes from the intersections of the main routes, i.e., the shared resource of multiple unicast sessions' path, it is upper bounded by the cardinality of the intersecting node set $\mid \mathcal{B} \mid$. Thus, the task of finding network coding gain is equiv-

alent to checking whether the network coding saving can exceed the minimum possible network coding penalty around the intersections, i.e., the minimum possible number of nodes in the opportunistic paths which are not used by the existing paths $\min(\mid \mathcal{W} \setminus \mathcal{R} \mid)$.

The network coding solution can be easily obtained after finding the generalized butterfly networks with network coding gain. The network coding operations should be done at the starting node of the intersections in order to mix the messages, and at the ending nodes of the opportunistic routes in order to remove the interference.

Our goal is to design a distributed and easy to implement systematic scheme to crawl all the physical connected links of the network with multiple unicast sessions, using some probe information to find all the coding opportunities. Moreover, we aim to develop a powerful system, yet to be practical, without global knowledge of network topology or overwhelming overhead.

Our distributed four-way handshaking coding opportunity detection algorithm is completed detailed as follows:

1. Let each source node send a probe packet in order to form the intersections among the multiple unicast sessions' paths.
   (a) Each probe packet should contain the unicast session's identification and its source and destination addresses.
   (b) Each intermediate node forwards its received probe packets using its own local forwarding table and the destination addresses in the received packets.
   (c) Each intermediate node only forwards the probe packets that are intended to it.
   (d) When an intermediate node detects a new unicast session traveling through it, it stores the unicast flow's identification in its own local buffer.
2. After each destination node receives the probe packet intended to it, it sends a detection packet back to the corresponding source node to detect the intersections.
   (a) Each detection packet should contain the unicast session's identification and its source and destination addresses. Moreover, it also contains a designated

field, called intersection field, to record the intersections involving this unicast session.

(b) Each detection packet was sent along the reverse direction of the unicast session's original path.

(c) When a detection packet arrives at an intermediate node, the router checks whether it serves other unicast sessions by examining its local buffer to search other unicast sessions' identifications.

i. If the router serves a new unicast session which is not present in the detection packet's intersection field, it updates the detection packet by adding an entry in the intersection field containing this intersecting unicast session's identification, current router's address and number of intersection times with this intersecting unicast session (1 in this case).

ii. If the router serves a unicast flow which is not new to the detection packet, it updates the detection packet by adding current router's address to the intersection field and increasing the number of intersection times with this intersecting unicast session by 1.

3. After each source node receives the detection packet, it knows exactly the intersecting unicast sessions' identifications, their intersections' addresses and the number of intersection times along their paths. If a source node sees there are other intersecting sessions along its paths, it sends an opportunistic packet to form the opportunistic paths.

(a) Each opportunistic packet should contain the unicast session's identification, all the intersecting unicast flows' identifications and their corresponding intersections' addresses. Moreover, it also contains a designated field, called path field, to record the path the opportunistic packet has traveled.

(b) When an intermediate node receives an opportunistic packet, it adds an entry containing current router's address and which unicast flows it serves to the path field, and then forwards it to its next hops.

(c) Each intermediate node saves a copy of its received opportunistic packets in its own local buffer.

4. After each destination node receives the opportunistic packets, it sends back a collection packet to detect the network coding penalty.

   (a) Each collect packet should contain the unicast session's identification, all the intersecting unicast sessions' identifications and their intersections' addresses. Moreover, it also contains a designated field, called penalty field, to record the minimum possible network coding penalty and the corresponding opportunistic paths for each intersecting unicast session.

   (b) Each collect packet was sent along the reverse direction of the unicast session's original path.

   (c) When an intermediate node receives a collect packet, it checks whether it serves any intersecting unicast session in the collect packet. If it finds any such intersecting unicast session, it adds a freeze flag to the collect packet to indicate that the penalty field information associated with this intersecting unicast flow is unchangeable anymore.

   (d) For the intersecting unicast sessions it does not serve, it checks for each of these intersecting unicast sessions in the collect packet, whether it has any path from collect packet's source or the intersecting unicast session's source to itself which does not travel through their intersections along its original path by examining the saved opportunistic packets in its own buffer.

      i. If it can find one or more such paths, it counts how many nodes in each such path that serve neither the received collect packet's unicast session nor the intersecting unicast session.

      ii. It compares the smallest number with the number of penalty in the collect packet, and replaces the number of penalty and the corresponding opportunistic path in the collect packet with this smallest number and its corresponding path if the smallest number is smaller than the number of

penalty in the collect packet. If there is no number of penalty and the corresponding opportunistic path in the collect packet, it simply adds an entry with this smallest number and its corresponding path to the penalty field.

5. When the starting node of some intersections receives the collect packets for all of the intersecting unicast sessions it serves, it calculated the saving of transmissions by using network coding. If there is any saving of transmissions offered by network coding, it adds a coding flag to the collect packet to indicate the network coding decision for this intersecting unicast session.

6. After each source node receives the collect packets, it checks whether there is a coding flag for each of the intersecting unicast flows. If it finds a coding flag, it sends a short control information along the opportunistic route for this intersecting unicast flow to indicate this coding decision.

Network coding solution can be easily obtained in a distributed way by the four-way handshaking algorithm described above. Network coding operations should be done at the starting node of the intersections and at the ending nodes of the opportunistic paths.

Bibliography

[1] 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.

[2] ——, "User cooperation diversity-part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[3] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[4] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

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

[7] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[8] S. Yang and J.-C. Belfiore, "Towards the optimal amplify-and-forward cooperative diversity scheme," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3114–3126, Sep. 2007.

[9] H. Ning, C. Ling, and K. Leung, "Generalized sequential slotted amplify and forward strategy in cooperative communications," *IEEE Trans. Inf. Theory*, to appear.

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

[11] A. A. E. Gamal and T. M. Cover, "Multile user information theory," *Proceedings of The IEEE*, vol. 68, no. 12, pp. 1466–1483, Dec. 1980.

[12] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[13] E. C. Van Der Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, no. 1, pp. 120–154, Spring 1971.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.

[15] A. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 569–570, Sep. 1975.

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

[17] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[18] R. Etkin and E. Ordentlich, "On the degrees-of-freedom of the k-user gaussian interference channel," *CoRR*, vol. abs/0901.1695, 2009.

[19] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[20] A. Host-Madsen and A. Nosratinia, "The multiplexing gain of wireless networks," in *Proceedings of IEEE International Symposium on Information Theory, 2005*, Sep. 2005, pp. 2065–2069.

[21] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[22] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[23] S.-Y. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[24] S. Jaggi, P. Sanders, P. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1973–1982, Jun. 2005.

[25] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[26] S. Riis, "Linear versus non-linear boolean functions in network flow," in *38th Annual Conference on Information Sciences and Systems*, Mar. 2004.

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

[28] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[29] A. R. Lehman and E. Lehman, "Complexity classification of network information flow problems," in *Proceedings of the fifteenth annual ACM-SIAM symposium on discrete algorithms*, Jan. 2004, pp. 142–150.

[30] Y. Wu, P. A. Chou, and S. Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," *Technical Report MSR-TR-2004-78*, Aug. 2004.

[31] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," in *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM Press, 2006, pp. 243–254.

[32] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proceedings of the 12th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 2006, pp. 358–365.

[33] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM Press, 2007, pp. 397–408.

[34] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[35] ——, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[36] L. L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[37] J. Laneman and G. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[38] Y. Wu, "On constructive multi-source network coding," *IEEE International Symposium on Information Theory*, pp. 1349–1353, July 2006.

[39] M. Yang and Y. Yang, "A linear inter-session network coding scheme for multicast," *Seventh IEEE International Symposium on Network Computing and Applications*, pp. 177–184, July 2008.

[40] C.-C. Wang and N. B. Shroff, "Beyond the butterfly - a graph-theoretic characterization of the feasibility of network coding with two simple unicast sessions," *IEEE International Symposium on Information Theory*, pp. 121–125, June 2007.

[41] ——, "Intersession network coding for two simple multicast sessions," *Proeedings of Annual Allerton Conference on Communications, Control, and Computing*, Sep. 2007.

[42] N. Ratnakar, D. Traskov, and R. Koetter, "Approaches to network coding for multiple unicasts," *International Zurich Seminar on Communications*, pp. 70–73, Feb. 2006.

[43] D. Traskov, N. Ratnakar, D. Lun, R. Koetter, and M. Medard, "Network coding for multiple unicasts: An approach based on linear optimization," *IEEE International Symposium on Information Theory*, pp. 1758–1762, July 2006.

[44] K. Gomadam, V. Cadambe, and S. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *IEEE Global Telecommunications Conference, 2008*, Nov. 30-Dec. 4 2008.

[45] S. Peters and R. Heath, "Interference alignment via alternating minimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009*, Apr. 2009, pp. 2445–2448.

[46] I. Thukral and H. Bolcskei, "Interference alignment with limited feedback," in *IEEE International Symposium on Information Theory, 2009*, Jun. 28-Jul. 3 2009, pp. 1759–1763.

[47] H. Yu, J. Park, Y. Sung, and Y. Lee, "A least squares approach to joint beam design for interference alignment in multiuser interference channels," in *IEEE 10th Workshop on Signal Processing Advances in Wireless Communications, 2009*, Jun. 2009, pp. 593–597.

[48] A. Khandani, S. Motahari, and B. Nourani, "Relay-aided interference alignment for the quasi-static x channel," in *IEEE International Symposium on Information Theory, 2009*, Jun. 28-Jul. 3 2009, pp. 1764–1768.

[49] B. Nazer, S. Jafar, M. Gastpar, and S. Vishwanath, "Ergodic interference alignment," in *IEEE International Symposium on Information Theory, 2009*, Jun. 28-Jul. 3 2009, pp. 1769–1773.

[50] A. Ozgur and D. Tse, "Achieving linear scaling with interference alignment," in *IEEE International Symposium on Information Theory, 2009*, Jun. 28-Jul. 3 2009, pp. 1754–1758.

[51] H. Ning, C. Ling, and K. Leung, "Relay-aided interference alignment: Feasibility conditions and algorithm," in *IEEE International Symposium on Information Theory, 2010*, Jun. 2010.

[52] R. Tresch and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," in *IEEE International Conference on Communications Workshops, 2009*, Jun. 2009.

[53] Y. Wu and A. Dimakis, "Reducing repair traffic for erasure coding-based storage via interference alignment," in *IEEE International Symposium on Information Theory, 2009*, Jun. 28-Jul. 3 2009, pp. 2276–2280.

[54] C. Yetis, T. Gou, S. Jafar, and A. Kayran, "Feasibility conditions for interference alignment," in *IEEE Global Telecommunications Conference, 2009*, Nov. 30-Dec. 4 2009.

[55] H. Ning, C. Ling, and K. Leung, "Near-optimal relaying strategy for cooperative broadcast channels," in *IEEE International Conference on Communications, 2009*, Jun. 2009.

[56] S. Katti, D. Katabi, H. Balakrishnan, and M. Medard, "Symbol-level network coding for wireless mesh networks," vol. 38, no. 4. New York, NY, USA: ACM, 2008, pp. 401–412.

[57] Z. Ding, K. Leung, D. Goeckel, and D. Towsley, "On the study of network coding with diversity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1247–1259, Mar. 2009.

[58] T. L. L. Hanzo and B. Yeap, *Turbo Coding, Turbo Equalisation, and Spacetime Coding: For Transmission Over Fading Channels*. Wiley, 2002.

[59] K. Jamieson and H. Balakrishnan, "PPR: Partial packet recovery for wireless networks," in *Proceedings of the 2007 conference on applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM Press, 2007, pp. 409–420.

[60] A. Argyriou and A. Pandharipande, "Collision recovery in distributed wireless networks with opportunistic cooperation," *IEEE Commun. Lett.*, vol. 14, no. 4, pp. 300–302, Apr. 2010.

[61] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over mimo x channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[62] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *Asilomar Conference on Signals, Systems, and Computers*, Nov. 2009.

[63] J. B. L. D. A. Cox and D. B. O'Shea, *Using Algebraic Geometry*. 2nd ed., New York: Springer-Verlag, 2005.

[64] Z. Wang and G. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389–1398, Aug. 2003.

[65] H. Ning, M. Estela, and C. Ling, "Noncoherent interference alignment: Trade signal power for diversity towards multiplexing," *IEEE Trans. Inf. Theory*, submitted.

[66] A. Sezgin, S. Jafar, and H. Jafarkhani, "Optimal use of antennas in interference networks: A tradeoff between rate, diversity and interference alignment," in *IEEE Global Telecommunications Conference, 2009*, Nov. 30-Dec. 4 2009.

[67] S. Lang, *Algebraic Number Theory, Graduate Texts in Mathematics*. 2nd ed., New York: Springer-Verlag, 1994.

[68] R. Fischer and J. Huber, “A new loading algorithm for discrete multitone transmission,” in *IEEE Global Telecommunications Conference, 1996.*, Nov. 1996, pp. 724–728.

[69] D. M. Chiu, R. Yeung, J. Huang, and B. Fan, “Can network coding help in p2p networks?” *4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, pp. 1–5, Apr. 2006.