COOPERATIVE CODING AND ROUTING IN MULTIPLE-TERMINAL WIRELESS NETWORKS

LAWRENCE ONG
(B.Eng.(1st Hons.),NUS; M.Phil.,Cambridge)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2007

To Theresa, mum, dad, and Jennie

Acknowledgements

I am specially grateful to Prof Mehul Motani, for his guidance, advice, encouragement, and criticism.

I would like to thank Chong Hon Fah and Yap Kok Kiong for many productive and inspiring discussions.

I thank my parents and my sister, Jennie, for their support throughout the course of my PhD study.

Last but not least, I would like to express my appreciation to my wife, Theresa, for her love and patience, which made this PhD journey a happy one.

Abstract

In this thesis, we take an information-theoretic view of the multiple-terminal wireless network. We investigate achievable rates, in the Shannon sense, and study how to achieve them through cooperative coding and routing. Our work takes an information-theoretic approach, bearing in mind the practical side of the wireless network. First, we find the best way to route data from the source to the destination if each relay must fully decode the source message. We design an algorithm which finds a set of routes, containing a rate-maximizing one, without needing to optimize the code used by the nodes. Under certain network topologies, we achieve *complete* routing and coding separation, i.e., the optimizations for the route and the code can be totally separated. In addition, we propose an algorithm with polynomial running time that finds an optimal route with high probability, without having to optimize the code. Second, we study the trade-off between the level of node cooperation and the achievable rates of a coding strategy. Local cooperation brings a few practical advantages like simpler code optimization, lower computational complexity, lesser buffer/memory requirements, and it does not require the whole network to be synchronized. We find that the performance of local cooperation is close to that of whole-network cooperation in the low transmit-power-to-receiver-noise-ratio regime. We also show that when each node has only a few cooperating neighbors, adding one node into the cooperation increases the transmission rate significantly. Last, we investigate achievable rates for networks where the source data might be correlated, e.g., sensor networks, through

different coding strategies. We study how different coding strategies perform in different channel settings, i.e., varying node position and source correlation. For special cases, we show that some coding strategies actually approach the capacity. Overall, our work highlights the value of cooperation in multiple-terminal wireless networks.

Contents

List of Tables

List of Figures

Nomenclature

Roman Symbols

d_{ij} The distance between nodes i and j.

$E[X]$ The expectation of the random variable X.

h_{ij} The large scale fading component.

$H(S)$ The entropy of random variable S.

$I(X;Y)$ The mutual information between random variables X and Y.

N_i The receiver noise variance at node i, $E[Z_i^2]$.

p An input probability density function.

P_e The average error probability.

P_i The power constraint of node i.

$P_{\text{out}}(\mathcal{M}, p, R)$ The outage probability of DF on route $\mathcal{M}$ with input distribution p at rate R.

R Achievable rate.

R_{DF} The maximum rate achievable by DF.

$R_{k-\text{hop}}$ The maximum rate achievable by a k-hop myopic coding strategy.

$R^E_{\mathcal{M}}(p)$ The ergodic rate using DF on route $\mathcal{M}$ with input distribution p.

$R_{\mathcal{M}}(p)$ The rate achievable on route $\mathcal{M}$ with input distribution p.

$R_m(\mathcal{M}, p)$ The reception rate of node m on route $\mathcal{M}$ with input distribution p.

$R_{\text{omniscient}}$ The maximum rate achievable by an omniscient coding strategy.

w The source message.

x_{ij} The j-th input from node i into the channel.

$\mathbf{x}_{ij}$ The j-th block of inputs from node i into the channel.

x_t^i $x_{t1}, x_{t2}, \ldots, x_{ti}$.

$X_{\mathcal{T}}$ $(X_{t_1}, X_{t_2}, \ldots, X_{t_{|\mathcal{T}|}})$, where $\mathcal{T} = \{t_1, t_2, \ldots, t_{|\mathcal{T}|}\}$.

y_{ij} The j-th output from the channel to node i.

$\mathbf{y}_{ij}$ The j-th block of outputs from the channel to node i.

z_i The receiver noise at node i.

Script Symbols

$\mathcal{A}_\epsilon^n(X_1, X_2, \ldots, X_k)$ The set of ϵ-typical n-sequences $(x_1^n, x_2^n, \ldots, x_k^n)$.

$\mathcal{M}$ A route.

$\mathcal{M}^{\mathrm{SOR}}(p)$ A shortest optimal route for DF with input distribution p.

$\mathcal{NNSA}(\mathcal{T})$ The NNSA candidates of network $\mathcal{T}$.

$\mathcal{NNSA}^{\mathrm{opt}}(\mathcal{T}, p)$ The optimal NNSA candidate set for for input distribution p.

$\mathcal{P}$ A set of probability density functions.

$\mathcal{Q}_{\mathrm{DF}}(\mathcal{P})$ The set of optimal routes for DF with respect to a set of input distributions $\mathcal{P}$.

$\mathcal{R}$ The set of all relays in an MRC.

Greek Symbols

η The large scale fading exponent.

γ_{ij} The received-signal-to-noise ratio (rSNR) of a pair of transmitter i and receiver j.

κ A large scale fading component.

λ_{ij} The channel gain from node i to node t.

ν_{ij} The small scale fading component.

Ω_{ij} The average small scale fading power, $E[\nu_{ij}]$.

$\Pi(\mathcal{T})$ The set of all possible routes from the source to the destination in network $\mathcal{T}$.

ψ_{ij} The transmitted-signal-to-noise ratio (tSNR) of a pair of transmitter i and receiver j.

Abbreviations

AF Amplify-forward.

CF Compress-forward.

CS-OB Cut-set outer bound.

DF Decode-forward.

FDS-DF Full decoding at sources with DF channel coding.

LDPC Low-density parity-check.

MACCS Multiple-access channel with correlated sources.

MACFCS Multiple-relay channel with feedback and correlated sources.

MACF Multiple-access channel with feedback.

MAC Multiple-access channel.

MH-DA Multi-hop coding with data aggregation.

MRC Multiple-relay channel.

MSPA Maximum sum-of-received-power algorithm.

NNA Nearest neighbor algorithm.

NNSA Nearest neighbor set algorithm.

NNSPA Nearest neighbor set pruning algorithm.

p.d.f. Probability density function.

rSNR Received-signal-to-noise ratio.

SC-CF Source coding for correlated sources and CF channel coding for the MACF.

SC-MAC Source coding for correlated sources and the MAC channel coding.

SOR Shortest optimal route.

SPC Single-Peak Condition.

SRC Single-relay channel.

tSNR Transmitted-signal-to-noise ratio.

Chapter 1

Introduction

1.1 Cooperation in Multiple-Terminal Wireless Networks

Multi-terminal wireless networks have been finding more applications and receiving much attention recently by both researchers and industry. Common wireless applications include cellular mobile networks, Wi-Fi networks, ad-hoc networks, and sensor networks. The main advantage of wireless technology to users is the seamless access to the network whenever and wherever they are; to service providers, easier deployment, as no cable laying is required.

A large amount of research has been carried out recently on various aspects of wireless networks, including how to achieve power saving for energy limited nodes (Younis & Fahmy, 2004; Yu *et al.*, 2004), how to route data from the source to the destination with minimum delay or using minimum power (Fang *et al.*, 2004; Shakkottai, 2004; Zhao *et al.*, 2003), how to determine the rate per unit distance supported by the network (Gopala & El Gamal, 2004; Gupta & Kumar, 2003), and how to ensure that all the nodes are connected, i.e., within communication range (Shakkottai *et al.*, 2003).

In this thesis, we investigate transmission rates achievable by cooperative routing and coding for multiple-terminal networks through an information-theoretic

approach. High data rate is desirable for many wireless applications, e.g., wireless Internet access, mobile video conferencing, and mobile TV on buses and trains. Some of these applications would have been impossible without transmission links that provide a certain quality of service, in terms of, for example, transmission rate, delay, and error rate. One way to increase transmission rates is through cooperative routing and coding.

Wireless networks are inherently broadcast, in that messages sent out by a node are heard by all nodes listening in the same frequency band and in communication range. This opens up opportunities for rich forms of cooperation among the wireless nodes. Instead of the traditional multi-hop data transmission where a node only forwards data to another node, i.e., from the source to a relay, from the relay to another relay, and so on until the destination, data transmission in the cooperative wireless network can be from multiple nodes to multiple nodes. This changes the way we think of routing (the sequence of nodes in which data propagate from the source to the destination) and coding (how the nodes encode and decode). We need a new definition of a route and routing algorithms for cooperative networks. We also need to re-think coding and construct cooperative coding strategies to tap the advantage of the multiple-node-to-multiple-node communication.

With an almost unlimited number of ways of interacting and cooperating, analyzing of these multiple-terminal networks is difficult. To date, the capacity of even the simple three-node channel (van der Meulen, 1971) is not known, except for special cases, e.g., the multiple-access channel (MAC) (Ahlswede, 1974; Liao, 1972), the degraded relay channel (Cover & El Gamal, 1979), the degraded broadcast channel (Bergmans, 1973), and the mesh network (Ong & Motani, 2006a, 2007c). However, this did not hinder research in channels with more nodes. A deeper understanding of multiple-terminal networks can help us to design more efficient protocols and algorithms for these networks.

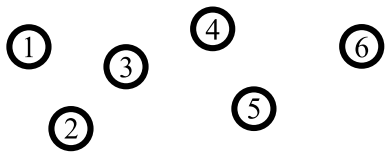


Figure 1.1: A multiple-terminal network.

1.2 Problem Areas

Now, we identify three problem areas that we will study in this thesis. We use the wireless network in Fig. 1.1 for illustration. Nodes 1–6 are users in the wireless network, equipped with transceivers. They can be sources which have data to be sent, relays which themselves have no data, or destinations where data from sources are to be decoded at. The nodes are operating in the same frequency range, and hence every node can receive the transmissions from all other nodes.

1. *Cooperative routing*: Let node 1 be the source, nodes 2–5 relays, and node 6 the destination. When the nodes cooperate (e.g., node 1 can transmit to nodes 2–6 simultaneously) to transmit data from the source to the destination, what do we mean by a route? How do we find an optimal (rate-maximizing) route?

2. *Myopic cooperation*: Consider the same setting. What rates are achievable when the nodes can only cooperate partially (e.g., node 1 knows the presence of only nodes 2 and 3)? What is the trade-off between partial cooperation and achievable rates?

3. *Correlated sources*: Consider only nodes 1–3, and let nodes 1 and 2 be the sources with correlated messages and node 3 the destination for both the sources. Since nodes 1 and 2 can receive each other's transmissions, they are said to receive *feedback* from the channel. For this channel, we are interested in the following: What are the different ways (coding strategies) for the nodes to cooperate to send correlated data to the destination? What are the achievable rate regions of these coding strategies?

These questions will be made more precise in the sequel.

1.3 Motivations and Contributions

Now, we motivate these three problems. We base our analyses on simple networks, e.g., the single-source single-destination network, as having too many parameters to analyze in the multiple-source multiple-relay multiple-destination network may hinder our understanding of the network and may obscure certain observations.

1.3.1 Cooperative Routing

First of all, we study how to optimally route data from the source to the destination in cooperative multiple-terminal wireless networks, i.e., finding a rate-maximizing route, through relays, for a source-destination pair.

In multiple-terminal wireless networks, two important factors that determine the transmission rate are *who* participate in the cooperation and *how* they facilitate data transmission between a source and destination pair. The former leads to the *routing* problem and the latter the *coding* problem. These two problems are often intertwined, i.e., the choice of code (and hence the transmission rate) depends on the route chosen. From an information-theoretic view, the problem can be translated to finding the optimal route and the optimal channel input probability density function (or input distribution).

With rich forms of cooperation among the nodes to transport data from the source to the destination, it is difficult to describe data paths using the traditional notion of a route in which data *hops* from one node to another. Hence, we propose a new definition for a route. Unfortunately, routing algorithms designed for the conventional non-cooperative data transmission are no longer optimal (rate-maximizing) when the nodes are allowed to cooperate.

A brute force way to determine the optimal route and the optimal input distribution is by finding the rates of all possible routes with all possible input distribu-

tions, and selecting the pair that gives the highest rate. This combined optimization is certainly not efficient. These optimizations can be much simplified if they can be separated.

We investigate if the optimization of the route can be separated from the optimization of the input distribution, and how to find an optimal route. As a first step toward understanding the problem, we consider the single-flow network, modeled by the multiple-relay channel (MRC) (Gupta & Kumar, 2003; Xie & Kumar, 2005), i.e., a single-source single-destination network with many relays. We choose the MRC to investigate the routing problem as it contains relays through which different routes can be compared. We study the routing problem for a class of coding strategies: decode-forward (DF) (Cover & El Gamal, 1979; Xie & Kumar, 2005), which achieves the capacity of the MRC when each relay must fully decode the source messages.

Our contributions are as follows:

1. We construct an algorithm, the nearest neighbor set algorithm (NNSA) (Ong & Motani, 2007a,b), which outputs a set of routes that contains an optimal route for the static Gaussian MRC without having to optimize the input distribution.

2. We show that a shortest route that can achieve the maximum rate is contained in at least one of the outputs of the NNSA.

3. We show that the NNSA is optimal in fading channels in the sense that it finds a route that maximizes the ergodic rate.

4. We construct a heuristic algorithm, the the maximum sum-of-received-power algorithm (MSPA), which disregards the input distribution and finds near-optimal routes in polynomial time.

5. We show by numerical calculations that the MSPA is able to find an optimal route with high probability.

The advantage of these routing algorithms is two-fold. Firstly, they show that routing and coding optimizations can be separated under certain conditions, e.g., when the NNSA outputs one route or when the MSPA finds an optimal route. Secondly, the algorithms enable us to find an optimal route without going through the complex brute force search.

1.3.2 Myopic Cooperation

Secondly, we investigate how to code and what rates are achievable in cooperative multiple-terminal wireless networks when every node is only allowed to *partially* cooperate with only a few nodes.

In the information theoretic literature, limits to transmission rates are found assuming that all nodes can fully cooperate, in both encoding and decoding. We term this *omniscient coding*. We often assume ideal operating conditions, e.g., unlimited processing powers at the nodes, perfect synchronization among all transmitters and receivers. This full cooperation makes practical code design in a large network difficult. Hence, we investigate how much worse (in terms of the transmission rate) if we allow only partial cooperation among the nodes, which we term *myopic coding* (Ong & Motani, 2005a,b, 2008).

In terms of code design, utilizing local information leads to a relatively simpler optimization. In terms of operation, myopic coding provides more robustness to topology changes and does not require the whole network to be synchronized. It also mitigates the high computational complexity and large buffer/memory requirements of processing under omniscient coding.

We choose the MRC to investigate partial cooperation in multiple-terminal networks as it contains relays through which we can compare different levels of cooperation. Our contributions are as follows:

1. We construct random codes for the myopic version of DF (Ong & Motani, 2005a,b, 2008) for the MRC with different levels of cooperation.

2. We derive achievable rates of myopic DF for or the discrete memoryless, the

static Gaussian, and the fading MRC.

3. We show that including a few nodes into the cooperation increases the transmission rate significantly, often making it close to that under full cooperation.

4. We show that achievable rates of myopic coding may be as large as that of omniscient coding in the low transmitted-signal-to-noise ratio regime.

5. We show that in the MRC, myopic DF can achieve rates bounded away from zero even as the network size grows to infinity.

1.3.3 Correlated Sources

Lastly, we investigate how to code and what rates are achievable in cooperative multiple-terminal wireless networks where the sources have correlated data. One example of networks with correlated sources is the wireless sensor network, where multiple sensors measure the environment and send possibly correlated data to their respective destinations. The sensors' measurements are possibly correlated as they are located in close proximity and are measuring the same environment.

To study networks with correlated sources, we need a network with more than one source. In addition, to study cooperation among the sources, we allow them to receives different feedback from the channel. We consider the simplest case, where there are two correlated sources and one destination. We term this channel the three-node *multiple-access channel with feedback and correlated sources* (MACFCS) (Ong & Motani, 2005c, 2006b, 2007d). We construct different coding strategies for this channel, showing different ways in which the nodes can cooperate, and explore the pros and cons of these strategies.

Our contributions are as follows:

1. We derive an outer bound on the capacity of the MACFCS (Ong & Motani, 2005c, 2006b, 2007d).

2. We construct two new coding strategies for the MACFCS, where the nodes cooperate by either fully decoding or compressing each other's data.

3. We derive achievable rate regions of these coding strategies for the discrete memoryless and the static Gaussian MACFCS.

4. We compare achievable rate regions of these strategies to that of existing strategies, e.g., channel coding for the MAC and the multi-hop strategy, and discuss the pros and cons of different coding strategies in different channel conditions.

5. We show that the outer bound on the capacity of the MACFCS is achievable under certain source correlation structures and channel topologies.

1.4 List of Publications

Part of the material in this thesis was published in the following journals:

1. Ong L. & Motani M., "Myopic Coding in Multiterminal Networks", *IEEE Transactions on Information Theory*, Volume 54, Number 7, pages 3295–3314, July 2008.

2. Ong L. & Motani M., "Coding Strategies for Multiple-Access Channels with Feedback and Correlated Sources ", *IEEE Transactions on Information Theory, Special Issue on Models, Theory & Codes for Relaying & Cooperation in Communication Networks*, Volume 53, Number 10, pages 3476–3497, October 2007.

and was presented at the following conferences:

1. Ong L. & Motani M., "Optimal Routing for the Decode-and-Forward Strategy in the Gaussian Multiple Relay Channel", *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Acropolis Congress and Exhibition Center, Nice, France, pages 1061–1065, June 24–29 2007.

2. Ong L. & Motani M., "Optimal Routing for Decode-and-Forward based Cooperation in Wireless Networks", *Proceedings of the Fourth Annual IEEE*

Communications Society Conference on Sensor, Mesh, and Ad Hoc Communications and Networks (SECON 2007), San Diego, California, pages 334–343, June 18-21 2007.

3. Ong L. & Motani M., "The Multiple Access Channel with Feedback and Correlated Sources", *Proceedings of the 2006 IEEE International Symposium on Information Theory (ISIT 2006)*, The Westin Seattle, Seattle, Washington, pages 2129–2133, July 9–14 2006.

4. Ong L. & Motani M., "Achievable Rates for the Multiple Access Channels with Feedback and Correlated Sources", *Proceedings of the 43rd Annual Allerton Conference on Communication, Control, and Computing*, Allerton House, the University of Illinois, September 28–30 2005.

5. Ong L. & Motani M., "Myopic Coding in Multiple Relay Channels", *Proceedings of the 2005 IEEE International Symposium on Information Theory (ISIT 2005)*, Adelaide Convention Centre, Adelaide, Australia, pages 1091-1095, September 4–9 2005.

6. Ong L. & Motani M., "Myopic Coding in Wireless Networks", *Proceedings of the 39th Conference on Information Sciences and Systems (CISS 2005)*, John Hopkins University, Baltimore, MD, March 16–18 2005.

1.5 Organization

The structure of this thesis is depicted in Fig. 1.2. In this chapter, we have given a brief introduction to the three problem areas that we will be investigating and motivated them. We have also included our main contributions of this thesis in this chapter. In Chapter 2, we review the definition of the MRC and rates achievable by DF for the MRC, and define what a route is in the cooperative scenario.

In Chapters 3–5, we present the main findings of this thesis in the following areas respectively: cooperative routing, myopic cooperation, and correlated sources. In

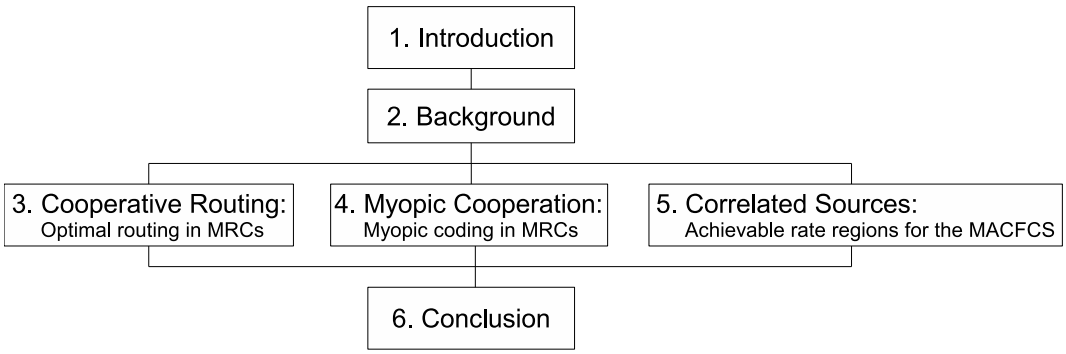


Figure 1.2: The structure of this thesis.

Chapter 3, we construct the NNSA to find optimal routes for DF for the static Gaussian MRC. We show that a shortest rate-maximizing route is contained in one of the routes output by the NNSA. Under certain conditions, the NNSA outputs a large set of routes, and this makes the route optimization runs in factorial time. Hence, we propose a heuristic algorithm, the MSPA that runs in polynomial time and finds an optimal route with high probability. In Chapter 4, we first define myopic coding, in which the communication of the nodes is constrained in such a way that a node communicates with only a few other nodes in the network. We discuss a few advantages of myopic coding over omniscient coding. We construct random codes for the myopic version of DF for the MRC with different levels of cooperation. We derive achievable rates of myopic DF for the discrete memoryless, the static Gaussian, and the fading MRC. We compare the rates achievable via different levels of cooperation, and investigate the rates achievable by myopic DF when the number of nodes in the channel grows large. In Chapter 5, we derive an outer bound on the capacity of the MACFCS. We then construct a few coding strategies for the MACFCS and derive achievable rate regions for these coding strategies. We combine existing coding strategies for other channels and see how it can be used in the MACFCS. We compare the rate regions of different coding strategies under different channel conditions and source correlation structures.

We conclude the thesis in Chapter 6.

Chapter 2

Background

We mentioned in the previous chapter that as analyzing multiple-source multiple-destination (multiple-flow) networks is difficult, we attempt to understand the problem better by focusing on simpler networks: the multiple-relay channel (MRC) and the multiple-access channel with feedback and correlated sources (MACFCS). In this chapter, we review the definition of the discrete memoryless MRC and the Gaussian channel, propose a new definition of a route, and review the decode-forward coding strategy (DF) for the MRC. DF is used to illustrate many concepts in this thesis. We present the rates achievable by DF for the discrete memoryless and the static Gaussian MRC in this chapter and extend the concept of DF to the MACFCS in Chapter 5.

2.1 The Multiple-Relay Channel (MRC)

The single-relay channel (SRC) (first introduced by van der Meulen (1971)) consists of three nodes: the source, the relay, and the destination. The source sends data to the destination with the help of the relay. To date, the largest achievable region for the SRC is due to Cover & El Gamal (1979), who constructed two coding strategies, commonly referred to as *decode-forward* (DF) and *compress-forward* (CF). Chong *et al.* (2007) recently introduced a different decoding technique to give a potentially larger achievable region for the SRC. The SRC was extended to

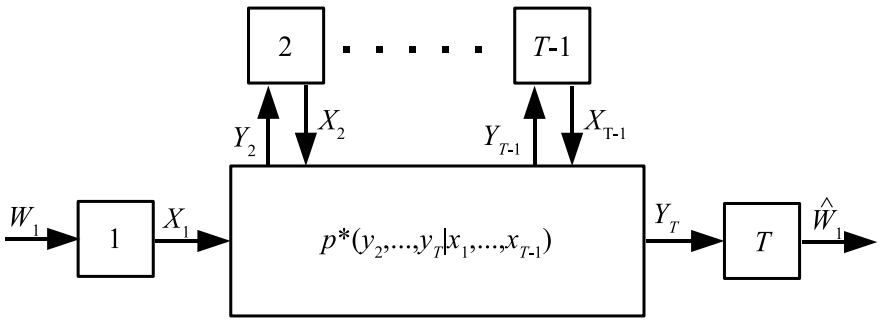


Figure 2.1: The T-node MRC.

the MRC by Gupta & Kumar (2003) and Xie & Kumar (2005), who presented an achievable rate region based on DF. The capacity of the MRC is not known except for special cases, including the degraded MRC (Xie & Kumar, 2005) (achievable by DF), the phase fading MRC where the relays are within a certain distance from the source (Kramer *et al.*, 2005) (achievable by DF), and the mesh network (Ong & Motani, 2006a, 2007c) (achievable by CF). The terms "coding" and "coding strategy" are used interchangeably in this thesis.

The MRC captures the single-flow scenario in the multiple-source multiple-destination network. The relevance of the MRC in multiple-flow networks is as follows:

1. In a multiple-flow network where the flows are allocated orthogonal channels: Each flow can be modeled as an independent MRC.

2. In a multiple-flow network with existing flows: If we wish to add a new flow, this new flow can be modeled by an MRC with the interference from other flows included in the receiver noise.

2.1.1 The Discrete Memoryless MRC

Now, we review the the definition of MRC. Fig. 2.1 depicts the T-node MRC, with node 1 being the source and node T the destination. Nodes 2 to $T-1$ are purely relays. Message W is generated at node 1 and is to be sent to node T. A MRC

can be completely described by the channel distribution

$$p^*(y_2, y_3, \ldots, y_T | x_1, x_2, \ldots, x_{T-1}) \tag{2.1}$$

on $\mathcal{Y}_2 \times \mathcal{Y}_3 \times \cdots \times \mathcal{Y}_T$, for each $(x_1, x_2, \ldots, x_{T-1}) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_{T-1}$. $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{T-1}, \mathcal{Y}_2, \mathcal{Y}_3, \ldots, \mathcal{Y}_T$ are finite sets. In this thesis, we only consider memoryless and time invariant channels (Kramer *et al.*, 2005), which means

$$p(y_{2i}, \ldots, y_{Ti} | x_1^i, \ldots, x_{T-1}^i, y_2^{i-1}, \ldots, y_T^{i-1}) = p^*(y_{2i}, \ldots, y_{Ti} | x_{1i}, \ldots, x_{(T-1)i}), \tag{2.2}$$

for all i.

We use the following notation: x_i denotes an input from node i into the channel, x_{ij} denotes the j-th input from node i into the channel, y_{ij} denotes the j-th output from the channel to node i, and $x_t^i = x_{t1}, x_{t2}, \ldots, x_{ti}$. We denote a block of n inputs from node i by $\mathbf{x}_i$. Similarly, $\mathbf{y}_t$ is a block of n channel outputs to node t. In addition, $\mathbf{x}_{ij}$ and $\mathbf{y}_{tj}$ denote the j-th block of inputs from node i and the j-th block of channel outputs to node t respectively.

We denote the T-node MRC by the tuple

$$\Big(\mathcal{X}_1 \times \cdots \times \mathcal{X}_{T-1}, p^*(y_2, \ldots, y_T | x_1, \ldots, x_{T-1}), \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T\Big). \tag{2.3}$$

In the MRC, the information source at node 1 emits random letters W, each taking on values from a finite set of size M, that is $w \in \{1, ..., M\} \triangleq \mathcal{W}$. We consider each n uses of the channel as a block.

Definition 1 *A sequence of codes* $\Big\{f_1, \{f_{ti}, 2 \le t \le T-2\}_{i=1}^n, g_T, n\Big\}$ *for a* T*-node MRC comprises of an integer* n*,*

- *An encoding function at node 1,* $f_1 : \mathcal{W} \to \mathcal{X}_1^n$*, which maps a source letter to a codeword of length* n*.*
- *Encoding functions at node* t*,* $f_{ti} : \mathcal{Y}_t^{i-1} \to \mathcal{X}_t, i = 1, 2, \ldots, n$ *and* $t = 2, 3, \ldots,$

$T-1$, *such that* $x_{ti} = f_{ti}(y_{t1}, y_{t2}, \ldots, y_{t(i-1)})$, *which map past received signals to the signal to be transmitted into the channel.*

- *A decoding function at the destination,* $g_T : \mathcal{Y}_T^n \to \mathcal{W}$, *such that* $\hat{w} = g_T(y_T^n)$, *which maps received signals of length* n *to a source letter estimate* $\hat{W}$.

Definition 2 *On the assumption that the source letter* W *is uniformly distributed over* $\{1, ..., M\}$, *the average error probability is defined as*

$$P_e = \Pr\{\hat{W} \neq W\}. \tag{2.4}$$

Definition 3 *The rate*

$$R \leq \frac{1}{n} \log M \tag{2.5}$$

is achievable if, for any $\epsilon > 0$, *there is at least a sequence of codes* $\left\{f_1, \{f_{ti}, 2 \leq t \leq T-2\}_{i=1}^n, g_T, n\right\}$ *such that* $P_e < \epsilon$.

For a set of nodes $\mathcal{T} = \{t_1, t_2, \ldots, t_{|\mathcal{T}|}\}$, we define $X_{\mathcal{T}} = (X_{t_1}, X_{t_2}, \ldots, X_{t_{|\mathcal{T}|}})$. We denote the set of all relays in the MRC by $\mathcal{R} = \{2, 3, \ldots, T-1\}$.

2.2 More Definitions

The following definition and lemma are taken from Cover & Thomas (1991, p. 384 & 386).

Definition 4 *Consider a finite collection of random variables* $(X_1, X_2, \ldots, X_k)$ *with some fixed joint distribution* $p(x_1, x_2, \ldots, x_k)$. *Let* S *denote an arbitrarily ordered subset of these random variables, and consider* n *independent copies of* S.

$$\Pr\{\mathbf{S} = \mathbf{s}\} = \prod_{i=1}^{n} \Pr\{S_i = s_i\}. \tag{2.6}$$

The set $\mathcal{A}_\epsilon^n$ *of* ϵ*-typical* n*-sequences* $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ *is defined as*

$$\mathcal{A}_\epsilon^n(X_1, X_2, \ldots, X_k) = \left\{ (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) : \left| -\frac{1}{n} \log p(\mathbf{s}) - H(S) \right| < \epsilon, \right.$$
$$\left. \forall S \subseteq \{X_1, X_2, \ldots, X_k\} \right\}. \quad (2.7)$$

Lemma 1 *For any* $\epsilon > 0$ *and for sufficiently large* n, $|\mathcal{A}_\epsilon^n(S)| \leq 2^{n(H(S)+\epsilon)}$.

2.3 The Gaussian Channel

We consider a flat fading Gaussian channel $\mathcal{T}$ with

$$Y_t = \sum_{i \in \mathcal{T} \setminus \{t\}} \sqrt{\lambda_{it}} X_i + Z_t, \quad (2.8)$$

where X_i is a random variable with *per-block* energy constraint, meaning,

$$\frac{1}{n} \sum_{k=1}^{n} E[X_{ik}^2] \leq P_i. \quad (2.9)$$

Z_t, the receiver noise at node t, is an independent zero-mean Gaussian random variable with variance N_t, i.e., $E[Z_t] = 0$ and $E[Z_t^2] = N_t$. λ_{it} is the channel gain from node i to node t. In this thesis, we consider both large scale fading and small scale fading (Sklar, 1997):

$$\lambda_{ij} = \nu_{ij} h_{ij}, \quad (2.10)$$

where h_{ij} are large scale fading components due to signal attenuation or path loss. We assume that the large scale fading components are constants in the network. This is applicable when the nodes are stationary. We assume that all h_{ij} are known to all transmitters and receivers. $r_{ij} = \sqrt{\nu_{ij}} \geq 0$ are small scale fading envelopes due to multi-path. Also, we assume that all r and Z are independent.

Definition 5 *We define the received-signal-to-noise ratio (rSNR) of a pair of*

transmitter i and receiver j as

$$\gamma_{ij} = \frac{E[\lambda_{ij}]E[X_i^2]}{E[Z_j^2]}. \tag{2.11}$$

Definition 6 *We define the transmitted-signal-to-noise ratio (tSNR) of a pair of transmitter i and receiver j as*

$$\psi_{ij} = \frac{E[X_i^2]}{E[Z_j^2]}. \tag{2.12}$$

2.3.1 Large Scale Fading Model

Let us now investigate large scale fading. Consider a point-to-point noiseless static channel from node i to node j, i.e., $N_j = 0$, and $r_{ij} = 1$. Using Friis free space path loss model, the channel gain is given by

$$h_{ij} = \frac{P_{rj}}{P_i} = \frac{G}{(4\pi f d_{ij})^2}, \tag{2.13}$$

where P_{rj} is the received power, P_i the transmit power, G the antenna gain, f the carrier frequency, d_{ij} the Euclidean distance between the transmitter and the receiver. In non-free space, different models are used to model the signal propagation attenuation in different environments. However, in most models, h_{ij} is inversely proportional to $d_{ij}^{-\eta}$, where η ranges from 2 to 8. Capturing the main characteristic of how h_{ij} varies with distance, one can simplify these path loss models to the following standard path loss model for the channel from node i to node j.

$$h_{ij} = \kappa d_{ij}^{-\eta}, \tag{2.14}$$

where η is the path loss exponent, and $\eta \geq 2$ with equality for free space transmission. κ is a positive constant as far as the analyses in this section are concerned. In this thesis, we set $\eta = 2$ and $\kappa = 1$. The standard path loss model is a widely ac-

cepted model and commonly used in the information theoretic literature (Gatspar & Vetterli, 2005; Gupta & Kumar, 2000; Kramer *et al.*, 2005; Toumpis & Goldsmith, 2003).

2.3.2 Small Scale Fading Model

In wireless channels, even when the nodes are stationary, the channel gains vary due to changes in the environment. These are captured in the small scale fading components. In this thesis, we consider multi-path fading. The received signal at node j from node i is subject to *fading envelope* $r_{ij} = \sqrt{\nu_{ij}} \geq 0$. In other words, the received power at node j from node i is subject to *fading power* ν_{ij}. We denote the *average fading power* by $\Omega_{ij} = E[\nu_{ij}]$.

Example 1 *For Rayleigh fading, the fading power is a random variable with the following probability density function (p.d.f.).*

$$p(\nu_{ij}) = \begin{cases} \frac{1}{\Omega_{ij}} \exp\left(\frac{-\nu_{ij}}{\Omega_{ij}}\right) & , \nu_{ij} \geq 0 \\ 0 & , \text{otherwise} \end{cases} . \tag{2.15}$$

In Section 3.7, we consider fading channels where ν_{ij} are random variables. In the rest of the thesis, we consider static channels, i.e., ν_{ij} are constants. Without loss of generality, we assume $\Omega_{ij} = E[\nu_{ij}] = 1$ for all channels. To model channels with different fading power, we can always normalize them to 1 by changing d_{ij} accordingly.

2.4 Definition of a Route

Now, we define what we mean by a route in a network. Kurose & Ross (2003) define a route as "the *path* taken by a datagram between source and destination". The datagram *hops* from one node to the next node, capturing the scenario in which a node receives data only from a node *behind* (or *upstream*) and forwards data

only to the node *in front* (or *downstream*). However, in the cooperative coding paradigm, data do not flow from one node to another; rather they "travel" from many to many in a complex cooperative way. To describe the flow of information in these new modes of cooperation, we re-define a route as follows.

Definition 7 *The route taken by a message from the source to the destination is an ordered set of nodes involved in encoding/transmitting of the message. The sequence of the nodes in the route is determined by the order in which nodes' transmit signals first depend on the message. The node that the message is intended for (the destination), though does not transmit, is the last node in the route.*

We define the route with respect to the encoding sequence rather than the decoding sequence in order to capture the *active participation* of the nodes. Consider a 4-node network with node 1 being the source and node 4 the destination. Node 1 sends a message w. Node 2 and 3 both fully decode the message. But only node 2 forwards the message w to node 4. In this case, the route taken is $\{1, 2, 4\}$ according to our definition, but not $\{1, 2, 3, 4\}$. This agrees with our common notion of a route, as node 3 does not participate in aiding the message forwarding and shall not be considered as part of the route. However, if node 3 is also a destination of another *flow*, then the route for that flow is $\{1, 3\}$.

This new definition of a route generalizes the usual notion of a multi-hop route to the multiple-terminal cooperative scenario, where nodes cooperate intimately with each other. It is clear that this definition reduces to the usual notion of a route in the multi-hop case. Note that this definition is applicable in the network coding (Ahlwsede *et al.*, 2000) scenario, where a node forwards functions of previously received data.

Remark 1 *If a group of nodes transmit simultaneously, then they can be ordered arbitrarily within the group. For example, consider a four-node network, in which node 1 first broadcasts the message, and then nodes 2 and 3 listen and simultaneously transmit to node 4. The route here can be described by* $\{1, 2, 3, 4\}$ *or* $\{1, 3, 2, 4\}$.

Refer to Example 2 for a route for DF.

We define the set of all possible routes from the source (node 1) to the destination (node T) by $\Pi(\mathcal{T}) = \Big\{\{m_1, m_2, \ldots, m_{|\mathcal{M}|}\} : m_2, \ldots, m_{|\mathcal{M}|-1}$ are all possible selections and permutations of the relays (including the empty set), $m_1 = 1, m_{|\mathcal{M}|} = T\Big\}$.

2.5 The Decode-Forward Coding Strategy (DF)

2.5.1 DF for the Discrete Memoryless MRC

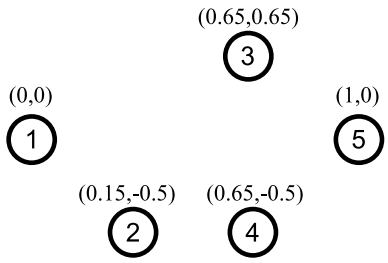
In DF (first introduced for the SRC by Cover & El Gamal (1979)), the source transmits to all relays and the destination. The relays fully decode the data sent by the source, and help it to forward the data to the destination. It is also known as the *decode-and-forward* strategy. DF for the MRC can achieve rates up to that given in the following theorem.

Theorem 1 *DF for the MRC achieves any rate up to*

$$R_{\mathrm{DF}} = \max_{\mathcal{M}\in\Pi(\mathcal{T})} \max_{p(x_1,\ldots,x_{T-1})} \min_{m_t\in\mathcal{M}\setminus\{1\}} I(X_{m_1}, \ldots, X_{m_{t-1}}; Y_{m_t} | X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c}), \tag{2.16}$$

where $\mathcal{M}^c = \mathcal{T} \setminus \mathcal{M}$.

Proof 1 (Proof of Theorem 1) *The DF rate in Theorem 1 follows that by Xie & Kumar (2005, Thm 3.1) and Kramer* et al. *(2005, Thm 1) with some modifications. Achievable rates of DF for the MRC was first derived by Xie & Kumar (2005, Thm 3.1) by assuming that data flow from node 1 to node 2, and so on until node T, the destination. Kramer* et al. *(2005, Thm 1) noted that higher achievable rates are possible by choosing the best permutation of the nodes through which the data flow. In this thesis, we argue that the achievable rates can be further increased by selecting which nodes to participate in data forwarding as well as permutating the selected nodes, which we call a route. The latter is a more relaxed constraint as it*



(a) Network topology

Route	DF Rate
$\{1,2,3,4,5\}$	2.40029
$\{1,2,4,3,5\}$	2.58613
$\{1,3,2,4,5\}$	1.84097
$\{1,3,4,2,5\}$	1.84097
$\{1,4,2,3,5\}$	1.99411
$\{1,4,3,2,5\}$	1.99411
$\{1,2,4,5\}$	2.61819

(b) Routes and Rates

Figure 2.2: Comparing DF rates on different routes. $P_i = 10, i \in \mathcal{M} \setminus \{5\}, N_j = 1, j \in \mathcal{M} \setminus \{1\}$, $\kappa = 1$ $\eta = 2$ and $r_{ij} = 1, \forall i, j$..

	Block 1	Block 2	Block 3	Block 4
Node 1	$x_1(w_1,0,0)$	$x_1(w_2,w_1,0)$	$x_1(w_3,w_2,w_1)$	$x_1(w_4,w_3,w_2)$
Node 2		$x_2(w_1,0)$	$x_2(w_2,w_1)$	$x_2(w_3,w_2)$
Node 3			$x_3(w_1)$	$x_3(w_2)$

Figure 2.3: An example of the DF encoding function.

does not require all relays to be in the route. When $|\mathcal{M}| < T$, the minimization is taken over a smaller set of nodes, and the maximum DF rate could be higher.

Now, we show by using an example that using forcing all nodes to be in the route is not optimal. Refer to the Gaussian MRC depicted in Fig. 2.2(a). We compute DF rates for different routes. The first six routes include all relays and all possible relay permutations. The last route $\{1,2,4,5\}$, which omits node 3, achieves DF rate higher than any other route that includes all relays.

For DF, the route is also the order for which the messages are decoded at the relays. By definition, node 1 is the first node in the route. Let us see an example of a route in the four-node MRC, and the encoding and decoding steps.

Example 2 *Consider DF for the four-node MRC. One way of encoding and decoding is as follows (refer Fig. 2.3). We use w_i to denote the i-th source message.*

1. *At the beginning of block 1, node 1 receives the first source message, w_1. It transmits, $\mathbf{x}_1(w_1)$.*
2. *At the end of block 1, all nodes receive a noisy version of $\mathbf{x}_1(w_1)$. But only node 2 decodes w_1.*
3. *In block 2, node 2 sends $\mathbf{x}_2(w_1)$. Receiving a new source message w_2, node 1 sends $\mathbf{x}_1(w_2, w_1)$, which is a function of the new and the old source message.*
4. *At the end of block 2, node 3 decodes w_1 over two blocks of received signal, i.e., $\mathbf{y}_{3,1}$ in block 1 and $\mathbf{y}_{3,2}$ in block 2.*
5. *Similarly, node 4 decodes w_1 over 3 blocks of received signal.*

Looking at how the transmitted signals first depend on w_1, the route for this code is $\{1, 2, 3, 4\}$. By definition, node 4 is the last node in the route, though it does not transmit.

Definition 8 *For a certain input distribution $p = p(x_1, \ldots, x_{T-1})$, we define the rate* supported *by route $\mathcal{M}$ as:*

$$R_{\mathcal{M}}(p) = \min_{m_t \in \mathcal{M} \setminus \{1\}} R_{m_t}(\mathcal{M}, p), \tag{2.17}$$

where $R_{m_t}(\mathcal{M}, p)$ is the reception rate *at node m_t given by*

$$R_{m_t}(\mathcal{M}, p) = I(X_{m_1}, \ldots, X_{m_{t-1}}; Y_{m_t} | X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c}). \tag{2.18}$$

So, the maximum DF rate can also be written as

$$R_{\mathrm{DF}} = \max_{\mathcal{M} \in \Pi(\mathcal{T})} \max_{p} R_{\mathcal{M}}(p). \tag{2.19}$$

2.5.2 DF with Gaussian Inputs for the Static Gaussian MRC

In DF for the Gaussian channel, a node splits its total transmission power between sending new information and repeating what the relays *in front* (or *downstream*, i.e., toward the destination) send. Every node decodes signals from all the nodes *behind* (or *upstream*, i.e., toward the source). At the same time, it cancels interfering transmissions from all the downstream nodes.

Using DF with jointly Gaussian inputs and route $\mathcal{M}$, node m_i transmits

$$X_{m_i} = \sum_{j=i+1}^{|\mathcal{M}|} \sqrt{\alpha_{m_i m_j} P_{m_i}} U_{m_j}, \tag{2.20}$$

for $\sum_{j=i+1}^{|\mathcal{M}|} \alpha_{m_i m_j} \leq 1$ and $\alpha_{m_i m_j} \geq 0$, $\forall i = 1, \ldots, |\mathcal{M}| - 1$. U_{m_j} are independent Gaussian random variables with unit variance. $\{\alpha_{m_i m_j} | j = i+1, \ldots, |\mathcal{M}|\}$ are the power splits of node m_i, allocating portions of its transmit power to transmit independent *sub-codewords* U_{m_j}.

In the static Gaussian MRC, DF on route $\mathcal{M}$ can achieve rates up to

$$R_{\mathcal{M}}(p) = \min_{m_t \in \mathcal{M} \setminus \{m_1\}} R_{m_t}(\mathcal{M}, p), \tag{2.21}$$

where $R_{m_t}(\mathcal{M}, p)$ is given by

$$R_{m_t}(\mathcal{M}, p) = \frac{1}{2} \log \left(1 + \sum_{j=2}^{t} \left(\sum_{i=1}^{j-1} \sqrt{\alpha_{m_i m_j} \gamma_{m_i m_t}} \right)^2 \right). \tag{2.22}$$

Any Gaussian distribution p can be completely determined by $\{\alpha_{m_i m_j}\}$ and $\{P_i\}$. Throughout this thesis, logarithm base 2 is used and hence the units of rate are bits per channel use.

Remark 2 *It has been shown by Kramer et al. (2005) that jointly Gaussian inputs*

achieve R_{DF} in the static Gaussian MRC, i.e.,

$$\max_{\mathcal{M}\in\Pi(\mathcal{T})}\max_{p\in\mathcal{P}_{\text{Gauss}}} R_{\mathcal{M}}(p) = \max_{\mathcal{M}\in\Pi(\mathcal{T})}\max_{p\in\mathcal{P}_{\text{All}}} R_{\mathcal{M}}(p) = R_{\text{DF}}, \tag{2.23}$$

where $\mathcal{P}_{\text{All}}$ is the set of all possible input distributions and $\mathcal{P}_{\text{Gauss}}$ the set of all jointly Gaussian distributions.

If the nodes transmit independent Gaussian inputs, i.e., we set $\alpha_{m_i m_t} = 1, \forall m_i \in \mathcal{M}\backslash\{1\}, \forall t = i{+}1$ and $\alpha_{m_i m_t} = 0, \forall t \neq i{+}1$, the reception rate at node $m_t \in \mathcal{M}\backslash\{1\}$ is

$$R_{m_t}(\mathcal{M}, p') = \frac{1}{2}\log\left(1 + \sum_{i=1}^{t-1}\gamma_{m_i m_t}\right). \tag{2.24}$$

where $p' = p(x_1)\cdots p(x_{T-1})$ and $X_i \sim \mathcal{N}(0, P_i)$. The rate supported by route $\mathcal{M}$ is thus

$$R_{\mathcal{M}}(p') = \min_{m_t\in\mathcal{M}\backslash\{1\}} R_{m_t}(\mathcal{M}, p'). \tag{2.25}$$

2.5.3 Why DF?

In the chapters on cooperative routing and myopic cooperation, we base our discussions on DF. In the chapter on correlated sources, we derive a DF-based coding strategy for the MACFCS. DF is an important coding strategy for the reasons given below.

1. DF achieves the capacity of the degraded MRC (Xie & Kumar, 2005).
2. DF achieves the capacity of the phase fading MRC when the relays are positioned within a certain distance from the source (Kramer *et al.*, 2005).
3. DF achieves the capacity of the MRC where all relays must fully decode all source messages.
4. Rates achievable by DF are lower bounded by that achievable by point-to-point multi-hop strategy (Ong & Motani, 2007a).

5. A DF-based coding strategy (derived in this thesis) achieves the capacity of MACFCS in several conditions.

6. There exist many DF-based low-density parity-check (LDPC) codes (Chakrabarti *et al.*, 2007; Ezri & Gastpar, 2006; Khojastepour *et al.*, 2004; Razaghi & Yu, 2006) and Turbo codes (Zhang *et al.*, 2004; Zhao & Valenti, 2003) which perform close to the information-theoretic DF rate. Analyses of DF may be applied directly or indirectly to these codes.

Chapter 3

Optimal Routing in Multiple-Relay Channels

Consider the multiple-terminal wireless network in which a node can overhear the transmissions of all other nodes transmitting in the same frequency band and in communication range. The nodes can *cooperate* to send information from sources to destinations, e.g., via cooperative relaying (Cover & El Gamal, 1979; Kramer *et al.*, 2005; Xie & Kumar, 2005) and opportunistic routing (Biswas & Morris, 2004). The gain from cooperation has been shown in information theoretic analyses (Ong & Motani, 2005a,b, 2008) and demonstrated in practical implementations (Lim *et al.*, 2006; Sendonaris *et al.*, 2003a,b). As data paths in this cooperative environment are difficult to describe using the traditional notion of a route, we proposed a new definition for a route (see Section 2.4). Unfortunately, routing algorithms designed for conventional non-cooperative multi-hop routing are no longer optimal (rate-maximizing) when the nodes are allowed to cooperate, e.g., via the decode-forward coding strategy (DF) (Cover & El Gamal, 1979; Kramer *et al.*, 2005; Xie & Kumar, 2005), which promises a higher transmission rate compared to multi-hop and even achieves the capacity of a few classes of networks. In this chapter, we propose new routing algorithms to find rate-maximizing routes for DF.

3.1 Problem Statement

We consider the T-node multiple-relay channel (MRC) (Gupta & Kumar, 2003; Xie & Kumar, 2005), in which the nodes are restricted to transmit in certain ways, or more precisely, the input distribution of the nodes are restricted to $p \in \mathcal{P}$. The maximum rate achievable by DF under this restriction is thus

$$R_{\mathrm{DF}}(\mathcal{P}) = \max_{\mathcal{M}\in\Pi(\mathcal{T})} \max_{p\in\mathcal{P}} R_{\mathcal{M}}(p), \tag{3.1}$$

where $\mathcal{T}$ is the set of all the nodes in the channel, and $\Pi(\mathcal{T})$ is the set of all possible routes from the source to the destination. Recall that the rate supported by route $\mathcal{M} = \{m_1 = 1, m_2, \ldots, m_{|\mathcal{M}|} = T\}$, where node 1 is the source and node T is the destination, is

$$R_{\mathcal{M}}(p) = \min_{m_t\in\mathcal{M}\setminus\{1\}} R_{m_t}(\mathcal{M}, p), \tag{3.2}$$

where $R_{m_t}(\mathcal{M}, p)$ is the *reception rate* at node m_t given by

$$R_{m_t}(\mathcal{M}, p) = I(X_{m_1}, \ldots, X_{m_{t-1}}; Y_{m_t} | X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c}). \tag{3.3}$$

Let $(\mathcal{M}^*, p^*), \mathcal{M}^* \in \Pi(\mathcal{T}), p^* \in \mathcal{P}$ be a pair of route and input distribution that achieves the maximum DF rate. We are interested in the following.

1. How to find $\mathcal{M}^*$, and subsequently $R_{\mathrm{DF}}(\mathcal{P})$?
2. Can $\mathcal{M}^*$ and p^* be optimized separately?

Definition 9 *We define the optimal route set for DF with respect to a set of input distributions as*

$$\mathcal{Q}_{\mathrm{DF}}(\mathcal{P}) \triangleq \left\{ \mathcal{M}^* \in \Pi(\mathcal{T}) : \max_{p\in\mathcal{P}} R_{\mathcal{M}^*}(p) = \max_{\mathcal{M}\in\Pi(\mathcal{T})} \max_{p'\in\mathcal{P}} R_{\mathcal{M}}(p') \right\},$$

We define the optimal route set because the rate-maximizing route may not be unique. Then the optimal DF routing problem for a network $\mathcal{T}$ is to find an $\mathcal{M}^* \in \mathcal{Q}_{\mathrm{DF}}(\mathcal{P})$.

We say that the routing and coding optimizations can be *completely* separated if we can find an $\mathcal{M}^* \in \mathcal{Q}_{\text{DF}}(\mathcal{P})$ without needing to consider the input distribution. On the other hand, we say that the routing and coding optimizations are *partially* separated if we can find a set of routes $\mathcal{B}$, without considering the input distribution, that contains at least one optimal route, i.e., $\exists \mathcal{M}^* \in \mathcal{B}$ where $\mathcal{M}^* \in \mathcal{Q}_{\text{DF}}(\mathcal{P})$ and $1 < |\mathcal{B}| < |\Pi(\mathcal{T})|$.

It has been noted by Xie & Kumar (2005) and Kramer *et al.* (2005) that the DF rate depends on the route selected (known as node order/permutation in the papers). Besides the *degraded* case, finding the optimal route is not easy in general. We refer to the strategy of testing all possible routes with all possible input distributions as brute force.

3.2 Contributions

In this chapter, we construct an algorithm, the nearest neighbor set algorithm (NNSA), to find a set of routes that contains an optimal route for DF, without needing to optimize the input distribution. From the set of routes, we then optimize the input distribution for each route to find an optimal route. So, instead of optimizing the input distribution for all possible routes (i.e., combined routing and coding using brute force), we only need to search for an optimal route in the set output by the NNSA. Under certain conditions, the NNSA outputs a single route, which is an optimal route. In this case, we can completely separate routing and coding.

However, the NNSA might output a large set of routes. In the worst case, it outputs all possible routes. This reverts the problem back to combined routing and coding optimization, which runs in factorial time. To solve this problem, we exploit the properties of the algorithm to design a heuristic algorithm, the maximum sum-of-received-power algorithm (MSPA), which disregards the input distribution and outputs a near optimal (or under certain conditions, optimal) route in polynomial time. We show, empirically, that the MSPA finds an optimal route with high

probability. This means, with high probability we can completely separate routing and coding.

A near-optimal route is useful as follows. Firstly, it can be used to calculate a lower bound on the maximum DF rate. Secondly, it can be used in a network to achieve rates close to DF.

We summarize our contributions in this chapter as follows.

1. We construct an algorithm, the nearest neighbor set algorithm (NNSA) (Ong & Motani, 2007a,b), which outputs a set of routes that contains an optimal route for the static Gaussian MRC without having to optimize the input distribution.

2. We show that the optimizations of routing and coding (input distribution) can be *completely* separated under certain scenarios, or *partially* separated.

3. We show that a shortest optimal route is contained in at least one of the outputs of the NNSA.

4. We show that the NNSA is optimal in fading channels with no delay constraint in the sense that it finds a route that maximizes the ergodic rate.

5. We show that the NNSA is not optimal in fading channels with delay constraints in the sense that it does not always find a route that minimizes the outage probability.

6. We construct a heuristic algorithm, the maximum sum-of-received-power algorithm (MSPA), which disregards the input distribution and finds near-optimal routes in polynomial time.

7. We show by numerical calculations that the MSPA finds an optimal route with high probability.

3.2.1 Organization

The rest of this chapter is organized as follows. Before jumping into the algorithms, we derive a few theorems and lemmas in Section 3.3. In Section 3.4, we present two algorithms that find optimal routes, namely the nearest neighbor algorithm (NNA) and the NNSA. We discuss a few interesting observations on the NNSA in Section 3.5. We show that the NNSA contains the shortest optimal route in Section 3.6, and that NNSA is also optimal in fading channels in the sense that it maximizes the ergodic rate in Section 3.7. In Section 3.8, we present a heuristic algorithm that finds an optimal route (with high probability) in polynomial time. We conclude this chapter in Section 3.9.

In the next section, we derived a few theorems and lemmas which will be used in the sequel.

3.3 A Few Theorems and Lemmas

First, we study how channel gains affect the mutual information between the channel inputs and the channel outputs.

Lemma 2 *Consider a Gaussian point-to-point channel with interference and noise:*

$$Y_2 = \sqrt{\lambda} X_1 + Z_2 + V_2, \tag{3.4}$$

where X_1 is the input with power constraint $E[X_1^2] \leq P_1$, Z_2 an independent zero-mean Gaussian random variable with power $E[Z_2^2] = N_2$, V_2 an independent and arbitrarily-distributed interference component with power $E[V_2^2] = P_V$, and $\lambda > 0$. Increasing λ does not necessarily increase $I(X_1; Y_2)$.

Proof 2 (Proof of Lemma 2) *Let us consider the case $N_2 = 0, P_1 = 1, P_V = 4$.*

Let the p.d.f. of X_1 and V_2 be

$$p(x_1) = \begin{cases} \frac{1}{2} & \text{, if } x_1 = 1 \\ \frac{1}{2} & \text{, if } x_1 = -1 \\ 0 & \text{, otherwise} \end{cases} \tag{3.5a}$$

$$p(v_2) = \begin{cases} \frac{1}{2} & \text{, if } v_2 = 2 \\ \frac{1}{2} & \text{, if } v_2 = -2 \text{ .} \\ 0 & \text{, otherwise} \end{cases} \tag{3.5b}$$

We can show that

$$\textit{If } \lambda = 4, I(X_1; Y_2) = H(Y_2) - H(Y_2|X_1) = 1.5 - 1 = 0.5 \text{ bit.} \tag{3.6a}$$

$$\textit{If } \lambda = 1, I(X_1; Y_2) = H(Y_2) - H(Y_2|X_1) = 2 - 1 = 1 \text{ bit.} \tag{3.6b}$$

We see, in this example, that increasing λ decreases $I(X_1; Y_2)$.

Lemma 2 says that for a point-to-point Gaussian channel with non-Gaussian interference, increasing the channel gain does not necessarily increase the mutual information between the channel input and the channel output. However, under the condition described below, we can show that increasing the channel gain increases the mutual information between the channel input and the channel output.

Definition 10 *Consider a function that maps a real number to a real number. We say that the function is* single-peak *if the support of the set of local maxima of the function is a convex set.*

Definition 11 *Consider a memoryless channel with channel inputs $\{x_i\}$ and channel outputs $\{y_j\}$. We say that the* single-peak condition *(SPC) is satisfied iff the output p.d.f. $p(y_j|x_i)$ at every receiver j conditioned on every transmitter i is single-peak.*

Lemma 3 *Consider a point-to-point channel with interference and noise:*

$$Y_2 = \sqrt{\lambda} X_1 + Z_2 + V_2, \tag{3.7}$$

where X_1 is the input with power constraint $E[X_1^2] \leq P_1$, Z_2 an independent and arbitrarily-distributed noise with power $E[Z_2^2] = N_2$, V_2 an independent and arbitrarily-distributed interference component with power $E[V_2^2] = P_V$, and $\lambda > 0$. If the SPC is satisfied, then $I(X_1; Y_2)$ increases monotonically with λ. Furthermore, if the SPC is satisfied and if Z_2 is zero-mean Gaussian, then $I(X_1; Y_2)$ increases strictly with λ. For this channel with one input x_1 and one output y_2, the SPC is satisfied iff $p(y_2|x_1)$ is single-peak.

Proof 3 (Sketch of proof of Lemma 2) *See Appendix A.1.*

For the rest of this chapter, we assume Gaussian channels and we operate the networks under the SPC.

For the general T-node Gaussian channel, we have the following theorem.

Theorem 2 *Consider a Gaussian channel $\mathcal{T}$. Consider a set of nodes $\mathcal{S} \subset \mathcal{T}$ and its complement in the network $\mathcal{S}^c = \mathcal{T} \setminus \mathcal{S}$. For a pair of nodes $j, k \in \mathcal{S}^c$, if*

$$\gamma_{ij} \geq \gamma_{ik}, \forall i \in \mathcal{S}, \tag{3.8}$$

and if the SPC is satisfied, then

$$I(X_{\mathcal{S}}; Y_j | X_{\mathcal{S}^c}) \geq I(X_{\mathcal{S}}; Y_k | X_{\mathcal{S}^c}). \tag{3.9}$$

Here, $\gamma_{ij} = \frac{E[\lambda_{ij}]E[X_i^2]}{E[Z_j^2]}$ is the received-signal-to-noise ratio (rSNR) of a pair of transmitter i and receiver j.

Proof 4 (Proof of Theorem 2) *See Appendix A.2.*

Remark 3 *Choosing the input of all nodes to be jointly Gaussian satisfies the SPC.*

Now, we derive two lemmas which we will need in the later part of the chapter.

Lemma 4

$$I(X_{\mathcal{T}}, X_A; Y | X_B) \geq I(X_A; Y | X_{\mathcal{T}}, X_B). \tag{3.10}$$

Proof 5 (Proof for Lemma 4)

$$I(X_{\mathcal{T}}, X_A; Y|X_B) = H(Y|X_B) - H(Y|X_{\mathcal{T}}, X_A, X_B) \tag{3.11a}$$
$$\geq H(Y|X_{\mathcal{T}}, X_B) - H(Y|X_{\mathcal{T}}, X_A, X_B) \tag{3.11b}$$
$$= I(X_A; Y|X_{\mathcal{T}}, X_B). \tag{3.11c}$$

Definition 12 *For two routes $\mathcal{M}_1$ and $\mathcal{M}_2$, $\mathcal{M}_1 \cup \mathcal{M}_2$ means concatenating route $\mathcal{M}_2$ to the end of route $\mathcal{M}_1$, while preserving the order of nodes in both routes.*

For a route and its extension, we have the following lemma.

Lemma 5 *Consider a route $\mathcal{M}_1$ and its* extension *$\mathcal{M}_3 = \mathcal{M}_1 \cup \mathcal{M}_2$. It follows that for DF with any input distribution p,*

$$R_{\mathcal{M}_1}(p) \geq R_{\mathcal{M}_3}(p). \tag{3.12}$$

Proof 6 (Proof of Lemma 5)

$$R_{\mathcal{M}_3}(p) = \min_{i \in \mathcal{M}_3 \setminus \{1\}} R_i(\mathcal{M}_3, p) \tag{3.13a}$$
$$= \min \left[\min_{i \in \mathcal{M}_1 \setminus \{1\}} R_i(\mathcal{M}_3, p), \min_{j \in \mathcal{M}_2 \setminus \{1\}} R_j(\mathcal{M}_3, p) \right] \tag{3.13b}$$
$$= \min \left[\min_{i \in \mathcal{M}_1 \setminus \{1\}} R_i(\mathcal{M}_1, p), \min_{j \in \mathcal{M}_2 \setminus \{1\}} R_j(\mathcal{M}_3, p) \right] \tag{3.13c}$$
$$\leq \min_{i \in \mathcal{M}_1 \setminus \{1\}} R_i(\mathcal{M}_1, p) \tag{3.13d}$$
$$= R_{\mathcal{M}_1}(p) \tag{3.13e}$$

Now, we are ready to present two algorithms that find optimal routes for DF for the static Gaussian MRC.

3.4 Finding an Optimal Route

In this section, we present two algorithms to find optimal route(s) for any input distribution (which satisfies the SPC) on the static Gaussian MRC.

3.4.1 Nearest Neighbor

First, we define a node (not in $\mathcal{M}$) with the highest received-signal-to-noise ratio (rSNR) from every node in $\mathcal{M}$.

Definition 13 *Node $i \notin \mathcal{M}$ is a* nearest neighbor *with respect to route $\mathcal{M}$ iff*

$$\gamma_{mi} \geq \gamma_{mj}, \quad \forall m \in \mathcal{M}, \forall j \in \mathcal{T} \setminus (\mathcal{M} \cup \{i\}). \tag{3.14}$$

Although the definition does not involve inter-node distance, we use the term nearest neighbor because for a system where all nodes transmit at the same power and are subject to the same receiver noise power, the closer two nodes are, the higher the rSNR between them, i.e., $\frac{1}{d_{ij}^{\eta}} \propto \gamma_{ij}$ as $\eta \geq 2$. Note that a nearest neighbor with respect to a route might not be unique, and might not even exist.

3.4.2 The Nearest Neighbor Algorithm

Now, we state the nearest neighbor algorithm (NNA).

Algorithm 1 (NNA)

1. *First, start with the source node, $\mathcal{M} = \{m_1\}$.*
2. *Find the unique nearest neighbor with respect to the current route $\mathcal{M}$. In other words, pick the unique node $i \notin \mathcal{M}$ such that*

$$\gamma_{mi} \geq \gamma_{mj}, \quad \forall m \in \mathcal{M}, \forall j \in \mathcal{T} \setminus (\mathcal{M} \cup \{i\}), \tag{3.15}$$

   with at least one strict inequality.
3. *If a unique nearest neighbor exists, append the nearest neighbor of $\mathcal{M}$ to the current $\mathcal{M}$, i.e., $\mathcal{M} \leftarrow \mathcal{M} \cup \{i\}$. Else if a unique nearest neighbor does not exist, the algorithm terminates prematurely.*
4. *Steps 2–3 are repeated until the destination, node T, is added into $\mathcal{M}$.*

The algorithm is said to terminate normally if node T is added to the route. At any time, if a unique nearest neighbor does not exists, the algorithm is said to terminate prematurely. If the NNA terminates normally, we have the following theorem.

Theorem 3 *Consider a static Gaussian MRC $\mathcal{T}$, in which node 1 is the source and node T is the destination. If the NNA terminates normally and outputs route $\mathcal{M}^*$, then for any input distribution of the form $p = p(x_1, \ldots, x_{T-1})$ satisfying the SPC,*

$$R_{\mathcal{M}^*}(p) = \max_{\mathcal{M} \in \Pi(\mathcal{T})} R_{\mathcal{M}}(p) \tag{3.16}$$

Proof 7 (Proof of Theorem 3) *See Appendix A.3.*

Remark 4 *We note that for the NNA to terminate normally, one unique nearest neighbor must exist after the addition of each node into the route. In the next section, we extend the NNA to an algorithm which terminates normally given any network topology.*

3.4.3 Nearest Neighbor Set

Before jumping into the algorithm, we define *nearest neighbor set*.

Definition 14 *The nearest neighbor set $\mathcal{N} = \{n_1, n_2, \ldots, n_{|\mathcal{N}|}\}$ with respect to route $\mathcal{M} = \{m_1, m_2, \ldots, m_{|\mathcal{M}|}\}$ is defined as the smallest non-empty set $\mathcal{N}$ where each $n \in \mathcal{N} \subseteq \mathcal{T} \setminus \mathcal{M}$ satisfies the following condition.*

$$\gamma_{mn} \geq \gamma_{ma}, \quad \forall m \in \mathcal{M}, \forall a \in \mathcal{T} \setminus (\mathcal{M} \cup \mathcal{N}), \tag{3.17}$$

with at least one strict inequality for every pair of $(n, a) \in \{(n, a) | n \in \mathcal{N}, a \in \mathcal{T} \setminus (\mathcal{M} \cup \mathcal{N})\}$.

In brief, any node in $\mathcal{M}$ must be closer or at least as close to all nodes in $\mathcal{N}$ than it is to all nodes not in $\mathcal{M} \cup \mathcal{N}$.

3.4.4 The Nearest Neighbor Set Algorithm

Having described the nearest neighbor set, we now present the nearest neighbor set algorithm (NNSA), which terminates normally given any network topology.

Algorithm 2 (NNSA)

1. *First, start with the source node,* $\mathcal{M} = \{m_1\}$.

2. *We find the nearest neighbor set* $\mathcal{N}$. *Now, each element in* $\mathcal{N}$ *is added to the end of* $\mathcal{M}$ *to form one new route. The original route* $\mathcal{M}$ *branches out to* $|\mathcal{N}|$ *routes, shown as follows.*

$$\mathcal{M}_i = \mathcal{M} \cup \{n_i\}, \quad i = 1, \ldots, |\mathcal{N}|. \tag{3.18}$$

3. *For each new route in* (3.18), *step 2 is repeated until the destination is added to all the routes.*

When the algorithm terminates, we end up with many routes. We term these routes *NNSA candidates* and denote the set of all NNSA candidates by $\mathcal{NNSA}(\mathcal{T})$. We calculate the supported rate of each candidate and choose one which gives the highest supported rate. In Appendix A.4, we show how to find NNSA candidates in three wireless networks.

The following theorem says that any NNSA candidate that gives the highest supported rate among all NNSA candidates is an optimal route for DF.

Theorem 4 *Consider a static Gaussian MRC* $\mathcal{T}$, *in which node* 1 *is the source and node* T *the destination. Let the set of NNSA candidates be* $\mathcal{NNSA}(\mathcal{T})$. *Then for any input distribution of the form* $p = p(x_1, \ldots, x_{T-1})$ *satisfying the SPC,*

$$\max_{\mathcal{M} \in \mathcal{NNSA}(\mathcal{T})} R_{\mathcal{M}}(p) = \max_{\mathcal{M}' \in \Pi(\mathcal{T})} R_{\mathcal{M}'}(p). \tag{3.19}$$

In other words, all NNSA candidates that give the highest supported rate among all the routes in $\mathcal{NNSA}(\mathcal{T})$ *are optimal routes for DF.*

Proof 8 (Proof of Theorem 4) *See Appendix A.5.*

Since the NNSA is optimal for any input distribution satisfying the SPC, we can show the following theorem.

Theorem 5 *Consider a static Gaussian MRC $\mathcal{T}$. Let the set of NNSA candidate set be $\mathcal{NNSA}(\mathcal{T})$. For any set of input distributions $\mathcal{P}$, with all $p \in \mathcal{P}$ satisfying the SPC, there exists at least one $\mathcal{M} \in \mathcal{NNSA}(\mathcal{T})$ such that $\mathcal{M} \in \mathcal{Q}_{\mathrm{DF}}(\mathcal{P})$. In other words, the NNSA candidate set contains an optimal route for DF over any input distribution set, when all the elements in the set satisfy the SPC.*

Proof 9 (Proof of Theorem 5) *Recall that $\mathcal{Q}_{\mathrm{DF}}(\mathcal{P}) \triangleq \{\mathcal{M} \in \Pi(\mathcal{T}) : \max_{p\in\mathcal{P}} R_{\mathcal{M}}(p) = R_{\mathrm{DF}}(\mathcal{P})\}$. Let $\mathcal{M}^* \in \Pi(\mathcal{T})$ and $p^* \in \mathcal{P}$ be a route and an input distribution for which $R_{\mathcal{M}^*}(p^*) = R_{\mathrm{DF}}(\mathcal{P})$. From Theorem 4, we know that there exists an $\mathcal{M}' \in \mathcal{NNSA}(\mathcal{T})$ for which $R_{\mathcal{M}'}(p^*) = R_{\mathcal{M}^*}(p^*) = R_{\mathrm{DF}}(\mathcal{P})$. Hence, $\mathcal{M}' \in \mathcal{Q}_{\mathrm{DF}}(\mathcal{P})$.*

We can easily show that an NNSA route is optimal for DF over all input distributions.

Theorem 6 *Consider a static Gaussian MRC $\mathcal{T}$. Let the set of NNSA candidates be $\mathcal{NNSA}(\mathcal{T})$. There exists at least one $\mathcal{M} \in \mathcal{NNSA}(\mathcal{T})$ and some p for which $R_{\mathcal{M}}(p) = R_{\mathrm{DF}}$. In other words, the NNSA candidate set contains an optimal route for DF over all input distribution set.*

Proof 10 (Proof of Theorem 6) *From Remark 2, there exist a route $\mathcal{M}^* \in \Pi(\mathcal{T})$ and some $p^* \in \mathcal{P}_{\mathrm{Gauss}}$ for which $R_{\mathcal{M}^*}(p^*) = \max_{\mathcal{M}\in\Pi(\mathcal{T})} \max_{p\in\mathcal{P}_{\mathrm{Gauss}}} R_{\mathcal{M}}(p) = \max_{\mathcal{M}'\in\Pi(\mathcal{T})} \max_{p'\in\mathcal{P}_{\mathrm{All}}} R_{\mathcal{M}'}(p') = R_{\mathrm{DF}}$. We can see that the optimal jointly Gaussian input p^* must satisfies the SPC. So, there exists some $\mathcal{M}' \in \mathcal{NNSA}(\mathcal{T})$ and some $p^* \in \mathcal{P}_{\mathrm{Gauss}}$ for which $R_{\mathcal{M}'}(p^*) = R_{\mathcal{M}^*}(p^*) = R_{\mathrm{DF}}$.*

3.4.5 Separating Coding and Routing

From (2.16), we see that achievable rates of DF depend on the route selected and the input distribution. At first sight, the problems of coding (input probability

density function) and routing seem intertwined. This makes the search for the best code and the best route (in the sense of maximizing the achievable rate) difficult.

In this section, we have presented an algorithm, NNSA, which outputs an NNSA candidate set $\mathcal{NNSA}(\mathcal{T})$ such that for any $\mathcal{P}$, with all $p \in \mathcal{P}$ satisfying the SPC, $\mathcal{NNSA}(\mathcal{T})$ contains at least one optimal route in $\mathcal{P}$:

$$\exists \mathcal{M}^* \in \mathcal{NNSA}(\mathcal{T}) \quad \text{s.t.} \quad \max_{p\in\mathcal{P}} R_{\mathcal{M}^*}(p) = \max_{\mathcal{M}'\in\Pi(\mathcal{T})} \max_{p'\in\mathcal{P}} R_{\mathcal{M}'}(p') = R_{\mathrm{DF}}(\mathcal{P}). \quad (3.20)$$

So, instead of using brute force to evaluate $R_{\mathcal{M}}(p)$ for every $\mathcal{M} \in \Pi(\mathcal{T})$ and for every $p \in \mathcal{P}$ to find the optimal route, we simply search in $\mathcal{NNSA}(\mathcal{T})$.

We see that under certain network topologies, $|\mathcal{NNSA}(\mathcal{T})| = 1$. That means we are able to find the optimal route without having to evaluate the input distribution. In other words, we have shown that we can achieve complete routing and coding separation under certain conditions. These are the scenarios when the NNA terminates normally.

Even if the NNA does not terminate normally for most network topologies, we could still achieve partial routing and coding separation using the NNSA. The NNSA candidate set $\mathcal{NNSA}(\mathcal{T})$ is obtained independent of the input distribution. In the next section, we will show that on average $|\mathcal{NNSA}(\mathcal{T})|$ is much smaller than $|\Pi(\mathcal{T})|$. This means we can discard "bad" routes without even evaluating codes on those routes.

3.5 Discussions on the NNSA

In this section, we discuss a few interesting properties of the NNSA.

3.5.1 Search Space Reduction

With the NNSA, we can now search for an optimal route in the NNSA candidate set $\mathcal{NNSA}(\mathcal{T})$, as compared to searching in $\Pi(\mathcal{T})$ using brute force. The number of candidates determines the number of routes whose rate we need to optimize in

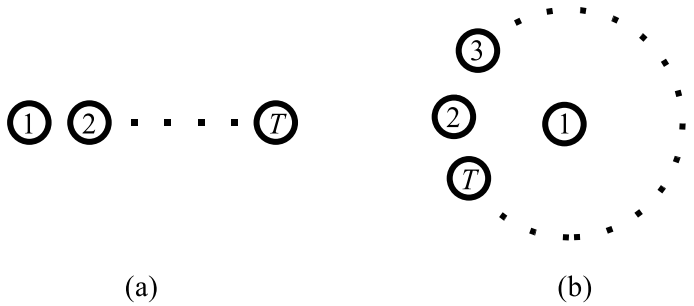


Figure 3.1: Two MRCs.

order to find an optimal route. Using brute force, the number of routes we need to check is

$$|\Pi(\mathcal{T})| = 1 + \binom{T-2}{1} + \binom{T-2}{2} + \cdots \binom{T-2}{T-2} \tag{3.21a}$$

$$= O((T-1)!), \tag{3.21b}$$

where T is the total number of nodes in the network and $\binom{n}{k} = \frac{n\times(n-1)\times\cdots\times1}{(n-k)\times(n-k-1)\times\cdots\times1}$. Using the NNSA, the number of possible routes that we need to check, which is $|\mathcal{NNSA}(\mathcal{T})|$, depends on the network topology.

We consider two extreme cases as shown in Fig. 3.1. Assume that the channels are static, all nodes transmit at the same power, and all nodes are subject to the same receiver noise power. In Fig. 3.1(a), the nodes are arranged in linear topology with the source at one end and the destination at the other end. In this case, there is only one NNSA candidate which is $\{1, 2, \ldots, T\}$. This means we can achieve complete coding and routing separation here. In Fig. 3.1(b), all relays and the destination are of equi-distance from the source. The NNSA candidates are $\{1, T\}, \{1, 2, T\}, \{1, 3, T\}, \ldots, \{1, T-1, T\}, \{1, 2, 3, T\}, \ldots$. Here, $\mathcal{NNSA}(\mathcal{T}) = \Pi(\mathcal{T})$. This means, we need to evaluate the rate supported by all possible routes to determine the one that gives the highest rate, i.e., we cannot separate coding and routing. These two examples give the best case and the worst case scenarios for the NNSA. We note that the size of the NNSA candidate set might, in the worst case, equal $|\Pi(\mathcal{T})|$.

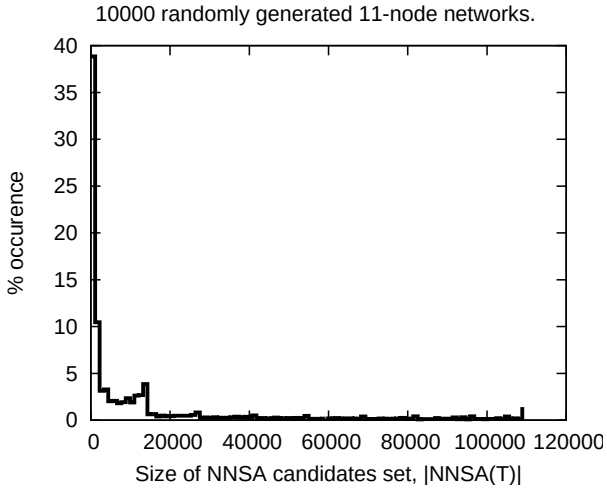


Figure 3.2: The p.d.f. of $|\mathcal{NNSA}(\mathcal{T})|$ for the 11-node network. $|\Pi(\mathcal{T})| = 986410$.

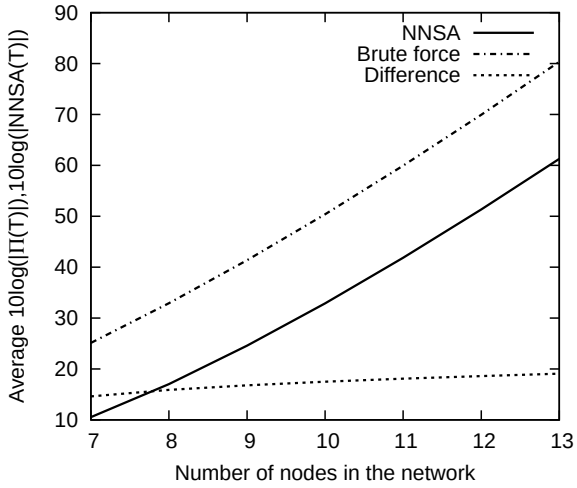


Figure 3.3: Average (over 10,000 random samples) $|\mathcal{NNSA}(\mathcal{T})|$ and $|\Pi(\mathcal{T})|$ for different $|\mathcal{T}|$.

More examples of how the NNSA reduces the search space of finding an optimal route compared to brute force can be found in Appendix A.4. In these examples, we see that the search space is very much reduced using the NNSA.

Now, we run simulations to obtain the average $|\mathcal{NNSA}(\mathcal{T})|$. For each network size $|\mathcal{T}| = T$, we randomly generated 10000 networks with nodes uniformly distributed in a 1m×1m square area. The coordinates of the source, the relays and the destination were randomly assigned. For each randomly generated network, we ran the NNSA to find out the number of NNSA candidates. Half the time, $|\mathcal{NNSA}(\mathcal{T})|$ was smaller than 0.715% of $|\Pi(\mathcal{T})|$ for the 8-node channel and smaller

than 0.253% for the 11-node channel. Fig. 3.2 shows the probability density function (p.d.f.) of $|\mathcal{NNSA}(\mathcal{T})|$ for the 11-node network.

Fig. 3.3 shows average $|\mathcal{NNSA}(\mathcal{T})|$ and $|\Pi(\mathcal{T})|$ for varying number of nodes in the network. We see that as the number of nodes increases, the number of routes increases exponentially for both the NNSA and brute force. However, on average, we get a 14.6–19.1dB reduction in the number of routes using the NNSA, which is a one to two order of magnitude reduction. Furthermore, the reduction in the number of routes using the NNSA increases as the number of nodes in the network increases. We bear in mind that the average values of $|\mathcal{NNSA}(\mathcal{T})|$ can be biased by extreme points (see Fig. 3.2).

We note that the average size of the NNSA candidate set grows factorially with the number of nodes in the network. However this does increase the range of finite size networks for which we can find optimal routes. Furthermore, the NNSA provides insights for designing heuristic algorithms to find good routes for DF-based codes.

3.5.2 The NNSA and the Shortest Optimal Route

We note that the total transmit power changes with the length of the route if each node in the route (except the destination) transmits at the same power. Consider two routes $\{1, 4\}$ and $\{1, 2, 3, 4\}$, where the latter supports a higher rate. One might argue that route $\{1, 4\}$, though supports a lower rate, is better as only 1/3 power is consumed compared to route $\{1, 2, 3, 4\}$. However, we stress that in this chapter, we find a route that maximizes the transmission rate given that each node must transmit within a given power constraint (and hence a constraint on the total power).

However, when two routes achieve the maximum DF rate, the shorter route might be preferred. This is of interest from a practical view point as it allows better utilization of the nodes. Nodes that are "redundant" can be put to sleep. This saves the total transmission power of the network, which may be the second

metric to be optimized after the rate. We term shortest routes which support the highest DF rate *shortest optimal routes* (SOR). In Section 3.6, we show that an SOR is actually contained in one or more of the NNSA candidates.

3.5.3 Non-Directional Routing

The NNSA is a non-directional routing algorithm, i.e., it does not find a route in the direction toward the destination. The NNSA builds routes by adding neighbors with respect to the partial routes. The destination is never in the picture; it only serves as the stopping criterion – the algorithm stops when all routes terminate at the destination.

This is counter-intuitive to how routing is commonly done in networking, in which a route is searched in the direction from the source to the destination. Using the NNSA, the route might go in the direction opposite to the destination. In Appendix A.6, we show by an example that routing backward can increase the transmission rate.

This observation provides insights for designing heuristic algorithms to find good routes for DF-based codes. In Section 3.8, we propose a heuristic algorithm that finds potentially good routes in polynomial time.

3.6 Finding a Shortest Optimal Route

Although the NNSA guarantees an optimal route, the algorithm does not attempt to find the shortest route that achieves the optimal rate. Now we define the shortest optimal route.

Definition 15 *A* shortest optimal route *(SOR) for an input distribution p is defined as a shortest route that achieves the highest DF rate, i.e.,*

$$\mathcal{M}^{\mathrm{SOR}}(p) \in \mathcal{Q}(p), \quad s.t.\ |\mathcal{M}^{\mathrm{SOR}}(p)| \leq |\mathcal{M}|, \forall \mathcal{M} \in \mathcal{Q}(p). \tag{3.22}$$

By definition and from Theorem 4,

$$R_{\mathcal{M}^{\mathrm{SOR}}}(p) = \max_{\mathcal{M}' \in \Pi(\mathcal{T})} R_{\mathcal{M}'}(p) = \max_{\mathcal{M}'' \in \mathcal{NNSA}(\mathcal{T})} R_{\mathcal{M}''}(p). \tag{3.23}$$

Next, we define a special subset of NNSA candidates for some input distribution p as follows.

Definition 16 *The* optimal NNSA candidate set *for input distribution p, $\mathcal{NNSA}^{\mathrm{opt}}(\mathcal{T}, p)$, contains all NNSA candidates that achieve the highest rate, meaning,*

$$\mathcal{NNSA}^{\mathrm{opt}}(\mathcal{T}, p) \triangleq \left\{ \mathcal{M}' \in \mathcal{NNSA}(\mathcal{T}) : R_{\mathcal{M}'}(p) = \max_{\mathcal{M} \in \mathcal{NNSA}(\mathcal{T})} R_{\mathcal{M}}(p) \right\}. \tag{3.24}$$

The routes in the optimal NNSA candidate set are called optimal NNSA candidates.

For two routes $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \subseteq \mathcal{B}$ means $\mathcal{A}$ is a subset of $\mathcal{B}$ and the node order in $\mathcal{A}$ follows that in $\mathcal{B}$. The following theorem says that some optimal NNSA candidates contain an SOR.

Theorem 7 *One (or more) of optimal NNSA candidates contains an (or more) SOR in the correct order, meaning,*

$$\mathcal{M}^{\mathrm{SOR}}(p) \subseteq \mathcal{M}, \quad \text{for some } \mathcal{M} \in \mathcal{NNSA}^{\mathrm{opt}}(\mathcal{T}, p). \tag{3.25}$$

Example 3 *Suppose an SOR for a network is*

$$\mathcal{M}^{\mathrm{SOR}}(p) = \{m_1^*, m_2^*, \ldots, m_{|\mathcal{M}^{\mathrm{SOR}}|-1}^*, m_{|\mathcal{M}^{\mathrm{SOR}}|}^*\}, \tag{3.26}$$

for some p that satisfies the SPC, then $\mathcal{M}^{\mathrm{SOR}}$ will be a subset of at least one optimal NNSA candidate for p with the same ordering, meaning m_2^* is in front of m_1^*, m_3^* is in front of m_2^*, but there might be other nodes in between them. For example, one of the optimal NNSA candidates might be $\{m_1^*, \ldots, m_{|\mathcal{M}^{\mathrm{SOR}}|-2}^*, a, m_{|\mathcal{M}^{\mathrm{SOR}}|-1}^*, m_{|\mathcal{M}^{\mathrm{SOR}}|}^*\}$.

Proof 11 (Proof of Theorem 7) *See Appendix A.7.*

With this theorem, we now describe an algorithm to find an SOR for some input distribution p. We term this algorithm the *Nearest Neighbor Set Pruning Algorithm* (NNSPA).

Algorithm 3 (NNSPA)

1. *Initialize an optimal sub-route set,* $\mathcal{OSR} = \mathcal{NNSA}^{\text{opt}}(\mathcal{T}, p)$.
2. *For each newly added* $\mathcal{M} \in \mathcal{OSR}$*, we prune one node from* $\{m_2, \ldots, m_{|\mathcal{M}|-1}\}$ *at a time to form* $|\mathcal{M}| - 2$ *sub-routes.*
3. *Find the rates supported by all sub-routes. Do the following for all sub-routes:*
   (a) *If the sub-route supports rates lower than the optimal rate, discard the sub-route.*
   (b) *Else if the sub-route* $\mathcal{M}_s$ *can support the optimal rate, add it into the* optimal sub-route set*, i.e.,* $\mathcal{OSR} \leftarrow \mathcal{OSR} \cup \mathcal{M}_s$.
4. *For each new sub-route formed in step 3b, repeat steps 2–3.*
5. *Select the shortest route(s) from* $\mathcal{OSR}$.

Theorem 8 *The shortest routes in the optimal sub-route set of the NNSPA are SORs for input distribution p.*

Proof 12 (Proof of Theorem 8) *Theorem 8 follows from the proof of Theorem 7. In the proof, we note that an SOR is contained in an optimal NNSA candidate, and removing* extra *nodes (nodes that are in the optimal NNSA candidate but not in the SOR) does not reduce the supported rate. Since the NNSPA removes nodes one by one from all optimal NNSA candidates, and checks the supported rate, one of the pruned routes must be an SOR.*

3.7 The NNSA on Fading Channels

In previous sections, we have seen that the NNSA is optimal in static channels in the sense that it finds a set of routes that contains at least one optimal route. Now, we investigate if the NNSA is optimal in fading channels, where ν_{ij} are random variables (see Section 2.3). Without loss of generality, we assume $E[\nu_{ij}] = 1, \forall i, j$, i.e., the power of the fading processes is unity. For channels where the fading processes have different power, we can always normalize the power to 1 by adjusting d_{ij} accordingly.

For fading channels, two measures of "rate" are often used, namely the *ergodic rate* and the *outage probability*. In this thesis, we consider cases where:

1. The transmitters are unaware of the fading processes and there is no feedback from the receivers back to the transmitters about the fading processes. Hence the codewords cannot be chosen as functions of the fading processes ν_{ij} (Tse & Hanly, 1998).

2. The receivers have perfect information regarding the fading processes.

We assume that all transmitters and receivers know the large scale fading components h_{ij} (see Section 2.3) *a priori*.

3.7.1 Ergodic Rate

In this section, we consider the *ergodic rate* of DF for the MRC. We assume that the fading processes ν_{ij} are i.i.d. stationary ergodic random processes. When there is no delay constraint for which the signals from the source must be decoded at the destination by a certain time, we can use codewords long enough for the fading processes to reflect their ergodic nature (Biglieri *et al.*, 1998). We can achieve transmission rates averaged over the fading processes. This is applicable for data applications that have large delay tolerance.

Since ν_{ij} are stationary ergodic processes, modifying the results for the point-to-point channel by Biglieri *et al.* (1998), the following ergodic rate is achievable

on the MRC using DF on route $\mathcal{M}$ with input distribution p.

$$R^E_{\mathcal{M}}(p) = E_\nu \left[\min_{m_t \in \mathcal{M}\setminus\{1\}} R_{m_t}(\mathcal{M}, p) \right] \tag{3.27a}$$

$$= \min_{m_t \in \mathcal{M}\setminus\{1\}} R^E_{m_t}(\mathcal{M}, p), \tag{3.27b}$$

where $R^E_{m_t}(\mathcal{M}, p)$ is the ergodic reception rate at node m_t given by

$$R^E_{m_t}(\mathcal{M}, p) = E_\nu I(X_{m_1}, \ldots, X_{m_{t-1}}; Y_{m_t} | X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c}). \tag{3.28}$$

We use superscript E to indicate ergodic rate. Eqn. (3.27b) is because the fading processes are independent. As a result, we can move the expectation operator into the individual reception rate expression in (3.27a).

We can view ergodic rate as follows. Consider an instance of the fading processes. The instantaneous reception rate at node m_t is

$$R_{m_t}(\mathcal{M}, p) = I\left(X_{m_1}, \ldots, X_{m_{t-1}}; \sum_{i=1}^{t-1} \sqrt{\nu_{m_i m_t} \kappa d^{-\eta}_{m_i m_t}} X_{m_i} + Z_{m_t} \,\middle|\, X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c} \right), \tag{3.29}$$

for some $\nu_{ij} \geq 0$. When the codeword length is long enough, we can achieve rate averaged over the fading processes. Since the fading processes are ergodic, we have the following ergodic reception rate, which is the average transmission rate over the fading of the channels.

$$R^E_{m_t}(\mathcal{M}, p) \tag{3.30a}$$

$$= \int \cdots \int_0^\infty I\left(X_{m_1}, \ldots, X_{m_{t-1}}; \sum_{i=1}^{t-1} \sqrt{\nu_{m_i m_t} \kappa d^{-\eta}_{m_i m_t}} X_{m_i} + Z_{m_t} \,\middle|\, X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c} \right) \times p(\nu_{m_1 mt}) \cdots p(\nu_{m_{t-1} mt}) \quad d\nu_{m_1 mt} \cdots d\nu_{m_{t-1} mt} \tag{3.30b}$$

$$= E_\nu I(X_{m_1}, \ldots, X_{m_{t-1}}; Y_{m_t} | X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c}). \tag{3.30c}$$

Applying this to all the nodes in the route, we get (3.27a).

Now, we show that all the lemmas in Appendix A.3 for the proof of Theorem 3 hold true for ergodic rate.

Lemma 6 *Consider a fading Gaussian MRC where the fading processes for all node pairs are i.i.d. stationary ergodic processes with unity power. Consider route* $\mathcal{M} = \{m_1, \ldots, m_{|\mathcal{M}|}\}$ *and assume that a unique nearest neighbor, node* a^**, exists. For any input distribution* p *satisfying the SPC, the ergodic rate supported by the new route* $\mathcal{M}_1 = \mathcal{M} \cup \{a^*\}$ *is greater or equal to the ergodic rate supported by the route adding any non-nearest neighbor, or* $\mathcal{M}_2 = \mathcal{M} \cup \{b\}$*, i.e.,*

$$R^E_{\mathcal{M}\cup\{a^*\}}(p) \geq R^E_{\mathcal{M}\cup\{b\}}(p), \quad \forall b \in \mathcal{T} \setminus (\mathcal{M} \cup \{a^*\}). \tag{3.31}$$

We use superscript * *to indicate a nearest neighbor.*

Proof 13 (Proof for Lemma 6) *Consider the following fading instance:* $\nu_{ia^*} = \nu_{ib}, \forall i \in \mathcal{M}$*. Since*

$$\gamma_{ma^*} \geq \gamma_{mb}, \quad \forall m \in \mathcal{M}, \forall b \in \mathcal{T} \setminus (\mathcal{M} \cup \{a^*\}), \tag{3.32}$$

we have

$$\gamma'_{ma^*} \geq \gamma'_{mb}, \quad \forall m \in \mathcal{M}, \forall b \in \mathcal{T} \setminus (\mathcal{M} \cup \{a^*\}), \tag{3.33}$$

where

$$\gamma'_{ij} = \frac{\lambda_{ij} E[X_i^2]}{E[Z_j^2]} = \frac{\nu_{ij} \kappa d_{ij}^{-\eta} E[X_i^2]}{E[Z_j^2]} \tag{3.34}$$

are the instantaneous rSNRs for a particular instance of the fading processes.

From Lemma 7 (in Appendix A.3), we can show that $R_{\mathcal{M}_1}(p) \geq R_{\mathcal{M}_2}(p)$ *for these fading process realizations. Since the fading processes are i.i.d.,* $p(\nu_{ia^*}) = p(\nu_{ib}), \forall i \in \mathcal{M}$*. Averaging over the fading processes, we get* $R^E_{\mathcal{M}_1}(p) \geq R^E_{\mathcal{M}_2}(p)$

We can apply the same technique to Lemmas 8–10 (in Appendix A.3) and prove the following theorems.

Theorem 9 *Consider a fading Gaussian MRC $\mathcal{T}$ where the fading processes are i.i.d., stationary, and ergodic. If the NNA terminates normally and outputs route $\mathcal{M}^*$, then for any input distribution of the form $p = p(x_1, \ldots, x_{T-1})$ satisfying the SPC,*

$$R^E_{\mathcal{M}^*}(p) = \max_{\mathcal{M} \in \Pi(\mathcal{T})} R^E_{\mathcal{M}}(p) \tag{3.35}$$

Theorem 10 *Consider a fading Gaussian MRC $\mathcal{T}$ where the fading processes are i.i.d., stationary, and ergodic. Let the set of NNSA candidates be $\mathcal{NNSA}(\mathcal{T})$. Then for any input distribution of the form $p = p(x_1, \ldots, x_{T-1})$ satisfying the SPC,*

$$\max_{\mathcal{M} \in \mathcal{NNSA}(\mathcal{T})} R^E_{\mathcal{M}}(p) = \max_{\mathcal{M}' \in \Pi(\mathcal{T})} R^E_{\mathcal{M}'}(p). \tag{3.36}$$

In other words, the NNSA candidates that give the highest supported ergodic rate are optimal routes for DF.

Proof 14 (Proofs of Theorems 9 & 10) *We apply the technique in the proof for Lemma 6 to Lemmas 8–10. Theorems 9 and 10 follow.*

Remark 5 *In Theorems 9 and 10, all we need is that the fading processes follow the same type of p.d.f. (e.g., all are Rayleigh fading processes). So, the results apply to various types of fading channel. As previously mentioned, in cases where the power of all fading processes $E[\nu_{ij}]$ is different, we can normalized $E[\nu_{ij}]$ to unity by changing d_{ij} accordingly.*

3.7.2 Supported Rate versus Outage Probability

In the derivation of ergodic rates, we assume that the fading processes are ergodic and the delay requirement is long enough to allow the long term rates to be averaged over the fading processes. However, this is not always true in a general communication system, as many applications are delay sensitive, e.g., telephony. In this section, we consider any of the following scenarios:

1. There is a stringent requirement on the delay constraint such that codewords cannot be chosen long enough to average the fading processes. Hence, the fading processes cannot reflect their ergodic behavior during data transmission, or

2. The fading processes are not ergodic.

We model these scenarios by letting all ν_{ij} be chosen at the beginning of time and be held fixed for all channel uses (Telatar, 1999). This model is also termed *quasi-static fading* (Kramer *et al.*, 2005). In this case, for any non-zero transmission rate chosen, there is a non-zero probability that the realization of ν_{ij} is not able to support it, and the achievable rate in the Shannon sense is zero. So, we investigate the probability that a given transmission rate cannot be supported (using some coding strategy), or *supported rate versus outage probability*. The definition follows that for the capacity versus outage by Ozarow *et al.* (1994).

Since $\nu_{ij}, \forall i, j$, are random variables, for a given route $\mathcal{M}$ and input distribution p, the reception rate at node m_t is a random variable. Recall that

$$R_{m_t}(\mathcal{M}, p) = I\left(X_{m_1}, \ldots, X_{m_{t-1}}; \sum_{\substack{i=1 \\ i \neq t}}^{i=t-1} \sqrt{\nu_{m_i m_t} \kappa d_{m_i m_t}^{-\eta}} X_{m_i} + Z_{m_t} \middle| X_{m_t}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{M}^c}\right). \tag{3.37}$$

We denote the outage probability at some transmission rate R by $P_{\text{out}}(\mathcal{M}, p, R)$. We note that

$$\begin{aligned} P_{\text{out}}(\mathcal{M}, p, R) &\triangleq \Pr\left(\min_{m_t \in \mathcal{M} \setminus \{1\}} R_{m_t}(\mathcal{M}, p) \leq R\right) && \text{(3.38a)} \\ &= 1 - \Pr\left(R_{m_t}(\mathcal{M}, p) > R, \forall m_t \in \mathcal{M} \setminus \{1\}\right) && \text{(3.38b)} \\ &= 1 - \Pi_{m_t \in \mathcal{M} \setminus \{1\}} \Pr\left(R_{m_t}(\mathcal{M}, p) > R\right). && \text{(3.38c)} \end{aligned}$$

In (3.38c), we assume that the fading processes are independent.

Under quasi-static fading, an optimal route is one which gives the lowest outage probability at all transmission rates. We can show that the NNSA is not optimal for DF under quasi-static fading, as given by the following theorem.

Theorem 11 *Consider a quasi-static fading Gaussian MRC, the NNSA does not always output routes that minimize the outage probability at all rates.*

Proof 15 (Proof of Theorem 11) *Consider a three-node Gaussian MRC with Raleigh fading, i.e.,*

$$p(\nu_{ij}) = \begin{cases} \frac{1}{\Omega_{ij}} \exp\left(\frac{-\nu_{ij}}{\Omega_{ij}}\right) & , \nu_{ij} \geq 0 \\ 0 & , \text{otherwise} \end{cases} \qquad i = 1, 2, j = 2, 3, \tag{3.39}$$

where $\Omega_{ij} = E[\nu_{ij}]$ is the fading power.

Suppose that the inter-node distances are $d_{12} = 0.9, d_{23} = 1.5, d_{13} = 1$. We consider the following parameters: node 1 and 2 send independent Gaussian codewords with power $P_1 = 5$ and $P_2 = 5$ respectively, $N_2 = N_3 = 1$, $\kappa = 1$, $\eta = 2$, $\Omega_{ij} = 1, \forall i, j$. The NNSA route is $\mathcal{M}_1 = \{1, 2, 3\}$, and a non-NNSA route is $\mathcal{M}_2 = \{1, 3\}$.

The outage probabilities at transmission rate R on the routes are

$$P_{\text{out}}(\mathcal{M}_1, p, R) = 1 - Pr(\{R_2(\mathcal{M}_1, p) > R\} \text{ AND } \{R_3(\mathcal{M}_1, p) > R\}) \tag{3.40a}$$

$$= 1 - Pr(R_2(\mathcal{M}_1, p) > R) \times Pr(R_3(\mathcal{M}_1, p) > R), \tag{3.40b}$$

$$P_{\text{out}}(\mathcal{M}_2, p, R) = 1 - Pr(R_3(\mathcal{M}_2, p) > R), \tag{3.40c}$$

where p here denotes a Gaussian input distribution with independent x_1 and x_2, with variances P_1 and P_2 respectively.

Fig. 3.4 shows supported rate versus outage probability for $\mathcal{M}_1$ and $\mathcal{M}_2$. We see that for a large range of R, i.e., $0.5 < R < 2.5$, $P_{\text{out}}(\mathcal{M}_1, p, R) > P_{\text{out}}(\mathcal{M}_2, p, R)$. Hence, the NNSA route does not always give a lower outage probability.

Remark 6 *In the three-node example, we note that the NNSA route $\{1, 2, 3\}$ gives*

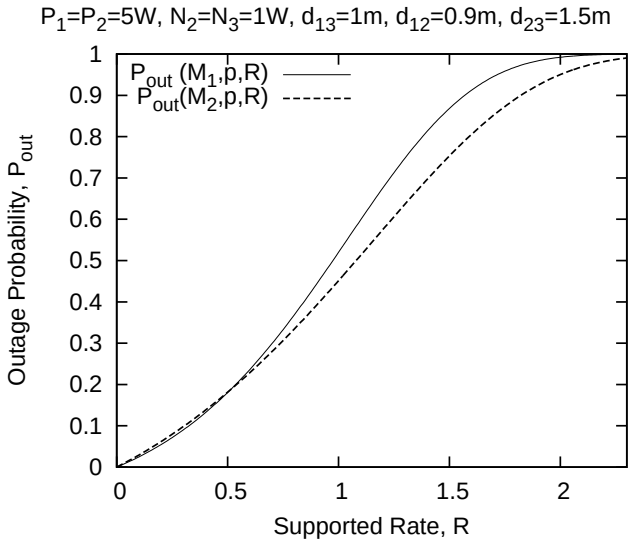


Figure 3.4: Supported rate versus outage probability for two routes.

a higher outage probability (at a certain supported rate) compared to the direct transmission $\{1, 3\}$ *because we need to ensure that this rate is supported at each relay and the destination. See* (3.40a) *that there are more probability terms compared to* (3.40c)*, and this increases the overall outage probability. Hence, when we consider the supported rate versus outage probability, the route length plays an important part besides rSNR. A longer route is likely to result in a higher outage probability.*

3.8 A Heuristic Algorithm for Routing

3.8.1 The Maximum Sum-of-Received-Power Algorithm

Using the NNSA, a route is constructed by adding the "next hop" node one by one to the *partial* route. The node to be added is from the nearest neighbor set. If the nearest neighbor set contains more than one node, the current route branches to more than one route, leading to a possibly large NNSA candidate set size. In this case, we will not be able to completely separate the optimizations of routing and coding, and a large NNSA candidate set size complicates the process of finding an optimal route.

To avoid searching for an optimal route in a large set of candidate routes, we consider a heuristic approach that starts from the source node and repeatedly adds only one "good" node to the partial until the destination is reached. For the choice of the next hop node, we consider the node that receives the largest sum of received power from all the nodes in the partial route. We call this the maximum sum-of-received-power algorithm (MSPA). By choosing only one node to be added to the partial route, we prevent the algorithm from branching out to multiple routes. This heuristic approach yields only one route, regardless of the network size and topology. We now explicitly describe the MSPA.

Algorithm 4 (MSPA)

1. *First, start with the source node,* $\mathcal{M} = \{m_1\}$.
2. *For every node* $t \in \mathcal{T} \setminus \mathcal{M}$*, find the sum of received power from all nodes in* $\mathcal{M}$ *to* t*,* $\sum_{i \in \mathcal{M}} \gamma_{it}$.
3. *Let* a^* *be* any *node with the highest sum of received power, i.e.,* $\sum_{i \in \mathcal{M}} \gamma_{ia^*} \geq \sum_{j \in \mathcal{M}} \gamma_{jt}, \forall t \in \mathcal{T} \setminus \mathcal{M}$*. Append node* a^* *to the route:* $\mathcal{M} \leftarrow \mathcal{M} \cup \{a^*\}$.
4. *Repeat steps 2–3 until the destination is added to the route.*

Remark 7 *Assuming that the value of the previous sum-of-received-power computations are cached, the complexity of step 2 in MSPA is* $O(T)$ *because there are at most* $(T-1)$ *nodes not in the route. The complexity of the comparisons in step 3 is* $O(T)$*. Steps 2–3 are repeated at most* $(T-1)$ *times, giving a worst case complexity of the MSPA of* $O(T^2)$.

3.8.2 Performance of the MSPA

It turns out that the MSPA is optimal if the nodes are restricted to sending independent codewords, as proven in the following theorem.

Table 3.1: Performance of the MSPA.

T	Average $\max_{p\in\mathcal{P}_{\text{Gauss}}} R_{\mathcal{M}_{\text{MSPA}}}(p)/R_{\text{DF}}$	$Pr\{\mathcal{M}_{\text{MSPA}}$ is optimal$\}$
3	1	1
4	0.9999951	0.99883
5	0.9999520	0.99557
6	0.9999168	0.99307

Theorem 12 *In a static Gaussian MRC, in which the nodes send independent Gaussian codewords, the MSPA route is optimal for DF.*

Proof 16 (Proof of Theorem 12) *See Appendix A.8.*

However, unlike the NNA and the NNSA, the MSPA might not output an optimal route when the nodes are allowed to send arbitrarily correlated codewords. We show this by the following example.

Example 4 *Consider a static four-node Gaussian MRC with node coordinates 1(0,0), 2(0.418,0), 3(0.209,0.6755), and 4(0.995,0), in which the nodes send arbitrarily correlated Gaussian signals. Assume* $P_i = 1, N_i = 1, \kappa = 1, \eta = 2$. *The MSPA route is* $\mathcal{M}_1 = \{1, 2, 4\}$. *The NNSA outputs* $\mathcal{M}_1$ *and* $\mathcal{M}_2 = \{1, 2, 3, 4\}$. *It is easy to compute that* $\max_{p\in\mathcal{P}_{\text{Gauss}}} R_{\mathcal{M}_1}(p) = 1.30826$ *and* $\max_{p\in\mathcal{P}_{\text{Gauss}}} R_{\mathcal{M}_2}(p) = 1.31576$.

We now investigate how well the MSPA route $\mathcal{M}_{\text{MSPA}}$ performs compared to the optimal route in the Gaussian MRC. Due to the complexity involved in optimizing the power splits, we only simulate MRCs up to 6 nodes, i.e., $T \leq 6$. For each T, we randomly place T nodes in a network area of size $(T-1)\text{m} \times (T-1)\text{m}$. The source and the destination are randomly chosen. We run the NNSA to find the optimal rate R_{DF}, and run the MSPA to find $\max_{p\in\mathcal{P}_{\text{Gauss}}} R_{\mathcal{M}_{\text{MSPA}}}(p)$. The results are shown in Table 3.1. With high probability, the MSPA is able to find an optimal route. Also, $\max_{p\in\mathcal{P}_{\text{Gauss}}} R_{\mathcal{M}_{\text{MSPA}}}(p)$ is a good indicator of R_{DF}.

Although the MSPA does not always guarantee an optimal route, we have shown that it is able to finds an optimal route with high probability. This means we can

completely separate routing and coding in many more cases, in addition to those where the NNA terminates normally.

Remark 8 *As argued earlier, the MSPA might not be useful when we want to calculate the exact value of* $R_{\mathrm{DF}}(\mathcal{P})$*. However, the algorithm can be used to find a route that can support transmission rate close to* $R_{\mathrm{DF}}(\mathcal{P})$*. This is of practical interest when we want to choose a "good" route to implement DF in a network.*

3.9 Conclusion

In this chapter, we presented an algorithm, the NNSA, that finds a set of routes that contains at least an optimal DF route, without having to consider the input distribution. This algorithm simplifies the process of finding the maximum DF rate and an optimal route that achieves the rate, without resorting to brute force. Through this algorithm, we showed that that optimizations of route and code can be separated under certain network topologies.

We also showed that the NNSA outputs route which contains the shortest optimal route. In addition, we found that the NNSA is optimal in fading channels in the sense that an NNSA route maximizes the ergodic rate.

However, under certain network topologies, the NNSA might output a large set of routes, which makes the search for an optimal route complicated. To tackle this, we constructed a heuristic algorithm that outputs an optimal route for DF with high probability, without having to consider the input distribution. This adds to the number of scenarios in which the optimizations for route and code can be totally separated.

Chapter 4

Myopic Coding in Multiple-Relay Channels

4.1 Introduction

Since the wireless medium is broadcast in nature, the transmission of one node can be received by all other nodes listening in the same frequency band. The simplest way of data transmission is for the source to transmit directly to the destination. However, direct transmission from the source to a far-situated destination may require high transmission power (due to the path loss of electromagnetic wave propagation). High-power transmission is not suitable for energy-limited nodes and it creates undesirable interference to other users. Transmitting data via intermediate relays, using multi-hop routing or cooperative relaying, can help to decrease the transmit power and reduce multi-user interference.

4.1.1 Point-to-Point Coding

A common approach to data transmission is to abstract the wireless network into a communication graph, with an edge connecting two nodes if they can communicate. Data communication happens by identifying a route, which is a sequence of nodes that connect the source to the destination. Each node sends data to the next node

in the route and decodes data from the previous node in the route. Transmissions of other nodes are treated as noise. We call this coding strategy *point-to-point coding* in a multi-terminal network. This way of transmitting data from the source to the destination is commonly called multi-hop routing in the communications and networking literature.

4.1.2 Omniscient Coding

Point-to-point coding ignores the inherent broadcast nature of the wireless channel, i.e., that a node can hear the transmissions meant for other nodes, and thus it can act as a relay for them. Clearly, the best thing to do is for all the nodes to cooperate, helping the source to send its data to the destination. This requires every node to be aware of the presence of other nodes and to have knowledge of the processing they do. We refer to coding strategies that utilize the global view and complete cooperation as *omniscient coding*. In the literature, omniscient coding strategies were investigated for multiple-terminal networks, e.g., the multiple-access relay channel, the broadcast relay channel (Kramer & Wijngaarden, 2000; Kramer *et al.*, 2004), and the multiple-relay channel (MRC) (Gupta & Kumar, 2003; Kramer *et al.*, 2005; Xie & Kumar, 2005). While the rates achievable by omniscient coding strategies are higher than that by point-to-point coding strategies in these channels, there are a number of practical difficulties in implementing complete cooperation, e.g., (i) designing codes based on omniscient coding is more difficult as it involves the optimization of the whole network, (ii) the failure of one node affects the decoding of all other nodes, and (iii) all nodes need to be synchronized (for some coding strategies).

4.1.3 Myopic Coding

In view of these practical issues, we investigate *myopic coding*, coding strategies with constrained communications, e.g., node have a local view of the network, and limited cooperation. Myopic coding positions itself between point-to-point

coding and omniscient coding. In myopic coding, communications of the nodes are constrained in such a way that a node communicates with more than two nodes (as opposed to point-to-point coding) but not with all the nodes (as opposed to omniscient coding) in the network. Myopic coding incorporates local cooperation. It allows cooperation among neighboring nodes to increase the transmission rate compared to point-to-point coding. On the other hand, it partially solves the practical difficulties encountered in omniscient coding.

Now, we define myopic coding (Ong & Motani, 2005a,b, 2008). This is an informal definition which will be made more precise later in the chapter

Informal Definition 1 *A myopic `X` coding strategy is a constrained version of the corresponding omniscient `X` coding strategy. The constraint in myopic coding is such that every node cooperates with only a few other nodes. This cooperation can be in the form of transmitting to another node, processing (e.g., decoding, amplifying, quantizing) or canceling the transmissions from another node.*

We note that a myopic coding strategy is defined with respect to an omniscient coding strategy. Though there is no fixed way of constraining an omniscient coding strategy, the idea is to limit the processing at the nodes by limiting the number of neighbors a node communicates and cooperates with. Myopic coding aims to achieve practical advantages, e.g., lower computational complexity, robustness to topology changes, and fewer storage/buffer requirements.

4.1.4 Problem Statement

We ask the following questions which we will partially answer in this chapter:

1. What rate regions are achievable in the MRC in which every node has only a localized or myopic view of the network?

2. What is the value of cooperation? In other words, what is the impact on the performance, in terms of transmission rates, when communications among the nodes are constrained compared to the case when it is unconstrained?

4.2 Contributions

We investigate myopic coding in the MRC based on the decode-forward coding strategy (DF) (Xie & Kumar, 2005). Answering the above questions leads to the main contributions of this chapter, which are:

1. We construct random codes for *myopic DF* (Ong & Motani, 2005a,b, 2008), i.e., DF with myopic outlook, for the MRC.

2. We derive achievable rates of DF for the discrete memoryless, the static Gaussian, and the fading MRC, with different levels of cooperation.

3. Comparing myopic DF and omniscient DF for the Gaussian MRC, we show that including a few nodes into the cooperation increases the transmission rate significantly, often making it close to that under full cooperation. In other words, sometimes more cooperation yields diminishing returns.

4. We show that achievable rates of myopic DF for the Gaussian MRC may be as large as that of omniscient DF in the low transmitted-signal-to-noise ratio (tSNR) regime.

5. We show that in the MRC, myopic DF can achieve rates bounded away from zero as the network size grows to infinity.

4.2.1 Organization

The rest of the chapter is organized as follows. In Section 4.3, we define myopic coding and give examples of two myopic coding strategies. We present advantages of myopic coding compared to omniscient coding in Section 4.4. In Section 4.5, we investigate myopic coding in the MRC, where we derive achievable rates of two-hop myopic DF. We then compare achievable rates of one-hop myopic DF, two-hop myopic DF, and omniscient DF for the Gaussian MRC in Section 4.6. In Section 4.7, we extend the code construction and achievable rate analyses to the general k-hop myopic coding for the T-node multiple-relay channel, where k can

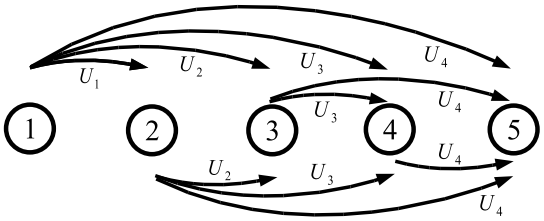


Figure 4.1: Omniscient DF for the five-node Gaussian MRC.

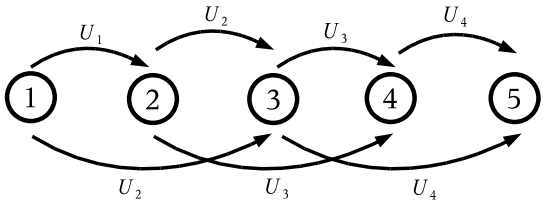


Figure 4.2: Two-hop myopic DF for the five-node Gaussian MRC.

be any positive integer from 1 to $T-1$ and T is the number of nodes (including the source, the relays, and the destination) in the channel. We briefly discuss myopic coding in the fading channel in Section 4.8. In Section 4.9, we investigate myopic coding in a large network, meaning that we see what happens to the achievable rates when the number of nodes grows to infinity. We conclude the chapter in Section 4.10.

4.3 Examples of Myopic Coding Strategies

Now, we discuss two myopic coding strategies for the MRC, namely myopic DF and myopic amplify-forward (AF). This illustrates that myopic coding is not restricted to only DF.

4.3.1 Myopic DF for the MRC

Let us consider DF for the MRC by Xie & Kumar (2005). In the five-node Gaussian MRC (see Section 2.5.2), a node transmits to all the downstream nodes. Fig. 4.1 depicts the transmissions of the nodes. Let all $U_i, i = 1, 2, 3, 4$, be independent random variables. When node 4 transmits U_4 to node 5, node 3 splits its power, transmitting new information (U_3) to node 4 and helping node 4 to transmit another copy of what node 4 transmits (U_4) to node 5. Similarly, nodes 1–3 split their power to transmit new information and old information (the same information of what the downstream nodes transmit). In decoding, a node decodes the transmissions from all the upstream nodes. For example, node 5 decodes all transmissions from

nodes 1–4. In addition, a node cancels the transmissions from the downstream nodes when it decodes. For example, when node 2 decodes U_1 from node 1, it cancels U_3 and U_4 from node 3, U_4 from node 4, as well as U_2, U_3, and U_4 from node 1. Here, we see how each node cooperates with all other nodes in omniscient DF.

Now, we consider a myopic version of DF in which the nodes are limited in how much information they can store and process. We define *k-hop myopic DF* for the MRC as follows.

Definition 17 *k-hop myopic DF for the MRC is a constrained version of omniscient DF by Xie & Kumar (2005), and the constraints are as follows.*

- *In encoding, a node can transmit messages that it has decoded from only the past k blocks of received signal.*
- *In decoding, a node must decode one message using at most k blocks of received signal.*
- *A node can store a decoded message in its memory over at most k blocks.*

At first glance, the above constraints for myopic DF do not seem to include the view of a node or how many other nodes a node can communicate with. However, these are embedded in the definition itself. The constraints automatically restrict the number of nodes a node can cooperate with. Furthermore, the restrictions stem from practical advantages of having fewer processing and storage requirements at the nodes, which are the motivations behind myopic coding.

Now, let us consider *two-hop myopic DF*. The encoding and the decoding processes at the nodes in the five-node MRC are as follows (refer to Fig. 4.2)

- Node 1 transmits U_1 and U_2, node 2 transmits U_2 and U_3, etc.
- Node 5 decodes U_3 and U_4, node 4 decodes U_2 and U_3, etc.
- During decoding, node 2 cancels U_2 and U_3, node 3 cancels U_3 and U_4, etc.

We note that this encoding technique is different from that by Gupta & Kumar (2003, Fig. 1), in which the source and the relay transmit independent signals (hence no coherent combining is possible) while the relays and the destination decode the transmissions from all upstream nodes, possibly over a large number of blocks. The decoding technique by Gupta & Kumar is only possible under omniscient coding.

For myopic DF for the MRC, we use the concept of regular block Markov encoding and sliding window decoding. However, the encoding and the decoding techniques differ from that found in the literature as the nodes have limited views. It is noted that myopic coding captures point-to-point coding and omniscient coding as special cases. In particular, k-hop myopic DF for the MRC where $k = 1$ is point-to-point (multi-hop) coding and $k = T - 1$ (T is the number of nodes in the channel) omniscient DF.

The reader is reminded that the term "hop" used here does not carry the same meaning as it does in multi-hop routing. The term hop is best understood by looking at the sequence in which messages are decoded, e.g., if the source messages are decoded by node i followed by node j, then node j is node i's next hop.

Definition 18 *We say that a set of nodes $\mathcal{V}$ are in the* view *of node i if node i processes (e.g., decodes, amplifies, or quantizes) or cancels the transmissions from all the nodes in $\mathcal{V}$.*

4.3.2 Myopic AF for the MRC

Next, let us consider AF for the MRC by Yuksel & Erkip (2003). We will use the one-source two-relay one-destination network as an example. Consider the "$S+R_1(S)+R_2(S,R_1)$" scheme (Yuksel & Erkip, 2003, Table I). In this scheme, the transmissions are split into three blocks. In block 1, the source transmits to both relays and the destination (hence the notation S). In block 2, relay 1 normalizes its received signal from the source in block 1 and forwards the normalized received signal to relay 2 and the destination (hence the notation $R_1(S)$). Relay 2 combines

the signals that it has received in blocks 1 and 2, normalizes to its own power value, and transmits the combined signal in block 3 (hence the notation $R_2(S, R_1)$). The destination then decodes using the three blocks of received signal (hence the notation $S + R_1(S) + R_2(S, R_1)$). We term this coding strategy omniscient AF, as each node cooperates with all other nodes.

Now, let us consider myopic AF for the MRC. Yuksel & Erkip noted that relay 2 can choose to listen to only relay 1 (which transmits in block 2) and forwards only this received signal to the destination (the notation used is $R_2(R_1)$). Instead of decoding over three blocks, the destination can choose to decode only from relay 2 (which transmits in block 3). We see that in this scheme, a node listens to only one node and forwards to another node. Hence, we term this strategy *one-hop* myopic AF. One can similarly construct two-hop myopic AF, and so on.

4.4 Practical Advantages of Myopic Coding

In this section, we discuss a few practical advantages of myopic coding compared to omniscient coding. These include simpler code design, increased robustness, reduced computation and memory requirements, and local synchronization. Though the analyses of myopic coding in this thesis are based on information-theoretic achievable rates (in Shannon's sense), the practical advantages here are relevant to codes designs based on these strategies (myopic or omniscient, decode-forward or amplify-forward, etc.). That researchers are interested in practical implementations of information-theoretic cooperative strategies is apparent in the recent work that has been proposed in this direction. There are various codes designed based on omniscient decode-forward for the single-relay channel (Chakrabarti *et al.*, 2007; Ezri & Gastpar, 2006; Khojastepour *et al.*, 2004; Razaghi & Yu, 2006) and the multiple-relay channel (Ong & Motani, 2007a,b; Yu, 2006). One may design myopic versions of these codes to tap the practical advantages discussed in this section.

Looking closely at the LDPC codes using parity forwarding (based on omni-

scient DF) for the MRC (Yu, 2006), we see that the complexity of designing codes grows with the number of relays. This means that constructing codes in which all nodes cooperate can be more difficult compared to designing codes in which nodes only cooperate with neighboring nodes. This technique of utilizing local knowledge (or limited cooperation) is prevalent in other wireless network problems, e.g., cluster-based routing (Jiang & Li, 1999), whereby nodes are split into clusters, and routes are optimized locally.

Myopic coding schemes are more robust to topology changes than the corresponding omniscient coding schemes. For example, consider cancellation of the interference from downstream nodes. In omniscient coding, a node needs to have the knowledge or an estimate of what every downstream node transmits in order to cancel it. Any error in the cancellation (due to topology changes or node failures not known to the decoder) will affect the decoding and thus the rate. In myopic coding, nodes only cancel the interference from a few neighboring nodes. This means that topology changes or node failures beyond a node's view are less likely to affect its decoding. In Appendix B.1, we give another example to show how node failures affect more nodes in myopic coding than in omniscient coding.

In addition, the encoding and decoding computations at each node under myopic coding can be less. Since a node only needs to transmit to and decode from a few nodes, the node encodes fewer data for its transmissions and decodes fewer data from the received signals.

Furthermore, since the nodes need to buffer fewer data for encoding, interference cancellation, and decoding, less memory is required for buffering and codebook storage. Consider the five-node Gaussian MRC. Using omniscient DF, node 1 encodes a message four times over four blocks, using different power splits. Node 5 buffers four blocks of its received signal to decode one message. The buffer grows as the number of nodes in the network increases. On the other hand, using myopic DF, the nodes buffer fewer blocks of received signal, and the buffer size for each node is independent of the number of nodes in the network.

Myopic coding mitigates the need for synchronization of the entire network. Under omniscient DF, all the nodes might need to be synchronized. On the other hand, under myopic coding, a node only needs to synchronize with a few neighboring nodes. Hence, synchronization can be done locally.

In brief, myopic coding can increase the robustness and scalability of the network. In the next section, we analyze the performance of myopic coding in the MRC using DF.

4.5 Achievable Rates of Myopic and Omniscient DF for the MRC

In this section, we construct random codes and derive achievable rates of myopic DF for the MRC. We include omniscient DF here for comparison.

4.5.1 Omniscient Coding

First, we recall that omniscient DF for the MRC achieves rates up to (see Section 2.5)

$$R_{\text{omniscient}} = \max_{\mathcal{M}\in\Pi(\mathcal{T})} \max_{p(x_1,\ldots,x_{T-1})} \min_{m_t\in\mathcal{M}\setminus\{1\}} I(X_{m_1},\ldots,X_{m_{t-1}};Y_{m_t}|X_{m_t},\ldots,X_{m_{|\mathcal{M}|}},X_{\mathcal{M}^c}). \tag{4.1}$$

Here, $\mathcal{M}$ is the route, $\mathcal{T} = \{1,2,\ldots,T\}$ the set of all nodes where node 1 is the source and node T the destination, and $\Pi(\mathcal{T})$ the set of all possible routes from the source to the destination.

Next, we investigate achievable rates of myopic DF. We note that using DF, all relays must fully decode the messages. We assume that the relays decode the messages sequentially.

4.5.2 One-Hop Myopic Coding (Point-to-Point Coding)

Under one-hop myopic DF, a relay node transmits what it has decoded from one block of received signal. This means a node transmits to only the node in the next hop. In decoding, a node decodes one message using one block of received signal. This means a node decodes from only one node behind. A node keeps its decoded message for one block, and it uses the last decoded message to cancel the effect of its own transmission. Using random coding (Shannon, 1948) on route $\mathcal{M} = \{m_1 = 1, m_2, \ldots, m_{|\mathcal{M}|} = T\}$, node m_t can reliably decode data up to the rate

$$R_{m_t} = I(X_{m_{t-1}}; Y_{m_t}|X_{m_t}), \tag{4.2}$$

for some $p(x_1)p(x_2)\cdots p(x_{T-1})$, $t \in \{2, \ldots, T\}$, and $X_{m_T} = 0$. Since all messages must pass through all the nodes in $\mathcal{M}$ in order to reach the destination, one-hop myopic DF can achieve rates up to

$$R_{\text{1-hop}} = \min_{m_t \in \mathcal{M}\setminus\{1\}} R_{m_t}. \tag{4.3}$$

Noting that the message can flow through the relays in any order (Kramer *et al.*, 2003) and maximizing over all input distributions, we have the following result.

Theorem 13 *Let*

$$\left(\mathcal{X}_1 \times \cdots \times \mathcal{X}_{T-1}, p^*(y_2, \ldots, y_T|x_1, \ldots, x_{T-1}), \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T\right)$$

be a memoryless MRC. One-hop myopic DF or point-to-point coding achieves rates up to

$$R_{\text{1-hop}} = \max_{\mathcal{M}\in\Pi(\mathcal{T})} \max_{p(\cdot)} \min_{m_t \in \mathcal{M}\setminus\{1\}} I(X_{m_{t-1}}; Y_{m_t}|X_{m_t}). \tag{4.4}$$

The outer maximization is over all possible routes and the inner maximization is taken over all joint distributions of the form

$$p(x_1, \ldots, x_{T-1}, y_2, \ldots, y_T) = p(x_1)p(x_2)\cdots p(x_{T-1})p^*(y_2, \ldots, y_T|x_1, \ldots, x_{T-1}).$$

4.5.3 Two-Hop Myopic Coding

Instead of just transmitting to only its immediate neighbor, a node might want to help the neighboring node to transmit to the neighbor's neighbor. Under two-hop myopic DF, a node can transmit messages that it has decoded in the past two blocks of received signals. That means in block i, a node transmits data that it has decoded in blocks $i-1$ and $i-2$. In decoding, it decodes one message using only two blocks of received signals. Two-hop myopic DF can achieve up to the rate given in the following theorem.

Theorem 14 *Let*

$$\left(\mathcal{X}_1 \times \cdots \times \mathcal{X}_{T-1}, p^*(y_2, \ldots, y_T | x_1, \ldots, x_{T-1}), \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T\right)$$

be a T-node memoryless MRC. Two-hop myopic DF achieves rates up to

$$R_{\text{2-hop}} = \max_{\mathcal{M}\in\Pi(\mathcal{T})} \max_{p(\cdot)} \min_{m_t \in \mathcal{M}\setminus\{1\}} I(U_{m_{t-2}}, U_{m_{t-1}}; Y_{m_t} | U_{m_t}, U_{m_{t+1}}), \tag{4.5}$$

where $U_{m_0} = U_{m_T} = U_{m_{T+1}} = 0$, for $m_0 = 0$ and $m_{T+1} = T+1$. The outer maximization is over all possible routes and the inner maximization is taken over all joint distributions of the form

$$\begin{aligned} &p(x_1, x_2 \ldots, x_{T-1}, u_1, u_2 \ldots, u_{T-1}, y_2, y_3 \ldots, y_T) \\ &= p(u_{m_1}) p(u_{m_2}) \cdots p(u_{m_{T-1}}) p(x_{m_1} | u_{m_1}, u_{m_2}) p(x_{m_2} | u_{m_2}, u_{m_3}) \cdots \\ &\quad p(x_{m_{T-1}} | u_{m_{T-1}}) p^*(y_2, \ldots, y_T | x_1, \ldots, x_{T-1}). \end{aligned}$$

Proof 17 (Proof of Theorem 14) *See Appendix B.2.*

4.6 Performance Comparison

In this section, we compare the achievable rates of two myopic DF and omniscient DF for the static Gaussian MRC. We first fix the route for all comparison. For

simplicity, we select $\mathcal{M} = \{1, 2, \ldots, T\}$.

In all analyses in this section, we use the following parameters: the channel input from node i, X_i, is a Gaussian random variable with fixed average power $E[X_i^2] = P_i$, the noise power at node $j, j = 2, \ldots, T-1$, is $N_j = N = 1\text{W}$, $\kappa = 1, \eta = 2$, and $\nu_{ij} = 1$.

We consider the Gaussian MRC with fixed average transmit power at the source and at all relays. We note that using DF under omniscient coding, having a maximum average power constraint on individual nodes is equivalent to having a fixed average transmit power constraint on each node, as the overall rate is a non-decreasing function of the average transmit power at any node, keeping the rest of the transmit powers constant. This is because a node decodes the transmissions from all nodes behind and cancels the transmissions from all nodes in front. So, the transmissions of all nodes are either used in decoding or canceled but are never treated as noise. However, under myopic coding, lowering the transmit power at certain nodes may help to reduce the interference at other nodes and increase the overall rate. Hence the achievable rate of the myopic DF with maximum average power constraints on individual nodes is lower bounded by that with fixed average powers.

The achievable rates of one-hop myopic DF and two-hop myopic DF for the Gaussian MRC can be found in Appendix B.3

We define the following efficiency term to benchmark the performance of k-hop myopic coding.

$$\rho_k = \frac{R_{k-\text{hop}}}{R_{\text{omniscient}}}, \tag{4.7}$$

where $k \in \{1, 2, \ldots, T-1\}$. It is the ratio of the maximum achievable rate of a k-hop myopic coding strategy to that of the corresponding omniscient coding strategy.

Figs. 4.3 and 4.4 show achievable rates for the five-node and six-node MRCs respectively, using DF under omniscient coding, two-hop myopic coding, and one-hop myopic coding.

The maximum rate achievable by myopic coding can never exceed that by the corresponding omniscient coding. This is because under myopic coding, every node treats transmissions from the nodes outside its view as noise. In addition, a node can only transmit limited messages. On the other hand, under omniscient coding, a node can decode the signals from all the nodes behind and cancel the transmissions of all the nodes in front. A node can also possibly transmit all previously decoded messages.

In Fig. 4.3, we see a seemingly strange result that the maximum achievable rate of two-hop myopic DF is as high as that of omniscient DF. This can happen in a five-node channel under certain circumstances. Using either omniscient DF or two-hop myopic DF, node 3 in the five-node MRC can communicate with all other nodes, i.e., it decodes from nodes 1 and 2, and cancels transmissions from node 4. So, when the overall transmission rates is constrained by R_3, the maximum achievable rate of the two-hop myopic coding is the same as that of the omniscient coding. This explains why $\rho_2 = 1$ at low transmitted-signal-to-noise ratio (tSNR) in Fig. 4.3. tSNR for a pair of transmitter i and receiver j is P_i/N_j.

However, as the number of relays increases, we expect achievable rates of the two-hop myopic coding to be strictly less than that of the corresponding omniscient coding because each node communicates with fewer neighbors in the former. We see that this is indeed the case from Fig. 4.4, in which ρ_2 is strictly less than 1.

Comparing achievable rates of one-hop myopic DF and two-hop myopic DF, the rates improve significantly when one more node is added into the nodes' view. This suggests that in a large network with many relays, k-hop myopic DF, where k need not be large, could achieve rates close to that of omniscient DF.

Furthermore, ρ_1 and ρ_2 are high in the low tSNR regime. The efficiency drops as the tSNR increases. To understand this phenomenon, we consider different types of noise, i.e., receiver noise and interference. The nodes in both omniscient coding and myopic coding experience the same receiver noise. So, in the low tSNR regime where the receiver noise is dominant, myopic coding performs close to omniscient

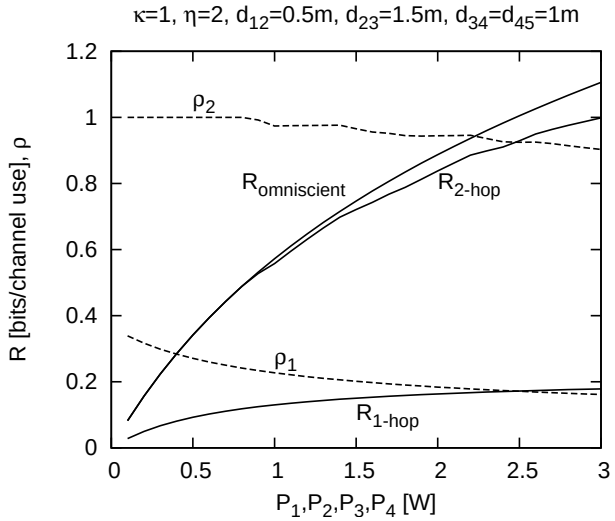


Figure 4.3: Achievable rates of different coding strategies for a five-node MRC.

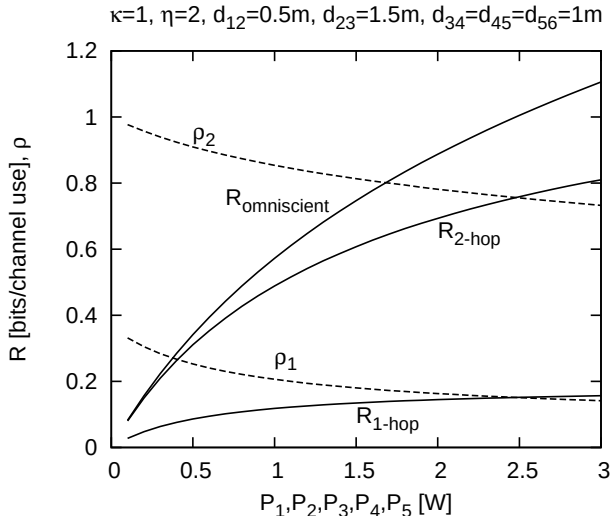


Figure 4.4: Achievable rates of different coding strategies for a six-node MRC.

coding, and the efficiency is higher. On the other hand, in the high tSNR regime, the interference (which a node cannot cancel in myopic coding but can in omniscient coding) is dominant. So, the efficiency of myopic coding drops.

4.7 Extending to k-Hop Myopic Coding

Now, we generalize two-hop myopic DF to k-hop myopic DF where $k \in \{1, \ldots, T-1\}$ and have the following theorem.

Theorem 15 *Let*

$$\left(\mathcal{X}_1 \times \cdots \times \mathcal{X}_{T-1}, p^*(y_2, \ldots, y_T | x_1, \ldots, x_{T-1}), \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T\right)$$

be a T-node memoryless MRC. k-hop DF achieves rates up to

$$R_{k\text{-hop}} = \max_{\mathcal{M} \in \Pi(\mathcal{T})} \max_{p(\cdot)} \min_{m_t \in \mathcal{M} \setminus \{1\}} I(U_{m_{t-k}}, \ldots, U_{m_{t-1}}; Y_{m_t} | U_{m_t}, \ldots, U_{m_{t+k-1}}). \quad (4.8)$$

Here, $U_{m_j} = 0$, for all $j = 2-k, 3-k, \ldots, 0, T, T+1, \ldots, T+k-1$. The outer maximization is over all routes and the inner maximization is taken over all joint distributions of the form

$$\begin{aligned}
&p(x_1, x_2 \ldots, x_{T-1}, u_1, u_2 \ldots, u_{T-1}, y_2, y_3 \ldots, y_T) \\
&= p(u_{m_1}) p(u_{m_2}) \cdots p(u_{m_{T-1}}) \\
&\quad \times p(x_{m_{T-1}} | u_{m_{T-1}}) p(x_{m_{T-2}} | u_{m_{T-2}}, u_{m_{T-1}}) \cdots p(x_{m_{T-k}} | u_{m_{T-k}}, u_{m_{T-k+1}} \ldots, u_{m_{T-1}}) \\
&\quad \times p(x_{m_{T-k-1}} | u_{m_{T-k-1}}, u_{m_{T-k}} \ldots, u_{m_{T-2}}) \cdots p(x_{m_1} | u_{m_1}, u_{m_2}, \ldots, u_{m_k}) \\
&\quad \times p^*(y_2, \ldots, y_T | x_1, \ldots, x_{T-1}).
\end{aligned}$$

The proof can be found in Appendix B.4. In the extreme case where $k = T-1$, we end up with omniscient DF.

4.8 On the Fading Gaussian MRC

In the analyses so far, we compared the performance of myopic coding in static Gaussian channels, i.e., without fading. Now, we explain how myopic coding is done in the Gaussian channel with phase fading or Rayleigh fading.

It has been shown by Kramer *et al.* (2005, Theorem 8) that under phase fading or Rayleigh fading, the maximum omniscient DF rate can be achieved by independent Gaussian input distributions. In this case, $X_i, i = 1, \ldots, T-1$, are independent Gaussian random variables. Under omniscient DF, node t decodes from all nodes

$i, i < j$, and cancels the transmissions of nodes $l, l \geq j$. In k-hop myopic DF, the nodes transmit independent Gaussian signals as they would under omniscient DF. However, in the decoding, node t decodes the signals only from k nodes behind, i.e., nodes $i, i = \max\{1, t-k\}, \ldots, t-1$. It cancels the transmissions from only k nodes in front (including itself), i.e., nodes $l, l = t, \ldots, \min\{t+k-1, T-1\}$. It treats the rest of the transmissions as noise. The following theorem characterizes the performance of k-hop myopic DF in the Gaussian MRC with phase fading or Rayleigh fading.

Theorem 16 *Consider a T-node Gaussian MRC with phase fading or Rayleigh fading. Using k-hop DF, the rate in equation (4.8) is achievable, by setting $X_i = U_i, x_i = u_i, \forall i = 1, 2, \ldots, T-1$.*

The proof for the above theorem is straight forward given that the nodes transmit independent signals in the fading channel.

4.9 Myopic Coding on Large MRCs

One potential problem of myopic coding is whether the rate vanishes when the number of nodes in the network grows. This concern arises because in myopic DF, a node treats transmissions of nodes beyond its view as pure noise. As the number of transmitting nodes grows to infinity and each decoding node only has a limited view, the noise power might sum to infinity. The noise might overpower the signal power and drive the transmission rate to zero.

In this section, we scrutinize achievable rates of two-hop myopic DF for the T-node MRC when T grows to infinity. The rationale of studying the two-hop myopic coding is that we can always achieve higher transmission rates using k-hop myopic DF with $k > 2$.

Theorem 17 *Achievable rates of k-hop myopic DF for the T-node Gaussian MRC are bounded away from zero, for any $T \geq 3$.*

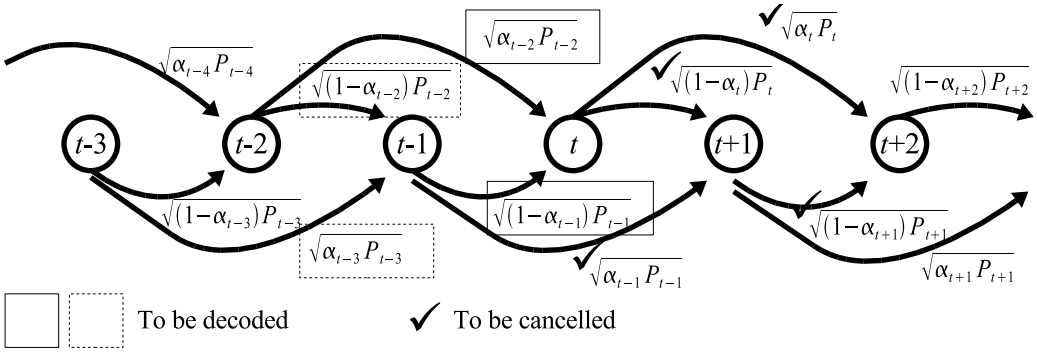


Figure 4.5: Power allocations for two-hop myopic DF for the Gaussian MRC.

Now, we prove Theorem 17. Using two-hop myopic DF on route $\mathcal{M} = \{1, 2, \ldots, T\}$ in the T-node Gaussian MRC (we shall extend T to infinity later), the transmission of each node is as follows.

- Node $t, t = 1, 2, \ldots, T-2$, sends $X_t = \sqrt{\alpha_t P_t} U_{t+1} + \sqrt{(1-\alpha_t)P_t} U_t$.
- Node $T-1$ sends $X_{T-1} = \sqrt{P_{T-1}} U_{T-1}$.

where $U_i, i = 1, 2, \ldots, T-1$, are independent Gaussian random variables with unit variances and $0 \leq \alpha_i \leq 1$. The transmissions of the nodes around node t are depicted in Fig. 4.5.

Assume that all the nodes are equally spaced at 1m apart and transmit at power P. Consider the received signal power at node t, we can always find a non-empty set $\{(\alpha_1, \ldots, \alpha_{T-2}) : 0 \leq \alpha_i \leq 1, i = 1, \ldots, T-2\}$ such that

$$\begin{aligned} P_{\text{sig}}(t) &= \left(\sqrt{3^{-\eta}\alpha_{t-3}\kappa P} + \sqrt{2^{-\eta}(1-\alpha_{t-2})\kappa P}\right)^2 \\ &\quad + \left(\sqrt{2^{-\eta}\alpha_{t-2}\kappa P} + \sqrt{1^{-\eta}(1-\alpha_{t-1})\kappa P}\right)^2 \qquad (4.10a) \\ &= \left(\sqrt{3^{-\eta}\alpha_{t-3}\kappa P} + \sqrt{2^{-\eta}(1-\alpha_{t-2})\kappa P}\right)^2 \\ &\quad + \left(\sqrt{2^{-\eta}\alpha_{t-2}\kappa P} + \sqrt{1^{-\eta}(1-\alpha_{t-1})\kappa P}\right)^2 \qquad (4.10b) \\ &> 0, \qquad (4.10c) \end{aligned}$$

for $t \geq 4$, and

$$P_{\text{sig}}(2) = (1-\alpha_1)\kappa P > 0 \tag{4.11a}$$

$$P_{\text{sig}}(3) = 2^{-\eta}(1-\alpha_1)\kappa P + \left(\sqrt{2^{-\eta}\alpha_1\kappa P} + \sqrt{1^{-\eta}(1-\alpha_2)\kappa P}\right)^2 > 0. \tag{4.11b}$$

Now we consider nodes $4 \leq t \leq T-3$, the noise power is $P_{\text{noise}}(t) = N_t < \infty$, and the interference power is given by

$$\begin{aligned} P_{\text{int}}(t) = &\left(\sqrt{3^{-\eta}(1-\alpha_{t-3})\kappa P} + \sqrt{4^{-\eta}\alpha_{t-4}\kappa P}\right)^2 \\ &+ \left(\sqrt{4^{-\eta}(1-\alpha_{t-4})\kappa P} + \sqrt{5^{-\eta}\alpha_{t-5}\kappa P}\right)^2 + \cdots \\ &+ \left(\sqrt{(t-2)^{-\eta}(1-\alpha_2)\kappa P} + \sqrt{(t-1)^{-\eta}\alpha_1\kappa P}\right)^2 + (t-1)^{-\eta}(1-\alpha_1)\kappa P \\ &+ \left(\sqrt{1^{-\eta}\alpha_{t+1}\kappa P} + \sqrt{2^{-\eta}(1-\alpha_{t+2})\kappa P}\right)^2 \\ &+ \left(\sqrt{2^{-\eta}\alpha_{t+2}\kappa P} + \sqrt{3^{-\eta}(1-\alpha_{t+3})\kappa P}\right)^2 + \cdots \\ &+ \left(\sqrt{(T-t-3)^{-\eta}\alpha_{T-3}\kappa P} + \sqrt{(T-t-2)^{-\eta}(1-\alpha_{T-2})\kappa P}\right)^2 \\ &+ \left(\sqrt{(T-t-2)^{-\eta}\alpha_{T-2}\kappa P} + \sqrt{(T-t-1)^{-\eta}\kappa P}\right)^2, \end{aligned} \tag{4.12a}$$

$$\begin{aligned} \frac{P_{\text{int}}(t)}{\kappa P} = &\, 3^{-\eta}\alpha_{t-3} + 4^{-\eta} + 5^{-\eta} + \cdots + (t-1)^{-\eta} + 2\sqrt{3^{-\eta}4^{-\eta}(1-\alpha_{t-3})\alpha_{t-4}} \\ &+ 2\sqrt{4^{-\eta}5^{-\eta}(1-\alpha_{t-4})\alpha_{t-5}} + \cdots + 2\sqrt{(t-2)^{-\eta}(t-1)^{-\eta}(1-\alpha_2)\alpha_1} \\ &+ 1^{-\eta}\alpha_{t+1} + 2^{-\eta} + 3^{-\eta} + \cdots + (T-t-1)^{-\eta} + 2\sqrt{1^{-\eta}2^{-\eta}\alpha_{t+1}(1-\alpha_{t+2})} \\ &+ 2\sqrt{2^{-\eta}3^{-\eta}\alpha_{t+2}(1-\alpha_{t+3})} + \cdots \\ &+ 2\sqrt{(T-t-3)^{-\eta}(T-t-2)^{-\eta}\alpha_{T-3}(1-\alpha_{T-2})}. \end{aligned} \tag{4.13a}$$

Simplifying, we get

$$\frac{P_{\text{int}}(t)}{\kappa P} = 3^{-\eta}\alpha_{t-3} + \sum_{j=4}^{t-1}\frac{1}{j^\eta} + 1^{-\eta}\alpha_{t+1} + \sum_{j=2}^{T-t-1}\frac{1}{j^\eta} + 2\sum_{j=3}^{t-2}\sqrt{\frac{(1-\alpha_{t-j})\alpha_{t-(j+1)}}{j^\eta(j+1)^\eta}} + 2\sum_{j=1}^{T-t-3}\sqrt{\frac{\alpha_{t+j}(1-\alpha_{t+j+1})}{j^\eta(j+1)^\eta}} \tag{4.14a}$$

$$< \sum_{j=3}^{t-1}\frac{1}{j^\eta} + \sum_{j=1}^{T-t-1}\frac{1}{j^\eta} + 2\sum_{j=3}^{t-2}\frac{1}{j^\eta} + 2\sum_{j=1}^{T-t-3}\frac{1}{j^\eta} \tag{4.14b}$$

$$< 6\sum_{j=1}^{T}\frac{1}{j^\eta} < 6\zeta(\eta). \tag{4.14c}$$

Here $\zeta(\eta) = \sum_{j=1}^{\infty}\frac{1}{j^\eta}$ is the Riemann zeta function. It has been calculated that $\zeta(2) = \frac{\pi^2}{6}$, $\zeta(3) = 1.202057...$ etc. It is easily seen that the Riemann zeta function is a decreasing function of η. Since, $\eta \geq 2$, $P_{\text{int}}(t) < \pi^2\kappa P$ for $4 \leq t \leq T-3$. We can also show that $P_{\text{int}}(t)/(\kappa P)$ for $t = 2, 3, T-2, T-1, T$ are bounded. Hence, we can always find a non-empty set $\{(\alpha_1, \ldots, \alpha_{T-2})\}$ such that the reception rate at every node t, $\forall t \in \{2, 3, \ldots, T\}$, is

$$R_t = \frac{1}{2}\log\left[1 + \frac{P_{\text{sig}}(t)}{P_{\text{int}(t)} + N_t}\right] > 0, \tag{4.15}$$

which is bounded away from zero. This means the maximum achievable rate

$$R_{\text{2-hop}} = \max_{\{\alpha_1,\ldots,\alpha_{T-2}\}} \min_{t\in\{2,3,\ldots,T\}} R_t > 0 \tag{4.16}$$

is bounded away from zero.

When more nodes are included in the view in the myopic coding, P_{sig} increases and P_{int} decreases. In general, assuming that the nodes are roughly equally spaced, the maximum achievable rate of myopic DF is bounded away from zero even when the network size grows to infinity.

4.10 Conclusion

In this chapter, we compared achievable rates of myopic DF and omniscient DF for the MRC.

We have shown that in the low tSNR regime, achievable rates of two-hop myopic DF are as large as that of omniscient DF in a five-node MRC, and close to that of the omniscient coding in a six-node MRC. Comparing the one-hop myopic coding and the two-hop myopic coding, we see that adding a node into the nodes' view improves the achievable rate significantly.

Chapter 5

Achievable Rate Regions for the Multiple-Access Channel with Feedback and Correlated Sources

In this chapter, we study achievable rates for the multiple-access channel with feedback and correlated sources (MACFCS) (Ong & Motani, 2005c, 2006b, 2007d). The MACFCS is a combination of the multiple access channel with correlated sources (MACCS) and the multiple access channel with feedback (MACF). The MACFCS serves as a model for the wireless sensor network in which multiple sources send possibly correlated data to a single destination. At the same time, each source receives feedback from the channel and we allow each node to receive different feedback. First of all, we define the MACFCS.

5.1 Introduction

5.1.1 The MACFCS

Fig. 5.1 depicts the three-node MACFCS, with nodes $\{1, 2, 3\}$. Nodes 1 and 2 are the sources (which can also act as relays), and node 3 the destination. Message $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$ are generated at nodes 1 and 2 respectively, and are

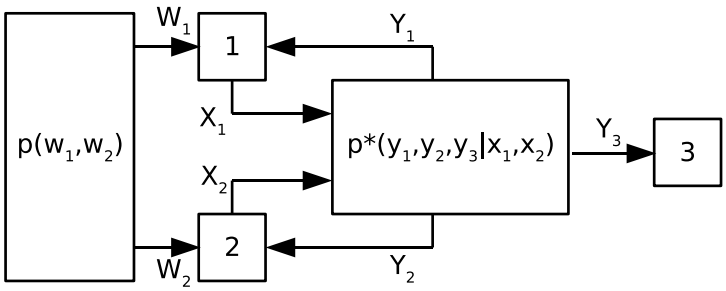


Figure 5.1: The three-node MACFCS.

to be sent to node 3. They are drawn from some discrete bivariate distribution $p(s_1, s_2)$, meaning that the source messages can be arbitrarily correlated. The three-node discrete memoryless MACFCS can be completely described by $\big(\mathcal{W}_1 \times \mathcal{W}_2, p(w_1, w_2), \mathcal{X}_1 \times \mathcal{X}_2, p^*(y_1, y_2, y_3|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3\big)$. $\mathcal{W}_1, \mathcal{W}_2, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2$, and $\mathcal{Y}_3$ are seven finite sets. $p^*(y_1, y_2, y_3|x_1, x_2)$ defines the channel transition probability on $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3$ for each $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. x_1 and x_2 are the inputs into the channel from nodes 1 and 2 respectively. y_1, y_2, and y_3 are the channel outputs to nodes 1, 2, and 3 (the destination) respectively. We consider memoryless channels, i.e., the current outputs (y_{1i}, y_{2i}, y_{3i}) depend on the past inputs (x_1^i, x_2^i) only through the current transmitted symbols (x_{1i}, x_{2i}).

Definition 19 *A sequence of codes* $\Big\{\{f_{1i}, f_{2i}\}_{i=1}^n, g, n\Big\}$ *for the three-node MACFCS comprises of an integer* n*, two sets of encoding functions* $\{f_{1i}, f_{2i}\}_{i=1}^n$ *at nodes 1 and 2, where*

$$x_{1i} = f_{1i}(w_1^n, y_{11}, y_{12}, \ldots, y_{1(i-1)}) \tag{5.1a}$$

$$x_{2i} = f_{2i}(w_2^n, y_{21}, y_{22}, \ldots, y_{2(i-1)}), \tag{5.1b}$$

and a decoding function at node 3, $g_3 : \mathcal{Y}_3^n \to \mathcal{W}_1^n \times \mathcal{W}_2^n$*, such that*

$$(\hat{w}_1^n, \hat{w}_2^n) = g_3(y_3^n) \tag{5.2}$$

where $\hat{w}_1^n$ *and* $\hat{w}_2^n$ *are estimates of* w_1^n *and* w_2^n *respectively.*

Without loss of generality, we assume that each encoder knows all n source messages before the encoding of each block. We can also think of the n source messages as one combined message from the source at the beginning of each block of the encoding. Hence, for any choice of codes, the joint probability density function (p.d.f.) on $(\mathcal{W}_1^n, \mathcal{W}_2^n, \mathcal{X}_1^n, \mathcal{X}_2^n, \mathcal{Y}_1^n, \mathcal{Y}_2^n, \mathcal{Y}_3^n)$ is given by

$$p(w_1^n, w_2^n, x_1^n, x_2^n, y_1^n, y_2^n, y_3^n) = \prod_{i=1}^{n} p(w_{1i}, w_{2i}) \prod_{i=1}^{n} p(x_{1i}|w_1^n, y_{11}, y_{12}, \ldots, y_{1i-1}) \cdot p(x_{2i}|w_1^n, y_{21}, y_{22}, \ldots, y_{2i-1}) \cdot p^*(y_{1i}, y_{2i}, y_{3i}|x_{1i}, x_{2i}). \tag{5.3a}$$

Definition 20 *The error probability is defined as*

$$P_e = \Pr\{(\hat{W}_1^n, \hat{W}_2^n) \neq (W_1^n, W_2^n)\} \tag{5.4a}$$

$$= \sum_{(w_1^n, w_2^n) \in \mathcal{W}_1^n \times \mathcal{W}_2^n} p(w_1^n, w_2^n) \Pr\{(\hat{W}_1^n, \hat{W}_2^n) \neq (w_1^n, w_2^n) | (W_1^n, W_2^n) = (w_1^n, w_2^n)\}. \tag{5.4b}$$

Definition 21 *We say that (W_1, W_2) can be reliably transmitted to the destination per channel use if for any $\epsilon > 0$, there exists a sequence of block codes $\left\{\{f_{1i}, f_{2i}\}_{i=1}^n, g, n\right\}$ such that $P_e < \epsilon$.*

We define an *achievable region* of the MACFCS as a set of triplets $[H(W_1|W_2), H(W_2|W_1), H(W_1, W_2)]$ for which we can reliably send (W_1, W_2) to the destination per channel use, for some $p^*(y_1, y_2, y_3|x_1, x_2)$. The *capacity region* is the closure of the set of all achievable regions.

In Section 5.7, where we compare different strategies for the static three-node Gaussian MACFCS, we use an alternative but useful definition of achievable region. The reason is that regions in the three dimensional space are difficult to plot and compare. Hence we use the following version of achievable region for the static Gaussian MACFCS. With a fixed correlation structure $H(W_1|W_2)$, $H(W_2|W_1)$, and $H(W_1, W_2)$, and node positions, an achievable region is the set of average transmit

power pairs $[P_1, P_2]$ for which we can reliably send (W_1, W_2) to the destination. Similarly, the capacity is defined as the closure of the convex hull of the set of all achievable regions. Note that we can convert an achievable region for the discrete memoryless MACFCS to that for the Gaussian MACFCS.

5.1.2 Problem Statement

For the three-node MACFCS, we are interested in investigating:

1. What is the tightest upper bound on the capacity of the MACFCS?
2. What are the achievable regions of various coding strategies for the MACFCS?
3. What are the characteristics of different coding strategies for the MACFCS?
4. How well do the coding strategies perform under different channel settings?
5. How does the study of the MACFCS help us to better understand coding and cooperation in sensor networks?

5.2 Related Work

The MACFCS is a combination of the multiple-access channel with correlated sources (MACCS) and the multiple-access channel with feedback (MACF). One practical setup of the MACFCS is the sensor network.

The MACCS (with a common part) was studied by Slepian & Wolf (1973b), who derived an achievable region. In their paper, separate source coding and channel coding are used, where the source coding is first performed to remove the correlation among the sources. The channel coding for the multiple-access channel (MAC) with independent sources is then employed. The MACCS (with possibly no common part) was considered by Cover *et al.* (1980). They showed, by using a simple example, that separating source and channel coding is not optimal. They derived an achievable region for the MACCS using a combined source-channel coding strategy

to preserve the correlation among the channel inputs. Outer bounds on the capacity of the MACCS were derived with infinite letter characterization by Cover *et al.* and later improved by Kang & Ulukus (2006) to finite-letter expressions. While Slepian & Wolf (1973b) assumes a certain structure for the correlation among the sources, we study arbitrarily correlated sources in this thesis.

The MACF (with independent sources) was investigated by Cover & Leung (1981), who derived an achievable region assuming all nodes receive common feedback. Ozarow (1984) found the capacity of the Gaussian MACF with common feedback and derived a capacity outer bound for the discrete memoryless MACF with common feedback. King (1978) investigated the MACF with all sources receiving common feedback, which is possibly different from what the destination receives, and derived an achievable region for the channel. Willems (1982) and Carleial (1982) further generalized the MACF with common channel feedback to the case where each node receives possibly different channel feedback, and derived achievable regions of the channel. Sendonaris *et al.* (2003a,b) considered the Gaussian MACF with different feedback to different nodes. They derived an information theoretic achievable region based on cooperation among the source nodes, and showed how the cooperation scheme can be implemented in a practical code-division multiple-access system.

Combining the MACF and the MACCS, we arrive at the MACFCS. One practical system modeled by the MACFCS is a sensor network in which every sensor is capable of transmitting as well as receiving, and each sensor collects data and aims to send them to a single destination. We note that the data collected by the sensor nodes might be correlated, e.g., if they are located close to one another.

Applying coding strategies designed for the MACF or the MACCS might be suboptimal for the MACFCS. Coding strategies for the MACF ignore the correlation among the sources, while coding strategies for the MACCS ignore the feedback from the channel to the sources. Taking both these extra pieces of information into account can help to enlarge the achievable region.

Murugan *et al.* (2004) investigated the Gaussian MACFCS with a total average power constraint on the sources. Their coding approach is based on time division multiple-access with the nodes operating in half-duplex. Our work differs from Murugan *et al.* in that we consider a general MACFCS (including both discrete memoryless channels and Gaussian channels) with full-duplex nodes, in which the source nodes can transmit and receive simultaneously. For the Gaussian case, we impose average power constraints on individual sources, rather than a total average power constraint. King (1978) considered the MACFCS with each source observing an independent private message, all sources observing a common message, and all nodes (all sources and the destination) receiving the same feedback. In this thesis, we consider arbitrary source correlation and possibly different feedback to all nodes.

5.3 Contributions

Our main contributions in this chapter are:

1. We derive an outer bound on the capacity of the MACFCS, which turns out to be the cut-set bound (Gastpar, 2003, 2004).

2. We construct a new coding strategy for the MACFCS, where the source nodes first exchange information and then cooperate to send full information to the destination. We term this strategy *full decoding at sources with DF channel coding* (FDS-DF) (Ong & Motani, 2005c, 2006b, 2007d). We derive an achievable region using this strategy.

3. We construct a CF-based coding strategy for the MACF, with each node receiving possibly different channel feedback. King (1978) derived an achievable region for the MACF with all sources receiving common feedback using combined DF and CF coding strategies.

4. We combine source coding for correlated sources (Slepian & Wolf, 1973b)

and our newly constructed CF for the MACF, and arrive at a new achievable region for the MACFCS. We term this strategy *source coding for correlated sources and CF channel coding for the MACF* (SC-CF).

5. We combine existing schemes, i.e., source coding for correlated sources (Slepian & Wolf, 1973b) with the MAC channel coding (Ahlswede, 1974; Liao, 1972) to arrive at another achievable region for the MACFCS. We term this strategy *source coding for correlated sources and MAC channel coding* (SC-MAC).

6. We find another achievable region of the MACFCS using a multi-hop coding strategy.

7. We compute achievable regions of the different strategies on the Gaussian MACFCS.

8. We show that certain strategies perform better under certain source correlation structures and channel topologies. More specifically, we observe the following for the symmetrical MACFCS (where the sources are of equi-distant from the destination, and they have the same amount of private information to send):

   (a) When the inter-source links get better than the source-destination links, FDS-DF approaches the capacity outer bound.

   (b) When the correlation among the sources gets higher, FDS-DF approaches the capacity outer bound.

   When one source is far away from the destination and another source is closer to the destination, SC-CF gives a better performance compared to FDS-DF and SC-MAC.

9. By comparing different coding strategies for the MACFCS, we show the value of cooperation in the multiple-source single-sink sensor network.

5.3.1 Organization

The rest of the chapter is organized as follows. In Section 5.4, we briefly mention the coding strategies for the MACFCS discussed in this thesis. In Section 5.5, we derive a capacity outer bound for the MACFCS. These will serve as benchmarks for the coding strategies constructed in Section 5.6. We compare the performance of different coding strategies on the Gaussian MACFCS in Section 5.7. This is followed by a discussion of the results in Section 5.8 and conclusions in Section 5.9.

Now, we briefly describe the different coding strategies investigated for the MACFCS.

5.4 Coding Strategies for the MACFCS

There are numerous coding strategies which we can apply to the MACFCS. The aim of this thesis is not to list all of them, but to compare different strategies and to study their strengths and weaknesses. In this chapter, we study the following coding strategies for the MACFCS.

1. *Full Decoding at Sources with Decode-Forward Channel Coding (FDS-DF):* In FDS-DF, the general idea is for the sources to communicate so that every source has the complete data of the other sources. They then cooperate to send the combined data to the destination. Since the data of different nodes are correlated, a node does not need to send all its data to other nodes for them to fully decode the data.

2. *Source Coding for Correlated Sources and Compress-Forward Channel Coding for the MACF (SC-CF):* Source coding for correlated sources (Slepian & Wolf, 1973b) is first performed at every source node to remove the correlation among the sources. At this point, we have turned the problem into that of channel coding for the MACF with independent sources. We then construct a coding strategy for the MACF based on CF to transport the independent data to the destination.

3. *Source Coding for Correlated Sources and MAC Channel Coding (SC-MAC):* Source coding with correlated sources is performed at individual nodes. Then we use channel coding for the MAC (Ahlswede, 1974; Liao, 1972) to send the independent data to the destination. In this case, we disregard the feedback from the channel to the source nodes.

4. *Multi-Hop Coding with Data Aggregation (MH-DA):* The nodes are sequenced (with the last node being the destination) to form a route. Each node (except the first node) decodes the data from the previous node in the route, combines it with its own data, and forwards all data (data that it decodes from the previous node, plus its own data, less the correlated part of the data which the next node already has) to the next node in the route. This continues until the second last node sends all aggregated data to the destination.

Remark 9 *The first two strategies, i.e., FDS-DF and SC-CF, use coding ideas for the relay channel, in which the relay helps the source to send data to the destination. These two strategies exploit the fact that there is an embedded relay channel in the MACFCS.*

Remark 10 *In SC-MAC, the sources ignore feedback from the channel. Feedback certainly has the potential to increase rates, but taking it into account carries with it a certain amount of complexity, both from a hardware and processing viewpoint. This is the motivation for SC-MAC and we find that this simple strategy can actually be better under certain topologies.*

Remark 11 *The first three strategies mentioned above involve multi-user coding (e.g., multi-point-to-multi-point), which requires a certain amount of coordination for synchronization and cooperation. In MH-DA, all transmissions are single-point-to-single-point, i.e., a node only decodes from a node behind it, treating all other transmissions as noise. We note that there are many practical coding schemes available for single-point-to-single-point communication. Through MH-DA, we can*

study the loss in performance of single-point-to-single-point coding in a multiple-terminal network.

Remark 12 *Barros & Servetto (2006) consider the problem of communicating correlated sources over a network of independent point-to-point links. The strategy by Barros & Servetto includes MH-DA as a special case and can be used for the MACFCS with appropriate modifications.*

5.4.1 The Value of Cooperation in the MACFCS

In the wireless channel, which is broadcast in nature, every node hears the transmissions of other nodes. It can treat the transmissions as pure noise, or make use of the received transmissions for cooperation. In the coding strategies described above, the nodes *cooperate* in the encoding and decoding of the data. Across the strategies, we find different levels of cooperation.

In all the strategies, we see nodes cooperate in the source coding, i.e., a node takes into account of other nodes (their data or correlation structure) during its data encoding. In FDS-DF, all the nodes send cooperative data (of all sources) to the destination. In SC-CF and SC-MAC, source coding for correlated sources is performed prior to channel coding. We can view this as a form of cooperation in the encoding. In MH-DA, a node receives data from the previous node, combines them with its own data, and sends the aggregated data to the next node. Again, we see cooperation in the encoding of data.

Now, we see how the nodes cooperate in the channel coding, e.g., multiple nodes decode the transmission of a node, and a node decodes the transmissions of multiple nodes. In FDS-DF, when the sources are exchanging data, the destination, overhearing these transmissions, makes use of the transmissions to aid its decoding of the data. In SC-CF, each source hears the transmissions of other sources, quantizes them, bins them, and sends them to the destination. In SC-MAC, though the sources ignore the transmissions of other nodes, the destination listens to all the source nodes. The coding strategies above involve channel coding for multiple

users. In contrast, MH-DA only considers node pairs, i.e., point-to-point coding. Encoding and decoding are done only for two nodes. Each node only transmits to one node (down the route) and each node only decodes from one node, ignoring all other transmissions. Hence, we see minimum cooperation in MH-DA.

5.5 Capacity Outer Bound

In this section, we derive an outer bound on the capacity of the MACFCS.

Theorem 18 (Cut-Set Outer Bound) *Let* $\left(\mathcal{W}_1 \times \mathcal{W}_2, p(w_1, w_2), \mathcal{X}_1 \times \mathcal{X}_2, p^*(y_1, y_2, y_3|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3\right)$ *be a discrete memoryless three-node MACFCS. The source symbols* (W_1, W_2) *can be reliably transmitted to the destination per channel use only if*

$$[H(W_1|W_2), H(W_2|W_1), H(W_1, W_2)] \in \mathbf{R}, \tag{5.5}$$

where

$$\mathbf{R} = \left\{ [R_1, R_2, R_3] : \left\{ \begin{array}{l} R_1 \leq I(X_1; Y_2, Y_3|X_2) \\ R_2 \leq I(X_2; Y_1, Y_3|X_1) \\ R_3 \leq I(X_1, X_2; Y_3) \end{array} \right\} \right\}, \tag{5.6}$$

for some

$$p(x_1, x_2). \tag{5.7}$$

In other words, an outer bound on the capacity of the MACFCS is given by $\mathbf{R}_{\text{OB}} = \bigcup_{\mathcal{P}} \mathbf{R}$, *where* $\mathcal{P}$ *is the set of all distributions satisfying* (5.7), *and* $\bigcup$ *is the union of sets operator.*

Proof 18 (Proof of Theorem 18) *See Appendix C.1.*

Remark 13 *We call the above outer bound the cut-set outer bound (CS-OB) as it turns out to be a special case of the cut-set argument by Gastpar (2003, 2004). Now, we start with the cut-set argument and see how it simplifies to the CS-OB. We partition the network into two sets, with a* cut *separating the sets. We assume that all nodes in each set can fully cooperate. We obtain bounds by as-*

sociating each cut with a corresponding point-to-point system. Consider the cut separating the sets $\{1\}$ *and* $\{2,3\}$*. The transmission rate from node 1 to nodes 2 and 3 is bounded by the corresponding point-to-point system* $X_1 \rightarrow (Y_2, Y_3)|X_2$ *(using the notation by Gastpar (2003)). In this point-to-point channel, node 3 receives side information* W_2 *from node 2. For node 3 to reliably decode* W_1*, node 1 needs to transmit at least* $H(W_1|W_2)$ *bits across the cut, to node 3. Hence we get* $H(W_1|W_2) \leq \max_{p(x_1,x_2)} I(X_1; Y_2, Y_3|X_2)$ *(cf. Gastpar (2003, eq. (3.9))). Applying this argument to the cut separating* $\{2\}$ *and* $\{1,3\}$*, we obtain the second inequality in* (5.6)*. Consider the cut separating* $\{1,2\}$ *and* $\{3\}$*. We need to transmit* (W_1, W_2) *across the cut, and the transmission rate is bounded by the corresponding point-to-point system* $(X_1, X_2) \rightarrow (Y_3)$*. Hence we get* $H(W_1, W_2) \leq \max_{p(x_1,x_2)} I(X_1, X_2; Y_3)$ *(cf. Gastpar (2003, eq. (3.2))). Note that for the point-to-point system, feedback does not increase the capacity, and can be ignored.*

Remark 14 *In the Gaussian MACFCS,* $\mathbf{R}_{\mathrm{OB}}$ *can be found by considering only jointly Gaussian input distributions. We can show that choosing Gaussian input distributions maximizes every mutual information expression in* (5.6) *(Kramer et al., 2005, Proposition 2). Hence, in the Gaussian MACFCS,* $\bigcup_{p(x_1,x_2)} \mathbf{R} = \bigcup_{\text{jointly Gaussian } x_1,x_2} \mathbf{R}$.

5.6 Achievability

Now, we present four achievable regions for the three-node MACFCS using four different coding strategies.

5.6.1 Full Decoding at Sources with Decode-Forward Channel Coding (FDS-DF)

In this strategy, every node decodes the data from all other nodes, and all nodes cooperate to send combined data to the destination. We note that for the nodes to

cooperate, they must first agree on the data to be sent. In order to do this, each of them must first decode the data from all other nodes.

In brief, this strategy does the following. Since W_1 and W_2 are correlated, using the method by Slepian & Wolf (1973a, Theorem 2), node 1 only needs to send $H(W_1|W_2)$ compressed bits to node 2 for it to decode W_1. Node 2 does the same. Now, both nodes have W_1 and W_2. They then cooperate to transmit the full information, i.e., (W_1, W_2), to the destination. At the same time, nodes 1 and 2 send the next (new) message to each other.

Murugan *et al.* (2004) proposed a similar coding scheme where the transmissions are split into two phases. In the first phase, the source nodes communicate with each other using time division multiple-access. At the end of the first phase, each source has the data of all nodes. In the second phase, all sources cooperate to transmit to the destination. In this thesis, we offer a more general coding scheme. Each source node transmits cooperative information of the previous block (data that it decodes from other nodes together with its own data) and new information (which is to be decoded by other sources and the destination) simultaneously. Since all nodes agree on the same fully decoded information of the previous block, *coherent combining* can be achieved in the Gaussian channel. We show that the coding strategy proposed by Murugan *et al.* is a special case of ours.

Using FDS-DF, we can show that the region given in the following theorem is achievable.

Theorem 19 (FDS-DF) *Let* $\left(\mathcal{W}_1 \times \mathcal{W}_2, p(w_1, w_2), \mathcal{X}_1 \times \mathcal{X}_2, p^*(y_1, y_2, y_3|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3\right)$ *be a discrete memoryless three-node MACFCS.* (W_1, W_2) *can be reliably*

transmitted to the destination per channel use if the following conditions hold.

$$H(W_1|W_2) < \min[I(X_1;Y_2|Q,V_0,V_1,V_2,X_2), I(V_1;Y_3|Q,V_0,V_2) + I(X_1;Y_3|Q,V_0,V_1,V_2,X_2)], \tag{5.8a}$$

$$H(W_2|W_1) < \min[I(X_2;Y_1|Q,V_0,V_1,V_2,X_1), I(V_2;Y_3|Q,V_0,V_1) + I(X_2;Y_3|Q,V_0,V_1,V_2,X_1)], \tag{5.8b}$$

$$I(W_1;W_2) < I(V_0;Y_3|Q,V_1,V_2), \tag{5.8c}$$

$$H(W_1) < I(V_0,V_1;Y_3|Q,V_2) + I(X_1;Y_3|Q,V_0,V_1,V_2,X_2), \tag{5.8d}$$

$$H(W_2) < I(V_0,V_2;Y_3|Q,V_1) + I(X_2;Y_3|Q,V_0,V_1,V_2,X_1), \tag{5.8e}$$

$$H(W_1|W_2) + H(W_2|W_1) < I(V_1,V_2;Y_3|Q,V_0) + I(X_1,X_2;Y_3|Q,V_0,V_1,V_2), \tag{5.8f}$$

$$H(W_1,W_2) < I(X_1,X_2;Y_3|Q), \tag{5.8g}$$

where

$$p(q,x_1,x_2,y_1,y_2,y_3,v_0,v_1,v_2) = p(q)p(v_0|q)p(v_1|q)p(v_2|q)p(x_1|q,v_0,v_1,v_2) \cdot p(x_2|q,v_0,v_1,v_2)p^*(y_1,y_2,y_3|x_1,x_2,x_3). \tag{5.9a}$$

$V_0 \in \mathcal{V}_0$, $V_1 \in \mathcal{V}_1$ *and* $V_2 \in \mathcal{V}_2$ *are auxiliary random variables with cardinalities* $|\mathcal{V}_0| \times |\mathcal{V}_1| \times |\mathcal{V}_2| \leq \min\{|\mathcal{X}_1| \times |\mathcal{X}_2|, |\mathcal{Y}_1|, |\mathcal{Y}_2|, |\mathcal{Y}_3|\}$. $Q \in \mathcal{Q}$ *is a time sharing variable which determines the portion of time we use a particular distribution* $p_1(x_1,x_2,y_1,y_2,y_3,v_0,v_1,v_2)$, $p_2(x_1,x_2,y_1,y_2,y_3,v_0,v_1,v_2)$ *and so on. Here,* $|\mathcal{Q}| \leq 3$.

Remark 15 *We note that by setting*

$$Q = 0 : V_0 = 0, V_1 = 0, V_2 = 0, X_2 = 0, \tag{5.10a}$$

$$Q = 1 : V_0 = 0, V_1 = 0, V_2 = 0, X_1 = 0, \tag{5.10b}$$

$$Q = 2 : X_1 = f(V_0,V_1,V_2), X_2 = f(V_0,V_1,V_2), \tag{5.10c}$$

for some deterministic function $f(\cdot)$*, we end up with the half-duplex coding scheme*

proposed by Murugan et al. (2004). At time $Q = 0$, node 1 transmits, and at time $Q = 1$, node 2 transmits. After both nodes fully decode the messages from each other, they coherently transmit at time $Q = 2$. However, in the coding scheme by Murugan et al., the destination only decodes at time $Q = 2$. In other words, the terms $I(\cdots; Y_3|Q = i, \cdots)$ for $i = 1, 2$ are excluded. In Theorem 19, the destination decodes at all $Q = 0, 1, 2$ and hence the achievable region can be larger.

Proof 19 (Outline of the proof of Theorem 19) *Now, we present an outline of the proof of Theorem 19. The complete proof can be found in Appendix C.2. We ignore Q in the following discussion to simplify the expressions.*

The codebook generation is as follows:

1. *Fix the p.d.f. $p(v_0), p(v_1), p(v_2), p(x_1|v_0, v_1, v_2)$, and $p(x_2|v_0, v_1, v_2)$.*
2. *Generate $2^{n[I(W_1;W_2)+\epsilon]}$ i.i.d. sequences $\mathbf{v_0}$ according to $\prod_{i=1}^{n} p(v_{0i})$. Index them $\mathbf{v}_0(i)$, $i \in \left\{1, 2, \ldots, 2^{n[I(W_1;W_2)+\epsilon]}\right\}$.*
3. *Generate $2^{n[H(W_1|W_2)+\epsilon]}$ i.i.d. sequences $\mathbf{v_1}$ according to $\prod_{i=1}^{n} p(v_{1i})$. Index them $\mathbf{v}_1(j)$, $j \in \left\{1, 2, \ldots, 2^{n[H(W_1|W_2)+\epsilon]}\right\}$.*
4. *Generate $2^{n[H(W_2|W_1)+\epsilon]}$ i.i.d. sequences $\mathbf{v_2}$ according to $\prod_{i=1}^{n} p(v_{2i})$. Index them $\mathbf{v}_2(k)$, $k \in \left\{1, 2, \ldots, 2^{n[H(W_2|W_1)+\epsilon]}\right\}$.*
5. *Define $h' = (i', j', k')$. For each $(\mathbf{v}_0(i'), \mathbf{v}_1(j'), \mathbf{v}_2(k'))$, generate $2^{n[H(W_1|W_2)+\epsilon]}$ sequences $\mathbf{x}_1$ according to $\prod_{i=1}^{n} p(x_{1i}|v_{0i}(i'), v_{1i}(j'), v_{2i}(k'))$. Index them $\mathbf{x}_1(j, h')$, $j \in \left\{1, 2, \ldots, 2^{n[H(W_1|W_2)+\epsilon]}\right\}$.*
6. *Again for each $(\mathbf{v}_0(i'), \mathbf{v}_1(j'), \mathbf{v}_2(k'))$, independently generate $2^{n[H(W_2|W_1)+\epsilon]}$ sequences $\mathbf{x}_2$ according to $\prod_{i=1}^{n} p(x_{2i}|v_{0i}(i'), v_{1i}(j'), v_{2i}(k'))$. Index them $\mathbf{x}_2(k, h')$, $k \in \left\{1, 2, \ldots, 2^{n[H(W_2|W_1)+\epsilon]}\right\}$.*

The encoding steps (refer to Fig. 5.2) are as follows:

1. *Slepian & Wolf (1973b, Theorem 2) showed that when node 1 only knows $\mathbf{w}_1$ and node 2 knows $\mathbf{w}_2$, node 1 can encode $\mathbf{w}_1$ using $n[H(W_1|W_2) + \epsilon]$ bits*

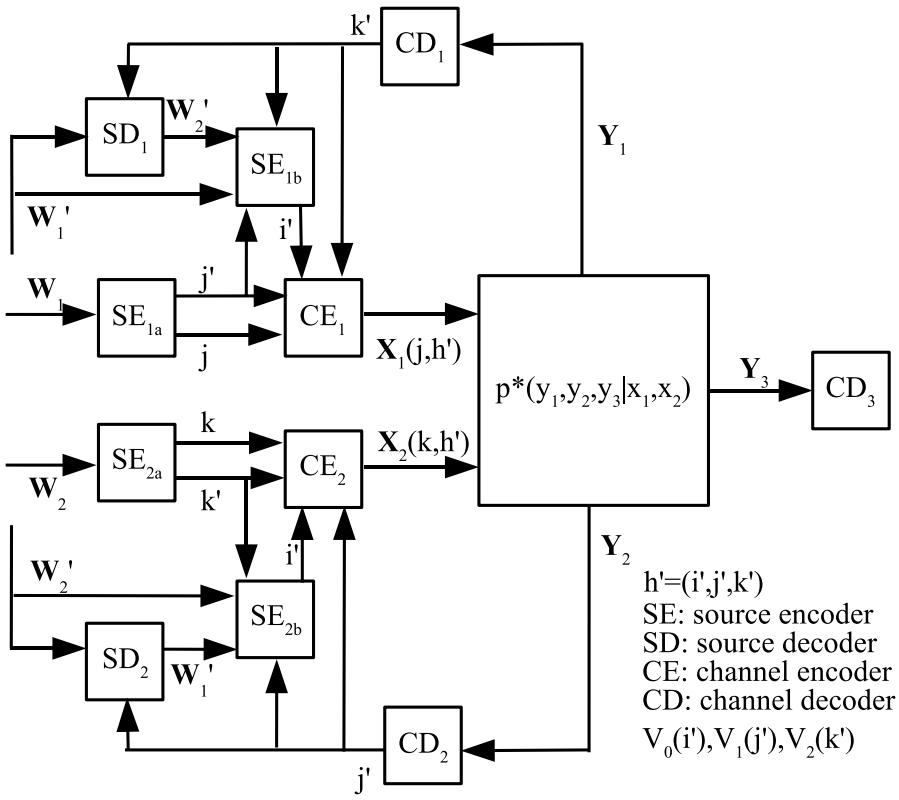


Figure 5.2: The encoding of FDS-DF.

(indexed by j) and it can be decoded by node 2. Similarly, node 2 can use $n[H(W_2|W_1)+\epsilon]$ bits (indexed by k) to encode $\mathbf{w}_2$. Node 1 transmits $\mathbf{x}_1(j,h')$, and node 2 transmits $\mathbf{x}_2(k,h')$, where h' is the cooperative information from the previous block. We use prime to indicate the index from the previous block.

2. *At the beginning of the new block, assume that node 1 correctly estimates k' sent by node 2. Using $\mathbf{w}_1'$, it can decode $\mathbf{w}_2'$. Node 2 does likewise to decode $\mathbf{w}_1'$.*

3. *Both sources now compress $(\mathbf{w}_1', \mathbf{w}_2')$ down to $n[H(W_1,W_2)+3\epsilon]$ bits and index it by $h' \in \{1,\ldots,2^{n[H(W_1,W_2)+3\epsilon]}\}$. Now, create $2^{n[H(W_1|W_2)+H(W_2|W_1)+2\epsilon]}$ bins and index each bin by a unique (j',k'). Assign h' to the bins so that each bin contains $2^{n[I(W_1;W_2)+\epsilon]}$ entries. Index the entries $i' \in \{1,\ldots,2^{n[I(W_1;W_2)+\epsilon]}\}$. Hence, each h' can be represented by a unique triplet (i',j',k').*

4. *In the new block, node 1 sends $\mathbf{x}_1(j,i',j',k')$ and node 2 sends $\mathbf{x}_2(k,i',j',k')$.*

The decoding steps are as follows:

1. *Upon observing the sequence* $\mathbf{y}_1$*, node 1 declares* $\hat{k}$ *has been sent by node 2 if there exists a unique* $\hat{k}$ *such that* $\left(\mathbf{x}_1(j,h'),\mathbf{v}_0(i'),\mathbf{v}_1(j'),\mathbf{v}_2(k'),\mathbf{x}_2(\hat{k},h'),\mathbf{y}_1\right)$ $\in \mathcal{A}_\epsilon$*. We use hat to indicate the estimate. Here,* $\mathcal{A}_\epsilon$ *is the set of jointly typical sequences (Cover & Thomas, 1991, pg. 195). We note that node 1 knows* $h' = (i', j', k')$*, which is the full information from the previous block, and its own information* j*. It can determine the correct* k *with diminishing error probability if*

$$H(W_2|W_1) < I(X_2; Y_1|V_0, V_1, V_2, X_1). \tag{5.11}$$

2. *Similarly, observing the sequence* $\mathbf{y}_2$*, node 2 declares* $\hat{j}$ *has been sent by node 1 if there exists a unique* $\hat{j}$ *such that* $\left(\mathbf{x}_1(\hat{j},h'),\mathbf{v}_0(i'),\mathbf{v}_1(j'),\mathbf{v}_2(k')\mathbf{x}_2(k,h'),\mathbf{y}_2\right)$ $\in \mathcal{A}_\epsilon$*. Node 2 can determine the correct* j *with diminishing error probability if*

$$H(W_1|W_2) < I(X_1; Y_2|V_0, V_1, V_2, X_2). \tag{5.12}$$

3. *Node 3 decodes* $(\hat{i},\hat{j},\hat{k})$ *over two blocks. In the first block, assuming that it has already correctly decoded* $h' = (i', j', k')$ *from the previous block, it finds a set of* $(\hat{j},\hat{k}) \in \mathcal{L}_1$ *where* $\left(\mathbf{x}_1(\hat{j},h'),\mathbf{x}_2(\hat{k},h'),\mathbf{v}_0(i'),\mathbf{v}_1(j'),\mathbf{v}_2(k'),\mathbf{y}_3\right) \in \mathcal{A}_\epsilon$*. In the second block, it then finds another set of* $(\hat{j},\hat{k}) \in \mathcal{L}_2$ *and a unique* $\hat{i}$ *where* $\left(\mathbf{v}_0(\hat{i}),\mathbf{v}_1(\hat{j}),\mathbf{v}_2(\hat{k}),\mathbf{y}_3\right) \in \mathcal{A}_\epsilon$*.* $\mathcal{L}_i, i = 1, 2$*, contains node 3's roughly estimates of* $(\hat{j},\hat{k})$ *from the* i*-th block of received signals. It declares* $(\hat{i},\hat{j},\hat{k})$ *has been sent if there is a unique* $\hat{i}$ *and a unique pair of* $(\hat{j},\hat{k})$ *in* $\mathcal{L}_1 \cap \mathcal{L}_2$*.*

This can be done with diminishing error probability if

$$I(W_1;W_2) < I(V_0;Y_3|V_1,V_2), \tag{5.13a}$$

$$H(W_1|W_2) < I(V_1;Y_3|V_0,V_2) + I(X_1;Y_3|V_0,V_1,V_2,X_2), \tag{5.13b}$$

$$H(W_2|W_1) < I(V_2;Y_3|V_0,V_1) + I(X_2;Y_3|V_0,V_1,V_2,X_1), \tag{5.13c}$$

$$H(W_1) < I(V_0,V_1;Y_3|V_2) + I(X_1;Y_3|V_0,V_1,V_2,X_2), \tag{5.13d}$$

$$H(W_2) < I(V_0,V_2;Y_3|V_1) + I(X_2;Y_3|V_0,V_1,V_2,X_1), \tag{5.13e}$$

$$H(W_1|W_2) + H(W_2|W_1) < I(V_1,V_2;Y_3|V_0) + I(X_1,X_2;Y_3|V_0,V_1,V_2), \tag{5.13f}$$

$$H(W_1,W_2) < I(X_1,X_2;Y_3). \tag{5.13g}$$

We consider all possible error combinations. Assuming that (i,j,k) *were sent,* (5.13a) *guarantees that the* $\Pr(\hat{i} \neq i, \hat{j} = j, \hat{k} = k) < \epsilon$ *for any* $\epsilon > 0$. (5.13b) *guarantees that* $\Pr(\hat{i} = i, \hat{j} \neq j, \hat{k} = k) < \epsilon$, (5.13c) *guarantees that* $\Pr(\hat{i} = i, \hat{j} = j, \hat{k} \neq k) < \epsilon$, (5.13d) *guarantees that* $\Pr(\hat{i} \neq i, \hat{j} \neq j, \hat{k} = k) < \epsilon$, (5.13e) *guarantees that* $\Pr(\hat{i} \neq i, \hat{j} = j, \hat{k} \neq k) < \epsilon$, (5.13f) *guarantees that* $\Pr(\hat{i} = i, \hat{j} \neq j, \hat{k} \neq k) < \epsilon$, *and* (5.13g) *guarantees that* $\Pr(\hat{i} \neq i, \hat{j} \neq j, \hat{k} \neq k) < \epsilon$.

4. *With* $(\hat{i},\hat{j},\hat{k})$, *node 3 can determine* $\hat{h}$ *and decode* $(\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2)$.

The total probability of error can be bounded, for large n, *if* (5.11), (5.12), *and* (5.13a)–(5.13g) *hold.*

For the cardinality of the auxiliary random variables, using the method by Salehi (1999), we can show that $|\mathcal{V}_0| \times |\mathcal{V}_1| \times |\mathcal{V}_2| \leq \min\{|\mathcal{X}_1| \times |\mathcal{X}_2|, |\mathcal{Y}_1|, |\mathcal{Y}_2|, |\mathcal{Y}_3|\}$. *Since the achievable region of FDS-DF can be plotted in the 3-dimensional space,* $|\mathcal{Q}| \leq 3$ *is sufficient.*

Hence, we have Theorem 19.

The probability error analysis can be found in Appendix C.2. The achievable region of FDS-DF on the Gaussian MACFCS can be found in Appendix C.3.

Remark 16 *In FDS-DF, the sources only need to exchange* $nH(W_1|W_2)$ $+nH(W_2|W_1)$ *bits in order for them to know the full information* $(\mathbf{w}_1, \mathbf{w}_2)$. *When both sources know the full information, they then cooperate (achieving coherent combining in the Gaussian channel) to send the full information to the destination.*

Under certain channel conditions, that all nodes fully decode the data of all other nodes might not be desirable. One example is when node 1 is far from the destination and node 2 is close to the destination. In this case, it is not necessary for node 1 to decode all of node 2's data. We note that if the sources only exchange partial information, they are not able to cooperate to send the full information to the destination. They can only cooperate to send the data that they exchange (in contrast with FDS-DF in which the sources can cooperate to send more information than what they exchange). Without full decoding at the sources, we study a few other types of coding strategies where full decoding of all messages (w_1, w_2) only occurs at the destination. We use the following method. First, source coding is performed at each individual source node to remove the correlation among the sources (see Section 5.6.2). At this point, we have turned the problem into that of channel coding for the MACF with independent sources. Then we apply a channel coding strategy for the MACF to transmit independent information to the destination.

5.6.2 Source Coding for Correlated Sources

Source coding for correlated sources is first performed at every source node. This removes correlation between the sources. This does not require physical communication among the sources. Each source node forms independent inputs to its channel encoder.

Recall that nodes 1 and 2 receive w_1 and w_2 from their respective sources. The data are correlated and drawn according to $p(w_1, w_2)$. First, we consider a noiseless channel. With node 1 knowing only w_1 and node 2 knowing only w_2, the destination can reconstruct (w_1, w_2) reliably if node 1 encodes w_1 with rate R_1 and

node 2 encodes w_2 with rate R_2 (Slepian & Wolf, 1973b), where

$$R_1 \geq H(W_1|W_2), \tag{5.14a}$$

$$R_2 \geq H(W_2|W_1), \tag{5.14b}$$

$$R_1 + R_2 \geq H(W_1, W_2). \tag{5.14c}$$

Figure 5.3 shows independent data (j, k) after source coding. After receiving n source messages, $\mathbf{w}_1$, encoder 1 encodes the data to $j \in \{1, 2, \ldots, 2^{nR_1}\}$. Encoder 2 receives $\mathbf{w}_2$ and encodes the data to $k \in \{1, 2, \ldots, 2^{nR_2}\}$. R_1 and R_2 are within the constraints (5.14a)–(5.14c).

Now, we consider an unreliable channel and explore how channel coding can help the destination to recover j and k. With these, it can recover $\mathbf{w}_1$ and $\mathbf{w}_2$.

5.6.3 Source Coding for Correlated Sources and Compress-Forward Channel Coding for the MACF (SC-CF)

In this section, we derive an achievable region for the MACF based on CF. Combining this with the source coding rate constraints in Section 5.6.2, we derive another achievable region for the MACFCS. We term this coding strategy source coding for correlated sources and CF channel coding for the MACF (SC-CF). To the best of our knowledge, CF has not been studied on the MACF where each node receives possibly different channel feedback. CF was first introduced by Cover & El Gamal (1979) for the single relay channel. It was subsequently extended to the MRC by Kramer *et al.* (2005) in which the strategy is termed the compress-and-forward strategy. King (1978) derived an achievable region for the MACF, with all sources receiving common feedback, using combined DF and CF coding strategies. In this thesis, we construct a CF for the MACF with possibly different feedback to every node. Here, we do not combine CF with DF as we want to compare the performance of different strategies. With the different strategies described in this thesis, we can easily pick and combine different strategies to get another achievable region.

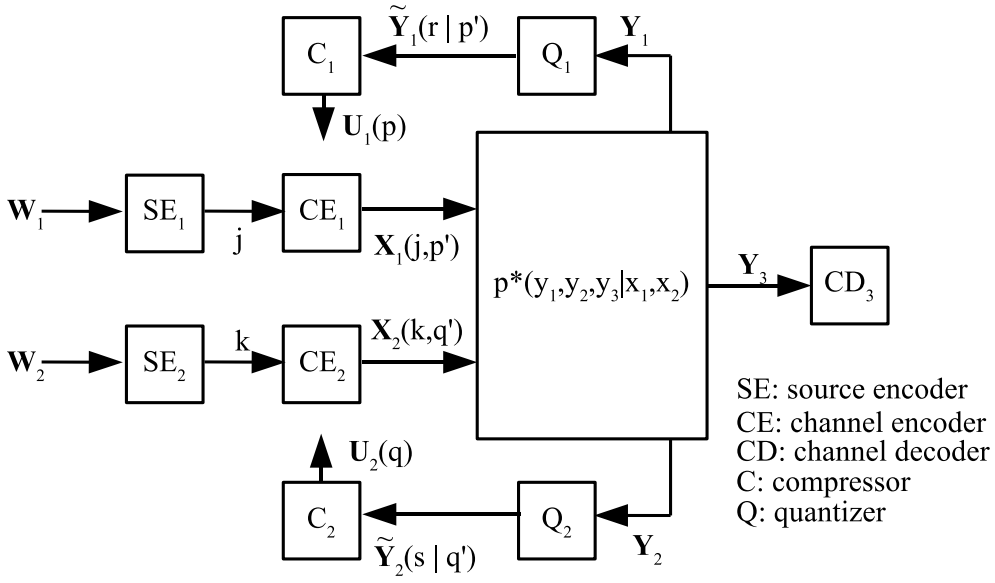


Figure 5.3: The encoding of SC-CF.

Remark 17 *It has been shown by Kramer et al. (2005) that in relay channels with different topologies, DF or CF can achieve higher rates. For the Gaussian relay channel, a rough guide is that when the relay is closer to the source, DF achieves higher rates; while when the relay is closer to the destination, CF achieves higher rates. This suggests that SC-CF might give larger achievable regions on the MACFCS compared to FDS-DF under different topologies.*

Using CF, each node transmits independent information as well as a quantized and binned version of its received signal. Referring to Figure 5.3, j and k are independent information after performing source coding on a block of n correlated source messages $(\mathbf{w}_1, \mathbf{w}_2)$. Consider node 1 as an example first. From the received signal $\mathbf{y}_1$, it produces a quantized version $\tilde{\mathbf{y}}_1$. It then bins $\tilde{\mathbf{y}}_1$ to $\mathbf{u}_1$. In the next block, it sends new information j as well as $\mathbf{u}_1$. We can view this as node 1 helping node 2 to send a noisy, quantized, and binned version of node 2's signal, k, without needing to fully decode k. Node 2 does likewise.

Using SC-CF, we show that the following region is achievable.

Theorem 20 (SC-CF) *Let* $\left(\mathcal{W}_1 \times \mathcal{W}_2, p(w_1, w_2), \mathcal{X}_1 \times \mathcal{X}_2, p^*(y_1, y_2, y_3|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3\right)$ *be a discrete memoryless three-node MACFCS. The source messages*

(W_1, W_2) can be reliably transmitted to the destination per channel use if

$$H(W_1|W_2) < I(X_1; \tilde{Y}_1, \tilde{Y}_2, Y_3|Q, U_1, U_2, X_2), \tag{5.15a}$$

$$H(W_2|W_1) < I(X_2; \tilde{Y}_1, \tilde{Y}_2, Y_3|Q, U_1, U_2, X_1), \tag{5.15b}$$

$$H(W_1, W_2) < I(X_1, X_2; \tilde{Y}_1, \tilde{Y}_2, Y_3|Q, U_1, U_2), \tag{5.15c}$$

where the mutual information is taken over all joint p.d.f.

$$\begin{aligned} p(q, u_1, u_2, x_1, x_2, \tilde{y}_1, \tilde{y}_2, y_1, y_2, y_3) = {} & p(q)p(u_1|q)p(x_1|q, u_1)p(u_2|q)p(x_2|q, u_2) \\ & \cdot p(\tilde{y}_1|q, y_1, x_1, u_1)p(\tilde{y}_2|q, y_2, x_2, u_2) \\ & \cdot p^*(y_1, y_2, y_3|x_1, x_2), \end{aligned} \tag{5.16a}$$

subject to the following constraints

$$I(U_1; Y_3|Q, U_2) > I(\tilde{Y}_1; Y_1|Q, X_1, U_1) - I(\tilde{Y}_1; Y_3|Q, \tilde{Y}_2, U_1, U_2), \tag{5.17a}$$

$$I(U_2; Y_3|Q, U_1) > I(\tilde{Y}_2; Y_2|Q, X_2, U_2) - I(\tilde{Y}_2; Y_3|Q, \tilde{Y}_1, U_1, U_2), \tag{5.17b}$$

$$I(U_1, U_2; Y_3|Q) > I(\tilde{Y}_1; Y_1|Q, X_1, U_1) + I(\tilde{Y}_2; Y_2|Q, X_2, U_2) - I(\tilde{Y}_1, \tilde{Y}_2; Y_3|Q, U_1, U_2). \tag{5.17c}$$

Here, $U_1 \in \mathcal{U}_1, U_2 \in \mathcal{U}_2, \tilde{Y}_1 \in \tilde{\mathcal{Y}}_1$, and $\tilde{Y}_2 \in \tilde{\mathcal{Y}}_2$ are auxiliary random variables. $|\mathcal{U}_1| \times |\mathcal{U}_2| \leq \min\{|\mathcal{X}_1| \times |\mathcal{X}_2|, |\mathcal{Y}_3|\}$, $|\tilde{\mathcal{Y}}_1|$ and $|\tilde{\mathcal{Y}}_2|$ are finite. $Q \in \mathcal{Q}$ is the time sharing variable, and $|\mathcal{Q}| \leq 3$.

Proof 20 (Outline of the proof of Theorem 20) *Now, we give a brief outline of the proof of Theorem 20. The error probability analysis can be found in Appendix C.4. We ignore Q in the following discussion to simplify the expressions.*

Figure 5.3 shows independent data (j, k) after source coding. Channel encoder 1 receives $j \in \{1, 2, \ldots, 2^{nR_1}\}$ and channel encoder 2 receives $k \in \{1, 2, \ldots, 2^{nR_2}\}$ for every n source messages. Now, we study a channel coding scheme to ensure that the independent data after source coding can be reliably transmitted to the destination. The codebook generation is as follows.

1. *Fix* $p(u_1), p(x_1|u_1), p(u_2), p(x_2|u_2), p(\tilde{y}_1|y_1, x_1, u_1)$ *and* $p(\tilde{y}_2|y_2, x_2, u_2)$.

2. *Generate* $2^{nR_1'}$ *i.i.d. sequences* $\mathbf{u}_1$ *according to* $\prod_{i=1}^n p(u_{1i})$. *Index them* $\mathbf{u}_1(p)$, $p \in \left\{1, \ldots, 2^{nR_1'}\right\}$. *Generate* $2^{nR_2'}$ *i.i.d. sequences* $\mathbf{u}_2$ *according to* $\prod_{i=1}^n p(u_{2i})$. *Index them* $\mathbf{u}_2(q)$, $q \in \left\{1, \ldots, 2^{nR_2'}\right\}$.

3. *For each* $\mathbf{u}_1(p')$, *generate* 2^{nR_1} *sequences* $\mathbf{x}_1$ *according to* $\prod_{i=1}^n p(x_{1i}|u_{1i}(p'))$. *Index them* $\mathbf{x}_1(j, p')$, $j \in \left\{1, \ldots, 2^{nR_1}\right\}$. *For each* $\mathbf{u}_2(q')$, *generate* 2^{nR_2} *sequences* $\mathbf{x}_2$ *according to* $\prod_{i=1}^n p(x_{2i}|u_{2i}(q'))$. *Index them* $\mathbf{x}_2(k, q')$, $k \in \left\{1, \ldots, 2^{nR_2}\right\}$.

4. *For each* $\mathbf{u}_1(p')$, *generate* $2^{n\tilde{R}_1}$ *sequences* $\tilde{\mathbf{y}}_1$ *according to* $\prod_{i=1}^n p(\tilde{y}_{1i}|u_{1i}(p'))$. *We define*

$$p(\tilde{y}_1|u_1) = \frac{\sum_{u_2,x_1,x_2,\tilde{y}_2,y_1,y_2,y_3} p(u_1, u_2, x_1, x_2, \tilde{y}_1, \tilde{y}_2, y_1, y_2, y_3)}{\sum_{u_2,x_1,x_2,\tilde{y}_1,\tilde{y}_2,y_1,y_2,y_3} p(u_1, u_2, x_1, x_2, \tilde{y}_1, \tilde{y}_2, y_1, y_2, y_3)}, \tag{5.18a}$$

where $p(u_1, u_2, x_1, x_2, \tilde{y}_1, \tilde{y}_2, y_1, y_2, y_3)$ *is defined in* (5.16). *Index them* $\tilde{\mathbf{y}}_1(r|p')$, $v \in \left\{1, \ldots, 2^{n\tilde{R}_1}\right\}$.

5. *Similarly, for each* $\mathbf{u}_2(q')$, *generate* $2^{n\tilde{R}_2}$ *sequences* $\tilde{\mathbf{y}}_2$ *according to* $\prod_{i=1}^n p(\tilde{y}_{2i}|u_{2i}(q'))$. *We define*

$$p(\tilde{y}_2|u_2) = \frac{\sum_{u_1,x_1,x_2,\tilde{y}_1,y_1,y_2,y_3} p(u_1, u_2, x_1, x_2, \tilde{y}_1, \tilde{y}_2, y_1, y_2, y_3)}{\sum_{u_1,x_1,x_2,\tilde{y}_1,\tilde{y}_2,y_1,y_2,y_3} p(u_1, u_2, x_1, x_2, \tilde{y}_1, \tilde{y}_2, y_1, y_2, y_3)}. \tag{5.19a}$$

Index them $\tilde{\mathbf{y}}_2(s|q')$, $w \in \left\{1, \ldots, 2^{n\tilde{R}_2}\right\}$.

6. *Randomly partition the set* $\{1, 2, \ldots, 2^{n\tilde{R}_1}\}$ *into* $2^{nR_1'}$ *cells* S_p, $p \in \{1, \ldots, 2^{nR_1'}\}$; *and partition the set* $\{1, \ldots, 2^{n\tilde{R}_2}\}$ *into* $2^{nR_2'}$ *cells* S_q, $q \in \{1, \ldots, 2^{nR_2'}\}$.

The encoding steps are as follows. Basically, node 1 quantizes its received signal from the previous block and bins it. It sends the binned information together with new information from the source in the new block. Node 2 does likewise.

1. *In the beginning of block* $t+1$, *remembering its previous transmission in block* t, $\mathbf{x}_1(j^t, q^{t-1})$ *and* $\mathbf{u}_1(q^{t-1})$, *and observing its received signal in block* t, $\mathbf{y}_1(t)$,

it finds a unique r^t *for which* $(\mathbf{x}_1(j^t, p^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{y}_1(t), \tilde{\mathbf{y}}_1(r^t|p^{t-1})) \in \mathcal{A}_\epsilon$. *Berger (1977, Lemma 2.1.3) showed that node 1 can find such a* r^{t-1} *with probability tending to 1, with a large enough* n, *if*

$$\tilde{R}_1 > I(\tilde{Y}_1; Y_1|X_1, U_1). \tag{5.20}$$

Here, r^t *is the quantized version of* $\mathbf{y}_1(t)$.

2. *Now, node 1 bins* r^t *to* p^t. *It finds* p^t *for which* $r^t \in S_{p^t}$. *It then sends* $\mathbf{x}_1(j^{t+1}, p^t)$ *in block* $t+1$, *where* j^{t+1} *is the new message from the source. Here,* p^t *is to be decoded and used by the destination to estimate* r^t. *We see here that node 1 helps node 2 to send a noisy, quantized, and binned version of node 2's signal to the destination.*

3. *In block* $t+1$, *node 2 quantizes* $\mathbf{y}_2(t)$ *to* s^t. *It can find a unique* s^t *with probability tending to 1 if*

$$\tilde{R}_2 > I(\tilde{Y}_2; Y_2|X_2, U_2). \tag{5.21}$$

It bins s^t *to* q^t, *where* $s^t \in S_{q^t}$. *It then sends* $\mathbf{x}_2(k^{t+1}, q^t)$ *in block* t, *where* k^{t+1} *is the new information.*

The decoding steps are as follows. The destination first decodes the quantized and binned information from nodes 1 and 2. It then estimates the quantized information. Using its received signal and the estimated quantized information, it decodes the messages from nodes 1 and 2.

1. *At the end of block* $t+1$, *the destination receives* $\mathbf{y}_3(t+1)$. *It declares* $(\hat{p}^t, \hat{q}^t)$ *were sent by nodes 1 and 2 if it can find a unique pair of* $(\hat{p}^t, \hat{q}^t)$ *for which* $(\mathbf{u}_1(\hat{p}^t), \mathbf{u}_2(\hat{q}^t), \mathbf{y}_3(t+1)) \in \mathcal{A}_\epsilon$. *This can be done with an arbitrarily small*

error probability if the following inequalities hold.

$$R_1' < I(U_1; Y_3|U_2), \tag{5.22a}$$

$$R_2' < I(U_2; Y_3|U_1), \tag{5.22b}$$

$$R_1' + R_2' < I(U_1, U_2; Y_3). \tag{5.22c}$$

2. *At the end of block $t+1$, assume that the destination has correctly decoded (p^{t-1}, q^{t-1}) and (p^t, q^t). It uses its received signal in block t to find a set $\mathcal{L}(t)$ of (r^t, s^t) such that $\left(\tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \mathbf{y}_3(t)\right) \in \mathcal{A}_\epsilon$. It declares that $(\hat{r}^t, \hat{s}^t)$ were sent if it can find a unique $(\hat{r}^t, \hat{s}^t) \in \{(\hat{r}^t, \hat{s}^t) : \hat{r}^t \in S_{p^t}$ and $\hat{s}^t \in S_{q^t}\} \cap \mathcal{L}(t)$. This can be done reliably if*

$$\tilde{R}_1 < I(\tilde{Y}_1; Y_3|\tilde{Y}_2, U_1, U_2) + R_1', \tag{5.23a}$$

$$\tilde{R}_2 < I(\tilde{Y}_2; Y_3|\tilde{Y}_1, U_1, U_2) + R_2', \tag{5.23b}$$

$$\tilde{R}_1 + \tilde{R}_2 < I(\tilde{Y}_1, \tilde{Y}_2; Y_3|U_1, U_2) + R_1' + R_2'. \tag{5.23c}$$

3. *At the end of block $t+1$, assume that the destination has correctly decoded (r^t, s^t) and (p^{t-1}, q^{t-1}). It uses $\tilde{\mathbf{y}}_1(r^t|p^{t-1})$, $\tilde{\mathbf{y}}_2(s^t|q^{t-1})$, and $\mathbf{y}_3(t)$. It declares $(\hat{j}^t, \hat{k}^t)$ were sent if $\Big(\mathbf{x}_1(\hat{j}^t, p^{t-1}), \mathbf{x}_2(\hat{k}^t, q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}),$ $\tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(\hat{s}^t|q^{t-1}), \mathbf{y}_3(t)\Big) \in \mathcal{A}_\epsilon$. This can be done with diminishing error probability if*

$$R_1 < I(X_1; \tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2, X_2), \tag{5.24a}$$

$$R_2 < I(X_2; \tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2, X_1), \tag{5.24b}$$

$$R_1 + R_2 < I(X_1, X_2; \tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2). \tag{5.24c}$$

We see that node 3 decodes (j^t, k^t) at the end of block $t+1$.

Combining these rate constraints for the MACF using CF and the constraints for the source coding, (5.14a)-(5.14c), we get Theorem 20.

The probability error analysis can be found in Appendix C.4. The achievable region of the Gaussian MACFCS using SC-CF can be found in Appendix C.5.

5.6.4 Source Coding for Correlated Sources and the MAC Channel Coding (SC-MAC)

Now, we consider a coding strategy for the MACF that ignores the feedback from the channel to the source nodes. Each source now simply sends independent messages as it would in the MAC. We call this strategy SC-MAC, and we will see later that it actually does well in certain network topologies. A coding strategy that achieves the capacity of the MAC was found by Liao (1972) and Ahlswede (1974). Combining source coding for correlated sources and this channel coding for the MAC, we have the following theorem.

Theorem 21 (SC-MAC) *Let* $\left(\mathcal{W}_1\times\mathcal{W}_2, p(w_1,w_2), \mathcal{X}_1\times\mathcal{X}_2, p^*(y_1,y_2,y_3|x_1,x_2), \mathcal{Y}_1 \times\mathcal{Y}_2 \times \mathcal{Y}_3\right)$ *be a discrete memoryless three-node MACFCS. The source messages* (W_1, W_2) *can be reliably transmitted to the destination per channel use if the following inequalities hold.*

$$H(W_1|W_2) \leq I(X_1; Y_3|X_2), \tag{5.25a}$$

$$H(W_2|W_1) \leq I(X_2; Y_3|X_1), \tag{5.25b}$$

$$H(W_1, W_2) \leq I(X_1, X_2; Y_3), \tag{5.25c}$$

where $p(x_1, x_2) = p(x_1)p(x_2)$.

5.6.5 Combination of Other Strategies

There is a multitude of ways which we can combine a coding strategy for the MACCS with that for the MACF to arrive at a coding strategy for the MACFCS. The aim of this thesis is not to list all of them. In this section, we briefly mention a few combinations.

1. Combining source coding for correlated sources and the *partial DF* channel coding for the MACF by Carleial (1982, Theorem 1): After source coding, each source node has independent data. Each source now exchanges part of their data with other source nodes. They then cooperate to send the exchanged data to the destination. We call that partial DF as every source only decodes part of the data of other sources. An achievable region for the MACFCS can be derived by combining the source coding constraints for correlated sources (constraint (5.14a)–(5.14c) in Section 5.6.2) and the channel coding constraints of the partial DF for the MACF (Carleial, 1982, constraints (3a), (3b), (7a)–(7q)).

2. Combining source coding for correlated sources and the partial DF channel coding for the MACF by Willems (1982, Theorem 7.1): Similar to that by Carleial (1982), the sources exchange part of their data through the channel feedback link. They then cooperate to send the exchanged data to the destination. An achievable region for the MACFCS can be derived by combining the source coding constraints for correlated sources (constraint (5.14a)-(5.14c) in Section 5.6.2) and the channel coding constraints of the partial DF for the MACF (Willems, 1982, Theorem 7.1).

3. Combining coding strategy for MACCS without common part by Cover *et al.* (1980, constraints (3)) and CF for the MACF (that we derived in Section 5.6.3): Each node sends information encoded directly from the source (so that correlation is preserved among the transmitted signals) as well as the received (via the feedback links), quantized, and binned signals from other nodes.

4. Combining coding strategy MACCS without common part by Cover *et al.* (1980, constraints (3)) and the partial DF by Carleial (1982, Theorem 1) or Willems (1982, Theorem 7.1): Each node sends information encoded directly from the source (so that correlation is preserved among the transmitted sig-

nals). At the same time, the source nodes partially decode the data from other nodes, and cooperate to send the exchanged data to the destination.

Remark 18 *The strategy mentioned in (i) above (in Section 5.6.5) is different from FDS-DF (in Theorem 19). In the former, the channel encoders at nodes 1 and 2 receive independent data stream, after performing source coding for correlated sources. Then, Carleial's technique for MACF is applied directly. Hence, if we want the nodes to cooperatively send the full information,* (W_1^n, W_2^n)*, they must exchange at least* $nH(W_1, W_2)$ *bits. In FDS-DF, however, the channel encoders receives correlated data from the sources (so, we do not apply Carleial's technique directly here), and so they only need to exchange* $nH(W_1|W_2) + nH(W_2|W_1)$ *bits to be able to cooperatively send the full information.*

Remark 19 *We note that FDS-DF, SC-CF, SC-MAC, and strategies (i) and (ii) in Section 5.6.5 are based on separate source and channel coding. Strategies (iii) and (iv) in Section 5.6.5 are based on combined source and channel coding. Evaluating the performance of the combined source and channel coding strategies in the Gaussian channel is difficult as it involves discrete and continuous variables.*

Remark 20 *The achievable regions for FDS-DF, SC-CF, and SC-MAC are derived assuming that the number of source symbols received per unit time equals the number of channel transmissions per unit time. However, using these separate source and channel coding strategies, we can easily match the source symbol rate to the channel usage rate, without re-deriving the coding strategies. Considering a general case when we wish to send* k *pairs of source symbols using* n *channel transmissions, the achievable regions can be found by simply replacing the mutual information expressions by* $kH(\cdot) \leq nI(\cdot)$*. In this way, the achievability question for a particular MACFCS is no longer just "*whether we can reliably transmit a pair of (W_1, W_2) per channel use*", but more generally, "*at what rate, k/n, we can reliably transmit k pairs of (W_1, W_2) per n channel uses*". However, using combined source and channel coding strategies, we need to modify the coding strategies*

such that the probability distributions involve k source symbols and n channel input symbols (e.g., see Cover et al. (1980, equation (87)). Doing so, the achievable region will no longer be a single letter characterization.

In Sections 5.6.1 to 5.6.5, we investigated coding strategies for the MACFCS where the nodes exploit the broadcast/multiple-access nature of the channel. They cooperate in the sense that either the transmission from a node is decoded/processed by more than one node (broadcast nature) or a node decodes/processes the transmissions from more than one node (multiple-access nature). In the following section, we study a strategy in which the network is abstracted to a set of point-to-points links. A collection of links forms a route, and data are then passed down the route from the source to the destination using point-to-point coding.

5.6.6 Multi-Hop Coding with Data Aggregation (MH-DA)

In the multi-hop coding with data aggregation strategy (MH-DA), data are passed from a node to another, until they reach the destination. First, we number the nodes in a sequence, which we call a route. The last node in the route is the destination. We consider a *combine-forward* multi-hop coding where each node decodes the data from the previous node in the route, combines that with its own data, and forwards the aggregated data (those that it decodes from the previous node, plus its own data, less the correlated part with the data at the next node) to the next node in the route. In the three-node MACFCS, assuming that node 1 receives $\mathbf{w}_1$ from its source and node 2 receives $\mathbf{w}_2$ from its source, they do the following:

1. Node 1 compresses $\mathbf{w}_1$ down to $nH(W_1|W_2)$ bits, indexes it by $j \in \{1, \ldots, 2^{nH(W_1|W_2)}\}$, and sends it to node 2.

2. We know that upon receiving j, node 2 can decode $\mathbf{w}_1$.

3. Node 2 compresses $(\mathbf{w}_1, \mathbf{w}_2)$ to $k \in \{1, \ldots, 2^{nH(W_1,W_2)}\}$, using $nH(W_1, W_2)$ bits, and sends it to the destination.

In this multi-hop coding scheme, a node only decodes from the node behind in the route. The achievable region of the MACFCS using MH-DA is given in the following theorem.

Theorem 22 (MH-DA) *Let* $\left(\mathcal{W}_1\times\mathcal{W}_2, p(w_1,w_2), \mathcal{X}_1\times\mathcal{X}_2, p^*(y_1,y_2,y_3|x_1,x_2), \mathcal{Y}_1\times\mathcal{Y}_2\times\mathcal{Y}_3\right)$ *be a discrete memoryless three-node MACFCS.* (W_1,W_2) *can be reliably transmitted to the destination per channel use if the following holds.*

$$H(W_1|W_2) < I(X_1;Y_2|X_2,Q), \tag{5.26a}$$

$$H(W_1,W_2) < I(X_2;Y_3|Q), \tag{5.26b}$$

where

$$p(q,x_1,x_2,y_1,y_2,y_3) = p(q)p(x_2|q)p(x_1|x_2,q)p^*(y_1,y_2,y_3|x_1,x_2). \tag{5.27}$$

$Q\in\mathcal{Q}$ *is the time sharing variable and* $|\mathcal{Q}|\leq 2$.

The proof of Theorem 22 is straightforward and is omitted.

Now we consider a time division MH-DA for the three-node Gaussian MACFCS. By time division, we mean that only one source transmits at a time, i.e., for fraction $(1-f)$ of the time $(0<f<1)$, node 1 transmits and node 2 does not transmit; for fraction f of the time, node 2 transmits and node 1 does not transmit. This might be done to reduce interference among the nodes.

The achievable region of the Gaussian MACFCS using time division MH-DA is

$$H(W_1|W_2) < \frac{1-f}{2}\log\left(1+\frac{\lambda_{12}P_1}{(1-f)N_2}\right), \tag{5.28a}$$

$$H(W_1,W_2) < \frac{f}{2}\log\left(1+\frac{\lambda_{23}P_2}{fN_3}\right). \tag{5.28b}$$

We have presented four achievable regions for the MACFCS using different coding strategies, and suggested a few coding strategies for the MACFCS by combining

coding strategies for the MACCS and the MACF. In the next section, we compare the four achievable regions to the CS-OB.

5.7 Comparison of Coding Strategies

In this section, we plot and compare achievable regions for the different strategies in the three-node static Gaussian MACFCS. We consider *symmetrical topologies*, i.e., both sources are of equi-distant from the destination, and also linear asymmetrical topologies where the three nodes form a straight line, with node 2 placed in between node 1 and the destination. Although Gaussian input distributions may not be optimal, we choose X_1, X_2, and the auxiliary random variables to be Gaussian for the sake of comparison. We use the following parameters: $\kappa = 1, \eta = 2$, and $\nu_{ij} = 1$.

5.7.1 Design Methodology

We perform numerical calculations to compare the achievable regions of different coding strategies and the CS-OB to gain insights into how node position and data correlation affect performance. First, we study the effect of node position. For this analysis, we assume *symmetrical source data*, meaning $H(W_1|W_2) = H(W_2|W_1)$. This is a reasonable assumption for sensor networks when homogeneous sensors are deployed, and each sensor is sensing the environment at the same rate. For the computations, we fix $H(W_1|W_2) = 0.5, H(W_2|W_1) = 0.5$, and $I(W_1; W_2) = 0.5$. Although there are many combinations of node positions that one can study, we group them into three main categories:

1. Symmetrical topology with the sources closer to the destination than they are to one another.

2. Symmetrical topology with the sources closer to one another than they are to the destination.

3. Asymmetrical topology. Without loss of generality, we assume that node 1 is further away from the destination than node 2 is from the destination. For simplicity, we study linear topologies where the three nodes form a straight line.

Taking a closer look at MH-DA, we note that this strategy is more suitable for asymmetrical topologies. This is because in symmetrical topologies, there is no reason why we would arrange the nodes in a route and "load" the node at the end of route. This strategy makes sense in the asymmetrical topology where some nodes are nearer to the destination. Hence we analyze the performance of MH-DA only in the asymmetrical topology.

After investigating the effect of node position, we study the effect of varying the correlation between the sources on the performance of the various coding strategies. As rationalized above, we still keep the source data symmetrical, i.e. $H(W_1|W_2) = H(W_2|W_1)$. We vary $I(W_1; W_2)$ while keeping one of the following constant:

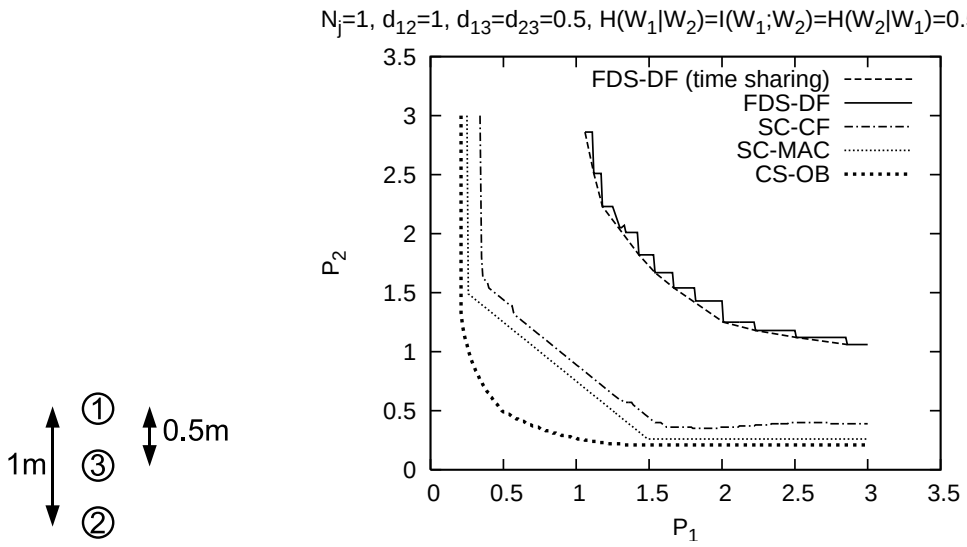
1. The information of each source, $H(W_1)$ and $H(W_2)$, is constant.

2. The total information $H(W_1, W_2)$ is constant.

5.7.2 The Effect of Node Position

Figs. 5.4 and 5.5 show the minimum average transmit powers (energy per channel use) required for nodes 1 and 2 to reliably transmit a pair of (W_1, W_2) to the destination per channel use. The achievable region is the region above the line. Note that we plot average transmit powers on both axes. So, if the nodes transmit with an average power pair in the achievable region, the nodes can reliably send (W_1, W_1) to the destination per channel use. We denote the average power of nodes 1 and 2 by P_1 and P_2 respectively.

We consider symmetrical source data with the following values: $H(W_1|W_2) = 0.5, H(W_2|W_1) = 0.5$, and $I(W_1; W_2) = 0.5$. First, we compare the two symmetrical topologies: (1) when the sources are further away from each other than they



(a) Network topology

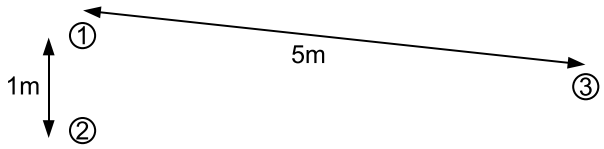
(b) P_1 versus P_2.

Figure 5.4: Minimum power required to transmit (W_1, W_2) to the destination per channel use, with weak inter-source link.

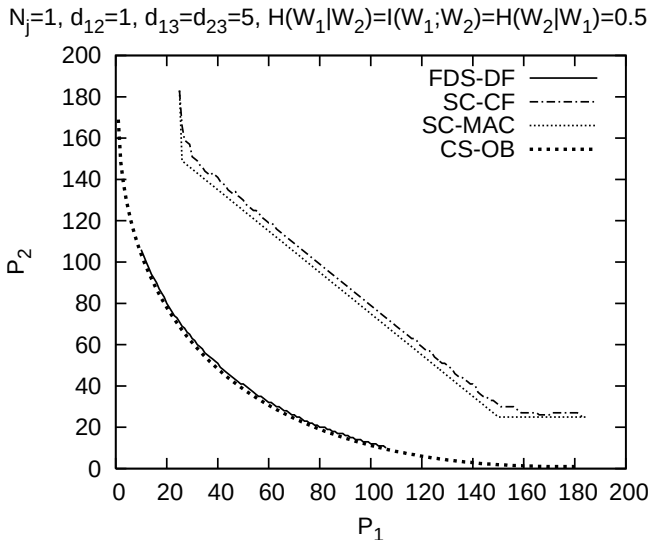
are from the destination, and (2) when the sources are further away from the destination than they are from each other. The first setup studies the case where the source-destination links are better than the inter-source link while the second setup studies the case where the inter-source link is better than the source-destination links.

When the inter-source link is weak, Fig. 5.4 shows that SC-CF and SC-MAC perform better than FDS-DF, i.e., the achievable regions for the SC-CF and SC-MAC contain that of FDS-DF. FDS-DF performs worst among the three strategies as the strategy requires each source node to get all the data from other nodes. This imposes an extra constraint on the average transmit power of the source nodes. When the source-destination link is stronger, a better strategy is to send the signals directly to the destination than to seek help from other sources.

On the other hand, when the inter-source link is strong, Fig. 5.5 shows that FDS-DF performs better than SC-CF and SC-MAC. The transmission bottleneck is now at the source-destination link. A good inter-source link lets each source node fully decode the messages from other nodes using little transmit power. In FDS-DF, the sources then use most of the transmit power to send the full information



(a) Network topology



(b) P_1 versus P_2.

Figure 5.5: Minimum power required to transmit (W_1, W_2) to the destination per channel use, with weak source-destination links.

coherently to the destination. Coherent combining makes a significant gain in transmission rate on the source-destination link. Also, we see that the achievable region of FDS-DF comes very close to the CS-OB when the inter-source link is much better than the source-destination link.

Remark 21 *In these two scenarios, we consider symmetrical topologies* $(d_{13} = d_{23})$ *and symmetrical source data* $(H(W_1|W_2) = H(W_2|W_1))$. *We term the channel with symmetrical topology and symmetrical source data,* symmetrical MACFCS. *In the symmetrical MACFCS, using FDS-DF, the total average transmit power is minimized when the nodes transmit at the same average power. In other words, it is more efficient for the nodes to share the load in transmitting data than for one to transmit at higher power. We can see this from the non-linearity in the coherent combining term in* (C.40), *or from Figs. 5.4 and 5.5 that the power curves are convex. However, for SC-CF and SC-MAC, there is a range for which individual source nodes can vary their transmit power while maintaining the minimum total*

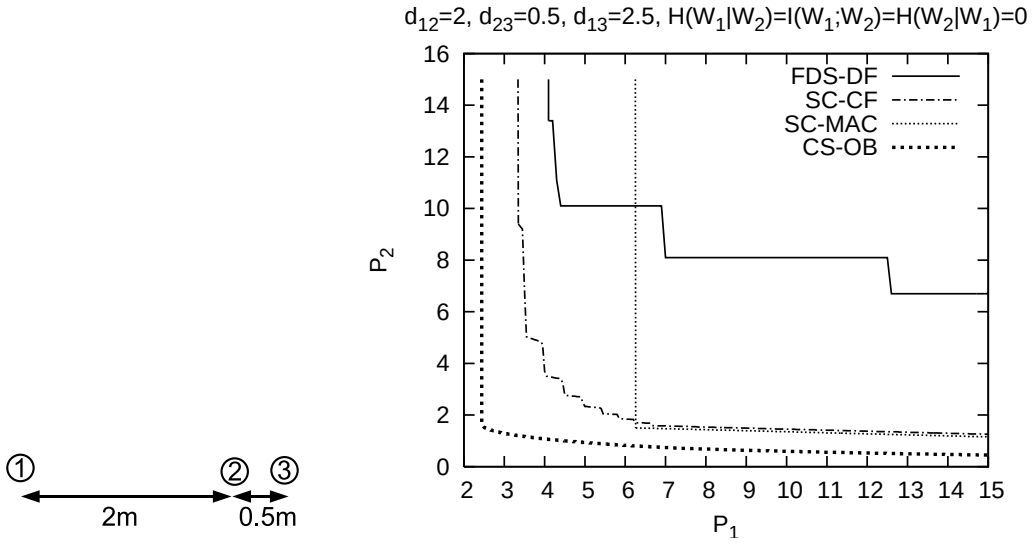
average transmit power. We can see this from the mutual information expression in (C.58), (C.59), *and* (C.63a) *that the relationship between* P_1 *and* P_2 *is linear, keeping* $P_1 + P_2$ *constant, or from Figs. 5.4 and 5.5 that there are portions of the SC-CF and SC-MAC curves where the slope is -1.*

Remark 22 *The staircase behavior of the FDS-DF curve in Fig. 5.4 is caused by the optimization involving different inequalities in Theorem 19 and the finite step size of* $\{\alpha_{ij}\}$ *in* (C.34a) *and* (C.34b). $\{\alpha_{ij}\}$ *are the power splits of node* i *used to carry different messages. The definition can be found in Appendix C.3. With time sharing, reliable transmission can be achieved using an average transmit power above the dotted line. Hence time sharing enlarges the achievable region of FDS-DF. This explains why the time sharing random variable* Q *is included in Theorem 19.*

Remark 23 *In the symmetrical MACFCS, SC-MAC performs better than SC-CF. This means after we remove the correlation among the sources, using the feedback of the channel via CF is worse than not using the feedback at all. This can be explained as follows. When the nodes are of same distance from the destination and have same amount of information to send, it is better for each of them to send their own message to the destination directly. It does not help when they try to help other nodes by sending a noisy, compressed, and binned version of what they received. The power can be better used to send their own uncorrupted data.*

That SC-MAC always outperforms SC-CF is no longer true in the asymmetrical topology. Fig. 5.6 shows the minimum power curves without time sharing. For illustration, we choose $d_{12} = 2, d_{23} = 0.5$, and $d_{13} = 2.5$. We note that choosing node 2 close to node 1 resembles a symmetrical topology. From the graph, we see that using SC-MAC, the minimum power required at node 1 is 6.25W. Using FDS-DF, the minimum power required at node 1 is 4.1W. We can further reduce the power at node 1 to 3.35W by increasing the power at node 2 by using SC-CF.

Using SC-MAC, we ignore the feedback in the channel. Hence, node 1 needs to transmit at least $H(W_1|W_2)$ bits to the destination, which is situated 2.5m away.



(a) Network topology

(b) P_1 versus P_2.

Figure 5.6: Minimum power required to transmit (W_1, W_2) to the destination per channel use, in a linear topology.

Note that without using feedback, there is no way node 2 can help in sending this portion of the message. However, we can reduce node 1's transmit power by using FDS-DF. Now, node 1 needs to send at least $H(W_1|W_2)$ bits to node 2, which is 2m away. Node 2, which is nearer to the destination can help node 1 to relay the message to the destination. If we wish to further reduce the transmission power of node 1, we use SC-CF. Using this strategy, node 2 does not need to fully decode W_1. Node 2 acts as an additional (but noisy) antenna for the destination. Hence this further enhances the "reception" of node 1's message. So node 1 needs to send at least $H(W_1|W_2)$ bits to the destination, equipped with an additional (noisy) antenna at node 2.

Remark 24 *From Fig. 5.6, we see the staircase behavior of the SC-CF curve. It shows that time sharing increases the achievability region of SC-CF. This accounts for the use of the time sharing auxiliary random variable Q in Theorem 20.*

5.7.3 The Effect of Source Correlation

Now, we study how correlation among the source data affects the different coding strategies for the MACFCS. We consider symmetrical topologies. Figs. 5.7 and

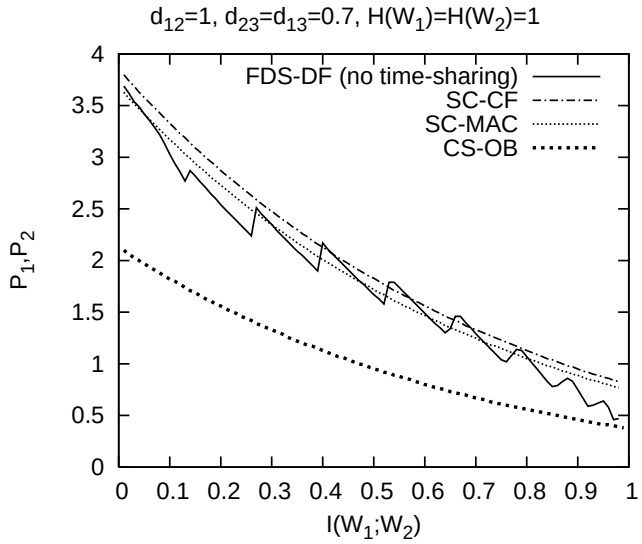


Figure 5.7: Minimum power required to transmit (W_1, W_2) to the destination per channel use, with different message correlation but constant $H(W_1)$ and $H(W_2)$.

5.8 depict achievable regions of the different coding strategies. We plot the equal power point for different correlation values.

From both graphs, we see that all three strategies perform either better or do not change when the data are more correlated. This make sense since if each node knows a larger portion of other nodes' data, it is easier for the nodes to cooperate.

When the nodes transmit at equal power, the achievable regions of SC-CF and SC-MAC do not vary with the correlation as long as the total information $H(W_1, W_2)$ remains constant and the correlation is symmetrical (Fig. 5.8). This is because using these two strategies, source coding is first performed. After that, the nodes send independent data to the destination. We know that the minimum total rate ($R_1 + R_2$ in (5.14c)) for which the nodes must transmit remains constant if $H(W_1, W_2)$ is constant.

In the same graph, although the total information $H(W_1, W_2)$ stays constant, increasing the correlation of the data enlarges the achievable region of FDS-DF. The reason is that when the correlation is higher, more power can be used for coherent transmission. The nodes need less power for inter-source communication.

When the sources are fully correlated, i.e., $H(W_1|W_2)$ and $H(W_2|W_1)$ approach zero, the achievable region of FDS-DF approaches the CS-OB. This does not come

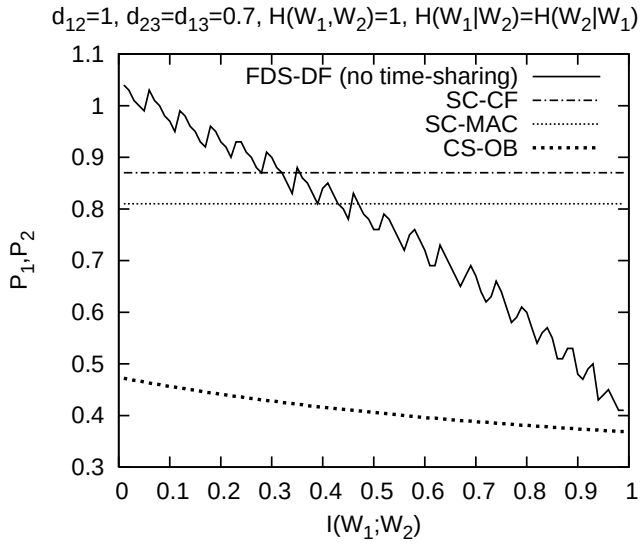


Figure 5.8: Minimum power required to transmit (W_1, W_2) to the destination per channel use, with different message correlation but constant $H(W_1, W_2)$

as a surprise as when $H(W_1|W_2) = H(W_2|W_1) = 0$, every source node has complete knowledge of other nodes' data. They can cooperate to form a multiple-transmit antenna without wasting any power to exchange data. Hence it achieves the CS-OB.

The achievable regions of all the three strategies are far from the CS-OB when the inter-source distance is large compared to the source-destination distance and the correlation between the sources are low. To achieve the CS-OB, all sources need to cooperate to send full information. When the correlation is low and the inter-source link is weak, the sources "waste" a larger portion of the transmit power to communicate among themselves in FDS-DF. For SC-CF and SC-MAC, as no coherent combining is possible, the achievable regions are far from the CS-OB. This highlights the value of cooperation in the MACFCS.

Remark 25 *We notice that in Figs. 5.7 and 5.8, the FDS-DF curves are zig-zag. This is because we plot the equal power point ($P_1 = P_2$) for the non-time-sharing FDS-DF. As can be seen from Fig. 5.4, time sharing might improve the FDS-DF region at the equal power point. The non-time-sharing FDS-DF curve in Fig. 5.4 coincides with the time-sharing curve at equal power point only at certain correlation levels. Hence the deviation from the time-sharing line for different*

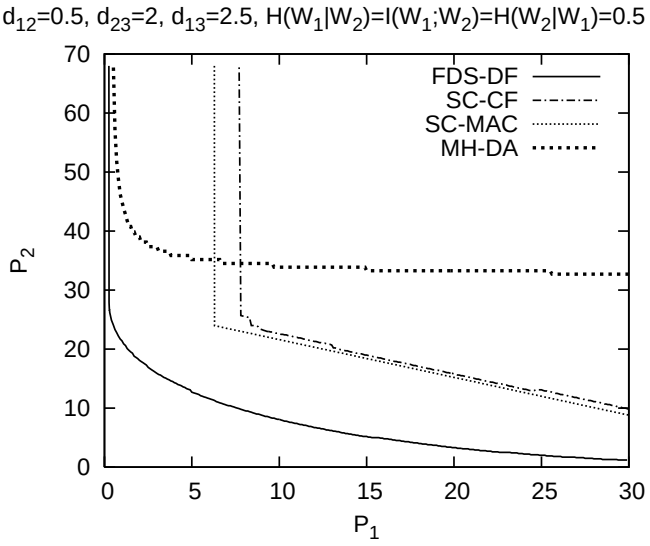


Figure 5.9: Minimum power required to transmit (W_1, W_2) to the destination per channel use, with node 2 closer to node 1.

correlation levels accounts for the zig-zag behavior of the FDS-DF curve in Figs. 5.7 and 5.8 when we change the correlation level.

5.7.4 Comparing MH-DA with other strategies

Figs. 5.9 and 5.10 compare MH-DA with other strategies in a three-node Gaussian MACFCS. As explained in Section 5.7.1, we will only consider the linear topology when comparing MH-DA with other strategies. We consider the cases when node 2 is closer to node 1 and when node 2 is closer to the destination. We show that in both cases, we can always find a strategy with multi-user coding (FDS-DF, SC-CF, or SC-MAC) that outperforms MH-DA.

Using MH-DA, we penalize the nodes toward the end of the route as they need to send more information. In this example, node 2 needs to send full information, which is at least $H(W_1, W_2)$ bits, "alone" to the destination as the destination only decodes from node 2. Hence the minimum required P_2 is high. In other strategies, node 1 helps node 2 to transmit to the destination and hence a lower P_2 is possible. An exception is FDS-DF when node 2 is closer to the destination. Here, node 2 needs to transmit at high power to ensure that node 1 (which is situated further away) can fully decode its transmission.

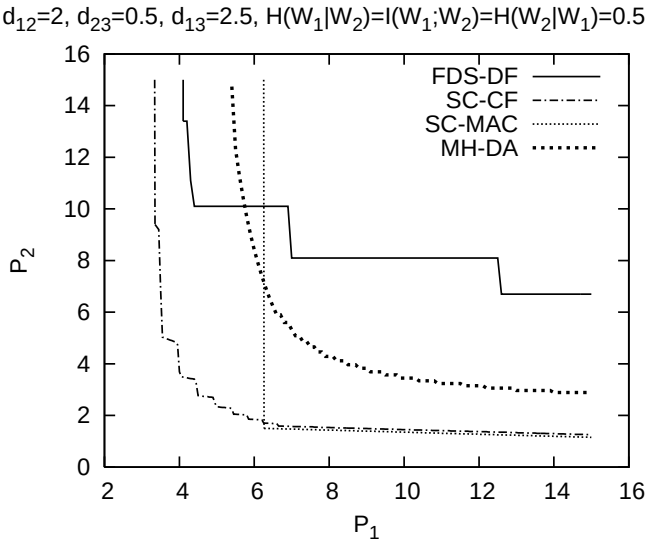


Figure 5.10: Minimum power required to transmit (W_1, W_2) to the destination per channel use, with node 2 closer to the destination.

Table 5.1: Node positioning, correlation, and coding strategies for the symmetrical Gaussian MACFCS.

Inter-source[♯] link	Low correlation	High correlation
Good	FDS-DF approaches CS-OB	FDS-DF approaches CS-OB
Poor	FDS-DF, SC-CF, SC-MAC far from CS-OB	

[♯]Relative to the source-destination links.

Nodes in the beginning of the route benefit from MH-DA, but only when the "next hop" is near. From Fig. 5.9, we see that P_1 can be low when node 2 is closer to node 1. When node 2 is further away (Fig. 5.10), node 1 suffers. FDS-DF and SC-CF help node 1 to lower its transmit power as both node 2 and the destination are listening.

5.8 Reflections

The analyses of the different coding strategies for the Gaussian MACFCS help us to understand better how sensor nodes can cooperate in a network given node positions and correlation structures. We summarize the results from the numerical computations for the symmetrical Gaussian MACFCS in Table 5.1.

Remark 26 *When every source node can fully decode from other sources using little power (i.e., when the inter-source link is good or when the data are highly*

correlated), FDS-DF is a good choice of coding strategy. The sources can coherently transmit to the destination with little inter-node communication.

Remark 27 *For the symmetrical topology when the inter-sources links are poor (e.g., when sources surround the destination), SC-MAC proves to be useful. A complicated scheme like SC-CF does not improve the achievable region.*

Remark 28 *For the asymmetrical topology, we note that SC-CF gives a better performance compared to FDS-DF and SC-MAC. SC-CF allows the furthest node to transmit at lower power as other source nodes now act as additional antennas for the destination.*

Remark 29 *For the linear asymmetrical topology with symmetrical source data, Figs. 5.9 and 5.10 show that we can always find a multi-user coding strategy that outperforms MH-DA. The problem with MH-DA is that it uses point-to-point coding and unfairly loads nodes nearer the end of the route. Multi-user coding strategies mitigate this by allowing richer forms of cooperation between nodes. This highlights the value of cooperative coding in a multiple-source network.*

Remark 30 *We investigated the three-node MACFCS in this chapter. This simple example enabled us to demonstrate the characteristics of different coding strategies. Consider a sensor network. The correlation between the measured data often depends on the inter-sensor positions. A shorter inter-sensor distance usually results in a higher correlation between the data of the two sensors. Hence, the upper right cell and the lower left cell in Table 5.1 are of greater interest. If the sensors are closer to one another than they are from the sink, which normally results in a high correlation among the data, they should fully decode the data from all sensors and transmit coherently to the sink. This can be done by using FDS-DF based coding schemes. If the sensors are scattered around the sink, which normally results in a lower correlation among the data, a simple coding strategy like SC-MAC might be sufficient.*

Remark 31 *In a network with more nodes, mixed coding strategies can be used. Here, we give an example of how the results in this thesis could help us to design a coding scheme for sensor network with more nodes. If there is a group of sensors situated further away from the destination and another group closer to the destination, we suggest that sensors that are further from the destination form a group and fully decode the data from each other. They, as a group, then cooperate with sensors nearer to the destination via SC-CF.*

5.9 Conclusion

In this chapter, we presented four achievable regions for the MACFCS. In addition, we derived an outer bound on the capacity of the MACFCS, which turned out to be the cut-set bound. Using Gaussian channels as examples, we compared the achievable regions of different strategies to the cut-set outer bound. We showed that FDS-DF, SC-CF, and SC-MAC can each give superior performance in certain channel settings. From the comparison, we found that when the inter-source links are better than the source-destination links, the achievable region of FDS-DF approaches the cut-set outer bound as the inter-source link gets better. The same strategy also approaches the cut-set outer bound when the correlation between the sources gets higher. In symmetrical topologies when the inter-source links are weak but the source-destination links are good, SC-MAC proves to be useful. In asymmetrical topologies, SC-CF can give better performance compared to the other strategies.

Chapter 6

Conclusion

The analyses in this thesis help us to understand cooperative coding and routing in multiple-terminal wireless networks better. We investigated three areas with practical interest, namely cooperative routing, myopic cooperation, and correlated sources.

From an information-theoretic view point, we presented algorithms to find optimal routes for DF for the MRC potentially without needing to optimize the channel input probability density functions. This saves computations in finding the optimal (rate maximizing) route for DF and calculating the maximum rate achievable by DF for an MRC. We derived achievable rates of DF for the MRC with different levels of cooperation, and showed that rates bounded away from zero are achievable with partial cooperation among the nodes even on large networks with the number of nodes growing infinitely large. We derived new achievable rate regions for the multiple-access channel with feedback and correlated sources using different types of cooperative coding strategies. We found the capacity of certain classes of the multiple-access channel with feedback and correlated sources.

From a practical view point, our routing algorithms can be used to find a rate-maximizing route (using the optimal routing algorithm) or to find a rate-maximizing route with high probability (using the heuristic routing algorithm) on which DF-based codes can be designed. The analyses and comparison on different cooperative coding strategies in this thesis help us to determine which form of

cooperation is useful for a particular network setting. Our work on myopic coding showed that we can even do a localized version of these codes, which reduces the complexity of the communication system, without compromising much on the transmission rates.

The cooperative (in the information-theoretic sense) relaying strategies discussed in this thesis provide a framework on which practical codes can be designed for multiple-terminal networks. A point to note is that these rates are derived based on information-theoretic calculations using codes with possibly infinitely long block length. It is for future research to design practical codes that can achieve rates. In fact, a few recent code designs were based on various information-theoretic coding strategies (see Razaghi & Yu (2006) for decode-forward and Hu & Li (2006) for compress-forward). These practical codes are able to push the transmission rates closer to those promised by the respective information-theoretic limits.

Appendix A

Appendices to Chapter 3

A.1 Sketch of Proof for Lemma 3

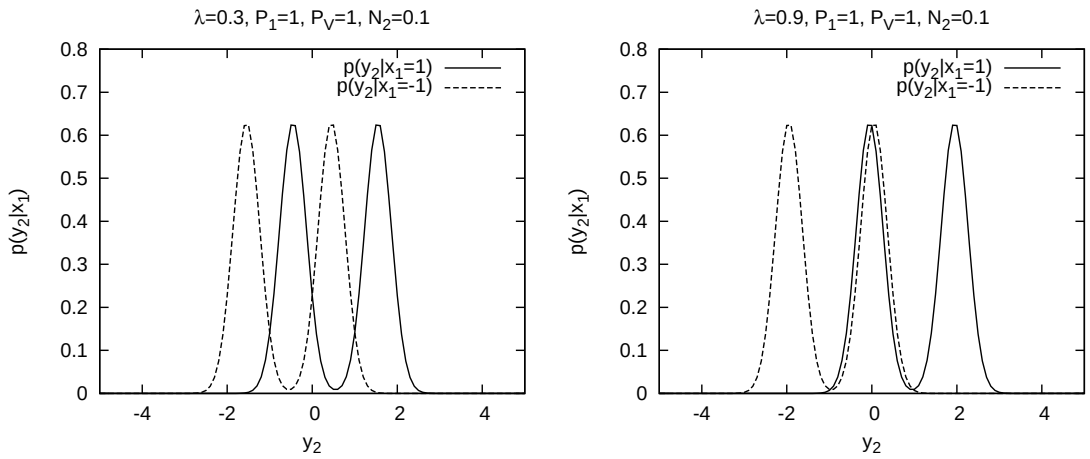
First, we investigate how $p(y_2)$ determines how $I(X_1;Y_2)$ varies with λ. We know that $I(X_1;Y_2) = H(Y_2) - H(Y_2|X_1)$. For Gaussian channels in the form $Y_2 = \sqrt{\lambda}X_1 + Z_2 + V_2$, and λ known to node 2, $H(Y_2|X_1) = H(Z_2) + H(V_2)$ is a constant. So, $H(Y_2|X_1)$ does not depend on λ. $H(Y_2)$ depends on λ, as λ controls the spread of $p(y_2)$. The spread of $p(y_2)$ is indicative of the level of randomness, i.e., the entropy $H(Y_2)$, of Y_2. The more $p(y_2)$ is spread, the higher $H(Y_2)$ is, and so the higher $I(X_1;Y_2)$ is.

Let $P_1 = 1, P_V = 1$, and the p.d.f. of X_1 and V_2 be

$$p(x_1) = \begin{cases} \frac{1}{2} & \text{, if } x_1 = 1 \\ \frac{1}{2} & \text{, if } x_1 = -1 \\ 0 & \text{, otherwise} \end{cases} \tag{A.1a}$$

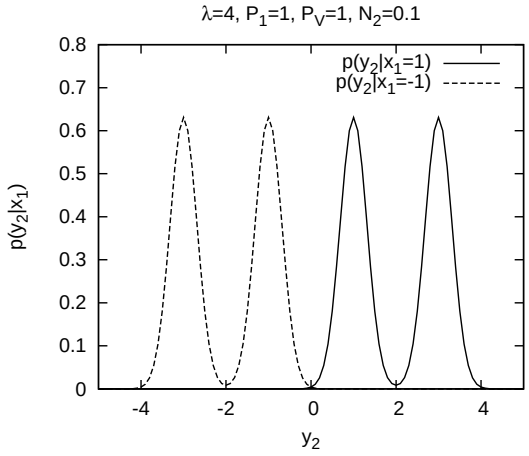
$$p(v_2) = \begin{cases} \frac{1}{2} & \text{, if } v_2 = 1 \\ \frac{1}{2} & \text{, if } v_2 = -1 \\ 0 & \text{, otherwise} \end{cases} . \tag{A.1b}$$

Let Z_2 be zero-mean Gaussian.



(a) $\lambda = 0.3$.

(b) $\lambda = 0.9$.

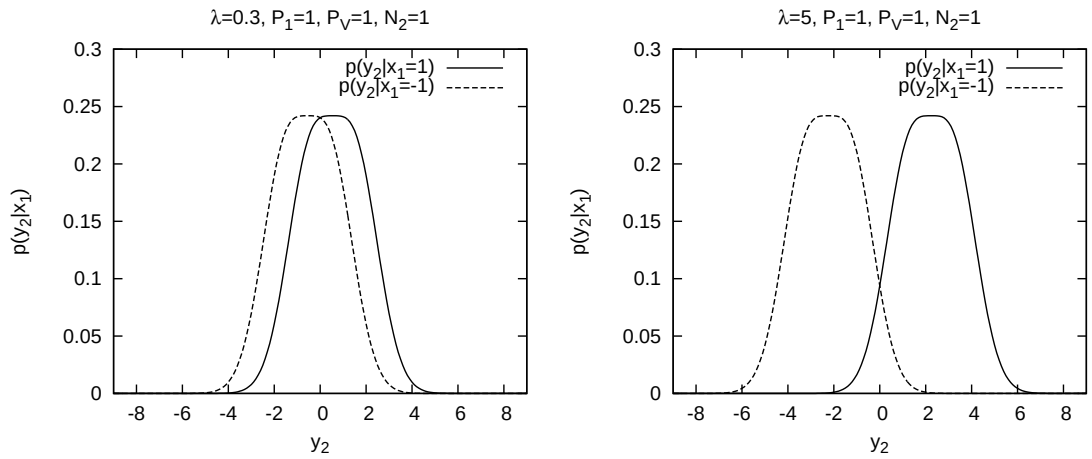


(c) $\lambda = 4$.

Figure A.1: Conditional channel output distribution for low receiver noise, $N_2 = 0.1$.

Let us consider low noise power case, $N_2 = 0.1$. Refer to Figs. A.1(a)–(c). When $\lambda = 0.3$ and $\lambda = 4$, $H(Y_2)$ is large as $p(y_2)$ are well spread. However, when $\lambda = 0.9$, $p(y_2|x_1 = 1)$ and $p(y_2|x = -1)$ overlap more and hence $H(Y_2)$ is low. As summarized in Fig. A.3, we see that increasing λ does not necessarily increase $I(X_1; Y_2)$ when the noise power is small. We see here that $p(y_2|x_1)$ have multiple peaks, and when we increase λ to a certain value, some peaks of $p(y_2|x_1 = 1)$ and $p(y_2|x = -1)$ overlap.

However, when the noise power is higher at $N_2 = 1$, Figs. A.2(a)–(b) show that the distribution of $p(y_2|x_1)$ is approximately Gaussian, which is single-peak. In this case, increasing λ always increases the spread of $p(y_2)$, leading to a higher $H(Y_2)$.



(a) $\lambda = 0.3$. (b) $\lambda = 5$.

Figure A.2: Conditional channel output distribution for higher receiver noise, $N_2 = 1$.

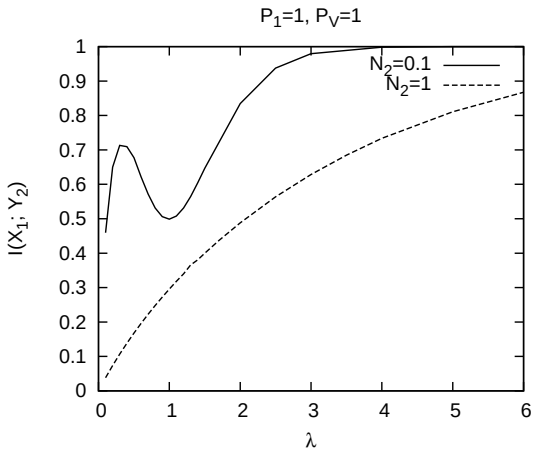


Figure A.3: Channel gain versus mutual information.

Fig. A.3 shows that increasing λ will only increase $I(X_1; Y_2)$.

The sketch of the proof for Lemma 3 is as follows. As $\sqrt{\lambda}x_1$ is added to $(z_2 + v_2)$ to form y_2, $p(y_2|x_1 = i)$ for all i have the same shape. If $p(y_2|x_1)$ is single-peak, when we increase λ, the peaks of $p(y_2|x_1 = i)$ for all i are further away from each other, i.e., less overlapping. Let $p(y_2)$ and $p'(y_2')$ correspond to the output p.d.f. for λ and λ' respectively, where $\lambda < \lambda'$. For all y_2 where $p(y_2) > 0$, we can always find a one-to-one mapping from y_2 to y_2' with $p(y_2) \geq p'(y_2')$. Hence $H(Y_2') \geq H(Y_2)$ and $I(X_1; Y_2') \geq I(X_1; Y_2)$. Furthermore, if $p(y_2|x_1)$ is single-peak and if Z_2 is Gaussian, then $p(y_2|x_1 = i) > 0, \forall y_2, i$. So, an increase in λ will always reduce the overlapping of $p(y_2|x_1 = i)$ and increase $I(X_1; Y_2)$.

A.2 Proof of Theorem 2

Without loss of generality, let $\mathcal{T} = \{1, 2, \ldots, |\mathcal{T}|\}$ and assume $E[Z_j^2] = E[Z_k^2]$. Hence, $\lambda_{ij} \geq \lambda_{ik}, \forall i \in \mathcal{T}$,

$$I(X_{\mathcal{T}}; Y_k | X_{\mathcal{T}^c}) \tag{A.2a}$$

$$= I(X_1, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1k}} X_1 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \tag{A.2b}$$

$$\begin{aligned} &= I(X_1; \sqrt{\lambda_{1k}} X_1 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \\ &\quad + I(X_2, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1k}} X_1 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_1, X_{\mathcal{T}^c}) \end{aligned} \tag{A.2c}$$

$$\begin{aligned} &= I(X_1; \sqrt{\lambda_{1k}} X_1 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \\ &\quad + I(X_2, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_1, X_{\mathcal{T}^c}) \end{aligned} \tag{A.2d}$$

$$\begin{aligned} &\leq I(X_1; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \\ &\quad + I(X_2, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_1, X_{\mathcal{T}^c}) \end{aligned} \tag{A.2e}$$

$$\begin{aligned} &= I(X_1; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \\ &\quad + I(X_2, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_1, X_{\mathcal{T}^c}) \end{aligned} \tag{A.2f}$$

$$= I(X_1, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \tag{A.2g}$$

$$\begin{aligned} &= I(X_2; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \\ &\quad + I(X_1, X_3 \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_2, X_{\mathcal{T}^c}) \end{aligned} \tag{A.2h}$$

$$\begin{aligned} &= I(X_2; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2k}} X_2 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \\ &\quad + I(X_1, X_3 \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{3k}} X_3 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_2, X_{\mathcal{T}^c}) \end{aligned} \tag{A.2i}$$

$$\begin{aligned} &\leq I(X_2; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{2j}} X_2 + + \sqrt{\lambda_{3k}} X_3 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \\ &\quad + I(X_1, X_3 \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}} X_1 + \sqrt{\lambda_{3k}} X_3 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_2, X_{\mathcal{T}^c}) \end{aligned} \tag{A.2j}$$

$$I(X_1, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}} X_1 + \lambda_{2j} X_2 + \sqrt{\lambda_{3k}} X_3 + \cdots + \sqrt{\lambda_{|\mathcal{T}|k}} X_{|\mathcal{T}|} + Z_k | X_{\mathcal{T}^c}) \tag{A.2k}$$

Applying the same technique down the route, we get

$$I(X_{\mathcal{T}}; Y_k|X_{\mathcal{T}^c}) \tag{A.3a}$$

$$\leq I(X_1, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}}X_1 + \cdots + \sqrt{\lambda_{(T-1)j}}X_{T-1} + \sqrt{\lambda_{|\mathcal{T}|k}}X_{|\mathcal{T}|} + Z_k|X_{\mathcal{T}^c}) \tag{A.3b}$$

$$= I(X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}}X_1 + \cdots + \sqrt{\lambda_{(|\mathcal{T}|-1)j}}X_{|\mathcal{T}|-1} + \sqrt{\lambda_{|\mathcal{T}|k}}X_{|\mathcal{T}|} + Z_k|X_{\mathcal{T}^c})$$
$$+ I(X_2, \ldots, X_{|\mathcal{T}|-1}; \sqrt{\lambda_{1j}}X_1 + \cdots + \sqrt{\lambda_{(|\mathcal{T}|-1)j}}X_{|\mathcal{T}|-1} + Z_k|X_{|\mathcal{T}|}, X_{\mathcal{T}^c}) \tag{A.3c}$$

$$\leq I(X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}}X_1 + \cdots + \sqrt{\lambda_{(|\mathcal{T}|-1)j}}X_{|\mathcal{T}|-1} + \sqrt{\lambda_{|\mathcal{T}|j}}X_{|\mathcal{T}|} + Z_k|X_{\mathcal{T}^c})$$
$$+ I(X_2, \ldots, X_{|\mathcal{T}|-1}; \sqrt{\lambda_{1j}}X_1 + \cdots + \sqrt{\lambda_{(|\mathcal{T}|-1)j}}X_{|\mathcal{T}|-1} + Z_k|X_{|\mathcal{T}|}, X_{\mathcal{T}^c}) \tag{A.3d}$$

$$= I(X_1, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}}X_1 + \cdots + \sqrt{\lambda_{|\mathcal{T}|j}}X_{|\mathcal{T}|} + Z_k|X_{\mathcal{T}^c}) \tag{A.3e}$$

$$= I(X_1, \ldots, X_{|\mathcal{T}|}; \sqrt{\lambda_{1j}}X_1 + \cdots + \sqrt{\lambda_{|\mathcal{T}|j}}X_{|\mathcal{T}|} + Z_j|X_{\mathcal{T}^c}) \tag{A.3f}$$

$$= I(X_{\mathcal{T}}; Y_j|X_{\mathcal{T}^c}) \tag{A.3g}$$

Eqns. (A.2e), (A.2j) and (A.3d) are due to Lemma 3 and when the SPC is satisfied.

A.3 Proof of Theorem 3

To prove Theorem 3, we need the following few lemmas.

Lemma 7 *Consider route $\mathcal{M} = \{m_1, \ldots, m_{|\mathcal{M}|}\}$ and assume that a unique nearest neighbor, node a^*, exists. For any input distribution p satisfying the SPC, the rate supported by the new route $\mathcal{M}_1 = \mathcal{M} \cup \{a^*\}$ is greater or equal than the rate supported by the route adding any non-nearest neighbor, or $\mathcal{M}_2 = \mathcal{M} \cup \{b\}$, i.e.,*

$$R_{\mathcal{M}\cup\{a^*\}}(p) \geq R_{\mathcal{M}\cup\{b\}}(p), \quad \forall b \in \mathcal{T} \setminus (\mathcal{M} \cup \{a^*\}). \tag{A.4}$$

We use superscript * to indicate a nearest neighbor.

Proof 21 (Proof for Lemma 7) *Since a^* is the unique nearest neighbor with*

respect to $\mathcal{M}$*, by definition,*

$$\gamma_{ma^*} \geq \gamma_{mb}, \quad \forall m \in \mathcal{M}, \forall b \in \mathcal{T} \setminus (\mathcal{M} \cup \{a^*\}). \tag{A.5}$$

From Theorem 2, it follows that $\forall b \in \mathcal{T} \setminus (\mathcal{M} \cup \{a^*\})$*,*

$$\begin{aligned} R_{a^*}(\mathcal{M}_1, p) &= I(X_{\mathcal{M}}; Y_{a^*} | X_{\mathcal{T}\setminus\mathcal{M}}) && \text{(A.6a)} \\ &\geq I(X_{\mathcal{M}}; Y_b | X_{\mathcal{T}\setminus\mathcal{M}}) && \text{(A.6b)} \\ &= R_b(\mathcal{M}_2, p). && \text{(A.6c)} \end{aligned}$$

In addition, for the first $|\mathcal{M}|$ *elements in both routes less the source, i.e.,* $2 \leq i \leq |\mathcal{M}|$*,*

$$\begin{aligned} R_{m_i}(\mathcal{M}_1, p) &= I(X_{m_1}, \ldots, X_{m_{i-1}}; Y_{m_i} | X_{m_i}, \ldots, X_{m_{|\mathcal{M}|}}, X_{\mathcal{T}\setminus\mathcal{M}}) && \text{(A.7a)} \\ &= R_{m_i}(\mathcal{M}_2, p). && \text{(A.7b)} \end{aligned}$$

Hence,

$$\begin{aligned} R_{\mathcal{M}_1}(p) &= \min_{j \in \mathcal{M}_1 \setminus \{1\}} R_j(\mathcal{M}_1, p) && \text{(A.8a)} \\ &\geq \min_{j \in \mathcal{M}_2 \setminus \{1\}} R_j(\mathcal{M}_2, p) && \text{(A.8b)} \\ &= R_{\mathcal{M}_2}(p) && \text{(A.8c)} \end{aligned}$$

We have proven that at any point of time during the NNA, in order to maximize the rate supported by the route, we must choose the nearest neighbor (assuming that it exists). Next, we show that choosing the nearest neighbor will not harm the rate supported by the route even when more nodes are added.

Lemma 8 *Let route* $\mathcal{M}$*, with all nodes chosen using the NNA, be*

$$\mathcal{M} = \{a_1^*, a_2^*, \ldots, a_{|\mathcal{M}|}^*\}. \tag{A.9}$$

Every node that has been added to route is the nearest neighbor with respect to the original route. Now, we arbitrarily add K nodes to $\mathcal{M}$. The first node b_1 is not a nearest neighbor and the rest may or may not be nearest neighbors, i.e.,

$$\mathcal{M}_1 = \{a_1^*, a_2^*, \ldots, a_{|\mathcal{M}|}^*, b_1, b_2, \ldots, b_K\}, \tag{A.10}$$

where b_1 is not a nearest neighbor with respect to $\mathcal{M}$. We can always replace b_1 by the nearest neighbor $a_{|\mathcal{M}|+1}^*$ (assuming that it exists) to obtain

$$\mathcal{M}_2 = \begin{cases} \{a_1^*, \ldots, a_{|\mathcal{M}|}^*, a_{|\mathcal{M}|+1}^*, b_1, \ldots, b_{K-1}\}, & \text{if } a_{|\mathcal{M}|+1}^* \notin \{b_1, \ldots, b_{K-1}\} \\ \{a_1^*, \ldots, a_{|\mathcal{M}|}^*, a_{|\mathcal{M}|+1}^*, b_1, \ldots, b_{k-1}, b_{k+1}, \ldots, b_K\}, & \text{if } a_{|\mathcal{M}|+1}^* = b_k, \end{cases} \tag{A.11}$$

for some $b_k \in \{b_1, \ldots, b_{K-1}\}$, and show that for any p satisfying the SPC,

$$R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}_1}(p). \tag{A.12}$$

Proof 22 (Proof for Lemma 8) *First, we prove the case when $a_{|\mathcal{M}|+1}^* \notin \{b_1, \ldots, b_{K-1}\}$. Lemma 7, it follows that*

$$R_{a_{|\mathcal{M}|+1}^*}(\mathcal{M}_2, p) \geq R_{b_1}(\mathcal{M}_1, p). \tag{A.13}$$

Now, for nodes $b_i, 1 \leq i \leq K-1$,

$$\begin{aligned} R_{b_i}(\mathcal{M}_2, p) &= I(X_{\mathcal{M}}, X_{m_{|\mathcal{M}|+1}^*}, X_{b_1} \ldots, X_{b_{i-1}}; Y_{b_i} | X_{b_i}, \ldots, X_{b_{K-1}}, X_{\mathcal{T}\setminus\mathcal{M}_2}) && \text{(A.14a)} \\ &\geq I(X_{\mathcal{M}}, X_{b_1} \ldots, X_{b_{i-1}}; Y_{b_i} | X_{b_i}, \ldots, X_{b_{K-1}}, X_{m_{|\mathcal{M}|+1}^*}, X_{\mathcal{T}\setminus\mathcal{M}_2}) && \text{(A.14b)} \\ &= I(X_{\mathcal{M}}, X_{b_1} \ldots, X_{b_{i-1}}; Y_{b_i} | X_{b_i}, \ldots, X_{b_{K-1}}, X_{b_K}, X_{\mathcal{T}\setminus\mathcal{M}_1}) && \text{(A.14c)} \\ &= R_{b_i}(\mathcal{M}_1, p). && \text{(A.14d)} \end{aligned}$$

Eqn (A.14b) is due to Lemma 4.

Clearly, for all nodes $m \in \mathcal{M} \setminus \{1\}$,

$$R_m(\mathcal{M}_2, p) = R_m(\mathcal{M}_1, p). \tag{A.15}$$

Hence,

$$R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}_1}(p). \tag{A.16}$$

Now, we study the case when $a^*_{|\mathcal{M}|+1} = b_k$ *for some* $b_k \in \{b_1, \ldots, b_{K-1}\}$. *Similar to the first case,* $R_{b_i}(\mathcal{M}_2, p) \geq R_{b_i}(\mathcal{M}_1, p)$ *for* $1 \leq i \leq k-1$ *and* $R_{a^*_{|\mathcal{M}|+1}}(\mathcal{M}_2, p) \geq R_{b_1}(\mathcal{M}_1, p)$. *It is easy to show that for any* $m \in \{a^*_2, a^*_3 \ldots, a^*_{|\mathcal{M}|}, b_k, b_{k+1}, \ldots, b_K\}$, $R_m(\mathcal{M}_2, p) = R_m(\mathcal{M}_1, p)$. *Hence,* $R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}_1}(p)$.

This proves Lemma 8.

Lemma 9 *For a route in which all nodes are formed by the NNA, the supported rate is always higher or equal to a route, of the same length, with one or more non-nearest neighbors in it.*

Proof 23 (Proof for Lemma 9) *Lemma 9 can be proven by applying Lemma 8 recursively. Starting from the second node onward, we replace the first non-nearest neighbor with a nearest neighbor. We remove the last node if the resulting route is longer. In each step, the supported rate can only increase. We do that until the entire route is replaced by nearest neighbor nodes. Hence, we get Lemma 9.*

Now we consider routes of different lengths but that end on the same node, which is the destination, T.

Lemma 10 *Consider an arbitrary route* $\mathcal{M}_1$ *from the source, node 1, to the destination* $m_{|\mathcal{M}_1|} = T$*:*

$$\mathcal{M}_1 = \{m^*_1, m_2, \ldots, m_{|\mathcal{M}_1|}\}. \tag{A.17}$$

Here, one or more nodes in $\{m_2, \ldots, m_{|\mathcal{M}_1|}\}$ *are not nearest neighbors. Consider the NNA route, assuming that the NNA terminates normally:*

$$\mathcal{M}_2 = \{m^*_1, m^*_2, \ldots, m^*_{|\mathcal{M}_2|}\}, \tag{A.18}$$

where $m^*_{|\mathcal{M}_2|} = T$ *and* $|\mathcal{M}_1|$ *not necessarily equals* $|\mathcal{M}_2|$. *It follows that, with any input distribution* p,

$$R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}_1}(p). \tag{A.19}$$

Proof 24 (Proof for Lemma 10) *First of all, we consider the case* $|\mathcal{M}_1| = |\mathcal{M}_2|$. *The results follows immediately from Lemma 9.*

Second, we consider the case $|\mathcal{M}_1| > |\mathcal{M}_2|$. *We consider first* $|\mathcal{M}_2|$ *nodes in* $\mathcal{M}_1$, *i.e.,*

$$\mathcal{M}'_1 = \{m^*_1, m_2, \ldots, m_{|\mathcal{M}_2|}\}. \tag{A.20}$$

From Lemma 5, $R_{\mathcal{M}'}(p) \geq R_{\mathcal{M}_1}(p)$. *From Lemma 9,* $R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}'}(p)$. *Hence,* $R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}_1}(p)$.

Lastly, consider the case $|\mathcal{M}_2| > |\mathcal{M}_1|$. *We replace all the nodes in* $\mathcal{M}_1$ *except the last node, with nearest neighbors, and obtain*

$$\mathcal{M}_3 = \{m^*_1, m^*_2, \ldots, m^*_{|\mathcal{M}_1|-1}, m_{|\mathcal{M}_1|} = T\}. \tag{A.21}$$

Note that the first $|\mathcal{M}_1| - 1$ *nodes of routes* $\mathcal{M}_2$ *and* $\mathcal{M}_3$ *are the same,*

Note that $m_{|\mathcal{M}_1|} = T$ *might not be the nearest neighbor with respect to* $\mathcal{M}_3 \setminus \{T\}$. *Clearly, using Lemma 8 recursively,*

$$R_{\mathcal{M}_3}(p) \geq R_{\mathcal{M}_1}(p). \tag{A.22}$$

Since $\mathcal{M}_2$ *is the output of the NNA, the following conditions are necessary.*

$$\gamma_{ma} \geq \gamma_{mT}, \quad \forall m \in \{m^*_1, \ldots, m^*_{|\mathcal{M}_1|-1}\}, \forall a \in \{m^*_{|\mathcal{M}_1|}, \ldots, m^*_{|\mathcal{M}_2|-1}\}. \tag{A.23}$$

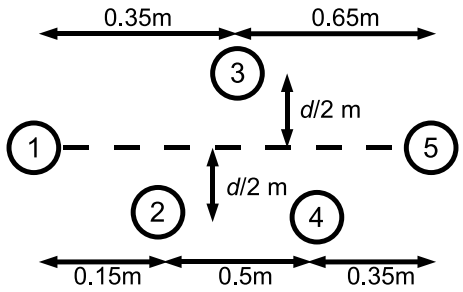


Figure A.4: A five-node MRC.

So, for $|\mathcal{M}_1| \leq i \leq |\mathcal{M}_2|$,

$$R_{m_i^*}(\mathcal{M}_2, p) = I(X_{m_1^*}, \ldots, X_{m_{i-1}^*}; Y_{m_i^*} | X_{m_i^*}, \ldots, X_{m^*_{|\mathcal{M}_2|}}, X_{\mathcal{T}\setminus\mathcal{M}_2}) \tag{A.24a}$$

$$\geq I(X_{m_1^*}, \ldots, X_{m^*_{|\mathcal{M}_1|-1}}; Y_{m_i^*} | X_{m^*_{|\mathcal{M}_1|}}, X_{m^*_{|\mathcal{M}_1|+1}}, \ldots, X_{m^*_{|\mathcal{M}_2|}}, X_{\mathcal{T}\setminus\mathcal{M}_2}) \tag{A.24b}$$

$$\geq I(X_{m_1^*}, \ldots, X_{m^*_{|\mathcal{M}_1|-1}}; Y_T | X_{m^*_{|\mathcal{M}_1|}}, X_{m^*_{|\mathcal{M}_1|+1}}, \ldots, X_{m^*_{|\mathcal{M}_2|}}, X_{\mathcal{T}\setminus\mathcal{M}_2}) \tag{A.24c}$$

$$= R_T(\mathcal{M}_3, p). \tag{A.24d}$$

Eqn (A.24b) *is due to Lemma 4, and* (A.24c) *is due to* (A.23) *and* $m^*_{|\mathcal{M}_2|} = T$. *Clearly, for the first* $|\mathcal{M}_1| - 1$ *nodes, i.e.,* $m \in \{m_2^*, m_3^*, \ldots, m^*_{|\mathcal{M}_1|-1}\}$, $R_m(\mathcal{M}_2, p) = R_m(\mathcal{M}_3, p)$. *Hence,* $R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}_3}(p) \geq R_{\mathcal{M}_1}(p)$.

Hence, we have Lemma 10.

From Lemma 10, we know that if the NNSA terminates normally, i.e., the nearest neighbor exists from the source to the destination, the route formed using the NNA can support transmission rates as high as any other route. In other words, the NNA finds a route that supports the highest achievable rate. Theorem 3 follows.

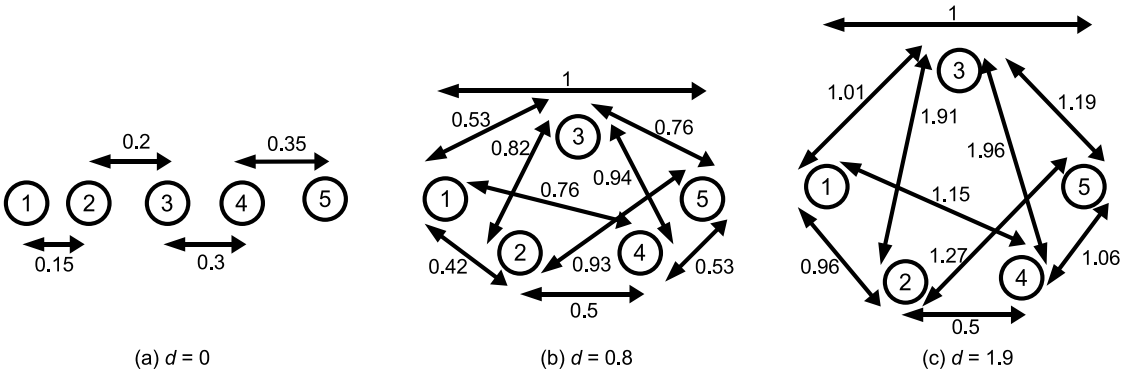


Figure A.5: Different topologies of the five-node MRC.

A.4 Examples of How the NNSA Reduces the Search Space for an Optimal Route

Now we provide an example to show that the search space of an optimum route can be very much reduced by using the NNSA.

We use the setup in Fig. A.4. We assume that $P_i = P, \forall i = 1,2,3,4$, $N_j = N, \forall j = 2,3,4,5$, $r_{ij} = 1, \kappa = 1, \eta = 2$. So, $\gamma_{ij} \propto \frac{1}{d_{ij}}$. We find the NNSA candidates for different three different d, as depicted in Figs. A.5(a)–(c).

Using brute force search,

$$|\Pi(\mathcal{T})| = 1 + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 16, \tag{A.25}$$

for all the three topologies.

For $d = 0$m, starting with $\mathcal{M} = \{1\}$, the nearest neighbor set contains only node 2. Adding node 2 to the transmitting set, we have $\mathcal{M} = \{1,2\}$. We find that after the addition of a node to the route, the nearest neighbor set contains only one node. Hence, the NNSA outputs only one route:

$$\{1\} \to \{1,2\} \to \{1,2,3\} \to \{1,2,3,4\} \to \{1,2,3,4,5\}. \tag{A.26}$$

Hence an optimal route for DF is $\{1,2,3,4,5\}$.

For $d = 0.8$m, starting with $\mathcal{M} = \{1\}$, the nearest neighbor is again node 2. At

this point, $\mathcal{M} = \{1, 2\}$. Now, node 3 is closer to node 1 than node 4 is to node 1. But node 4 is closer to node 2 than node 3 is to node 2. Node 5 is further away from node 1 (and 2) compared to 3 (and 4). We form a nearest neighbor set $\mathcal{N} = \{3, 4\}$. The transmitting set is split into two paths, i.e. $\{1, 2, 3\}$ and $\{1, 2, 4\}$. For route $\{1, 2, 3\}$ and nearest neighbor set is $\{4, 5\}$; and for route $\{1, 2, 4\}$, the nearest neighbor set is $\{3, 5\}$. In summary the NNSA yields the following NNSA candidates

$$\{1\} \to \{1, 2\} \to \{1, 2, 3\} \to \{1, 2, 3, 4\} \to \{1, 2, 3, 4, 5\} \tag{A.27a}$$

$$\{1\} \to \{1, 2\} \to \{1, 2, 3\} \to \{1, 2, 3, 5\} \tag{A.27b}$$

$$\{1\} \to \{1, 2\} \to \{1, 2, 4\} \to \{1, 2, 4, 3\} \to \{1, 2, 4, 3, 5\} \tag{A.27c}$$

$$\{1\} \to \{1, 2\} \to \{1, 2, 4\} \to \{1, 2, 4, 5\}. \tag{A.27d}$$

An optimal route is the one that supports the highest rate among these four NNSA candidates. Instead of finding the rates of all the 16 routes found by the brute force search, we only need to optimize the input distribution for these four routes to find the maximum DF rate.

For $d = 1.9$m, we start with $\mathcal{M} = \{1\}$. Adding the only node in the nearest neighbor set, node 2, we get $\mathcal{M} = \{1, 2\}$. Now, the nearest neighbor set is $\mathcal{N} = \{3, 4, 5\}$. $\{1, 2, 5\}$ is an NNSA candidate. From route $\{1, 2, 3\}$, we get the following NNSA candidates: $\{1, 2, 3, 5\}$ and $\{1, 2, 3, 4, 5\}$. From route $\{1, 2, 4\}$, the nearest neighbor set contains only node 5. So, we get $\{1, 2, 4, 5\}$. So, there are four NNSA candidates.

A.5 Proof of Theorem 4

The complete proof is very similar to that for Theorem 3 and will hence be omitted. Using the technique used in the proof of Theorem 3, we can show that adding a node that does not belong to the nearest neighbor set can only be suboptimal. We can always replace that node with one from the nearest neighbor set and obtain

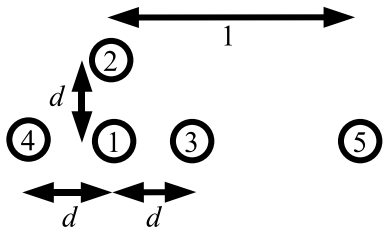


Figure A.6: An example to show that the NNSA routes backward.

an equal or higher rate. In other words, we can show that the supported rate of

$$\mathcal{M}_1 = \{m_1^*, m_2, \ldots, m_{|\mathcal{M}_1|}\}, \tag{A.28}$$

where $m_1^* = 1$ is the source, $m_{|\mathcal{M}_1|} = T$ is the destination, and one or more nodes in $\{m_2, \ldots, m_{|\mathcal{M}_1|}\}$ are not from the nearest neighbor set, is lower or equal to the supported rate of

$$\mathcal{M}_2 = \{m_1^*, m_2^*, \ldots, m_{|\mathcal{M}_2|}^*\}, \tag{A.29}$$

where $m_{|\mathcal{M}_2|}^* = T$ and all nodes in $\mathcal{M}_2$ are added according to the NNSA. In other words, for any input distribution p,

$$R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}_1}(p). \tag{A.30}$$

The NNSA algorithm finds all routes for which every node that is added to an route is from the nearest neighbor set. Hence one or more of the NNSA candidates must achieve the highest rate. With this, we have Theorem 4.

A.6 An Example Showing Routing Backward Can Improve Transmission Rates

We use the topology depicted in Fig. A.6 to show that routing backward can be desirable as it can improve the transmission rate. In this setting, nodes 2, 3, and 4 are each dm away from the source (node 1). The destination (node 5) is 1m

from the source. Clearly, the NNSA, which always add nearest neighbors, will add nodes 2, 3, and 4 before adding 5 when d is less than 1m. One NNSA candidate is $\mathcal{M}_1 = \{1, 2, 3, 4, 5\}$. However, it might not be intuitive that routing the data to node 4, which is "behind" node 1 can help in the transmission rate. We will compare the optimal rate on $\mathcal{M}_1$ with that on $\mathcal{M}_2 = \{1, 2, 3, 5\}$, which only routes data forward (to nodes nearer to the destination).

Table A.1: Achievable rates for different routes.

d	$\max_{p \in \mathcal{P}_{\text{All}}} R_{\mathcal{M}_1}(p)$	$\max_{p \in \mathcal{P}_{\text{All}}} R_{\mathcal{M}_2}(p)$
0.1	3.646	3.299
0.2	3.589	3.287
0.3	3.394	3.240
0.4	2.994	2.994
0.5	2.679	2.679

Table A.1 compares the rate of two routes $\mathcal{M}_1$ and $\mathcal{M}_2$ using DF, maximized over all input distributions. The second route chooses not to route the data backward. When d is small (0.1–0.3), the bottleneck of the transmission is the reception of node 5. Hence, adding node 4 into the route helps to increase the overall transmission rate. However, when d increases (≥ 0.4), the bottleneck shifts to the reception rate of nodes 2 and 3. Under this condition, adding node 4 to the route does not help in the overall transmission rate. However, adding node 4 does not harm the rate either.

A.7 Proof of Theorem 7

First, we show what an SOR is contained in at least one of the NNSA candidates.

Lemma 11 *If*

$$\mathcal{M} = \{m_1^*, m_2^*, \ldots, m_{|\mathcal{M}|-1}^*, m_{|\mathcal{M}|}^*\} \tag{A.31}$$

is an SOR for some p, then

$$\gamma_{am_{|\mathcal{M}|-1}^*} \geq \gamma_{am_{|\mathcal{M}|}^*} \tag{A.32}$$

for at least one $a \in \{m_1^*, m_2^*, \ldots, m_{|\mathcal{M}|-2}^*\}$.

Proof 25 (Proof of Lemma 11) *We prove this by contradiction. If*

$$\gamma_{am^*_{|\mathcal{M}|-1}} < \gamma_{am^*_{|\mathcal{M}|}}, \quad \forall a \in \{m^*_1, m^*_2, \ldots, m^*_{|\mathcal{M}|-2}\} \tag{A.33}$$

then from Lemma 7, for any p,

$$R_{\mathcal{M}_1}(p) \geq R_{\mathcal{M}}(p), \tag{A.34}$$

where

$$\mathcal{M}_1 = \{m^*_1, m^*_2, \ldots, m^*_{|\mathcal{M}|-2}, m^*_{|\mathcal{M}|}\}. \tag{A.35}$$

In this case, $\mathcal{M}_1$ a shorter route and can support rate as high as $\mathcal{M}$. So, the latter cannot be an SOR. Contradiction!

Lemma 11 says that the rSNR at the second last node from at least one of the first $|\mathcal{M}| - 2$ nodes must be higher than (or equal to) the rSNR at the last node. It does not mean that the second last node must be the from the nearest neighbor set with respect to the first $|\mathcal{M}| - 2$ nodes. However, when the partial route is $\{m^*_1, \ldots, m^*_{|\mathcal{M}|-2}\}$, the NNSA will eventually include $m^*_{|\mathcal{M}|-1}$ in one or more of of the NNSA candidates before it includes $m^*_{|\mathcal{M}|}$.

We now extend Lemma 11 to the third last node. Using a similar argument, we can show that

$$\gamma_{am^*_{|\mathcal{M}|-2}} \geq \gamma_{am^*_{|\mathcal{M}|}}, \tag{A.36}$$

for some $a \in \{m^*_1, m^*_2, \ldots, m^*_{|\mathcal{M}|-3}\}$. It then follows that if the partial route is $\{m^*_1, \ldots, m^*_{|\mathcal{M}|-3}\}$, the NNSA will definitely include $m^*_{|\mathcal{M}|-2}$ in one or more of the NNSA candidates before it includes $m^*_{|\mathcal{M}|}$.

We have shown that the NNSA will include $m^*_{|\mathcal{M}|-2}$ and $m^*_{|\mathcal{M}|-1}$ before $m^*_{|\mathcal{M}|}$ in one or more of its NNSA candidates. This does not guarantee that there exists an NNSA candidate that includes $m^*_{|\mathcal{M}|-2}, m^*_{|\mathcal{M}|-1}, m^*_{|\mathcal{M}|}$ in that order. For instance, if

$$\gamma_{am^*_{|\mathcal{M}|-1}} > \gamma_{am^*_{|\mathcal{M}|-2}}, \quad \forall a \in \{m^*_1, \ldots, m^*_{|\mathcal{M}|-3}\}, \tag{A.37}$$

then none of the NNSA candidate will contain $\mathcal{M}$ with the same order, but some will contain

$$\mathcal{M}' = \{m_1^*, m_2^*, \ldots, m_{|\mathcal{M}|-3}^*, m_{|\mathcal{M}|-1}^*, m_{|\mathcal{M}|-2}^*, m_{|\mathcal{M}|}^*\}, \tag{A.38}$$

However, in this case, $R_{\mathcal{M}'}(p) \geq R_{\mathcal{M}}(p)$ and hence $\mathcal{M}'$ is also an SOR.

This can be further extended to all nodes in $\mathcal{M}$. We can show that if $\mathcal{M}$ is an SOR, all the nodes in $\mathcal{M}$ must be in some NNSA candidates. If an NNSA candidate contains $\mathcal{M}'$ (which is $\mathcal{M}$ in a different order) but no NNSA candidate contains $\mathcal{M}$, then $\mathcal{M}'$ is also an SOR. This can be summarized in the following lemma.

Lemma 12 *One or more NNSA candidates contain an SOR, with the same node order as in the SOR.*

This means if $\mathcal{M} = \{m_1^*, m_2^*, \ldots, m_{|\mathcal{M}|-1}^*, m_{|\mathcal{M}|}^*\}$ is the SOR contained in one of the NNSA candidates, then m_2^* comes after m_1^*, m_3^* comes after m_2^*, etc, in the NNSA candidate. But there might be other nodes in between them. For example, the NNSA candidate might be $\{m_1^*, \ldots, m_{|\mathcal{M}|-2}^*, a, m_{|\mathcal{M}|-1}^*, m_{|\mathcal{M}|}^*\}$. Now, we will go further and prove that an SOR, $\mathcal{M}^{\text{SOR}}(p)$ is contained in one or more of the optimal NNSA candidates, $\mathcal{NNSA}^{\text{opt}}(\mathcal{T}, p)$, not just the NNSA candidates, $\mathcal{NNSA}(\mathcal{T})$.

Consider an SOR

$$\mathcal{M}^* = \{m_1^*, m_2^*, \ldots, m_K^*, m_{K+1}^*, \ldots, m_{|\mathcal{M}|-1}^*, m_{|\mathcal{M}|}^*\}, \tag{A.39}$$

and an NNSA candidate

$$\mathcal{M}_1 = \{m_1^*, m_2^*, \ldots, m_K^*, a, m_{K+1}^*, \ldots, m_{|\mathcal{M}|-1}^*, m_{|\mathcal{M}|}^*\}, \tag{A.40}$$

This is possible since $\mathcal{M}_1$ contains $\mathcal{M}^*$ in the correct order.

We term the set of the first K nodes $\mathcal{M}_1'$, i.e.,

$$\mathcal{M}_1' = \{m_1^*, \ldots, m_K^*\}. \tag{A.41}$$

Since $\mathcal{M}_1$ is formed by the NNSA algorithm, a must be from the nearest neighbor set with respect to $\mathcal{M}_1'$, and m_{K+1}^* must belong to one of these three cases:

1. m_{K+1}^* is not from the nearest neighbor set with respect to $\mathcal{M}_1'$.
2. $(a, m_{K+1}^*, m_{K+2}^*, \ldots, m_{K+i}^*)$ are from the nearest neighbor set with respect to $\mathcal{M}_1'$, and $m_{K+i}^* \neq m_{|\mathcal{M}|}^*$.
3. $(a, m_{K+1}^*, m_{K+2}^*, \ldots, m_M^*)$ are from the nearest neighbor set with respect to $\mathcal{M}_1'$.

Consider the first case, it means

$$\gamma_{ba} \geq \gamma_{bm_{K+1}^*}, \quad \forall b \in \mathcal{M}_1', \tag{A.42}$$

with at least one inequality. For any input distribution p,

$$R_{m_n^*}(\mathcal{M}_1, p) = R_{m_n^*}(\mathcal{M}^*, p), \quad \text{for } n = 2, 3, \ldots, K \tag{A.43a}$$

$$R_a(\mathcal{M}_1, p) \geq R_{m_{K+1}^*}(\mathcal{M}^*, p) \tag{A.43b}$$

$$R_{m_n^*}(\mathcal{M}_1, p) \geq R_{m_n^*}(\mathcal{M}^*, p), \quad \text{for } n = K+1, K+2, \ldots, |\mathcal{M}|. \tag{A.43c}$$

So, $R_{\mathcal{M}_1}(p) \geq R_{\mathcal{M}^*}(p) = \max_{\mathcal{M} \in \Pi(\mathcal{T})} R_{\mathcal{M}}(p)$, and hence $\mathcal{M}_1 \in \mathcal{NNSA}^{\text{opt}}(\mathcal{T}, p)$.

Consider the second case, since $(a, m_{K+1}^*, m_{K+2}^*, \ldots, m_{K+i}^*)$ are from the nearest neighbor set with respect to $\mathcal{M}_1'$, it follows that

$$\mathcal{M}_2 = \{m_1^*, m_2^*, \ldots, m_K^*, m_{K+1}^*, \ldots, m_{K+i}^*, a, m_{K+i+1}^*, \ldots, m_M^*\} \tag{A.44}$$

is also an NNSA candidate. This is because the NNSA tries all permutations of the nodes in the nearest neighbor set. Now, consider $\mathcal{M}_2' = \{m_1^*, m_2^*, \ldots, m_K^*, m_{K+1}^*, \ldots, m_{K+i}^*\}$.

If

$$\gamma_{ba} \geq \gamma_{bm_{K+i+1}^*}, \quad \forall b \in \mathcal{M}_2', \tag{A.45}$$

with at least one inequality, meaning a is the unique nearest neighbor of $\mathcal{M}_2'$, then using the same argument for case (i) above, we can show that for any p, $R_{\mathcal{M}_2}(p) \geq R_{\mathcal{M}^*}(p) = \max_{\mathcal{M}\in\Pi(\mathcal{T})} R_{\mathcal{M}}(p)$. So, $\mathcal{M}_2 \in \mathcal{NNSA}^{\text{opt}}(\mathcal{T}, p)$. We have shown that $\mathcal{M}^*$ is contained in another optimal NNSA candidate, i.e., $\mathcal{M}_2$.

Else if $(a, m^*_{K+i+1}, m^*_{K+i+j})$ are the nearest neighbors with respect to $\mathcal{M}_2'$, we rearrange and notice that

$$\mathcal{M}_3 = \{m^*_1, m^*_2, \ldots, m^*_{K+i+j}, a, m^*_{K+i+j+1}, \ldots, m^*_M\} \tag{A.46}$$

is also an NNSA candidate. We check if a is the unique nearest neighbor with respect to the nodes behind. We stop when node a is the only nearest neighbor. This is equivalent to case (i), i.e., we can show that $\mathcal{M}'$ is contained in an optimal NNSA candidate with the same nodes order. If a is not the unique nearest neighbor, we continue to "push" node a backwards until a is in the nearest neighbor set that includes the destination, which is case (iii).

Now, consider case (iii). $(a, m^*_{K+1}, m^*_{K+2}, \ldots, m^*_M)$ are from the nearest neighbor set with respect to $\mathcal{M}_1'$. Since the NNSA tries all possible permutations of nodes in the nearest neighbor set, omitting node a also yields an NNSA candidates. So, $\mathcal{M}^* \in \mathcal{NNSA}(\mathcal{T})$. Since $R_{\mathcal{M}^*}(p) = \max_{\mathcal{M}\in\Pi(\mathcal{T})} R_{\mathcal{M}}(p)$, $\mathcal{M}^* \in \mathcal{NNSA}^{\text{opt}}(\mathcal{T}, p)$.

Earlier, we have shown that an SOR, say $\mathcal{M}^*$, must be contained in some NNSA candidate. Let $\mathcal{M}_1$ be that NNSA candidate that contain $\mathcal{M}^*$. If $\mathcal{M}_1$ has more nodes compared to $\mathcal{M}^*$, we can use the above method to "push" the extra nodes backwards in the route until they are the only nearest neighbors. Let $\mathcal{M}_2$ be this new route. We can show that $\mathcal{M}_2$ is also an NNSA candidate, which contains $\mathcal{M}^*$. We have shown that the rate supported by $\mathcal{M}_2$ cannot be lower than the SOR and hence it must achieve the optimal rate. Hence $\mathcal{M}^*$ is contained in an optimal NNSA candidate.

So, we have Theorem 7.

A.8 Proof of Theorem 12

Recall that when all nodes are constrained to transmitting independent Gaussian codewords, the reception rate at node $m_t \in \mathcal{M} \setminus \{1\}$ is

$$R_{m_t}(\mathcal{M}, p) = \frac{1}{2} \log \left(1 + \sum_{i=1}^{t-1} \gamma_{m_i m_t} \right). \tag{A.47}$$

where $p = p(x_1) \cdots p(x_{T-1})$ and $X_i \sim \mathcal{N}(0, P_i)$. The rate supported by route $\mathcal{M}$ is thus

$$R_{\mathcal{M}}(p) = \min_{m_t \in \mathcal{M} \setminus \{1\}} R_{m_t}(\mathcal{M}, p). \tag{A.48}$$

Consider an optimal route $\mathcal{M}_1 = \{m_1^*, m_2^*, \ldots, m_k^*, m_{k+1}^*, \ldots, m_{|\mathcal{M}|}^*\}$. Suppose that the first k nodes of the MSPA route are the same as this optimal route but the $(k+1)$-th node is different, i.e., the MSPA route is $\mathcal{M}_2 = \{m_1^*, m_2^*, \ldots, m_k^*, a, \ldots\}$ where $a \neq m_{k+1}^*$. The rest of the nodes in the MSPA might or might not be the same as the optimal route.

Since a is added to the route by MSPA, a necessary condition is $\sum_{i=1}^{k} \gamma_{m_i^* a} \geq \sum_{j=1}^{k} \gamma_{m_j^* m_{k+1}^*}$. So,

$$R_a(\mathcal{M}_2, p) \geq R_{m_{k+1}^*}(\mathcal{M}_1, p). \tag{A.49}$$

Now, we will change the $(k+1)$-th node in $\mathcal{M}_1$ by node a, and show that the rate of the new route is still optimal. There are two cases to be considered.

First, consider the case where $a \neq m_i^*, \forall i = k+2, \ldots, |\mathcal{M}|$. We add a to $\mathcal{M}_1$ and obtain $\mathcal{M}_3 = \{m_1^*, m_2^*, \ldots, m_k^*, a, m_{k+1}^*, \ldots, m_{|\mathcal{M}|}^*\}$. Then,

$$R_{m_i^*}(\mathcal{M}_3, p) = R_{m_i^*}(\mathcal{M}_1, p), \quad i = 2, \ldots, k \tag{A.50a}$$

$$R_a(\mathcal{M}_3, p) \geq R_{m_{k+1}^*}(\mathcal{M}_1, p) \tag{A.50b}$$

$$R_{m_i^*}(\mathcal{M}_3, p) > R_{m_i^*}(\mathcal{M}_1, p), \quad i = k+1, \ldots, |\mathcal{M}|. \tag{A.50c}$$

So, $R_{\mathcal{M}_3}(p) \geq R_{\mathcal{M}_1}(p)$.

Second, suppose $a = m_n^*$, for some $n \in \{k+2, \ldots, |\mathcal{M}|\}$. We change the position

of a and obtain $\mathcal{M}_4 = \{m_1^*, m_2^*, \ldots, m_k^*, a, m_{k+1}^*, \ldots, m_{n-1}^*, m_{n+1}^*, m_{|\mathcal{M}|}^*\}$. Then,

$$R_{m_i^*}(\mathcal{M}_4, p) = R_{m_i^*}(\mathcal{M}_1, p), \quad i = 2, \ldots, k, n+1, \ldots, |\mathcal{M}| \tag{A.51a}$$

$$R_a(\mathcal{M}_4, p) \geq R_{m_{k+1}^*}(\mathcal{M}_1, p) \tag{A.51b}$$

$$R_{m_i^*}(\mathcal{M}_4, p) > R_{m_i^*}(\mathcal{M}_1, p), \quad i = k+1, \ldots, n-1. \tag{A.51c}$$

So, $R_{\mathcal{M}_4}(p) \geq R_{\mathcal{M}_1}(p)$.

In summary, we choose an optimal route. Starting from the second node, we compare the optimal route with the MSPA route. If the i-th position in the optimal route is different from that in the MSPA route, we insert the i-th node in the MSPA route to the i-th position in the optimal route (and remove the node if it appears in the optimal route at another position). Now, the first i nodes in both route are the same. From above, we know that the rate of the modified optimal route does not change. We move on to the $(i+1)$-th position and so on until the destination. We can show that the MSPA route achieves the highest DF rate, i.e., that of an optimal route. So, we have Theorem 12.

Appendix B

Appendices to Chapter 4

B.1 An Example to Show that Myopic Coding is More Robust

To illustrate the robustness of myopic coding, we consider DF in the seven-node Gaussian MRC in which node 4 fails. This means the signal contributed by node 4 will stop. We consider the following scenarios in myopic and omniscient coding:

1. Two-hop myopic DF:

   (a) When the overall transmission rate is not affected: Node 2 decodes only from node 1, and cancels the interference only from itself (echo cancellation) and node 3. So, the failure of node 4 does not affect the decoding at node 2. Node 7 will also not be affected as it decodes only from nodes 5 and 6. In brief, the failure of node t only affects nodes $t-1, t+1$, and $t+2$ in two-hop myopic DF.

   (b) When the overall transmission rate is affected: Suppose that upon node 4's failure, the overall transmission rate is lowered due to the change in the reception rate of node 5. Additional re-configuration at the source is required. Now, the source will have to transmit at a lower rate. One way of doing this is to use the existing code, but pad the lower rate

messages with zeros. With zero-padding, the encoding and decoding at nodes 2 and 7 need not be changed as the supported rates at these nodes are not affected.

2. Omniscient DF: Nodes 2 and 3, who presume that node 4 is still transmitting and attempt to cancel its transmissions, will introduce more noise to their decoders. Nodes 5 to 7, who use node 4's signal contribution in the decoding, will experience a lower rSNR. Hence the supported rates at these nodes will be lowered.

Using omniscient DF, any topology change in the network (e.g., node failure or relocation) requires re-configuration of more nodes compared to using myopic coding.

B.2 Proof of Theorem 14

In this appendix, we describe the encoding and decoding schemes, and prove the achievable rates of two-hop myopic DF for the MRC. We consider the route $\mathcal{M} = \{1, 2, \ldots, T\}$. We consider $B + T - 2$ transmission blocks, each of n uses of the channel. A sequence of independent B indices, $w_b \in \{1, 2, \ldots, 2^{nR}\}$, $b = 1, 2, \ldots, B$ are sent over $n(B + T - 2)$ uses of the channel. As $B \to \infty$, the rate $RnB/n(B + T - 2) \to R$ for any n.

Note: We use w and z to represent the source message. The notation w_j denotes the information which the source outputs at the j-th block. This means the source emits $w_1, w_2, \ldots$ in blocks $1, 2, \ldots$ respectively. The notation z_t denotes the new information which node t transmits. Since each node transmits codewords derived from the last two decoded messages, node 2 always transmits (z_2, z_3). These different notations are used at different instances for better illustration.

B.2.1 Codebook Generation

In this section, we see how the codebook at each node is generated.

- First, fix the p.d.f.

$$p(u_1, u_2, \ldots, u_{T-1}, x_1, x_2, \ldots, x_{T-1}) = \\ p(u_1)p(u_2)\cdots p(u_{T-1})p(x_1|u_1, u_2)p(x_2|u_2, u_3)\cdots p(x_{T-1}|u_{T-1})$$

  for each $u_i \in \mathcal{U}_i$.

- For each $t \in \{1, \ldots, T-1\}$, generate 2^{nR} independent and identically distributed (i.i.d.) n-sequences in $\mathcal{U}_t^n$, each drawn according to $p(\mathbf{u}_t) = \prod_{i=1}^n p(u_{ti})$. Index them as $\mathbf{u}_t(z_t)$, $z_t \in \{1, \ldots, 2^{nR}\}$.

- Define $\mathbf{x}_{T-1}(z_{T-1}) = \mathbf{u}_{T-1}(z_{T-1})$.

- For each $t \in \{1, \ldots, T-2\}$, define a deterministic function that maps $(\mathbf{u}_t, \mathbf{u}_{t+1})$ to $\mathbf{x}_t$:

$$\mathbf{x}_t(z_t, z_{t+1}) = f_t\big(\mathbf{u}_t(z_t), \mathbf{u}_{t+1}(z_{t+1})\big). \tag{B.1}$$

- Repeat the above steps to generate a new independent codebook (Xie & Kumar, 2005). These two codebooks are used in alternate block of transmission. The reason for using two independent codebooks will be clear in the error probability analysis section.

We see that in each transmission block, node t, $t \in \{1, \ldots, T-2\}$, sends messages of two blocks: z_t (new data) and z_{t+1} (old data). In the same block, node $t+1$ sends messages z_{t+1} and z_{t+2}. Note that a node cooperates with the node in the next hop by repeating the transmission z_{t+1}. We will see this clearer in the next section.

B.2.2 Encoding

Fig. B.1 shows the encoding process for two-hop myopic DF. The encoding steps are as follows:

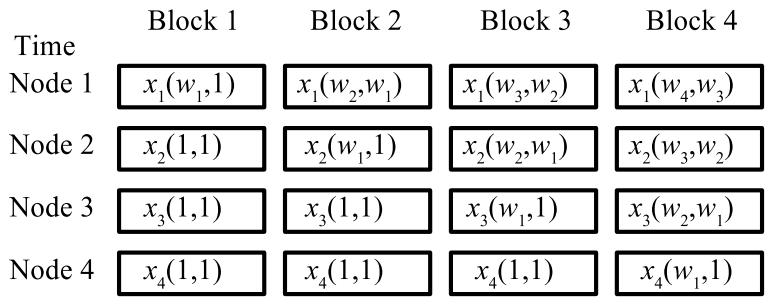


Figure B.1: The encoding scheme for two-hop myopic DF for the MRC.

- In the beginning of block 1, the source emits the first source letter w_1. Note that there is no new information after B blocks. We define $w_{B+1} = w_{B+2} = \cdots = w_{B+T-2} = 1$.

- In block 1, node 1 transmits $\mathbf{x}_1(w_1, w_0)$. Since the rest of the nodes have not received any information, they send dummy symbols $\mathbf{x}_i(w_{2-i}, w_{1-i})$, $i \in \{2, \ldots, T-1\}$. We define $w_b = 1$, for $b \leq 0$. In block 1, $z_1 = w_1, z_2 = w_0, \ldots$

- At the end of block 1, assume that node 2 correctly decodes the first signal w_1.

- In block 2, node 2 transmits $\mathbf{x}_2(w_1, w_0)$. Node 1 transmits $\mathbf{x}_1(w_2, w_1)$. It helps node 2 to re-transmit w_1 and sends w_2 (new information) at the same time. In block 2, $z_1 = w_2, z_2 = w_1, z_3 = w_0, \ldots$

- Generalizing, in block $b \in \{1, \ldots, B+T-2\}$, node t, $t \in \{1, \ldots, T-1\}$, has data $(w_1, w_2, \ldots, w_{b-t+1})$. Under two-hop myopic DF, it sends $\mathbf{x}_t(w_{b-t+1}, w_{b-t})$.

- We see that a node sends messages that it has decoded in the past two blocks. This adheres to the constraints of two-hop myopic DF.

B.2.3 Decoding

- Under the two-hop myopic DF constraints, a node can store a decoded message no longer than two blocks and can use two blocks of received signal to decode one message.

- Node 2's decoding is slightly different from the other nodes as there is only one upstream node. So it decodes every message using one block of received signal. We illustrate the decoding of message w_4 at node 2. At the end of block 4, assuming that node 2 has already decoded messages (w_1, w_2, w_3) correctly. However, due to the myopic coding constraint, it only has w_2 and w_3 in its memory. This is because w_1 was decoded at the end of block 1 and would have to be discarded at the end of block 3. So, it finds the a unique $\mathbf{u}_1(w_4)$ which is jointly typical with $\mathbf{u}_3(w_2), \mathbf{u}_2(w_3)$, and $\mathbf{y}_{2,4}$ (the received signal at node 2 in block 4). We write $\mathbf{y}_{2,4}$ instead of $\mathbf{y}_{24}$ to avoid the confusion with the received signal of node 24. An error is declared if there is no such w_4 or more than one unique w_4.

- Nodes 3 to T decode a message using two blocks of received signal. Consider node 3. At the end of block 4, assuming that node 3 has already decoded w_1 (decoded at the end of block 2) and w_2 (decoded at the end of block 3) correctly. Assume that it now correctly decodes w_3 using signals from blocks 3 and 4. At the end of block 4, it finds a set of $\mathbf{u}_1(w_4)$ which is jointly typical with $\mathbf{u}_4(w_1), \mathbf{u}_3(w_2), \mathbf{u}_2(w_3)$, and $\mathbf{y}_{3,4}$. We call this set $\mathcal{L}_1(w_4)$. Since it can only keep messages decoded over two blocks, it keeps w_2 and w_3 and discard w_1. At the end of block 5, node 3 finds a set of $\mathbf{u}_2(w_4)$ that is jointly typical with $\mathbf{u}_4(w_2), \mathbf{u}_3(w_3)$, and $\mathbf{y}_{3,5}$. We call this set $\mathcal{L}_2(w_4)$. It finds a unique w_4 that belong to both sets, that is $\hat{w}_4 \in \mathcal{L}_1(w_4) \cap \mathcal{L}_2(w_4)$. Here $\cap$ denotes intersection of sets. An error is declared when the intersection contains more than one index or the sets do not intersect.

- We now generalize the decoding process. Refer to Fig. B.2, at the end of block $b-1$, assuming that node t has correctly decoded $(w_1, \ldots, w_{b-t})$. Due to the myopic coding constraint, it has in its memory w_{b-t-1} and w_{b-t}. It decodes w_{b-t+1}. It then finds a set of $\mathbf{u}_{t-2}(w_{b-t+2})$ that is jointly typical with $(\mathbf{u}_{t-1}(w_{b-t+1}), \mathbf{u}_t(w_{b-t}), \mathbf{u}_{t+1}(w_{b-t-1}), \mathbf{y}_{t(b-1)})$. Call this set $\mathcal{L}_1(w_{b-t+2})$. It discards w_{b-t-1} from its memory. At the end of block b, it finds the set

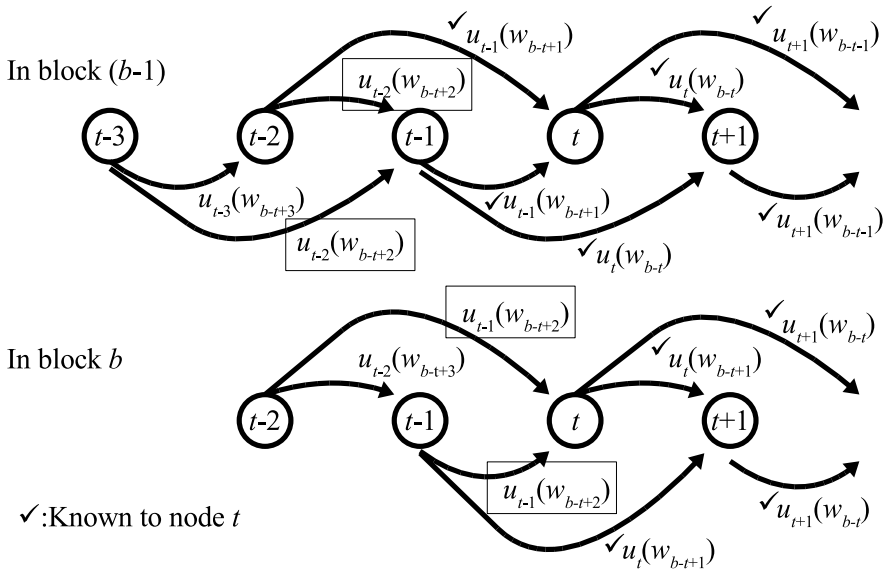


Figure B.2: Decoding at node t of message w_{b-t+2}.

of $\mathbf{u}_{t-1}(w_{b-t+2})$ that is jointly typical with $(\mathbf{u}_t(w_{b-t+1}), \mathbf{u}_{t+1}(w_{b-t}), \mathbf{y}_{tb})$. Call this set $\mathcal{L}_2(w_{b-t+2})$. It declare $\hat{w}_{b-t+2}$ if there is one and only one index in $\mathcal{L}_1(w_{b-t+2}) \cap \mathcal{L}_2(w_{b-t+2})$.

B.2.4 Achievable Rates and Probability of Error Analysis

In the previous section, we said that node t decodes message w_{b-t+2} in block b. We denote the event that no decoding error is made at all nodes in the first b block, $1 \leq b \leq B + T - 2$, by

$$\mathcal{C}(b) \triangleq \{\hat{w}_{t(k-t+2)} = w_{k-t+2} : \forall t \in [2, T] \text{ and } k \in [1, b]\} \tag{B.2}$$

where $\hat{w}_{t(b)}$ is node t's estimate of the message w_b. This means in the first b blocks, node 2 will have correctly decoded $(w_1, w_2, \ldots, w_b)$, node 3 will have correctly decoded $(w_0, w_1, \ldots, w_{b-1})$, and so on. We set $w_k = 1$ for $k \leq 0$. They are the dummy signals sent by the nodes.

We denote the probability that there is no decoding error up to block b as

$$P_c(b) \triangleq \Pr\{\mathcal{C}(b)\} \tag{B.3}$$

and $P_c(0) \triangleq 1$. The probability that one or more error occurs during block $b \in [1, B+T-2]$ at some node $t \in [2, T]$, given that there is no error in decoding at all nodes in all blocks up to $b-1$, is

$$\begin{aligned} P_e(b) &\triangleq \Pr\left\{\hat{w}_{t(b-t+2)} \neq w_{b-t+2} : \text{ for some } t \in \{2, \ldots, T\} \middle| \mathcal{C}(b-1)\right\} \\ &\leq \sum_{t=2}^{T} \Pr\left\{\hat{w}_{t(b-t+2)} \neq w_{b-t+2} | \mathcal{C}(b-1)\right\} \quad \text{(B.4a)} \\ &\triangleq \sum_{t=2}^{T} P_{et}(b) \quad \text{(B.4b)} \end{aligned}$$

where $P_{et}(b) \triangleq \Pr\left\{\hat{w}_{t(b-t+2)} \neq w_{b-t+2} | \mathcal{C}(b-1)\right\}$, which is the probability that node t wrongly decodes the latest letter w_{b-t+2} in block b, given that it has correctly decoded the past letters.

Now, we need to compute the error probability $P_{et}(b)$. As mentioned in the decoding section, the decoding of a message spans over two blocks. For example, let us look at the decoding of message w_{b-t+2} at node t, as depicted in Fig. B.2. The message to be decoded is boxed and the messages that node t has correctly decoded are marked with ✓. In block $b-1$, node t find a set of w_{b-t+2} for which

$$\begin{aligned} \Big(\mathbf{u}_{t-2}(w_{b-t+2}), \mathbf{u}_{t-1}(w_{b-t+1}), \mathbf{u}_t(w_{b-t}), \mathbf{u}_{t+1}(w_{b-t-1}), \mathbf{y}_{t(b-1)}\Big) \in \\ \mathcal{A}_\epsilon^n(U_{t-2}, U_{t-1}, U_t, U_{t+1}, Y_t) \triangleq \mathcal{A}_1. \quad \text{(B.5)} \end{aligned}$$

In block b, node t finds a set of w_{b-t+2} for which

$$(\mathbf{u}_{t-1}(w_{b-t+2}), \mathbf{u}_t(w_{b-t+1}), \mathbf{u}_{t+1}(w_{b-t}), \mathbf{y}_{tb}) \in \mathcal{A}_\epsilon^n(U_{t-1}, U_t, U_{t+1}, Y_t) \triangleq \mathcal{A}_2. \quad \text{(B.6)}$$

Node t then finds the intersection of the two sets to determine the value of w_{b-t+2}.

Assuming that node t has correctly decoded w_{b-t-1}, w_{b-t}, and w_{b-t+1}, we define

the following error events:

$$\mathcal{E}_1 \triangleq \left(\mathbf{u}_{t-2}(w_{b-t+2}), \mathbf{u}_{t-1}(w_{b-t+1}), \mathbf{u}_t(w_{b-t}), \mathbf{u}_{t+1}(w_{b-t-1}), \mathbf{y}_{t(b-1)}\right) \notin \mathcal{A}_1 \tag{B.7a}$$

$$\mathcal{E}_2 \triangleq \left(\mathbf{u}_{t-2}(v), \mathbf{u}_{t-1}(w_{b-t+1}), \mathbf{u}_t(w_{b-t}), \mathbf{u}_{t+1}(w_{b-t-1}), \mathbf{y}_{t(b-1)}\right) \in \mathcal{A}_1 \tag{B.7b}$$

$$\mathcal{E}_3 \triangleq \left(\mathbf{u}_{t-1}(w_{b-t+2}), \mathbf{u}_t(w_{b-t+1}), \mathbf{u}_{t+1}(w_{b-t}), \mathbf{y}_{tb}\right) \notin \mathcal{A}_2 \tag{B.7c}$$

$$\mathcal{E}_4 \triangleq \left(\mathbf{u}_{t-1}(v), \mathbf{u}_t(w_{b-t+1}), \mathbf{u}_{t+1}(w_{b-t}), \mathbf{y}_{tb}\right) \in \mathcal{A}_2 \tag{B.7d}$$

for some $v \in \left\{v \in [1, \ldots, 2^{nR}] : v \neq w_{b-t+2}\right\}$, and

$$\mathcal{E}_5 \triangleq \mathcal{E}_2 \cap \mathcal{E}_4. \tag{B.8}$$

$\mathcal{E}_5$ is the event where $v \neq w_{b-t+2}$ is found in the intersection of the decoding sets and is, therefore, wrongly decoded as the transmitted message. An error occurs during the decoding in block b at node t if events $\mathcal{E}_1$, $\mathcal{E}_3$, or $\mathcal{E}_5$ occurs. Now, we can rewrite

$$P_{et}(b) = \Pr\{\mathcal{E}_1 \cup \mathcal{E}_3 \cup \mathcal{E}_5\} \leq \Pr\{\mathcal{E}_1\} + \Pr\{\mathcal{E}_3\} + \Pr\{\mathcal{E}_5\}. \tag{B.9}$$

The last equation is due to the union bound of events.

From the definition of jointly typical sequences (see Definition 4), we know that

$$\Pr\{\mathcal{E}_1\} \leq \epsilon \tag{B.10a}$$

$$\Pr\{\mathcal{E}_3\} \leq \epsilon, \tag{B.10b}$$

for sufficiently large n.

Using Lemma 1, we derive the probability of a particular $v \neq w_{b-t+2}$ that

satisfies (B.7b):

$$\Pr\left\{(\mathbf{u}_{t-2}(v), \mathbf{u}_{t-1}(w_{b-t+1}), \mathbf{u}_t(w_{b-t}), \mathbf{u}_{t+1}(w_{b-t-1}), \mathbf{y}_{t(b-1)}) \in \mathcal{A}_1\right\}$$

$$= \sum_{(\mathbf{u}_{t-2},\mathbf{u}_{t-1},\mathbf{u}_t,\mathbf{u}_{t+1},\mathbf{y}_t)\in\mathcal{A}_1} p(\mathbf{u}_{t-2})p(\mathbf{u}_{t-1},\mathbf{u}_t,\mathbf{u}_{t+1},\mathbf{y}_t) \tag{B.11a}$$

$$\leq |\mathcal{A}_1| 2^{-n(H(U_{t-2})-\epsilon)} 2^{-n(H(U_{t-1},U_t,U_{t+1},Y_t)-\epsilon)} \tag{B.11b}$$

$$\leq 2^{n(H(U_{t-2},U_{t-1},U_t,U_{t+1},Y_t)+\epsilon)} 2^{-n(H(U_{t-2})-\epsilon)} \times$$

$$2^{-n(H(U_{t-1},U_t,U_{t+1},Y_t)-\epsilon)} \tag{B.11c}$$

$$= 2^{-n(H(U_{t-2})-H(U_{t-2}|Y_t,U_{t-1},U_t,U_{t+1})-3\epsilon)} \tag{B.11d}$$

$$\leq 2^{-n(I(U_{t-2};Y_t|U_{t-1},U_t,U_{t+1})-3\epsilon)}. \tag{B.11e}$$

The last equation is because $H(U_{t-2}) \geq H(U_{t-2}|U_{t-1}, U_t, U_{t+1})$.

By a similar method, we can calculate the probability of a particular $v \in \{v \in \{1, \ldots, 2^{nR}\} : v \neq w_{b-t+2}\}$ satisfies (B.7d):

$$\Pr\left\{(\mathbf{u}_{t-1}(v_2), \mathbf{u}_t(w_{b-t+1}), \mathbf{u}_{t+1}(w_{b-t}), \mathbf{y}_{tb}) \in \mathcal{A}_2\right\} \leq 2^{-n(I(U_{t-1};Y_t|U_t,U_{t+1})-3\epsilon)}. \tag{B.12}$$

Combining these two probabilities, we find the probability that node t wrongly decodes w_{b-t+2} to any $v \in \{v \in \{1, \ldots, 2^{nR}] : v \neq w_{b-t+2}\}$ to be

$$\Pr\{\mathcal{E}_5\} = \sum_{\substack{v\in\{1,\ldots,2^{nR}\}\\ v\neq w_{b-t+2}}} \Pr\{v \text{ satisfies (B.8)}\} \tag{B.13a}$$

$$= \sum_{\substack{v\in\{1,\ldots,2^{nR}\}\\ v\neq w_{b-t+2}}} \Pr\{v \text{ satisfies (B.7b)}\} \Pr\{v \text{ satisfies (B.7d)}\} \tag{B.13b}$$

$$\leq \left(2^{nR} - 1\right) \times 2^{-n(I(U_{t-2};Y_t|U_{t-1},U_t,U_{t+1})-3\epsilon)} \times 2^{-n(I(U_{t-1};Y_t|U_t,U_{t+1})-3\epsilon)} \tag{B.13c}$$

$$< 2^{-n(I(U_{t-2},U_{t-1};Y_t|U_t,U_{t+1})-6\epsilon-R)} \tag{B.13d}$$

$$\leq \epsilon. \tag{B.13e}$$

Here, (B.13b) is due to the use of independent codebooks for each alternating block.

The last equation is made possible for sufficiently large n and if

$$R < I(U_{t-2}, U_{t-1}; Y_t | U_t, U_{t+1}) - 6\epsilon. \tag{B.14}$$

With this rate constraint and large n, we see that the probability of error is

$$\begin{aligned} P_e(b) &= \sum_{t=2}^{T} P_{et}(b) && \text{(B.15a)} \\ &\leq \sum_{t=2}^{T} [\Pr\{\mathcal{E}_1\} + \Pr\{\mathcal{E}_3\} + \Pr\{\mathcal{E}_5\}] && \text{(B.15b)} \\ &\leq (T-1)3\epsilon, && \text{(B.15c)} \end{aligned}$$

which can be made arbitrarily small. Hence, the rate in (B.14) is achievable.

Equation (B.14) is only the rate constraint at one node. In two-hop myopic DF, each message must be fully decoded at each node, hence the overall rate is constrained by

$$R_{\text{2-hop}} = \min_{t \in \{2,\ldots,T\}} R_t \tag{B.16}$$

where

$$R_t = I(U_{t-2}, U_{t-1}; Y_t | U_t, U_{t+1}) \tag{B.17}$$

and $U_0 = U_T = U_{T+1} = 0$. Hence we arrive at Theorem 14.

B.3 Achievable Rates of Myopic DF for the Gaussian MRC

B.3.1 One-Hop Myopic DF

In one-hop myopic DF, node t transmits only to node $t+1$. Let us first consider node 1. It sends X_1 to node 2. Node 2 receives

$$Y_2 = \sqrt{\kappa d_{12}^{-\eta}} X_1 + \sqrt{\kappa d_{32}^{-\eta}} X_3 + \sqrt{\kappa d_{42}^{-\eta}} X_4 + Z_2. \tag{B.18}$$

Node 2 decodes new messages from node 1's transmission. From (4.2), the reception rate at node 2 is

$$\begin{aligned} R_2 &= I(X_1; Y_2|X_2) \\ &= \frac{1}{2}\log 2\pi e\left[\kappa d_{12}^{-\eta}P_1 + \kappa d_{23}^{-\eta}P_3 + \kappa d_{24}^{-\eta}P_4 + N_2\right] \\ &\quad - \frac{1}{2}\log 2\pi e\left[\kappa d_{23}^{-\eta}P_3 + \kappa d_{24}^{-\eta}P_4 + N_2\right] \qquad \text{(B.19a)} \\ &= \frac{1}{2}\log\left[1 + \frac{d_{12}^{-2}P_1}{1 + d_{23}^{-2}P_3 + d_{24}^{-2}P_4}\right]. \qquad \text{(B.19b)} \end{aligned}$$

Here, we have substituted $\kappa = 1$, $\eta = 2$, and $N_2 = 1$W. The reception rates at nodes 3, 4, and 5 can be computed in similar way. One-hop myopic DF achieves rates up to

$$R_{\text{1-hop}} = \min_{t\in\{2,3,4,5\}} R_t. \qquad \text{(B.20)}$$

B.3.2 Two-Hop Myopic DF

In two-hop myopic DF, node $t, t = 1, 2, 3$, allocates α_t of its power to transmit to node $t+2$ and $(1-\alpha_t)$ of its power to node $t+1$. Since there is only one node in front of node 4, it transmits only to node 5. The transmission by each node is listed as follows:

- Node 4 sends $X_4 = \sqrt{P_4}U_4$.
- Node 3 sends $X_3 = \sqrt{\alpha_3 P_3}U_4 + \sqrt{(1-\alpha_3)P_3}U_3$.
- Node 2 sends $X_2 = \sqrt{\alpha_2 P_2}U_3 + \sqrt{(1-\alpha_2)P_2}U_2$.
- Node 1 sends $X_1 = \sqrt{\alpha_1 P_1}U_2 + \sqrt{(1-\alpha_1)P_1}U_1$.

Here, $U_i, i = 1, 2, 3, 4$ are independent Gaussian random variables, each with unit variance, and $0 \le \alpha_1, \alpha_2, \alpha_3 \le 1$.

From (B.17), for fixed $\{\alpha_1, \alpha_2, \alpha_3\}$, the reception rate at node 2 is

$$R_2 = I(U_1; Y_2|U_2, U_3) \tag{B.21a}$$

$$= \frac{1}{2}\log 2\pi e\left[\kappa d_{12}^{-\eta}(1-\alpha_1)P_1 + \left(\sqrt{\kappa d_{23}^{-\eta}\alpha_3 P_3} + \sqrt{\kappa d_{24}^{-\eta} P_4}\right)^2 + N_2\right]$$

$$- \frac{1}{2}\log 2\pi e\left[\left(\sqrt{\kappa d_{23}^{-\eta}\alpha_3 P_3} + \sqrt{\kappa d_{24}^{-\eta} P_4}\right)^2 + N_2\right] \tag{B.21b}$$

$$= \frac{1}{2}\log\left[1 + \frac{d_{12}^{-2}(1-\alpha_1)P_1}{1 + \left(\sqrt{d_{23}^{-2}\alpha_3 P_3} + \sqrt{d_{24}^{-2} P_4}\right)^2}\right]. \tag{B.21c}$$

Here, we have substituted $\kappa = 1$, $\eta = 2$, and $N_2 = 1$W. The reception rates at nodes 3, 4, and 5 can be computed in a similar way.

Minimizing over all reception rates and maximizing over all possible power splits on route $\mathcal{M}$, two-hop myopic DF achieves rates up to

$$R_{\text{2-hop}} = \max_{\{\alpha_1,\alpha_2,\alpha_3\}} \min_{t\in\{2,3,4,5\}} R_t. \tag{B.22}$$

B.4 Proof of Theorem 15

Now, we prove Theorem 15. We start by describing the codebook generation. We send B blocks of information over $B + T - 2$ blocks of channel use.

B.4.1 Codebook Generation

The codebook generation for k-hop myopic DF for the MRC is as follows.

- Fix the p.d.f.

$$p(u_1, u_2, \ldots, u_{T-1}, x_1, x_2, \ldots, x_{T-1})$$

$$= p(u_1)p(u_2)\cdots p(u_{T-1})p(x_{T-1}|u_{T-1})$$

$$\times p(x_{T-2}|u_{T-2}, u_{T-1})\cdots \times p(x_{T-k}|u_{T-k}, u_{T-k+1}\ldots, u_{T-1})$$

$$\times p(x_{T-k-1}|u_{T-k-1}, u_{T-k}\ldots, u_{T-2})\cdots \times p(x_1|u_1, u_2, \ldots, u_k). \tag{B.23a}$$

- For each $t \in \{1, \ldots, T-1\}$, generate 2^{nR} independent and identically distributed (i.i.d.) n-sequences in $\mathcal{U}_t^n$, each drawn according to $p(\mathbf{u}_t) = \prod_{i=1}^{n} p(u_{ti})$. Index them as $\mathbf{u}_t(z_t)$, $z_t \in \{1, \ldots, 2^{nR}\}$.

- Define $\mathbf{x}_{T-1}(z_{T-1}) = \mathbf{u}_{T-1}(z_{T-1})$.

- For each $t \in [T-k, T-2]$, define a deterministic function that maps $(\mathbf{u}_t, \mathbf{u}_{t+1}, \ldots, \mathbf{u}_{T-1})$ to $\mathbf{x}_t$:

$$\mathbf{x}_t(z_t, z_{t+1}, \ldots, z_{T-1}) = f_t\big(\mathbf{u}_t(z_t), \mathbf{u}_{t+1}(z_{t+1}), \ldots, \mathbf{u}_{T-1}(z_{T-1})\big). \tag{B.24}$$

- For each $t \in [1, T-k-1]$, define a deterministic function that maps $(\mathbf{u}_t, \mathbf{u}_{t+1}, \ldots, \mathbf{u}_{t+k-1})$ to $\mathbf{x}_t$:

$$\mathbf{x}_t(z_t, z_{t+1}, \ldots, z_{t+k-1}) = f_t\big(\mathbf{u}_t(z_t), \mathbf{u}_{t+1}(z_{t+1}), \ldots, \mathbf{u}_{t+k-1}(z_{t+k-1})\big). \tag{B.25}$$

- Repeat the above steps to generate $k-1$ new independent sets of codebook. These k codebooks are used in cycle and reused after k blocks of n transmissions.

For the sake of illustration, we denote the code of node t, $t \in \{1, \ldots, T-1\}$ by $\mathbf{x}_t(z_t, z_{t+1}, \ldots, z_{t+k-1})$ where $z_j = 1$ for $j \geq T$. These are dummy symbols that do not affect the encoding process.

B.4.2 Encoding

We now describe the encoding process for k-hop myopic DF. It is depicted in Fig. B.3.

- In the beginning of block 1, the information source emits the first source letter w_1. Note that there is no new information in blocks b for $B+1 \leq b \leq B+T-2$. We assume that $w_{B+1} = w_{B+2} = \cdots = w_{B+T-2} = 1$.

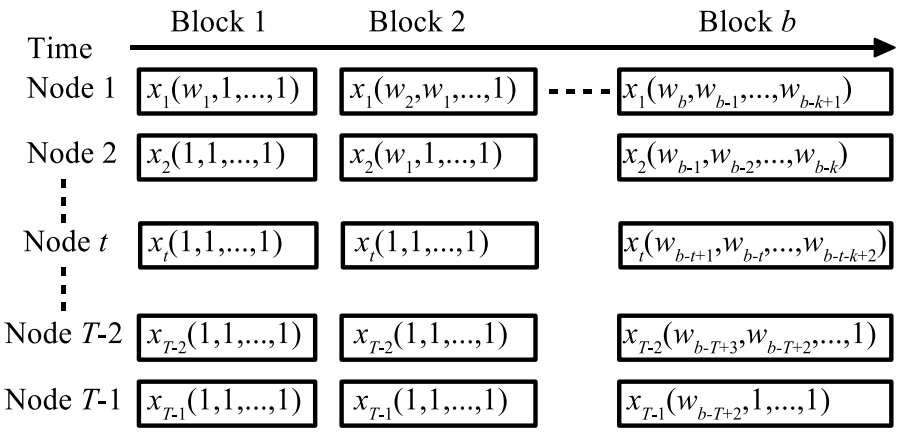


Figure B.3: The encoding scheme for k-hop myopic DF.

- In block 1, node 1 transmits $\mathbf{x}_1(w_1, w_0, \ldots, w_{2-k})$. Since the rest of the nodes have not received any information, they send dummy symbols $\mathbf{x}_i(w_{2-i}, w_{1-i}, \ldots, w_{3-k-i})$, $i \in \{2, \ldots, T-1\}$. We define $w_b = 1$, for $b \leq 0$.

- At the end of block $b-1$, $b \geq 2$, we assume that node t has correctly decoded messages up to w_{b-t+1}. Under the k-hop myopic constraints, a node can encode with at most k previously decoded messages in each block of transmission. So, in block b, node t encode $\min\{k, T-t\}$ previously decoded messages, i.e., it sends $\mathbf{x}_t(w_{b-t+1}, w_{b-t}, \ldots, w_{b-t-k+2})$. We note that there are only $T-t$ nodes in front of node t. For the case of $T-t < k$, node t sends $\mathbf{x}_t(w_{b-t+1}, w_{b-t}, \ldots, w_{b-T+2}, 1, \ldots, 1)$. This means, it sets $w_i = 1$ for $i \geq b-T+1$, which is equivalent to sending dummy symbols. This is because at the end of block $b-1$, node T will have already correctly decoded signals up to w_{b-T+1}. As this is the last node in the network, all other nodes will have had decoded those signals. Hence no node needs to transmit $w_i = 1$ for $i \geq b-T+1$ again. The dummy symbols are included so that the same transmit notation can be used for all the nodes.

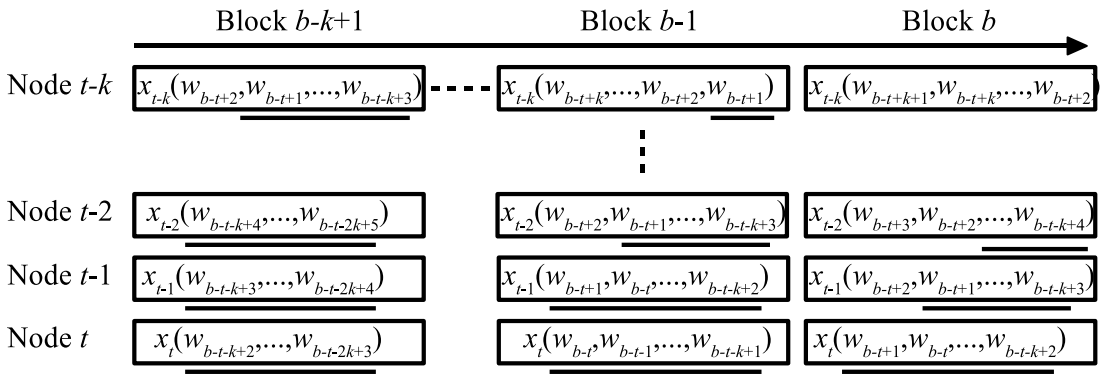


Figure B.4: The decoding scheme for k-hop myopic DF. Underlined symbols are those that has been decoded by node t prior to block b.

B.4.3 Decoding and Achievable Rates

We look at how node t, for $t \geq k+1$, decodes w_{b-t+2} at the end of block b. Fig. B.4 shows what the nodes transmit.

- During block b, there are k nodes that encode w_{b-t+2} in their transmission. These are nodes $\{t-k, \ldots, t-1\}$. Nodes $\{1, \ldots, t-k-1\}$ do not encode w_{b-t+2} in their transmission in block b as they have to discard the message due to the buffering constraint of k-hop myopic DF.

- At the end of block b, node t finds $\mathcal{L}_1(\hat{w}_{b-t+2})$ in which

$$\Big(\mathbf{u}_{t-1}(\hat{w}_{b-t+2}), \mathbf{u}_t(w_{b-t+1}), \ldots, \mathbf{u}_{t+k-1}(w_{b-t-k+2}), \mathbf{y}_{tb}\Big) \in \mathcal{A}_\epsilon^n. \tag{B.26}$$

  Here, we note that node t can store k old messages. Hence, during the decoding at the end of block b, it knows $(\mathbf{u}_t(w_{b-t+1}), \ldots, \mathbf{u}_{t+k-1}(w_{b-t-k+2}))$. The rate contribution from (B.26) is

$$R_t^{(1)} = I(U_{t-1}; Y_t | U_t, \ldots, U_{t+k-1}). \tag{B.27}$$

- Moving back one block, at the end block $b-1$, node t has messages $\big(\mathbf{u}_t(w_{b-t}), \ldots, \mathbf{u}_{t+k-1}(w_{b-t-k+1})\big)$ in its storage. After decoding $\mathbf{u}_{t-1}(w_{b-t+1})$, it then

forms the set $\mathcal{L}_2(\hat{w}_{b-t+2})$ which

$$\left(\mathbf{u}_{t-2}(\hat{w}_{b-t+2}), \mathbf{u}_{t-1}(w_{b-t+1}), \ldots, \mathbf{u}_{t+k-1}(w_{b-t-k+1}), \mathbf{y}_{t(b-1)}\right) \in \mathcal{A}_\epsilon^n. \quad \text{(B.28)}$$

The rate contribution from this is

$$R_t^{(2)} = I(U_{t-2}; Y_t | U_{t-1}, \ldots, U_{t+k-1}). \quad \text{(B.29)}$$

- Repeating this for blocks $(b-i+1)$, $3 \leq i \leq k$, node t find the set $\mathcal{L}_i(\hat{w}_{b-t+2})$, and the rate contribution is

$$R_t^{(i)} = I(U_{t-i}; Y_t | U_{t-i+1}, \ldots, U_{t+k-1}). \quad \text{(B.30)}$$

  The proof is similar to that for two-hop myopic DF and will be omitted here.

- Finally, node t finds $\hat{w}_{b-t+2} \in \bigcap_{i=1}^{k} \mathcal{L}_i(\hat{w}_{b-t+2})$, where $\bigcap$ denotes the intersection of sets. A unique $\hat{w}_{b-t+2}$ can be found if the reception rate at node t is not more than

$$R_t = \sum_{i=1}^{k} R_t^{(i)} = I(U_{t-k}, \ldots, U_{t-1}; Y_t | U_t, \ldots, U_{t+k-1}). \quad \text{(B.31)}$$

- Since all data must pass through every node, the overall rate is constrained by the node which has the lowest reception rate, that is

$$R_{k\text{-hop}} = \min_{t \in \{2,\ldots,T\}} R_t. \quad \text{(B.32)}$$

With this, we have Theorem 15.

Appendix C

Appendices to Chapter 5

C.1 Proof of Theorem 18

Given any code $\left\{\{f_{1i}, f_{2i}\}_{i=1}^{n}, g\right\}$ for the MACFCS, the p.d.f. on the joint ensemble $W_1^n, W_2^n, X_1^n, X_2^n, Y_1^n, Y_2^n, Y_3^n$ is given by

$$\begin{aligned} p(w_1^n, w_2^n, x_1^n, x_2^n, y_1^n, y_2^n, y_3^n) &= \prod_{i=1}^{n} p(w_{1i}, w_{2i}) \\ &\times \prod_{i=1}^{n} p(x_{1i}|w_1^n, y_{11}, y_{12}, \ldots, y_{1i-1}) \\ &\cdot p(x_{2i}|w_1^n, y_{21}, y_{22}, \ldots, y_{2i-1}) \cdot p^*(y_{1i}, y_{2i}, y_{3i}|x_{1i}, x_{2i}). \end{aligned} \tag{C.1a}$$

By Fano's inequality (Cover & Thomas, 1991),

$$H(W_1^n, W_2^n|Y_3^n) \leq n \log_2 |\mathcal{W}_1 \times \mathcal{W}_2| P_e + 1 \triangleq n\delta_n. \tag{C.2}$$

Now, we consider $H(W_1|W_2)$.

$$\begin{aligned} nH(W_1|W_2) &= H(W_1^n|W_2^n) && \text{(C.3a)} \\ &= I(W_1^n; Y_3^n|W_2^n) + H(W_1^n|Y_3^n, W_2^n) && \text{(C.3b)} \\ &\leq I(W_1^n; Y_3^n|W_2^n) + H(W_1^n, W_2^n|Y_3^n) && \text{(C.3c)} \\ &\leq I(W_1^n; Y_3^n|W_2^n) + n\delta_n. && \text{(C.3d)} \end{aligned}$$

Now,

$$
\begin{aligned}
& I(W_1^n; Y_3^n | W_2^n) \\
&= H(Y_3^n | W_2^n) - H(Y_3^n | W_1^n, W_2^n) && \text{(C.4a)} \\
&= \sum_{i=1}^{n} \left[H(Y_{3i} | W_2^n, Y_3^{i-1}) - H(Y_{3i} | W_1^n, W_2^n, Y_3^{i-1}) \right] && \text{(C.4b)} \\
&\leq \sum_{i=1}^{n} \left[H(Y_{3i}) - H(Y_{3i} | X_{1i}, X_{2i}, W_1^n, W_2^n, Y_3^{i-1}) \right] && \text{(C.4c)} \\
&= \sum_{i=1}^{n} \left[H(Y_{3i}) - H(Y_{3i} | X_{1i}, X_{2i}) \right] && \text{(C.4d)} \\
&= \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_{3i}), && \text{(C.4e)}
\end{aligned}
$$

where

- (C.4b) and (C.4e) are by the chain rule.
- (C.4c) is because conditioning reduces entropy.
- (C.4d) is because of the memoryless channel, Y_{3i} and $(Y_3^{i-1}, W_1^n, W_2^n)$ are independent given (X_{1i}, X_{2i}).

Also,

$$
\begin{aligned}
& I(W_1^n; Y_3^n | W_2^n) \\
& \leq I(W_1^n; Y_2^n, Y_3^n | W_2^n) && \text{(C.5a)} \\
& = \sum_{i=1}^{n} I(W_1^n; Y_{2i}, Y_{3i} | W_2^n, Y_2^{i-1}, Y_3^{i-1}) && \text{(C.5b)} \\
& = \sum_{i=1}^{n} I(Y_{2i}, Y_{3i}; W_1^n | X_{2i}, W_2^n, Y_2^{i-1}, Y_3^{i-1}) && \text{(C.5c)} \\
& = \sum_{i=1}^{n} \left[H(Y_{2i}, Y_{3i} | X_{2i}, W_2^n, Y_2^{i-1}, Y_3^{i-1}) - H(Y_{2i}, Y_{3i} | X_{2i}, W_1^n, W_2^n, Y_2^{i-1}, Y_3^{i-1}) \right] \\
& && \text{(C.5d)} \\
& \leq \sum_{i=1}^{n} \left[H(Y_{2i}, Y_{3i} | X_{2i}) - H(Y_{2i}, Y_{3i} | X_{1i}, X_{2i}, W_1^n, W_2^n, Y_2^{i-1}, Y_3^{i-1}) \right] && \text{(C.5e)} \\
& = \sum_{i=1}^{n} \left[H(Y_{2i}, Y_{3i} | X_{2i}) - H(Y_{2i}, Y_{3i} | X_{1i}, X_{2i}) \right] && \text{(C.5f)} \\
& = \sum_{i=1}^{n} I(X_{1i}; Y_{2i}, Y_{3i} | X_{2i}), && \text{(C.5g)}
\end{aligned}
$$

where

- (C.5b) is by the chain rule and the memoryless nature of the channel.
- (C.5c) is because X_{2i} is a function of (W_2^n, Y_2^{i-1}).
- (C.5e) is because conditioning reduces entropy.
- (C.5f) is because (Y_{2i}, Y_{3i}) and $(W_1^n, W_2^n, Y_2^{i-1}, Y_3^{i-1})$ are independent given (X_{1i}, X_{2i}).

In addition,

$$
\begin{aligned}
n H(W_1, W_2) &= H(W_1^n, W_2^n) && \text{(C.6a)} \\
&= I(W_1^n, W_2^n; Y_3^n) + H(W_1^n, W_2^n | Y_3^n) && \text{(C.6b)} \\
&\leq I(W_1^n, W_2^n; Y_3^n) + n\delta_n. && \text{(C.6c)}
\end{aligned}
$$

Consider $I(W_1^n, W_2^n; Y_3^n)$,

$$
\begin{aligned}
& I(W_1^n, W_2^n; Y_3^n) \\
&= \sum_{i=1}^{n} I(W_1^n, W_2^n; Y_{3i} | Y_3^{i-1}) && \text{(C.7a)} \\
&= \sum_{i=1}^{n} \left[H(Y_{3i} | Y_3^{i-1}) - H(Y_{3i} | W_1^n, W_2^n, Y_3^{i-1}) \right] && \text{(C.7b)} \\
&\leq \sum_{i=1}^{n} \left[H(Y_{3i}) - H(Y_{3i} | X_{1i}, X_{2i}, W_1^n, W_2^n, Y_3^{i-1}) \right] && \text{(C.7c)} \\
&= \sum_{i=1}^{n} \left[H(Y_{3i}) - H(Y_{3i} | X_{1i}, X_{2i}) \right] && \text{(C.7d)} \\
&= \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_{3i}). && \text{(C.7e)}
\end{aligned}
$$

Now we introduce a new variable Q independent of $W_1^n, W_2^n, X_1, X_2, Y_1, Y_2, Y_3$ (Cover & El Gamal, 1979) that takes values in the set $\{1, 2, \ldots, n\}$ with probability

$$
\Pr\{Q = i\} = \frac{1}{n}, \quad i \in \{1, 2, \ldots, n\}, \tag{C.8}
$$

and such that

$$
Q \to (W_1^n, W_2^n) \to (X_{1Q}, X_{2Q}) \to (Y_{1Q}, Y_{2Q}, Y_{3Q}) \tag{C.9}
$$

forms a Markov chain. We set

$$
X_1 \triangleq X_{1Q}, X_2 \triangleq X_{2Q}, Y_1 \triangleq Y_{1Q}, Y_2 \triangleq Y_{2Q}, Y_3 \triangleq Y_{3Q}. \tag{C.10}
$$

Now, (C.4e) becomes

$$
\begin{aligned}
\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_{3i}) &= \sum_{i=1}^{n} I(X_{1Q}, X_{2Q}; Y_{3Q} |, Q = i) && \text{(C.11a)} \\
&= nI(X_1, X_2; Y_3 | Q) && \text{(C.11b)} \\
&\leq nI(X_1, X_2; Y_3). && \text{(C.11c)}
\end{aligned}
$$

Similarly,

$$\sum_{i=1}^{n} I(X_{1i}; Y_{2i}, Y_{3i}|X_{2i}) \leq nI(X_1; Y_2, Y_3|X_2). \tag{C.12}$$

Taking the limit as $n \to \infty$ and $P_e \to 0$, and combining (C.3d), (C.4e), (C.5g), (C.11c), and (C.12), we have

$$H(W_1|W_2) \leq \min\{I(X_1, X_2; Y_3), I(X_1; Y_2, Y_3|X_2)\}. \tag{C.13}$$

By symmetry we can show that

$$H(W_2|W_1) \leq \min\{I(X_2, X_1; Y_3), I(X_2; Y_1, Y_3|X_1)\}. \tag{C.14}$$

Combining (C.6c), (C.7e), and (C.11c), we have

$$H(W_1, W_2) \leq I(X_1, X_2; Y_3). \tag{C.15}$$

Equation (C.15) guarantees

$$H(W_1|W_2) \leq I(X_1, X_2; Y_3) \tag{C.16a}$$

$$H(W_2|W_1) \leq I(X_1, X_2; Y_3). \tag{C.16b}$$

Hence, we have Theorem 18.

C.2 Proof of Theorem 19

In this section, we prove Theorem 19. We calculate the error probabilities and show that they diminish when certain conditions are satisfied.

In each block, node 1 encodes $\mathbf{w}_1$ to j. Similarly, node 2 encodes $\mathbf{w}_2$ to k. Assuming noiseless channel, node 1 receives k correctly and node 2 receives j

correctly. We define the following source coding error events.

$$\mathcal{E}_{0a} \triangleq \{\text{node 2 wrongly decodes } \mathbf{w}_1\}, \tag{C.17a}$$

$$\mathcal{E}_{0b} \triangleq \{\text{node 1 wrongly decodes } \mathbf{w}_2\}, \tag{C.17b}$$

$$\mathcal{E}_0 \triangleq \mathcal{E}_{0a} \cup \mathcal{E}_{0b}. \tag{C.17c}$$

Using the results by Slepian & Wolf (1973b, Theorem 2), $\Pr(\mathcal{E}_{0a})$ and $\Pr(\mathcal{E}_{0b})$ can be bounded by ϵ if j is encoded with no less than $n[H(W_1|W_2)+\epsilon]$ bits and k is encoded in no less than $n[H(W_2|W_1)+\epsilon]$ bits. Hence

$$\Pr(\mathcal{E}_0) \leq \Pr(\mathcal{E}_{0a}) + \Pr(\mathcal{E}_{0b}) < 2\epsilon. \tag{C.18}$$

Now, both sources have $(j, k, \mathbf{w}_1, \mathbf{w}_2)$. They compress $(\mathbf{w}_1, \mathbf{w}_2)$ to $h \in \{1, \ldots, 2^{nH(W_1,W_2)}\}$. We know that the destination can correctly decode $(\mathbf{w}_1, \mathbf{w}_2)$ from h if h is at least $nH(W_1, W_2)$ bits. Now, create $2^{n[H(W_1|W_2)+H(W_2|W_1)]}$ bins and index each bin by a unique (j, k). Assign h to the bins so that each bin contains $2^{nI(W_1;W_2)}$ entries. Index the entries $i \in \{1, \ldots, 2^{nI(W_1;W_2)}\}$. Hence, each h can be represented by a unique triplet (i, j, k).

Assume that in the beginning of block t, nodes 1 and 2 have correctly received (j^{t-1}, k^{t-1}) and determined i^{t-1}. They send $\mathbf{x}_1(j^t|h^{t-1})$ and $\mathbf{x}_2(k^t|h^{t-1})$ respectively, where $h^{t-1} = (i^{t-1}, j^{t-1}, k^{t-1})$. At the end of block t, nodes 1 and 2 received $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ respectively. We define the following error events at nodes 1 and 2.

$$\mathcal{E}_1 \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j^t|h^{t-1}), \mathbf{x}_2(k^t|h^{t-1}), \mathbf{y}_1(t)\right\} \notin \mathcal{A}_\epsilon, \tag{C.19a}$$

$$\mathcal{E}_2 \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j^t|h^{t-1}), \mathbf{x}_2(k|h^{t-1}), \mathbf{y}_1(t)\right\} \in \mathcal{A}_\epsilon, \tag{C.19b}$$

$$\mathcal{E}_3 \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j^t|h^{t-1}), \mathbf{x}_2(k^t|h^{t-1}), \mathbf{y}_2(t)\right\} \notin \mathcal{A}_\epsilon, \tag{C.19c}$$

$$\mathcal{E}_4 \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j|h^{t-1}), \mathbf{x}_2(k^t|h^{t-1}), \mathbf{y}_2(t)\right\} \in \mathcal{A}_\epsilon, \tag{C.19d}$$

for all $j \in \{1, 2, \ldots, 2^{n[H(W_1|W_2)+\epsilon]}\} \setminus \{j^t\}$ and $k \in \{1, 2, \ldots, 2^{n[H(W_2|W_1)+\epsilon]}\} \setminus \{k^t\}$.

By the AEP, for sufficiently large n,

$$\Pr(\mathcal{E}_1) < \epsilon, \tag{C.20a}$$

$$\Pr(\mathcal{E}_3) < \epsilon. \tag{C.20b}$$

The probability that error event $\mathcal{E}_2$ occurs for all $k \neq k^t$ is given by

$$\Pr(\mathcal{E}_2) = \sum_{k \neq k^t} \sum_{\{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1\} \in \mathcal{A}_\epsilon} p(\mathbf{x}_2|\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2) p(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_1, \mathbf{y}_1) \tag{C.21a}$$

$$< 2^{n[H(W_1|W_2)+\epsilon]} 2^{n[H(V_0,V_1,V_2,X_1,X_2,Y_1)+\epsilon]} 2^{-n[H(X_2|V_0,V_1,V_2)-\epsilon]} 2^{-n[H(V_0,V_1,V_2,X_1,Y_1)-\epsilon]} \tag{C.21b}$$

$$= 2^{n[H(W_2|W_1)+H(X_2|V_0,V_1,V_2,X_1,Y_1)-H(X_2|V_0,V_1,V_2)+4\epsilon]} \tag{C.21c}$$

$$= 2^{n[H(W_2|W_1)+H(X_2|V_0,V_1,V_2,X_1,Y_1)-H(X_2|V_0,V_1,V_2,X_1)+4\epsilon]} \tag{C.21d}$$

$$= 2^{n[H(W_2|W_1)-I(X_2;Y_1|V_0,V_1,V_2,X_1)+4\epsilon]}. \tag{C.21e}$$

This can be made arbitrarily small if

$$H(W_2|W_1) < I(X_2; Y_1|V_0, V_1, V_2, X_1) - 4\epsilon \tag{C.22}$$

holds and n is sufficiently large. Similarly, we can show that $\Pr(\mathcal{E}_4)$ can be made small if

$$H(W_1|W_2) < I(X_1; Y_2|V_0, V_1, V_2, X_2) - 4\epsilon. \tag{C.23}$$

Now we look at the error probability at the destination. Assume that nodes 1 and 2 send $\mathbf{x}_1(j^t|h^{t-1})$ and $\mathbf{x}_2(k^t|h^{t-1})$ respectively in block t; and $\mathbf{x}_1(j^{t+1}|h^t)$ and $\mathbf{x}_2(k^{t+1}|h^t)$ respectively in block $t+1$. Assume that the destination has correctly decoded $h^{t-1} = (i^{t-1}, j^{t-1}, k^{t-1})$. We define the following error events at the

destination.

$$\mathcal{E}_{5a} \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j^t|h^{t-1}), \mathbf{x}_2(k^t|h^{t-1}), \mathbf{y}_3(t)\right\} \notin \mathcal{A}_\epsilon, \quad \text{(C.24a)}$$

$$\mathcal{E}_{5b} \triangleq \left\{\mathbf{v}_0(i^t), \mathbf{v}_1(j^t), \mathbf{v}_2(k^t), \mathbf{y}_3(t+1)\right\} \notin \mathcal{A}_\epsilon, \quad \text{(C.24b)}$$

$$\mathcal{E}_5 \triangleq \mathcal{E}_{5a} \cap \mathcal{E}_{5b}, \quad \text{(C.24c)}$$

$$\mathcal{E}_6 \triangleq \left\{\mathbf{v}_0(i), \mathbf{v}_1(j^t), \mathbf{v}_2(k^t), \mathbf{y}_3(t+1)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24d)}$$

$$\mathcal{E}_{7a} \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j|h^{t-1}), \mathbf{x}_2(k^t|h^{t-1}), \mathbf{y}_3(t)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24e)}$$

$$\mathcal{E}_{7b} \triangleq \left\{\mathbf{v}_0(i^t), \mathbf{v}_1(j), \mathbf{v}_2(k^t), \mathbf{y}_3(t+1)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24f)}$$

$$\mathcal{E}_7 \triangleq \mathcal{E}_{7a} \cap \mathcal{E}_{7b}, \quad \text{(C.24g)}$$

$$\mathcal{E}_{8a} \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j^t|h^{t-1}), \mathbf{x}_2(k|h^{t-1}), \mathbf{y}_3(t)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24h)}$$

$$\mathcal{E}_{8b} \triangleq \left\{\mathbf{v}_0(i^t), \mathbf{v}_1(j^t), \mathbf{v}_2(k), \mathbf{y}_3(t+1)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24i)}$$

$$\mathcal{E}_8 \triangleq \mathcal{E}_{8a} \cap \mathcal{E}_{8b}, \quad \text{(C.24j)}$$

$$\mathcal{E}_{9a} \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j|h^{t-1}), \mathbf{x}_2(k^t|h^{t-1}), \mathbf{y}_3(t)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24k)}$$

$$\mathcal{E}_{9b} \triangleq \left\{\mathbf{v}_0(i), \mathbf{v}_1(j), \mathbf{v}_2(k^t), \mathbf{y}_3(t+1)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24l)}$$

$$\mathcal{E}_9 \triangleq \mathcal{E}_{9a} \cap \mathcal{E}_{9b}, \quad \text{(C.24m)}$$

$$\mathcal{E}_{10a} \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j^t|h^{t-1}), \mathbf{x}_2(k|h^{t-1}), \mathbf{y}_3(t)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24n)}$$

$$\mathcal{E}_{10b} \triangleq \left\{\mathbf{v}_0(i), \mathbf{v}_1(j^t), \mathbf{v}_2(k), \mathbf{y}_3(t+1)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24o)}$$

$$\mathcal{E}_{10} \triangleq \mathcal{E}_{10a} \cap \mathcal{E}_{10b}, \quad \text{(C.24p)}$$

$$\mathcal{E}_{11a} \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j|h^{t-1}), \mathbf{x}_2(k|h^{t-1}), \mathbf{y}_3(t)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24q)}$$

$$\mathcal{E}_{11b} \triangleq \left\{\mathbf{v}_0(i^t), \mathbf{v}_1(j), \mathbf{v}_2(k), \mathbf{y}_3(t+1)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24r)}$$

$$\mathcal{E}_{11} \triangleq \mathcal{E}_{11a} \cap \mathcal{E}_{11b}, \quad \text{(C.24s)}$$

$$\mathcal{E}_{12a} \triangleq \left\{\mathbf{v}_0(i^{t-1}), \mathbf{v}_1(j^{t-1}), \mathbf{v}_2(k^{t-1}), \mathbf{x}_1(j|h^{t-1}), \mathbf{x}_2(k|h^{t-1}), \mathbf{y}_3(t)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24t)}$$

$$\mathcal{E}_{12b} \triangleq \left\{\mathbf{v}_0(i), \mathbf{v}_1(j), \mathbf{v}_2(k), \mathbf{y}_3(t+1)\right\} \in \mathcal{A}_\epsilon, \quad \text{(C.24u)}$$

$$\mathcal{E}_{12} \triangleq \mathcal{E}_{12a} \cap \mathcal{E}_{12b}, \quad \text{(C.24v)}$$

for all $i \neq i^{t+1}$, $j \neq j^{t+1}$, and $k \neq k^{t+1}$.

By the AEP and for sufficiently large n, $\Pr(\mathcal{E}_{5a}) < \epsilon$ and $\Pr(\mathcal{E}_{5b}) < \epsilon$. Hence,

$\Pr(\mathcal{E}_5) \leq \Pr(\mathcal{E}_{5a}) \Pr(\mathcal{E}_{5b}) < \epsilon^2$.

$$\Pr(\mathcal{E}_6) = \sum_{i \neq i^t} \sum_{\{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{y}_3\} \in \mathcal{A}_\epsilon} p(\mathbf{v}_0) p(\mathbf{v}_1, \mathbf{v}_2, \mathbf{y}_3) \tag{C.25a}$$

$$< 2^{n[I(W_1;W_2)+\epsilon]} 2^{n[H(V_0,V_1,V_2,Y_3)+\epsilon]} 2^{-n[H(V_0)-\epsilon]} 2^{-n[H(V_1,V_2,Y_3)-\epsilon]} \tag{C.25b}$$

$$= 2^{n[I(W_1;W_2)+\epsilon]} 2^{-n[H(V_0|V_1,V_2)-\epsilon]} 2^{n[H(V_0|V_1,V_2,Y_3)-\epsilon]} \tag{C.25c}$$

$$= 2^{n[I(W_1;W_2)-I(V_0;Y_3|V_1,V_2)+4\epsilon]}. \tag{C.25d}$$

Hence $\Pr(\mathcal{E}_6)$ can be made small if

$$I(W_1; W_2) < I(V_0; Y_3 | V_1, V_2) - 4\epsilon \tag{C.26}$$

holds and n is sufficiently large.

$$\Pr(\mathcal{E}_7) = \sum_{j \neq j^t} \Bigg[\sum_{\{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_3\} \in \mathcal{A}_\epsilon} p(\mathbf{x}_1 | \mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2) p(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_2, \mathbf{y}_3) \times$$

$$\sum_{\{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{y}_3\} \in \mathcal{A}_\epsilon} p(\mathbf{v}_1) p(\mathbf{v}_0, \mathbf{v}_2, \mathbf{y}_3) \Bigg] \tag{C.27a}$$

$$< 2^{n[H(W_1|W_2)+\epsilon]} 2^{-n[I(X_1;Y_3|V_0,V_1,V_2,X_2)-3\epsilon]} 2^{-n[I(V_1;Y_3|V_0,V_2)-3\epsilon]} \tag{C.27b}$$

$$= 2^{n[H(W_1|W_2)-I(X_1;Y_3|V_0,V_1,V_2,X_2)-I(V_1;Y_3|V_0,V_2)+7\epsilon]}. \tag{C.27c}$$

Hence $\Pr(\mathcal{E}_7)$ can be made small if

$$H(W_1|W_2) < I(V_1; Y_3|V_0, V_2) + I(X_1; Y_3|V_0, V_1, V_2, X_2) - 7\epsilon \tag{C.28}$$

holds and n is sufficiently large.

Similarly, $Pr(\mathcal{E}_8)$ be made arbitrarily small if

$$H(W_2|W_1) < I(V_2; Y_3|V_0, V_1) + I(X_2; Y_3|V_0, V_1, V_2, X_1) - 7\epsilon. \tag{C.29}$$

$$\Pr(\mathcal{E}_9) = \sum_{(i,j)\neq(i^t,j^t)} \Bigg[\sum_{\{\mathbf{v}_0,\mathbf{v}_1,\mathbf{v}_2,\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_3\}\in\mathcal{A}_\epsilon} p(\mathbf{x}_1|\mathbf{v}_0,\mathbf{v}_1,\mathbf{v}_2)p(\mathbf{v}_0,\mathbf{v}_1,\mathbf{v}_2,\mathbf{x}_2,\mathbf{y}_3)\times \sum_{\{\mathbf{v}_0,\mathbf{v}_1,\mathbf{v}_2,\mathbf{y}_3\}\in\mathcal{A}_\epsilon} p(\mathbf{v}_0,\mathbf{v}_1)p(\mathbf{v}_2,\mathbf{y}_3) \Bigg] \tag{C.30a}$$

$$< 2^{n[I(W_1;W_2)+\epsilon+H(W_1|W_2)+\epsilon]} 2^{-n[I(X_1;Y_3|V_0,V_1,V_2,X_2)-3\epsilon]} 2^{-n[I(V_0,V_1;Y_3|V_2)-3\epsilon]} \tag{C.30b}$$

$$= 2^{n[H(W_1)-I(X_1;Y_3|V_0,V_1,V_2,X_2)-I(V_0,V_1;Y_3|V_2)+8\epsilon]}. \tag{C.30c}$$

Hence $\Pr(\mathcal{E}_9)$ can be made small if

$$H(W_1) < I(V_0,V_1;Y_3|V_2) + I(X_1;Y_3|V_0,V_1,V_2,X_2) - 8\epsilon \tag{C.31}$$

holds and n is sufficiently large.

Similarly, $Pr(\mathcal{E}_{10})$, $Pr(\mathcal{E}_{11})$, and $Pr(\mathcal{E}_{12})$ be made arbitrarily small if

$$H(W_2) < I(V_0,V_2;Y_3|V_1) + I(X_2;Y_3|V_0,V_1,V_2,X_1) - 8\epsilon, \tag{C.32a}$$

$$H(W_1|W_2) + H(W_2|W_1) < I(V_1,V_2;Y_3|V_0) + I(X_1,X_2;Y_3|V_0,V_1,V_2) - 8\epsilon, \tag{C.32b}$$

$$H(W_1,W_2) < I(X_1,X_2;Y_3) - 9\epsilon, \tag{C.32c}$$

hold respectively.

If all these constraints are satisfied and if n is large enough, the total probability of error can be bounded by

$$P_e = \bigcup_{i=0}^{12} \Pr(\mathcal{E}_i) < \sum_{i=0}^{12} \Pr(\mathcal{E}_i) < 12\epsilon + \epsilon^2, \tag{C.33}$$

for any $\epsilon > 0$.

Combining these rate constraints and adding the time sharing random variable

Q, we get Theorem 19.

C.3 Achievable Region of FDS-DF for the Gaussian MACFCS

On the static Gaussian channel, using FDS-DF, nodes 1 and 2 send the following respectively.

$$X_1 = \sqrt{\alpha_{10}P_1}V_0 + \sqrt{\alpha_{11}P_1}V_1 + \sqrt{\alpha_{12}P_1}V_2 + \sqrt{\alpha_{13}P_1}U_1, \tag{C.34a}$$

$$X_2 = \sqrt{\alpha_{20}P_2}V_0 + \sqrt{\alpha_{21}P_2}V_1 + \sqrt{\alpha_{22}P_2}V_2 + \sqrt{\alpha_{23}P_2}U_2, \tag{C.34b}$$

where V_i and U_j are independent Gaussian random variables with unit power $E[V_i^2] = [U_j^2] = 1$, $\forall i = 0, 1, 2$ and $\forall j = 1, 2$. $0 \leq \sum_{k=0}^{3} \alpha_{jk} \leq 1$ for $j = 1, 2$. Recall that the channel outputs are

$$Y_1 = \sqrt{\kappa d_{21}^{-\eta}}X_2 + Z_1 = \sqrt{\kappa d_{21}^{-\eta}P_1}(\sqrt{\alpha_{10}}V_0 + \sqrt{\alpha_{11}}V_1 + \sqrt{\alpha_{12}}V_2 + \sqrt{\alpha_{13}}U_1) + Z_1, \tag{C.35a}$$

$$Y_2 = \sqrt{\kappa d_{12}^{-\eta}}X_1 + Z_2 = \sqrt{\kappa d_{12}^{-\eta}P_2}(\sqrt{\alpha_{20}}V_0 + \sqrt{\alpha_{21}}V_1 + \sqrt{\alpha_{22}}V_2 + \sqrt{\alpha_{23}}U_2) + Z_2, \tag{C.35b}$$

$$\begin{aligned} Y_3 &= \sqrt{\kappa d_{13}^{-\eta}}X_1 + \sqrt{\kappa d_{23}^{-\eta}}X_2 + Z_3 \qquad \text{(C.35c)} \\ &= \left[\sqrt{\kappa d_{13}^{-\eta}\alpha_{10}P_1} + \sqrt{\kappa d_{23}^{-\eta}\alpha_{20}P_2}\right] V_0 + \left[\sqrt{\kappa d_{13}^{-\eta}\alpha_{11}P_1} + \sqrt{\kappa d_{23}^{-\eta}\alpha_{21}P_2}\right] V_1 \\ &\quad + \left[\sqrt{\kappa d_{13}^{-\eta}\alpha_{12}P_1} + \sqrt{\kappa d_{23}^{-\eta}\alpha_{22}P_2}\right] V_2 + \sqrt{\kappa d_{13}^{-\eta}\alpha_{13}P_1}U_1 + \sqrt{\kappa d_{23}^{-\eta}\alpha_{23}P_2}U_2 + Z_3. \end{aligned} \tag{C.35d}$$

Now, we calculate the mutual information terms in Theorem 19.

$$I(X_1; Y_2|V_0, V_1, V_2, X_2) = \frac{1}{2}\log\left(1 + \frac{\kappa d_{12}^{-\eta}\alpha_{13}P_1}{N_2}\right), \tag{C.36a}$$

$$I(X_2; Y_1|V_0, V_1, V_2, X_1) = \frac{1}{2}\log\left(1 + \frac{\kappa d_{21}^{-\eta}\alpha_{23}P_2}{N_1}\right), \tag{C.36b}$$

For $(a,b,c) \in \{\{0,1,2\}^3 : a \neq b \neq c\}$,

$$I(V_a;Y_3|V_b,V_c) = \frac{1}{2}\log\left(1+\frac{\kappa\left[\sqrt{d_{13}^{-\eta}\alpha_{1a}P_1}+\sqrt{d_{23}^{-\eta}\alpha_{2a}P_2}\right]^2}{\kappa d_{13}^{-\eta}\alpha_{13}P_1+\kappa d_{23}^{-\eta}\alpha_{23}P_2+N_3}\right), \tag{C.37a}$$

$$\begin{aligned} I(V_a,V_b;Y_3|V_c) = \frac{1}{2}\log\Bigg(1+&\frac{\kappa\left[\sqrt{d_{13}^{-\eta}\alpha_{1a}P_1}+\sqrt{d_{23}^{-\eta}\alpha_{2a}P_2}\right]^2}{\kappa d_{13}^{-\eta}\alpha_{13}P_1+\kappa d_{23}^{-\eta}\alpha_{23}P_2+N_3} \\ &+\frac{\kappa\left[\sqrt{d_{13}^{-\eta}\alpha_{1b}P_1}\sqrt{d_{23}^{-\eta}\alpha_{2b}P_2}\right]^2}{\kappa d_{13}^{-\eta}\alpha_{13}P_1+\kappa d_{23}^{-\eta}\alpha_{23}P_2+N_3}\Bigg). \end{aligned} \tag{C.37b}$$

Also, for $(a,b) \in \{\{1,2\}^2 : a \neq b\}$

$$I(X_a;Y_3|V_0,V_1,V_2,X_b) = \frac{1}{2}\log\left(1+\frac{\kappa d_{a3}^{-\eta}\alpha_{a3}P_a}{N_3}\right), \tag{C.38}$$

and,

$$I(X_1,X_2;Y_3|V_0,V_1,V_2) = \frac{1}{2}\log\left(1+\frac{\kappa d_{13}^{-\eta}\alpha_{13}P_1+\kappa d_{23}^{-\eta}\alpha_{23}P_2}{N_3}\right). \tag{C.39}$$

Finally,

$$\begin{aligned} &I(X_1,X_2;Y_3) = \\ &\frac{1}{2}\log\left(1+\frac{\sum_{i=0}^{2}\kappa\left[\sqrt{d_{13}^{-\eta}\alpha_{1i}P_1}+\sqrt{d_{23}^{-\eta}\alpha_{2i}P_2}\right]^2+\kappa d_{13}^{-\eta}\alpha_{13}P_1+\kappa d_{23}^{-\eta}\alpha_{23}P_2}{N_3}\right). \end{aligned} \tag{C.40}$$

C.4 Proof of Theorem 20

In this section, we prove Theorem 20. Node 1 receives $\mathbf{y}_1(t)$ in block t. It knows $\mathbf{x}_1(j^t,p^{t-1})$ and $\mathbf{u}_1(p^{t-1})$. It finds r^t such that $(\tilde{\mathbf{y}}_1(r^t|p^{t-1}),\mathbf{y}_1(t),\mathbf{x}_1(j^t,p^{t-1}),$ $\mathbf{u}_1(p^{t-1})) \in \mathcal{A}_\epsilon$. Berger (1977, Lemma 2.1.3) showed that node 1 can find such a r^t

with probability tends to 1 as $n \to \infty$ if

$$\tilde{R}_1 > I(\tilde{Y}_1; Y_1 | X_1, U_1). \tag{C.41}$$

By similar argument, node 2 can find s^t with probability tends to 1 as $n \to \infty$ such that $\left(\tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{y}_2(t), \mathbf{x}_2(k^t, q^{t-1}), \mathbf{u}_2(q^{t-1})\right) \in \mathcal{A}_\epsilon$ if

$$\tilde{R}_2 > I(\tilde{Y}_2; Y_2 | X_2, U_2). \tag{C.42}$$

Suppose that nodes 1 and 2 send $\mathbf{x}_1(j^{t+1}, p^t)$ and $\mathbf{x}_2(k^{t+1}, q^t)$ respectively in block $t+1$. Define the following event where the destination wrongly decodes the quantized and binned signal p^t or q^t.

$$\mathcal{E}_1 \triangleq \left(\mathbf{u}_1(p^t), \mathbf{u}_2(q^t), \mathbf{y}_3(t+1)\right) \notin \mathcal{A}_\epsilon, \tag{C.43a}$$

$$\mathcal{E}_2 \triangleq \left(\mathbf{u}_1(p), \mathbf{u}_2(q^t), \mathbf{y}_3(t+1)\right) \in \mathcal{A}_\epsilon, \tag{C.43b}$$

$$\mathcal{E}_3 \triangleq \left(\mathbf{u}_1(p^t), \mathbf{u}_2(q), \mathbf{y}_3(t+1)\right) \in \mathcal{A}_\epsilon, \tag{C.43c}$$

$$\mathcal{E}_4 \triangleq \left(\mathbf{u}_1(p), \mathbf{u}_2(q), \mathbf{y}_3(t+1)\right) \in \mathcal{A}_\epsilon, \tag{C.43d}$$

for all $p \in \{1, 2, \ldots, 2^{nR_1'}\} \setminus \{p^t\}$ and $q \in \{1, 2, \ldots, 2^{nR_2'}\} \setminus \{q^t\}$.

By the AEP, $\Pr(\mathcal{E}_1) < \epsilon$ for large n. We can show that $\Pr(\mathcal{E}_2), \Pr(\mathcal{E}_3)$, and $\Pr(\mathcal{E}_4)$ can be bounded by ϵ for large n if the following holds.

$$R_1' < I(U_1; Y_3 | U_2) - 3\epsilon, \tag{C.44a}$$

$$R_2' < I(U_2; Y_3 | U_1) - 3\epsilon, \tag{C.44b}$$

$$R_1' + R_2' < I(U_1, U_2; Y_3) - 3\epsilon. \tag{C.44c}$$

At the end of block t, assume that the destination has already correctly decoded the quantized and binned signals p^t, q^t, p^{t-1}, and q^{t-1}. Suppose that r^t and s^t are the quantized values of nodes 1 and 2 respectively. We define the following events where the destination decodes the estimates wrongly, for all $r \in \{1, 2, \ldots, 2^{n\tilde{R}_1}\} \setminus$

$\{r^t\}$ and $s \in \{1, 2, \ldots, 2^{n\tilde{R}_2}\} \setminus \{s^t\}$.

$$\mathcal{E}_5 \triangleq \left(\tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \mathbf{y}_3(t)\right) \notin \mathcal{A}_\epsilon, \tag{C.45a}$$

$$\mathcal{E}_{6a} \triangleq \left(\tilde{\mathbf{y}}_1(r|p^{t-1}), \tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \mathbf{y}_3(t)\right) \in \mathcal{A}_\epsilon, \tag{C.45b}$$

$$\mathcal{E}_6 \triangleq \mathcal{E}_{6a} \cap \{r \in S_{p^t}\}, \tag{C.45c}$$

$$\mathcal{E}_{7a} \triangleq \left(\tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(s|q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \mathbf{y}_3(t)\right) \in \mathcal{A}_\epsilon, \tag{C.45d}$$

$$\mathcal{E}_7 \triangleq \mathcal{E}_{7a} \cap \{s \in S_{q^t}\}, \tag{C.45e}$$

$$\mathcal{E}_{8a} \triangleq \left(\tilde{\mathbf{y}}_1(r|p^{t-1}), \tilde{\mathbf{y}}_2(s|q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \mathbf{y}_3(t)\right) \in \mathcal{A}_\epsilon, \tag{C.45f}$$

$$\mathcal{E}_8 \triangleq \mathcal{E}_{8a} \cap \{r \in S_{p^t}\} \cap \{s \in S_{q^t}\}. \tag{C.45g}$$

By the AEP, $\Pr(\mathcal{E}_5) < \epsilon$ for large n. The probability of the event $\mathcal{E}_6$ is as follows.

$$\Pr(\mathcal{E}_6) = \Pr\left(\mathcal{E}_{6a} \cap \{r \in S_{p^t}\}\right) \tag{C.46a}$$

$$= \sum_{\substack{r \neq r^t \\ r \in S_{p^t}}} \sum_{(\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{y}_3) \in \mathcal{A}_\epsilon} p(\tilde{\mathbf{y}}_1|\tilde{\mathbf{y}}_2, \mathbf{u}_1, \mathbf{u}_2) p(\tilde{\mathbf{y}}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{y}_3) \tag{C.46b}$$

$$< 2^{n(\tilde{R}_1 - R_1')} \times 2^{n[H(\tilde{Y}_1, \tilde{Y}_2, U_1, U_2, Y_3) + \epsilon]} \times 2^{-n[H(\tilde{Y}_1|\tilde{Y}_2, U_1, U_2) - \epsilon]} \times 2^{-n[H(\tilde{Y}_2, U_1, U_2, Y_3) - \epsilon]} \tag{C.46c}$$

$$= 2^{n(\tilde{R}_1 - R_1')} \times 2^{-n(I(\tilde{Y}_1; Y_3|\tilde{Y}_2, U_1, U_2) - 3\epsilon)}. \tag{C.46d}$$

This can be made small, for a large n, if

$$\tilde{R}_1 < I(\tilde{Y}_1; Y_3|\tilde{Y}_2, U_1, U_2) + R_1' - 3\epsilon. \tag{C.47}$$

Similarly $\Pr(\mathcal{E}_7) < \epsilon$ and $\Pr(\mathcal{E}_8) < \epsilon$ for large n if

$$\tilde{R}_2 < I(\tilde{Y}_2; Y_3|\tilde{Y}_1, U_1, U_2)) + R_2' - 3\epsilon, \tag{C.48a}$$

$$\tilde{R}_1 + \tilde{R}_2 < I(\tilde{Y}_1, \tilde{Y}_2; Y_3|U_1, U_2) + R_1' + R_2' - 3\epsilon. \tag{C.48b}$$

Now, supposed that nodes 1 and 2 send $\mathbf{x}_1(j^t, p^{t-1})$ and $\mathbf{x}_2(k^t, q^{t-1})$ respectively in block t. Assume that the destination has correctly estimated r^t, s^t, p^{t-1}, and q^{t-1}. It decodes (j^t, k^t) using $\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2$, as well as its received symbol $\mathbf{y}_3(t)$. The error events, where the destination wrongly decodes the source signal(s), are as follows.

$$\mathcal{E}_9 \triangleq \big(\mathbf{x}_1(j^t, p^{t-1}), \mathbf{x}_2(k^t, q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{y}_3(t)\big) \notin \mathcal{A}_\epsilon, \tag{C.49a}$$

$$\mathcal{E}_{10} \triangleq \big(\mathbf{x}_1(j, p^{t-1}), \mathbf{x}_2(k^t, q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{y}_3(t)\big) \in \mathcal{A}_\epsilon, \tag{C.49b}$$

$$\mathcal{E}_{11} \triangleq \big(\mathbf{x}_1(j^t, p^{t-1}), \mathbf{x}_2(k, q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{y}_3(t)\big) \in \mathcal{A}_\epsilon, \tag{C.49c}$$

$$\mathcal{E}_{12} \triangleq \big(\mathbf{x}_1(j, p^{t-1}), \mathbf{x}_2(k, q^{t-1}), \mathbf{u}_1(p^{t-1}), \mathbf{u}_2(q^{t-1}), \tilde{\mathbf{y}}_1(r^t|p^{t-1}), \tilde{\mathbf{y}}_2(s^t|q^{t-1}), \mathbf{y}_3(t)\big) \in \mathcal{A}_\epsilon. \tag{C.49d}$$

By the AEP, $\Pr(\mathcal{E}_9) < \epsilon$ for large n. Now,

$$\begin{aligned}
&\Pr(\mathcal{E}_{10}) \\
&= \sum_{j \neq j^t} \sum_{(\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2, \tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \mathbf{y}_3) \in \mathcal{A}_\epsilon} p(\mathbf{x}_1|\mathbf{u}_1) p(\mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2, \tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \mathbf{y}_3) && \text{(C.50a)} \\
&< 2^{nR_1} \sum_{(\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2, \tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \mathbf{y}_3) \in \mathcal{A}_\epsilon} p(\mathbf{x}_1|\mathbf{u}_1) p(\mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2) p(\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \mathbf{y}_3|\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_2) && \text{(C.50b)} \\
&= 2^{nR_1} 2^{n[H(U_1, U_2, X_1, X_2, \tilde{Y}_1, \tilde{Y}_2, Y_3) + \epsilon]} 2^{-n[H(X_1|U_1) - \epsilon]} 2^{-n[H(U_1, U_2, X_2) - \epsilon]} \\
&\quad \times 2^{-n[H(\tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2, X_2) - \epsilon]} && \text{(C.50c)} \\
&= 2^{n[R_1 + H(X_1, \tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2, X_2) - H(X_1|U_1, U_2, X_2) - H(\tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2, X_2) + 4\epsilon]} && \text{(C.50d)} \\
&= 2^{n[R_1 - I(X_1; \tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2, X_2) + 4\epsilon]}. && \text{(C.50e)}
\end{aligned}$$

$\Pr(\mathcal{E}_{10})$ can be made small if

$$R_1 < I(X_1; \tilde{Y}_1, \tilde{Y}_2, Y_3|U_1, U_2, X_2) - 4\epsilon. \tag{C.51}$$

Similarly, $\Pr(\{\mathcal{E}_{11})$ and $\Pr(\mathcal{E}_{12})$ can be bounded if

$$R_2 < I(X_2; \tilde{Y}_1, \tilde{Y}_2, Y_3 | U_1, U_2, X_1) - 4\epsilon, \tag{C.52a}$$

$$R_1 + R_2 < I(X_1, X_2; \tilde{Y}_1, \tilde{Y}_2, Y_3 | U_1, U_2) - 4\epsilon \tag{C.52b}$$

hold respectively.

Combining these rate constraints for the MACF using CF and the constraints for the source coding, (5.14a)-(5.14c), and adding the time sharing random variable Q, we get Theorem 20.

C.5 Achievable Region of SC-CF for the Gaussian MACFCS

On the static Gaussian channel, using SC-CF, nodes 1 and 2 send $X_1 = U_1 + V_1$ and $X_2 = U_2 + V_2$ respectively. Here U_1 (quantized and binned information of the previous block from Y_1), V_1 (new information from source 1), U_2 (old quantized and binned information of the previous block from Y_2), and V_2 (new information from source 2) are independent Gaussian random variables with power constraints $E[U_1^2] \leq P_{U1}$, $E[V_1^2] \leq P_{V1}$ $E[U_2^2] \leq P_{U2}$, and $E[V_2^2] \leq P_{V2}$ respectively. We note that $P_1 = P_{U1} + P_{V1}$ and $P_2 = P_{U2} + P_{V2}$.

The nodes receive

$$Y_1 = \sqrt{\kappa d_{21}^{-\eta}} X_2 + Z_1 = \sqrt{\kappa d_{21}^{-\eta}} (U_2 + V_2) + Z_1 \tag{C.53a}$$

$$Y_2 = \sqrt{\kappa d_{12}^{-\eta}} X_1 + Z_2 = \sqrt{\kappa d_{12}^{-\eta}} (U_1 + V_1) + Z_2 \tag{C.53b}$$

$$Y_3 = \sqrt{\kappa d_{13}^{-\eta}} (U_1 + V_1) + \sqrt{\kappa d_{23}^{-\eta}} (U_2 + V_2) + Z_3, \tag{C.53c}$$

where $Z_1 \sim \mathcal{N}(0, N_1)$, $Z_2 \sim \mathcal{N}(0, N_2)$, and $Z_3 \sim \mathcal{N}(0, N_3)$ are independent noise.

The quantized signals are

$$\tilde{Y}_1 = Y_1 + \tilde{Z}_1 = \sqrt{\kappa d_{21}^{-\eta}} X_2 + Z_1 = \sqrt{\kappa d_{21}^{-\eta}}(U_2 + V_2) + Z_1 + \tilde{Z}_1 \tag{C.54a}$$

$$\tilde{Y}_2 = Y_2 + \tilde{Z}_2 = \sqrt{\kappa d_{12}^{-\eta}} X_1 + Z_2 = \sqrt{\kappa d_{12}^{-\eta}}(U_1 + V_1) + Z_2 + \tilde{Z}_2, \tag{C.54b}$$

where $\tilde{Z}_1 \sim \mathcal{N}(0, \tilde{N}_1)$ and $\tilde{Z}_2 \sim \mathcal{N}(0, \tilde{N}_2)$ are independent quantization noise.

Now,

$$\begin{aligned}
&I(X_1; \tilde{Y}_1, \tilde{Y}_2, Y_3 | U_1, U_2, X_2) \\
&= H(\tilde{Y}_1, \tilde{Y}_2, Y_3 | U_1, U_2, X_2) - H(\tilde{Y}_2, Y_3 | U_1, U_2, X_1, X_2) && \text{(C.55a)} \\
&= H\Big(\sqrt{\kappa d_{21}^{-\eta}}(U_2 + V_2) + Z_1 + \tilde{Z}_1, \sqrt{\kappa d_{12}^{-\eta}}(U_1 + V_1) + Z_2 + \tilde{Z}_2, \sqrt{\kappa d_{13}^{-\eta}}(U_1 + V_1) \\
&\quad + \sqrt{\kappa d_{23}^{-\eta}}(U_2 + V_2) + Z_3 | \quad U_1, U_2, U_2 + V_2\Big) - H\Big(\sqrt{\kappa d_{21}^{-\eta}}(U_2 + V_2) + Z_1 + \tilde{Z}_1, \\
&\qquad \sqrt{\kappa d_{12}^{-\eta}}(U_1 + V_1) + Z_2 + \tilde{Z}_2, \quad \sqrt{\kappa d_{13}^{-\eta}}(U_1 + V_1) + \sqrt{\kappa d_{23}^{-\eta}}(U_2 + V_2) + Z_3 | \\
&\qquad U_1, U_2, U_1 + V_1, U_2 + V_2\Big) && \text{(C.55b)} \\
&= H\left(Z_1 + \tilde{Z}_1, \sqrt{\kappa d_{12}^{-\eta}} V_1 + Z_2 + \tilde{Z}_2, \sqrt{\kappa d_{13}^{-\eta}} V_1 + Z_3\right) - H\left(Z_1 + \tilde{Z}_1, Z_2 + \tilde{Z}_2, Z_3\right). \\
& && \text{(C.55c)}
\end{aligned}$$

The first term is

$$\begin{aligned}
&H\left(Z_1 + \tilde{Z}_1, \sqrt{\kappa d_{12}^{-\eta}} V_1 + Z_2 + \tilde{Z}_2, \sqrt{\kappa d_{13}^{-\eta}} V_1 + Z_3\right) \\
&= \frac{1}{2} \log(2\pi e)^3 \begin{vmatrix} N_1 + \tilde{N}_1 & 0 & 0 \\ 0 & \kappa d_{12}^{-\eta} P_{V1} + N_2 + \tilde{N}_2 & \kappa \sqrt{d_{12}^{-\eta} d_{13}^{-\eta}} P_{V1} \\ 0 & \kappa \sqrt{d_{12}^{-\eta} d_{13}^{-\eta}} P_{V1} & \kappa d_{13}^{-\eta} P_{V1} + N_3 \end{vmatrix} && \text{(C.56a)} \\
&= \frac{1}{2} \log(2\pi e)^3 \left[N_1 + \tilde{N}_1\right] \left[N_3(N_2 + \tilde{N}_2) + (\kappa d_{12}^{-\eta} N_3 + \kappa d_{13}^{-\eta}(N_2 + \tilde{N}_2)) P_{V1}\right]. \\
& && \text{(C.56b)}
\end{aligned}$$

The second term is

$$H\left(Z_1 + \tilde{Z}_1, Z_2 + \tilde{Z}_2, Z_3\right) = \frac{1}{2} \log(2\pi e)^3 \left[(N_1 + \tilde{N}_1)(N_2 + \tilde{N}_2) N_3\right]. \tag{C.57a}$$

Thus,

$$I(X_1;\tilde{Y}_1,\tilde{Y}_2,Y_3|U_1,U_2,X_2)=\frac{1}{2}\log\left[1+\frac{\kappa d_{12}^{-\eta}P_{V1}}{N_2+\tilde{N}_2}+\frac{\kappa d_{13}^{-\eta}P_{V1}}{N_3}\right]. \tag{C.58}$$

Similarly, we can show that

$$I(X_2;\tilde{Y}_1,\tilde{Y}_2,Y_3|U_1,U_2,X_1)=\frac{1}{2}\log\left[1+\frac{\kappa d_{21}^{-\eta}P_{V2}}{N_1+\tilde{N}_1}+\frac{\kappa d_{23}^{-\eta}P_{V2}}{N_3}\right]. \tag{C.59}$$

Now, we evaluate

$$I(X_1,X_2;\tilde{Y}_1,\tilde{Y}_2,Y_3|U_1,U_2)=H(\tilde{Y}_1,\tilde{Y}_2,Y_3|U_1,U_2)-H(\tilde{Y}_1,\tilde{Y}_2,Y_3|U_1,U_2,X_1,X_2). \tag{C.60}$$

The first term is

$$\begin{aligned}
&H(\tilde{Y}_1,\tilde{Y}_2,Y_3|U_1,U_2)\\
&=\frac{1}{2}\log(2\pi e)^3\begin{vmatrix}\kappa d_{21}^{-\eta}P_{V2}+N_1+\tilde{N}_1 & 0 & \kappa\sqrt{d_{21}^{\eta}d_{23}^{-\eta}}P_{V2}\\ 0 & \kappa d_{12}^{-\eta}P_{V1}+N_2+\tilde{N}_2 & \kappa\sqrt{d_{12}^{-\eta}d_{13}^{-\eta}}P_{V1}\\ \kappa\sqrt{d_{21}^{\eta}d_{23}^{-\eta}}P_{V2} & \kappa\sqrt{d_{12}^{-\eta}d_{13}^{-\eta}}P_{V1} & \kappa d_{13}^{-\eta}P_{V1}+\kappa d_{23}^{-\eta}P_{V2}+N_3\end{vmatrix} && \text{(C.61a)}\\
&=\frac{1}{2}\log(2\pi e)^3\Big[\kappa d_{12}^{-\eta}P_{V1}N_3(N_1+\tilde{N}_1)+\kappa d_{21}^{-\eta}P_{V2}N_3(N_2+\tilde{N}_2)\\
&\quad+\kappa d_{13}^{-\eta}P_{V1}(N_1+\tilde{N}_1)(N_2+\tilde{N}_2)+\kappa d_{23}^{-\eta}P_{V2}(N_1+\tilde{N}_1)(N_2+\tilde{N}_2)\\
&\quad+\kappa^2 d_{12}^{-\eta}d_{21}^{-\eta}P_{V1}P_{V2}N_3\kappa^2 d_{21}^{-\eta}d_{13}^{-\eta}P_{V1}P_{V2}(N_2+\tilde{N}_2)+\kappa^2 d_{12}^{-\eta}d_{23}^{-\eta}P_{V1}P_{V2}(N_1+\tilde{N}_1)\\
&\quad+(N_1+\tilde{N}_1)(N_2+\tilde{N}_2)N_3\Big] && \text{(C.61b)}\\
&\triangleq\frac{1}{2}\log(2\pi e)^3\mathcal{B}_1, && \text{(C.61c)}
\end{aligned}$$

and the second term is

$$H(\tilde{Y}_1,\tilde{Y}_2,Y_3|U_1,U_2,X_1,X_2)=\frac{1}{2}\log(2\pi e)^3(N_1+\tilde{N}_1)(N_2+\tilde{N}_2)N_3. \tag{C.62}$$

Hence,

$$
\begin{aligned}
&I(X_1, X_2; \tilde{Y}_1, \tilde{Y}_2, Y_3 | U_1, U_2) \\
&= \frac{1}{2} \log \left[1 + \frac{\kappa d_{12}^{-\eta} P_{V1}}{N_2 + \tilde{N}_2} + \frac{\kappa d_{21}^{-\eta} P_{V2}}{N_1 + \tilde{N}_1} + \frac{\kappa d_{13}^{-\eta} P_{V1}}{N_3} + \frac{\kappa d_{23}^{-\eta} P_{V2}}{N_3} + \frac{\kappa^2 d_{12}^{-\eta} d_{21}^{-\eta} P_{V1} P_{V2}}{(N_1 + \tilde{N}_1)(N_2 + \tilde{N}_2)} \right. \\
&\quad \left. + \frac{\kappa^2 d_{21}^{-\eta} d_{13}^{-\eta} P_{V1} P_{V2}}{(N_1 + \tilde{N}_1) N_3} + \frac{\kappa^2 d_{12}^{-\eta} d_{23}^{-\eta} P_{V1} P_{V2}}{(N_2 + \tilde{N}_2) N_3} \right]. \qquad \text{(C.63a)}
\end{aligned}
$$

We can show that

$$
I(U_1, Y_3 | U_2) = \frac{1}{2} \log \left[1 + \frac{\mathcal{C}_1}{\mathcal{B}_2} \right], \qquad \text{(C.64a)}
$$

$$
I(U_2, Y_3 | U_1) = \frac{1}{2} \log \left[1 + \frac{\mathcal{C}_2}{\mathcal{B}_2} \right], \qquad \text{(C.64b)}
$$

$$
I(U_1, U_2, Y_3) = \frac{1}{2} \log \left[1 + \frac{\mathcal{C}_1 + \mathcal{C}_2}{\mathcal{B}_2} \right]. \qquad \text{(C.64c)}
$$

where $\mathcal{B}_2 \triangleq \kappa d_{13}^{-\eta} P_{V1} + \kappa d_{23}^{-\eta} P_{V2} + N_3$, $\mathcal{C}_1 \triangleq \kappa d_{13}^{-\eta} P_{U1} = \kappa d_{13}^{-\eta}(P_1 - P_{V1})$, and $\mathcal{C}_2 \triangleq \kappa d_{23}^{-\eta} P_{U2} = \kappa d_{23}^{-\eta}(P_2 - P_{V2})$. Also,

$$
I(\tilde{Y}_1; Y_1 | X_1, U_1) = \frac{1}{2} \log \left[1 + \frac{\kappa d_{21}^{-\eta} P_2 + N_1}{\tilde{N}_1} \right] \triangleq \frac{1}{2} \log \left[1 + \frac{\mathcal{D}_1}{\tilde{N}_1} \right] \qquad \text{(C.65a)}
$$

$$
I(\tilde{Y}_2; Y_2 | X_2, U_2) = \frac{1}{2} \log \left[1 + \frac{\kappa d_{12}^{-\eta} P_1 + N_2}{\tilde{N}_2} \right] \triangleq \frac{1}{2} \log \left[1 + \frac{\mathcal{D}_2}{\tilde{N}_2} \right]. \qquad \text{(C.65b)}
$$

$$
\qquad \text{(C.65c)}
$$

We write $I(\tilde{Y}_1; Y_3 | \tilde{Y}_2, U_1, U_2) = H(Y_3 | \tilde{Y}_2, U_1, U_2) - H(Y_3 | \tilde{Y}_1, \tilde{Y}_2, U_1, U_2)$. Evaluating and simplifying, we get

$$
I(\tilde{Y}_1; Y_3 | \tilde{Y}_2, U_1, U_2) = \frac{1}{2} \log \left[1 + \frac{\kappa^2 d_{23}^{-\eta} d_{21}^{-\eta} P_{V2}^2 (\kappa d_{12} P_{V1} + N_2 + \tilde{N}_2)}{\mathcal{B}_1} \right]. \qquad \text{(C.66a)}
$$

So, constraint (5.17a) becomes

$$
\left(1 + \frac{\mathcal{D}_1}{\tilde{N}_1} \right) < \left(1 + \frac{\mathcal{C}_3}{\mathcal{B}_1} \right) \left(1 + \frac{\mathcal{C}_1}{\mathcal{B}_2} \right), \qquad \text{(C.67)}
$$

where $\mathcal{C}_3 \triangleq \kappa^2 d_{23}^{-\eta} d_{21}^{-\eta} P_{V2}^2 (\kappa d_{12}^{-\eta} P_{V1} + N_2 + \tilde{N}_2)$. Similarly, constraint (5.17b) becomes

$$\left(1 + \frac{\mathcal{D}_2}{\tilde{N}_2}\right) < \left(1 + \frac{\mathcal{C}_4}{\mathcal{B}_1}\right)\left(1 + \frac{\mathcal{C}_2}{\mathcal{B}_2}\right), \tag{C.68}$$

where $\mathcal{C}_4 \triangleq \kappa^2 d_{13}^{-\eta} d_{12}^{-\eta} P_{V1}^2 (\kappa d_{21}^{-\eta} P_{V2} + N_1 + \tilde{N}_1)$. Lastly, constraint (5.17c) becomes

$$\left(1 + \frac{\mathcal{D}_1}{\tilde{N}_1}\right)\left(1 + \frac{\mathcal{D}_2}{\tilde{N}_2}\right) < \left(1 + \frac{\mathcal{C}_3 + \mathcal{C}_4}{\mathcal{B}_1}\right)\left(1 + \frac{\mathcal{C}_1 + \mathcal{C}_2}{\mathcal{B}_2}\right). \tag{C.69}$$

We note that the achievability derived in Theorem 20 makes use of the Markov lemma (Berger, 1977, Lemma 4.1), which requires strong typicality. Though strong typicality does not extend to continuous random variables, we can generalize the Markov lemma for Gaussian inputs and thus show that the rate governed by (C.58), (C.59), and (C.63a) is achievable (Kramer *et al.*, 2005).

Bibliography

Ahlswede R. The capacity of a channel with two senders and two receivers. *Ann. Probab.*, 2:805–814, Oct. 1974.

Ahlwsede R., Cai N., Li S.R., & Yeung R.W. Network information flow. *IEEE Trans. Inf. Theory*, 46(4):1204–1216, Jul. 2000.

Barros J. & Servetto S.D. Network information flow with correlated sources. *IEEE Trans. Inf. Theory*, 52(1):155–170, Jan. 2006.

Berger T. Multiterminal source coding. In *Lecture notes presented at the 1977 CISM Summer School*, pages 171–231, Udine, Italy, Jul. 18-20 1977.

Bergmans P.P. Random coding theorem for broadcast channals with degraded components. *IEEE Trans. Inf. Theory*, IT-19(2):197–207, Mar. 1973.

Biglieri E., Proakis J., & Shamai S. Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inf. Theory*, 44(6):2619–2692, Oct. 1998.

Biswas S. & Morris R. Opportunistic routing in multi-hop wireless networks. *SIGCOMM Computer Commun. Review*, 34(1):69–74, 2004.

Carleial A.B. Multiple-access channels with different generalized feedback signals. *IEEE Trans. Inf. Theory*, IT-28(6):841–850, Nov. 1982.

Chakrabarti A., De-Baynast A., Sabharwal A., & Aazhang B. Low density parity check codes for the relay channel. *IEEE Sel. Areas Commun.: Special Issue on Cooperative Commun. and Netw.*, 25(2):280–291, Feb. 2007.

Chong H.F., Motani M., & Garg H.K. Generalized backward decoding strategies for the relay channel. *IEEE Trans. Inf. Theory*, 53(1):394–401, Jan. 2007.

Cover T.M. & El Gamal A.A. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, IT-25(5):572–584, Sep. 1979.

Cover T.M. & Leung C.S.K. An achievable rate region for the multiple-access channel with feedback. *IEEE Trans. Inf. Theory*, IT-27(3):292–298, May 1981.

Cover T.M. & Thomas J.A. *Elements of Information Theory*. John Wiley and Sons, 1991.

Cover T.M., El Gamal A.A., & Salehi M. Multiple access channels with arbitrarily correlated sources. *IEEE Trans. Inf. Theory*, IT-26(6):648–657, Nov. 1980.

Ezri J. & Gastpar M. On the performance of independently designed ldpc codes for the relay channel. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 977–981, Seattle, Washington, Jul. 9-14 2006.

Fang Q., Gao J., & Guibas L. Locating and bypassing routing holes in sensor networks. In *Proc. 23rd Annual Joint Conf. of the IEEE Computer and Commun. Societies (INFOCOM)*, pages 2458–2468, Hong Kong, Mar. 7-11 2004.

Gastpar M. On source-channel communication in networks. In P. Gupta A.J.v.W., G. Kramer, editor, *DIMACS Workshop on Network Information Theory, Piscataway, NJ*, pages 217–238. Mar. 17-19 2003.

Gastpar M. Cut-set arguments for source-channel networks. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, page 34, Chicago, IL, Jun. 27-Jul. 2 2004.

Gatspar M. & Vetterli M. On the capacity of large Gaussian relay networks. *IEEE Trans. Inf. Theory*, 51(3):765–779, Mar. 2005.

Gopala P.K. & El Gamal H. On the scaling laws of multi-modal wireless sensor networks. In *Proc. 23rd Annual Joint Conf. of the IEEE Computer and Commun. Societies (INFOCOM)*, pages 558–563, Hong Kong, Mar. 7-11 2004.

Gupta P. & Kumar P.R. The capacity of wireless networks. *IEEE Trans. Inf. Theory*, 46(2):388–404, Mar. 2000.

Gupta P. & Kumar P.R. Towards an information theory of large network: an achievable rate region. *IEEE Trans. Inf. Theory*, 49(8):1877–1894, Aug. 2003.

Hu R. & Li J. Practical compress-forward in user cooperation: Wyner-ziv cooperation. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 1673–1667, Seattle, Washington, Jul. 9-14 2006.

Jiang Y.T.M. & Li J. Cluster based routing protocol. IETF Internet Draft. Available on: http://www.comp.nus.edu.sg/~tayyc/cbrp/, Jul. 1999.

Kang W. & Ulukus S. An outer bound for multiple access channels with correlated sources. In *Proc. Conf. on Info. Science and Sys. (CISS)*, pages 7–12, Princeton, NJ, Mar. 22-24 2006.

Khojastepour M.A., Ahmed N., & Aazhang B. Code design for the relay channel and factor graph decoding. In *Proc. 38th Annual Asilomar Conf. on Signal, Syst., and Computers*, pages 2000–2004, Pacific Grove, CA, Nov. 7-10 2004.

King R.C. *Multiple access channels with generalized feedback*. PhD thesis, Stanford Univ., Stanford, CA, Mar. 1978.

Kramer G. & Wijngaarden J.V. On the white Gaussian multiple-access relay channel. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, page 40, Sorrento, Italy, Jun. 2000.

Kramer G., Gastpar M., & Gupta P. Capacity theorems for wireless relay channels. In *Proc. 41st Allerton Conf. on Commun., Control, and Comput.*, pages 1074–1083, Allerton, IL, Oct. 2003.

Kramer G., Gastpar M., & Gupta P. Information-theoretic multi-hopping for relay networks. In *Proc. Int. Zurich Seminar on Commun. (IZS)*, pages 192–195, Zurich, Switzerland, Feb. 18-20 2004.

Kramer G., Gastpar M., & Gupta P. Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inf. Theory*, 51(9):3037–3063, Sep. 2005.

Kurose J.F. & Ross K.W. *Computer Networking: A Top-Down Approach Featuring the Intnernet*. Pearson Education, 2003.

Liao H. *Multiple access channel*. PhD thesis, Univ. Hawaii Honolulu, HI, 1972.

Lim T.L., Srivastava V., & Motani M. Selective cooperation based on link distance estimations in wireless ad-hoc networks. In *Proc. IEEE Int. Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM)*, pages 391–400, Niagara-Falls, Buffalo-NY, Jun. 26-29 2006.

Murugan A.D., Gopala P.K., & El Gamal H. Correlated sources over wireless channels: Cooperative source-channel coding. *IEEE J. Sel. Areas Commun.*, 22(6):988–998, Aug. 2004.

Ong L. & Motani M. Myopic coding in wireless networks. In *Proc. Conf. on Info. Science and Sys. (CISS)*, Baltimore, MD, Mar. 16-18 2005a.

Ong L. & Motani M. Myopic coding in multiple relay channels. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 1091–1095, Adelaide, Australia, Sep. 4-9 2005b.

Ong L. & Motani M. Achievable rates for the multiple access channels with feedback and correlated sources. In *Proc. 43rd Allerton Conf. on Commun., Control, and Comput.*, Allerton, IL, Sep. 28-30 2005c.

Ong L. & Motani M. The capacity of the single source multiple relay single destination mesh network. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 1673–1667, Seattle, Washington, Jul. 9-14 2006a.

Ong L. & Motani M. The multiple access channel with feedback and correlated sources. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 2129–2133, Seattle, Washington, Jul. 9-14 2006b.

Ong L. & Motani M. Optimal routing for decode-and-forward based cooperation in wireless network. In *Proc. 1st Annu. IEEE Conf. on Sensor and Ad Hoc Commun. and Netw. (SECON)*, pages 334–343, San Diego, CA, Jun. 18-21 2007a.

Ong L. & Motani M. Optimal routing for the decode-and-forward strategy in the Gaussian multiple relay channel. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, pages 1061–1065, Nice, France, Jun. 24-29 2007b.

Ong L. & Motani M. On the capacity of the single source multiple relay single destination mesh network. *Ad Hoc Networks (Elsevier): Special Issue on Wireless Mesh Networks*, 5(6):786–800, Aug. 2007c.

Ong L. & Motani M. Coding strategies for multiple-access channels with feedback and correlated sources. *IEEE Trans. Inf. Theory, Special Issue on Models, Theory & Codes for Relaying & Cooperation in Communication Networks*, 53 (10):3476–3497, Oct. 2007d.

Ong L. & Motani M. Myopic coding in multiterminal networks. *IEEE Trans. Inf. Theory*, 54(7):3295–3314, Jul. 2008.

Ozarow L.H. The capacity of the white Gaussian multiple access channel with feedback. *IEEE Trans. Inf. Theory*, IT-30(4):623–629, Jul. 1984.

Ozarow L.H., Shamai S., & Wyner A.D. Information theoretic considerations for cellular mobile radio. *IEEE Trans. Veh. Technol.*, pages 359–378, May 1994.

Razaghi P. & Yu W. Bilayer ldpc codes for the relay channel. In *Proc. IEEE Int. Conf. on Commun. (ICC)*, pages 1574–1579, Istanbul, Turkey, Jun. 11-15 2006.

Salehi M. Improved computational efficiency in derivation of capacity regions in network information theory. *Computers and Electr. Eng. (Elsevier)*, 22(4): 247–256, 1999.

Sendonaris A., Erkip E., & Aazhang B. User cooperation diversity-part i: System description. *IEEE Trans. Commun.*, 51(11):1927–1938, Nov. 2003a.

Sendonaris A., Erkip E., & Aazhang B. User cooperation diversity-part ii: Implementation aspects and performance analysis. *IEEE Trans. Inf. Theory*, 51 (11):1939–1948, Nov. 2003b.

Shakkottai S. Asymptotics of query strategies over a sensor network. In *Proc. 23rd Annual Joint Conf. of the IEEE Computer and Commun. Societies (INFOCOM)*, pages 548–557, Hong Kong, Mar. 7-11 2004.

Shakkottai S., Srikant R., & Shroff N. Unreliable sensor grids: Coverage, connectivity and diameter. In *Proc. 22nd Annual Joint Conf. of the IEEE Computer and Commun. Societies (INFOCOM)*, pages 1073–083, San Francisco, CA, Mar. 30-Apr. 3 2003.

Shannon C.E. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423, 623–656, Jul. and Oct. 1948.

Sklar B. Rayleigh fading channels in mobile digital communication systems part i: Characterization. *IEEE Commun. Mag.*, pages 90–100, Jul. 1997.

Slepian D. & Wolf J.K. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, IT-19(4):471–480, Jul. 1973a.

Slepian D. & Wolf J.K. A coding theorem for multiple access channels with correlated sources. *Bell Syst. Tech. J.*, 52(7):1037–1076, Sep. 1973b.

Telatar E. Capacity of the multiple antenna Gaussian channel. *Europ. Trans. Telecommun.*, 10(6):585–595, Nov. and Dec. 1999.

Toumpis S. & Goldsmith A.J. Capacity regions for wireless ad hoc networks. *IEEE Trans. Wireless Commun.*, 2(4):736–748, Jul. 2003.

Tse D.N.C. & Hanly S.V. Multiaccess fading channels - part i: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Trans. Inf. Theory*, 44(7):2796–2815, Nov. 1998.

van der Meulen E.C. Three-terminal communication channels. *Adv. Appl. Prob.*, 3:120–154, 1971.

Willems F.M.J. *Information theoretical Results For the Discrete Memoryless Multiple Access Channel*. PhD thesis, Katholieke Universiteit Leuven, Belgium, 1982.

Xie L. & Kumar P.R. An achievable rate for the multiple level relay channel. *IEEE Trans. Inf. Theory*, 51(4):1348–1358, Apr. 2005.

Younis O. & Fahmy S. Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach. In *Proc. 23rd Annual Joint Conf. of the IEEE Computer and Commun. Societies (INFOCOM)*, pages 629–640, Hong Kong, Mar. 7-11 2004.

Yu W. Parity forwarding for the relay network. In *Proc. IEEE Commun. Theory Workshop (CTW)*, Dorado, Puerto Rico, May 21-24 2006.

Yu Y., Krishnamachari B., & Prasanna V. Energy-latency tradeoffs for data gathering in wireless sensor networks. In *Proc. 23rd Annual Joint Conf. of the IEEE Computer and Commun. Societies (INFOCOM)*, pages 244–255, Hong Kong, Mar. 7-11 2004.

Yuksel M. & Erkip E. Diversity in relaying protocols with amplify and forward. In *Proc. Global Commun. Conf. (GLOBECOM)*, pages 2025–2029, San Francisco, CA, Dec. 1-5 2003.

Zhang Z., Bahceci I., & Duman T.M. Capacity approaching codes for relay channels. In *Proc. IEEE Int. Symposium on Inf. Theory (ISIT)*, page 2, Chicago, IL, Jun. 27-Jul. 2 2004.

Zhao B. & Valenti M.C. Distributed turbo coded diversity for relay channel. *Electron. Lett.*, 39(10):786–787, May 2003.

Zhao F., Liu J., Liu J., Guibas L., & Reich J. Collaborative signal and information processing: An information directed approach. *Proc. IEEE*, 91(8): 1199–1209, Aug. 2003.