2011

Achievable rates for relay networks using superposition coding

Neevan Kakkan Ramalingam
Iowa State University

Achievable rates for relay networks using superposition coding

by

Neevan Ramalingam

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Program of Study Committee:

Zhengdao Wang, Major Professor

Aleksandar Dogandžić

Sang Wu Kim

Aditya Ramamoorthy

Sung-Yell Song

Iowa State University

Ames, Iowa

2011

Dedicated to my parents,

R.Kulanthaiammal and **Dr.K.Ramalingam**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

It has been a long and rewarding journey for my Ph.D. I learnt a lot of things about research and problem solving skills in general. I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis.

First and foremost, I thank Dr. Zhengdao Wang for his guidance, patience and support throughout this research work. He introduced me to the world of information theory and gave me the opportunity to work with him on many interesting problems over the years. Dr. Wang always provided lot of insights on the problems and inspired me to think in different ways. I thank him for his patience and hours of discussions we had during our meetings. These discussions were definitely useful in terms of my understanding of research methodology and information theory. He continually encouraged me to work on new problems and gave me freedom to work on the problems that intrigued me. His insights and encouragements have often inspired me and renewed my hopes for completing my graduate education. I want to express my gratitude to Dr. Wang for continuous support and help in every aspect of life over the past five years.

I would like to thank my committee members for their efforts and contributions to this work: Dr. Aditya Ramamoorthy, Dr. Sang W. Kim, Dr Aleksandar Dogandžić and Dr. Sung-Yell Song. They taught me lot of fundamental concepts which helped me with the basics of my research. I am grateful to the faculty of Mathematics department, Dr. Maria Axenovich, Dr. Justin Peters and specially Dr. Sung-Yell Song. They taught me advanced mathematical courses as part of my minor. The courses helped me a lot in the basic understanding of mathematical concepts and found significant application to my research problems.

I acknowledge the role of my father Dr.K.Ramalingam in this Ph.D dissertation. This work would not have been possible without his contribution and support. He inspired me to work towards a Ph.D. degree as a career goal and helped me make career choices wisely. He provided valuable feedbacks on all my papers, dissertation and presentations. He has always been a constant motivation to work on research problems and guided me through each and every stage of my career. I would also acknowledge the contributions of my mother who taught me the basic concepts of mathematics and laid a solid foundation early in my life. She spent hours encouraging me to study hard and helped me a lot during my study sessions. I thank my brother Naveen Rajkumar, my sister Sumathi Khanna and brother in law R.Sivakumar for their love, support and guidance. I wouldn't have pursued Ph.D. without my brother's support and inspiration. He always boosted my confidence and morale. He believed in my abilites to pursue research more than anyone else. I would not be at this stage of my life without his contribution. The contributions of my family in shaping my life has been immense and I owe my deepest gratitude to them.

I thank my lab mates and friends Naveen Kumar and Samarjit Das. I have spent hours of academic discussions with them. They would patiently listen to my research problems and guide me with new approaches to tackle the problems. I am also grateful to my labmates Lei Ke, Fardad Raisali, Peng Yu, Hakan Topakayya, Kun Qui, Amit Diwadkar and Vamsi Andalam for creating a wonderful environment for research.

I acknowledge all my wonderful friends. Specifically, Bhupendra Kumar takes care of me like an elder brother. My stay at Ames was memorable with my friends Hemant Bhanot, Priyanka Chand, Kanchan Guntupalli, Rucha Agnihotri, Ashutosh Bhanot, Abhishek Sinha, Abrar Hasan, Narasimha Rao, Anurag Sharda, Swaroop Rajendra, Nikhil Laghave, Saayan Mitra, Shamik Dasgupta, Prashant Karra and many others.

ABSTRACT

We investigate the superposition strategy and its usefulness in terms of achievable information theoretic rates. The achievable rate of the superposition of block Markov encoding (decode-forward) and side information encoding (compress-forward) for the three-node Gaussian relay channel is analyzed. It is generally believed that the superposition can out perform decode-forward or compress-forward due to its generality. We prove that within the class of Gaussian distributions, this is not the case: the superposition scheme only achieves a rate that is equal to the maximum of the rates achieved by decode-forward or compress-forward individually.

We use the insight gathered on superposition forward scheme and devise a new coding scheme. This new scheme is termed as superposition noisy network coding. Superposition noisy network coding combines partial decode-forward with noisy network coding. The novel coding scheme is designed and analyzed for a single relay channel, single source multicast network and multiple source multicast network. The special cases of Gaussian single relay channel and two way relay channel are analyzed for superposition noisy network coding. The achievable rate of the proposed scheme is higher than the existing schemes of noisy network coding, compress-forward and binning.

CHAPTER 1. INTRODUCTION

Transmission of smoke signals is one of the first known forms of human communication and the system used relays. Since then technology has improved so much that we send satellites in space for communication. An important question has been intriguing researchers all along the ages of technology advancements. What is the fastest rate of transmission possible for communication channels with relay? What strategies should be employed to achieve these rates?

Information theory provides vast range of tools to find the theoretical limits on the rates achievable for a given channel. Shannon [3] introduced the ideas of information theory and derived the capacity result for a simple point to point communication channel. The capacity of a point to point AWGN channel is

$$C = W \log \left(1 + \frac{P}{N_0 W}\right) \text{bits/Hz},$$

where W is the bandwidth of the channel, P is the power constraint at the source and N_0 is the noise power spectral density. $\frac{P}{N_0 W}$ is the Signal to Noise Ratio (SNR) of the channel.

An optimum strategy is designed to prove the capacity for a given channel. The strategy achieves a specific rate by maximizing the use of resources available. This is a lower bound on the capacity of the channel and is termed as "achievable rate". It is also shown that there exists no scheme which can achieve a rate higher than a specific rate. This would be an upper bound on the capacity. The upper bound and the lower bound should coincide to the capacity of the channel.

Relay channels are known to provide higher capacity than point to point channels. The relay node is an extra resource which facilitates transmission of information between the source and destination at a higher rate. A simple probabilistic model for a three node terminal was introduced as the relay channel in 1971 [4]. Cover and El Gamal [5], formalized the probabilistic model for the relay channel and introduced several achievable schemes using the random coding argument. These schemes form the basis of the achievable rates for a simple relay channel where the relay node tries to assist the communication between the source and the destination. The main schemes introduced are Decode-Forward (DF), Compress-Forward (CF) and Superposition-Forward (SF). The achievable rates for these schemes do not coincide with the upper bound known for the relay channel. The best known upper bound for the general relay channel is the cut-set bound [6]. The capacity of the relay channel is an open problem for more than forty years. Many advances are being made to understand the existing schemes for the relay channel and to propose novel strategies to achieve a higher transmission rate.

Relay networks are multi terminal systems with relay channel as their fundamental building block. Relay networks are an important system model in multi user information theory. The achievability schemes for relay networks are based on the schemes devised for simple three node relay channel. We look at the lower bounds on the capacity achieved by using superposition encoding. We analyze the advantages of the superposition coding scheme for simple relay channel and also propose new coding schemes based on superposition for relay networks. Let us understand the importance of relay networks and multi user information theory in the present context before we discuss the system model and contributions of this work.

1.1 Motivation

The relay channel has gained renowned interest in modern communication systems. The main characteristics of modern communication systems are

- An increasing number of users or nodes in a network.
- A requirement of higher and higher data rates with increasing application in mobile internet and media streaming.
- The transmission rates are to be improved with minimized power consumption.

The idea is to use the available resources efficiently and to optimally transmit information. Some of the strategies involve resource allocation, interference management and cooperative communication.

The relay channel is a three node communication channel. A source node transmits information to a destination node with the help of a relay. This is similar to a point to point communication channel except for the presence of a relay node. The relay node does not have any information of its own to transmit. It facilitates the transmission of information from source node to destination node possibly at a higher rate than the point to point channel. The strategy employed to make the most efficient and optimal use of the available resource (the relay node in this case) forms the basis for most research on relay channel.

The relay channel is a fundamental building block in network information theory. Recent surge of interest in sensor networks and ad-hoc networks has bought the relay channel problem back into the limelight. Any capacity achieving strategy for a large network would depend on the capacity of its building block, the relay channel. Many transmission schemes for modern networks are based on the strategies introduced for the simple relay networks.

The capacity of the general relay channel is not known completely even after 40 years of research. We improve the achievability rates for relay networks with the relay channel. The best known achievability scheme for the general discrete memoryless relay channel is the superposition-forward scheme. An analysis of superposition-forward is carried out. We propose a better achievability strategy for single source and multiple source discrete memoryless multicast networks. The strategy is termed as superposition noisy network coding. This strategy combines superposition-forward and noisy network coding ideas. The strategy achieves a rate better than the individual schemes of compress-forward, decode-forward and noisy network coding in general.

1.2 System model

We provide a formal definition of the relay channel. The relay channel is a three node network with a single source and a single destination. We also provide the system model for a more general network, the discrete memoryless multi-source multicast network.

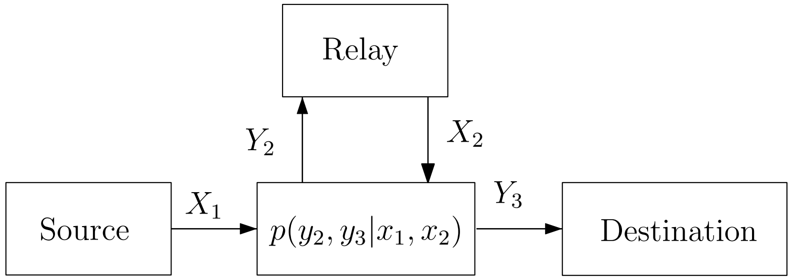

Figure 1.1 Discrete memoryless relay channel

The general discrete memoryless relay channel (DMRC) [5] shown in Fig. 1.1 is denoted by $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3)$, where $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_2, \mathcal{Y}_3$ are finite sets and $p(., .|x_1, x_2)$ is a collection of probability distributions on $\mathcal{Y}_2 \times \mathcal{Y}_3$, one for each $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$; x_1 and x_2 are the transmitted symbols at the source and the relay respectively; y_2 and y_3 are the received symbols at the relay and the destination terminal respectively.

The memoryless property defines the channel model such that the output signal is only dependent on the current channel input for each channel use. That is,

$$p(y_i|x_i, y^{i-1}) = p(y_i|x_i).$$

The system wants to reliably communicate the message $M \in [1, 2^{nR}]$ from source node 1 (X_1) to destination node 3 (Y_3) with the help of the relay node 2 (X_2, Y_2).
The relay can transmit a function of its past received symbols

$$x_{21} = f_i(y_{21}, y_{22}, \ldots, y_{2(i-1)})$$

The relay channel is the building block for large networks. For a more general example, consider an N node discrete memoryless network. A discrete memoryless network shown in Fig. 1.2 is represented by $p(y_1, \ldots, y_N|x_1, x_2, \ldots, x_N)$. x_k is the transmitted symbol and y_k is the received symbol respectively at node k. Each node wants to transmit its message to a set of destination nodes and also act as a relay to transmit messages from other nodes. The relay transmits a function of its past received symbol superimposed over its own message.

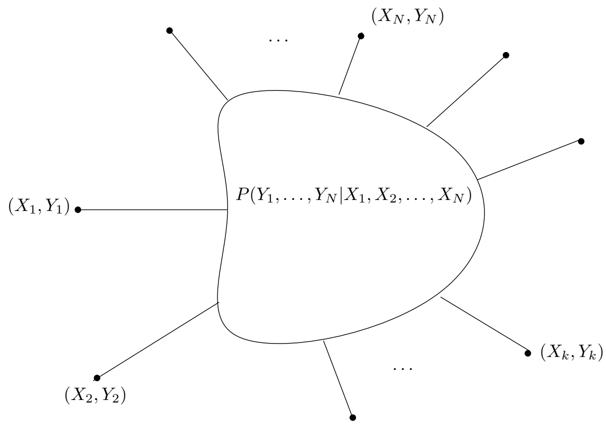

Figure 1.2 An N-node discrete memoryless network.

1.3 Contribution of thesis

We analyze the advantages of using superposition coding to achieve higher rates in relay channel and relay networks. We present analytic results on the scheme of superposition forward. There is a trade-off in the performance of the schemes decode forward and compress forward as a function of the relative position of the relay node. Superposition-forward scheme is a combination of these two schemes and has been thought to provide a potentially higher achievable rate than the existing schemes. The strategy of superposition forward is analyzed with sufficient insight into the fine details. The achievable rate is derived for a Gaussian relay channel model and compared to the existing schemes. Superposition forward schemes is shown to achieve the best of decode forward and compress forward rates for single relay channel. The scheme cannot provide higher rates than the individual schemes but a single unified strategy which achieves the advantages of both CF and DF depending on the position of the relay.

The insights gained from the superposition forward scheme is applied to design a novel coding scheme. This scheme is easily extended to networks which have relays as their building block. The superposition noisy network coding scheme is designed for the simple relay channel and then extended to single source and multiple source multicast networks. The achievable rates are derived for the Gaussian single and two-way relay channels. These rates are compared to the existing schemes numerically. Superposition noisy network coding is shown to outperform the existing schemes for networks with simple relay channels. The contributions of the work along with the results are presented in chapters 4 and 5.

1.4 Organization of dissertation

Literature review of the field of information theory specific to relay channels is provided in Chapter 2. The relay channel was conceptualized and formally defined in the 70's. After the early work where the cut set bound and main cooperation strategies for the relay channel were defined, there was very little work in the 90's. Interest in the problem and in general information theory resurged in 2000's. Many problems have been posed and partially solved, all centering around the idea of multi-user information theory and networks. A highlight of important works carried out in the field of relay Channel and their main results is provided.

The fundamental concepts and tools useful in information theory are introduced in Chapter 3. These are important tools needed for information theoretic analysis of communication channels. The ideas of encoding and decoding used in the random coding argument are also introduced. The concepts introduced here will be used extensively in all subsequent proofs and analysis.

An analysis of the superposition of block Markov encoding (decode-forward) and side information encoding (compress-forward) for the three-node Gaussian relay channel is provided in Chapter 4. We prove that within the class of Gaussian distributions, the superposition scheme achieves a rate that is at most equal to the maximum of the rates achieved by decode-and-forward or compress-and-forward individually. We also present a superposition scheme that combines broadcast with decode-forward, which even though does not achieve a higher rate than decode-and-forward, provides us the insight to the main result mentioned above.

We use our insight on superposition-forward scheme to propose a scheme which improves on the noisy network coding scheme for the Discrete Memoryless Multicast Network (DM-MN) in Chapter 5. The relay nodes decode partial information and use it to make a better compressed signal. The primary idea is not to send any redundant information and use the resources more efficiently. The superposition noisy network coding scheme is explained for a single relay channel and then extended to single source and multiple source multicast networks.

The extension of decode-forward scheme to discrete memoryless multiple source multicast networks is used as a backbone for superposition noisy network coding. This gives better achievable rates for many important special scenarios. Numerical results are provided verifying the advantages of superposition noisy network coding scheme for Gaussian single and two way relay channels.

Conclusion and future work are presented in Chapter 6.

CHAPTER 2. REVIEW OF LITERATURE

Claude E. Shannon founded the field of information theory in his landmark paper "A mathematical theory of communication" [3]. He introduced a mathematical model for a point to point communication channel. One of the major contribution of the work is to answer the question "what is the fundamental limit of communication ?" The idea of channel capacity is introduced and derived for a discrete memoryless channel and additive white Gaussian noise (AWGN) channel.

Channel capacity is the tightest upper bound on the amount of information that can be reliably transmitted over a communications channel. Encoding and decoding schemes are devised to get an achievable rate of information transfer. An achievable rate is termed as the capacity if we can prove that it is not possible to achieve any rate higher than this upper bound.

Various encoding and decoding techniques have been proposed in general for different channels in information theory. These techniques are block Markov encoding, superposition encoding, cumulative encoding, repetitive encoding, successive decoding, joint decoding and list decoding. These techniques will be explained briefly in the next chapter.

A $(2^{nR}, n)$ code for a point-to-point DMC consists of the following:

- An index set $\mathcal{W} := \{1, 2, \ldots, 2^{nR}\}$.
- An encoding function $f_n : \mathcal{W} \rightarrow X^n$.
- A decoding function $g_n : Y^n \rightarrow \mathcal{W}$.

In multi-user information theory, there are two fundamental nature of wireless communication. Broadcast channel where a single source transmits the same message to multiple destinations. Multiple access channel where many sources transmit to a single destination. Interference among different messages is a common phenomenon in multi-user networks. Simple building blocks of a network are analyzed for their capacity limit. The building blocks of a network are broadcast channel, multiple access channel, relay channel and interference channel.

Many coding schemes and achievable rates have been derived for the multiple access, broadcast, interference and the relay channel. The capacity region of the broadcast, interference and the relay channel remains unknown in general. We focus on the capacity of the relay channel and its subsequent application to multi terminal networks. A review of some of the important results on relay channel is presented next.

2.1 Early work

Relay channels were first encountered in the case of satellite communication around 1970. The information signals can be transmitted directly or via satellite. The satellite decodes the information and forwards it either to the destination or to another satellite.

No work on relay channel can go without reference to Van der Meulen's work on 3 terminal channels [4]. The relay channel was conceptualized and simple coding schemes were provided. Different channel conditions were analyzed to show the performance of the schemes. The schemes include multi-hopping, decode-forward, compress-forward and combination of these schemes.

A more comprehensive work on relay channel was provided by Cover and El Gamal in 1979 [5]. Cover and El Gamal derived the cut set upper bound on the capacity C of the relay channel. They introduced the strategies of decode-forward, compress-forward and superposition-forward. These strategies use ideas of superposition encoding, Wyner-Ziv compression, binning and list decoding.

In decode-forward, the relay decodes the message transmitted by the source. The source uses block Markov encoding to instruct the relay on what to transfer in the next block. In the next block, the relay and source coherently transmit the message to the destination. In compress-forward, the relay does not decode the message transmitted by the source. It compresses the received symbol and transmits to the destination. The destination uses the side information provided by the relay and the original message from the source to decode the information. Superposition-forward uses a combination of partial decode-forward and compress-forward to provide a generalized lower bound for the relay channel.

Decode-forward achieves the capacity of degraded relay channel and relay channel with feedback. A channel is degraded if one receiver is a degraded version of the other receiver. In degraded relay channel, the relay receiver y_2 is better than the ultimate receiver y_3. The relay can thus co-operate to send the source message. The relay channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3)$ is said to be degraded if

$$p(y_2, y_3|x_1, x_2) = p(y_2|x_1, x_2)p(y_3|y_2, x_1)$$

that is $X_1 \rightarrow (X_2, Y_2) \rightarrow Y_3$. There is also a case of reversely degraded channel where the relay y_2 is worse than y_3. This channel is less interesting since relay can contribute no new information to the receiver.

Cover and El Gamal also introduced Additive White Gaussian Noise (AWGN) relay channel [5]. The noise added at the relay and destination are independent of each other. The noise is additive Gaussian with zero mean and covariance N. The coding schemes namely decode-forward and compress-forward were derived for the AWGN relay channel under power constraints at the source and the relay.

Another important coding scheme for the relay channel is the Amplify-forward scheme [7]. In amplify-forward, the relay sends a scaled version of its previously received symbol. The amplification is adjusted according to the relay and the source power constraints.

In summary, decode-forward achieves within 1/2 bit of the cut set bound [8] and outperforms compress-forward when relay is close to the source. Compress-forward also achieves within 1/2 bit of cut set [9] and outperforms decode-forward when the relay is close to destination. Amplify-forward achieves within 1 bit of cut set [9] and compress-forward always outperforms amplify-forward.

El Gamal in 1981 [10], worked to generalize the cut set upper bound on the capacity to multi terminal networks. The cut set upper bound is given as

$$C \leq \max_{p(x^N)} \min_{k \in \mathcal{D}} \min_{\mathcal{S}: 1 \in \mathcal{S}, k \in \mathcal{S}^c} H(Y(\mathcal{S}^c)|X(\mathcal{S}^c))$$

Aref further derived the cutset bound for the relay networks and extended the Decode-and-Forward strategy to derive the capacity of a cascade of degraded relay channels [11]. The bound is tight for physically degraded network.

Willems and Carleial [12] [13] introduced new decoding strategies namely backward decoding and sliding window decoding. These strategies provide the same achievable rate as the Decode-and-Forward rate. The advantage of these strategies is ease of analysis and might provide higher rates when used in multiple source networks.

Zhang [14] partially established the capacity for a special class of relay channel where the channel from relay to destination is noiseless.

Liao derived the capacity region of the multiple access channel [15]. Results on the capacity of the broadcast channel and degraded broadcast channel were derived by Cover, Bergman and Gallager [16] [17] [18] [19]. Körner and Marton proposed new coding schemes for the discrete memoryless broadcast channel and derived the capacity of the general broadcast channel with degraded message sets [20] [21]. The main idea of the achievability scheme is to split the message into common message and independent message. The common message is decoded by both the receivers. The independent message is transmitted using superposition over the common message. One of the receiver decodes the superposed information after decoding the common information. This method is referred as successive interference cancellation.

2.2 Recent work

Recent developments in the field of networks has renewed interest in AWGN relay networks. Gupta and Kumar [22] derived fundamental information theoretic limits on traffic carrying capacity of multi hop wireless networks. The network studied is of arbitrary size and topology. Xie and Kumar [23] extended the generalized block Markov encoding scheme to a network of multiple relays in 2004. Different aspects of relaying and co-operation in wireless networks were introduced in 2005 by Kramer, Gastpar and Gupta [24]. One of the main results is the extension of compress-forward to networks.

El Gamal, Mohseni and Zahedi [25] worked on bounds on capacity and minimum energy per bit for AWGN relay channels in 2006. The main results were compress-forward with timesharing and an equivalent representation of the side information lower bound.

$$R_{CF} = \sup \min\{I(X_1; \hat{Y}_2, Y_3|X_2), I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_3)\}$$

This equivalent representation extends naturally to multiple relay networks. Another important result is the capacity of frequency division AWGN relay channels with linear relaying.

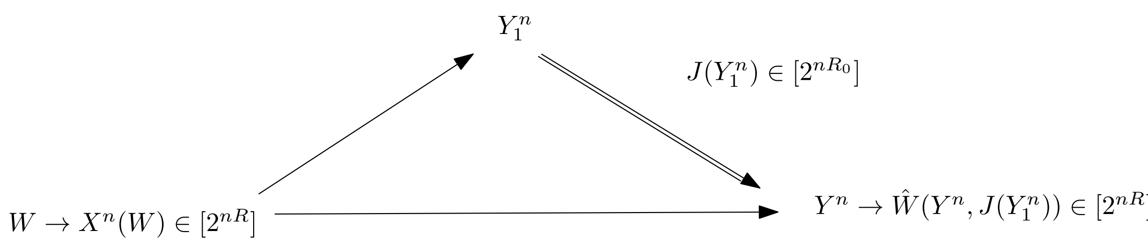

Figure 2.1 Primitive relay channel [1]

Young-Han Kim worked on the coding techniques for primitive relay channels [1]. The primitive relay channel shown in Fig. 2.1 is a simpler version of the more general relay channel. The constant capacity link between the relay and the destination makes it easier to understand and devise new capacity achieving coding schemes. In particular, the schemes of decode-forward, compress-forward and hash-forward are analyzed. The hash-forward scheme uses random binning at the relay as compared to random covering in compress-forward. In the hash-forward scheme, the relay information is summarized in a manner completely independent of geometry (random binning). The destination uses list decoding. In compress-forward, the covers are hamming spheres of radius zero. It is observed that both compress-forward and hash-forward achieve the capacity for the deterministic primitive relay channel where $Y_2 = f(X_1, Y_3)$ [26].

Tse et al. [8] introduced the idea of deterministic channels which provides an easy setup to deal with interference and cooperation. The deterministic channels provide ideas for schemes that achieve capacity within a gap from the cutset bound. The upper and lower bounds coincide for networks with no interference [27] and finite field networks [8].

In a seemingly unrelated but important work, Ahlswede, Cai, Li and Yeung [28] introduced the network coding theorem. This work extends the max-flow min-cut theorem for the noiseless unicast case. The noiseless unicast network is shown in Fig. 2.2 The cut set bound of the

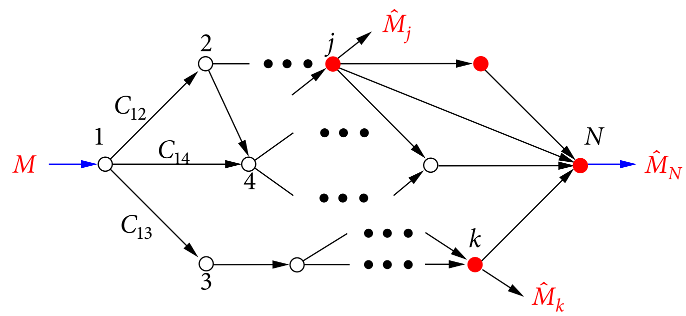

Figure 2.2 Noiseless unicast network [2, Page 16-2]

noiseless unicast network is achieved using forwarding. In general, network coding is required to achieve the cut set bound for noiseless multicast network. The butterfly network is a strong example of how network coding achieves capacity. The network coding theorem is extended to noiseless multi-source multicast networks, erasure multicast networks [29] and to deterministic networks in [8] [27].

L.L.Xie [30] extends the idea of network coding to noisy multiuser channels using the technique of random binning. In this scheme, the relay decodes the messages from more than one source nodes as a multiple access channel. The decoded messages are randomly binned and the index is transmitted to the destination. This scheme provides a general extension of the network coding scheme to noisy multi-terminal networks.

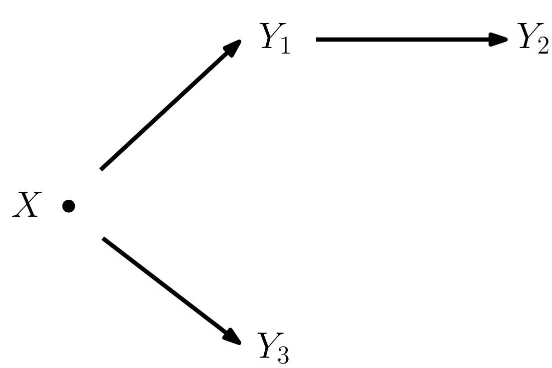

Figure 2.3 3 receiver broadcast channel with degraded message sets

A new concept for a better achievable rate is presented by Nair and El Gamal [31] [32] [33] in their work on broadcast channels. This new idea may be applicable to channels with relay. It is shown that decoding a part of information not required by a node can be useful in decoding the required message at a higher rate. The new achievable rate region for the general broadcast channel is termed as the Nair-Gamal region. The main results and key ideas are presented here.

A 3-receiver broadcast channel with degraded message sets is shown in Fig. 2.3. M_0 is the common message to be transmitted to all the receivers. M_1 is message intended for one of the receiver Y_1. It is shown that the general extension of the Körner-Marton region to more than two receivers is not optimal in general. A new coding scheme is developed which achieves a strictly larger achievable rate region. One of the receiver decodes the common message directly as the cloud center. The second receiver decodes indirectly with the satellite codeword and the third receiver uses joint decoding to decode the transmitted messages.
The Körner-Marton region for a general 3 receiver broadcast channel is

$$R_0 \leq \min\{I(U;Y_1), I(U;Y_2), I(U;Y_3)\} \tag{2.1}$$

$$R_1 \leq I(X;Y_1|U) \tag{2.2}$$

This region is optimal for degraded broadcast channels. It is also optimal for a special case of linear deterministic broadcast channels. For the case under consideration shown in Fig. 2.3, we have $I(U;Y_2) \leq I(U;Y_1)$. The Körner-Marton region reduces to

$$R_0 \leq \min I(U;Y_2), I(U;Y_3) \tag{2.3}$$

$$R_1 \leq I(X;Y_1|U) \tag{2.4}$$

The Nair-Gamal region achieves the following rate region

$$R_0 \leq \min\{I(U;Y_2), I(V;Y_3)\} \tag{2.5}$$

$$R_1 \leq I(X;Y_1|U) \tag{2.6}$$

$$R_0 + R_1 \leq I(V;Y_3) + I(X;Y_1|V) \tag{2.7}$$

The key idea is to still use the message M_0 as the cloud center U but split the message M_1 into two parts and form the satellite messages V and X.

- The receiver Y_1 decodes X to find the messages M_0 and M_1.

- The receiver Y_2 decodes U to find the message M_0.
- The receiver Y_3 decodes V to find the message M_0 indirectly.

The choice of U is such that $I(U;Y_3) \leq I(U;Y_2)$ which necessitates $R_0 \leq I(U;Y_3)$. If we can find a V such that $I(V;Y_3) \geq I(V;Y_2)$, R_0 can be increased to $I(U;Y_2)$. The observation follows from the Markov chains $U - V - X$ and $X - Y_1 - Y_2$.

$$I(V;Y_1) \geq I(V;Y_2) \geq I(U;Y_2)$$

An interesting observation is that decoding extra information at the receiver helps decoding its own information at a higher rate. This works in spite of the fact that the extra information is private information meant for the other receiver. The other interesting channel models analyzed are less noisy and essentially less noisy broadcast channels.

A new strategy of interference alignment [34] is presented next. This strategy can be useful in relay networks with multiple source nodes and managing interference becomes a key issue. The interference alignment scheme is constructed for a deterministic K-user interference channel. The idea is translated to a K-user fully connected real Gaussian interference network. The scheme is shown to achieve the degree of freedom outer bound of K/2. The Gaussian channel is assumed to be real non-zero and constant coefficients with no propagation delays. For the deterministic channel, the desired bits are received without a shift and the interfering bits are received with a one bit shift.

The key ideas of interference alignment are

- Channel coefficients are continuous but vary over time and frequency: The variations create a distinct linear transformation between each sender receiver pair. Thus the same set of signal can align as interference at one receiver but be distinct at the desired receiver.
- Channel coefficients are continuous and fixed: Each node has $m > 1$ antennas. Channel matrices provide distinct spatial rotation that align signal vectors at one receiver and not at another.

- K-user fully connected Gaussian channel with constant but complex coefficients: Interference is aligned in one dimension (imaginary) while the desired signal is received in the other dimension (real).

- Gaussian interference network with constant channel coefficients with propagation delays: The channels have even delays for the desired receiver transmitter pair and odd delays for the undesired receiver transmitter pairs. All transmitter send over even time slots and ensure desired signal is heard interference free at even time slots.

Kim et al [35] quantified the loss in compress-forward relaying without Wyner-Ziv coding at the relay. This provides a good insight on the importance of Wyner-Ziv encoding to achieve higher rates. The main problem addressed is the quantification of the loss in using standard source coding as compared to Wyner-Ziv encoding at the relay. The quantification is done in terms of Diversity Multiplexing Tradeoff (DMT). The channels are assumed such that the relay has perfect knowledge of all the three channel coefficients and that the relay makes use of Wyner-Ziv source coding with side information. Using source coding without side information results in a significant loss in terms of DMT. For a given constraint on multiplexing gain, the loss can be compensated using power control at the relay. Otherwise the loss remains significant.

Lim et al. [36] combined the compress-forward strategy with network coding to propose noisy network coding. This scheme uses repetition coding, no Wyner-Ziv binning and joint decoding as key concepts. The noisy network coding scheme is shown to achieve better rates for networks with more than one source. General achievability for multiple source multicast networks is provided using noisy network coding.

More recently, Xie [37] analyzed the Noisy Network coding scheme and showed that the advantage of joint decoding is useful only in network with more than one source. It is proved that successive decoding, joint decoding and backward decoding all provide the same rates as noisy network coding irrespective of repetition encoding or successive encoding.

In our next chapter, We introduce the methods and tools required for information theoretic analysis of channel models. The state of the art encoding and decoding schemes are introduced.

Important theorems and lemmas are introduced with their significance and relevance to common problems. These tools will be extensively used in the analysis of superposition forward and design of superposition noisy network coding schemes.

CHAPTER 3. TOOLS OF THE TRADE

The basic tools and concepts used in information theory are explained in brief here. The tools provide insight on the random coding argument and form the basis for all information theoretic analysis. They are necessary to understand the proofs and information theoretic analysis of relay networks. The reader is referred to [2] [6] [38] [39] for detailed proofs of the techniques introduced here.

3.1 Introduction

The information theoretic analysis of channel models relies on the typicality of sequences that are stochastically generated. The Asymptotic Equipartition Property (AEP) divides all stochastically generated sequences into typical and non typical sets. The AEP is a direct consequence of the law of large numbers. We begin with the law of large numbers and introduce the AEP and typical sets. Once the idea of typical sets is clear, important lemmas namely joint typicality, conditional and covering lemmas are introduced. These lemmas would be used extensively in probability of error analysis. We next look at basic techniques for encoding and decoding messages at different nodes.

3.2 Preliminaries

3.2.1 Typical sets

Consider a set of numbers $X_1, X_2, \ldots, X_n$ generated independently and identically distributed (i.i.d) with probability distribution $p(X)$. The Law of Large Numbers (LLN) states that the average of these numbers is close to the expected value for large n.

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E\{X\}$$

As a direct consequence of the law of large numbers, we have the AEP

$$\frac{1}{n}\log\frac{1}{p(X_1, X_2, \ldots, X_n)} \xrightarrow{p} H(X)$$

where $H(X)$ is the entropy of X. For large n, the AEP states that given a stochastic source, the sequence of numbers generated can be any one of the random possibilities but the one that is actually generated lies within a set where all the sequence have equal probability of being generated. It is with high probability that the sequence generated will lie within this set.

AEP divides all sequences into two sets, typical set and non-typical set. The typical set $\mathcal{T}_\epsilon^{(n)}$ has the following properties [2] [6] [38]

1. The probability of a sequence in the typical set is given by

$$2^{-n(H(X)+\delta(\epsilon))} \leq p(X_1, X_2, \ldots, X_n) \leq 2^{-n(H(X)-\delta(\epsilon))}$$

2. For sufficiently large n, the probability that a stochastically generated sequence will lie in this set is close to one.

$$\Pr\{X \in \mathcal{T}_\epsilon^{(n)}\} > 1 - \epsilon$$

3. For sufficiently large n, the number of sequences or the cardinality of the typical set is given by

$$(1-\epsilon)2^{n(H(X)-\epsilon)} \leq |\mathcal{T}_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$$

We next look at a similar property for a bivariate distribution

3.2.2 Joint typicality

The joint AEP for bivariate sequences X, Y distributed with probability distribution $p(x, y)$ states

$$\frac{1}{n}\log\frac{1}{p(X^n, Y^n)} \xrightarrow{p} H(X, Y).$$

The jointly typical set $\mathcal{T}_\epsilon^{(n)}$ for sequences (X^n, Y^n) drawn i.i.d. according to $p(x^n, y^n)$ has the following properties [2] [6] [38]

1. For sufficiently large n,

$$\Pr\{(X^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}\} > 1 - \epsilon$$

2. The cardinality of jointly typical set is given by

$$|\mathcal{T}_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$$

3. If X and Y are independent with same marginals as $p(x^n, y^n)$, then for large n

$$(1-\epsilon)2^{-n(I(X,Y)+3\epsilon)} \leq \Pr\{(X^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}\} \leq 2^{-n(I(X,Y)-3\epsilon)}$$

where $I(X;Y)$ is the mutual information between the variables X and Y.

Following the properties of a jointly typical sequences, we state two important lemmas which would be used extensively in all the proofs.

3.2.2.1 Conditional typicality lemma [2] [6] [38]:

The conditional typicality lemma states that for a given typical sequence x^n and another sequence Y^n generated for each sequence x^n according to probability distribution $\prod_{i=1}^{n} p_{Y|X}(y_i|x_i)$, the probability that x^n and Y^n are jointly typical is close to one for sufficiently large n.

$$\Pr\{(x^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X,Y)\} \to 1 \quad \text{as} \quad n \to \infty$$

and for all typical sequences x^n,

$$|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \geq (1-\epsilon)2^{n(H(Y|X)-\delta(\epsilon))}$$

for $Y^n \sim \prod_{i=1}^{n} p_{Y|X}(y_i|x_i)$.

The conditional typicality lemma states the number of jointly typical sequences Y^n which are generated given the typical sequences x^n gets close to $2^{n(H(Y|X))}$ as $n \to \infty$. This is shown graphically in Fig. 3.1. The result is useful in the channel coding argument and the proofs of achievability theorems.

We next state another lemma useful in the achievability proofs and follows from the joint typicality of random triples.

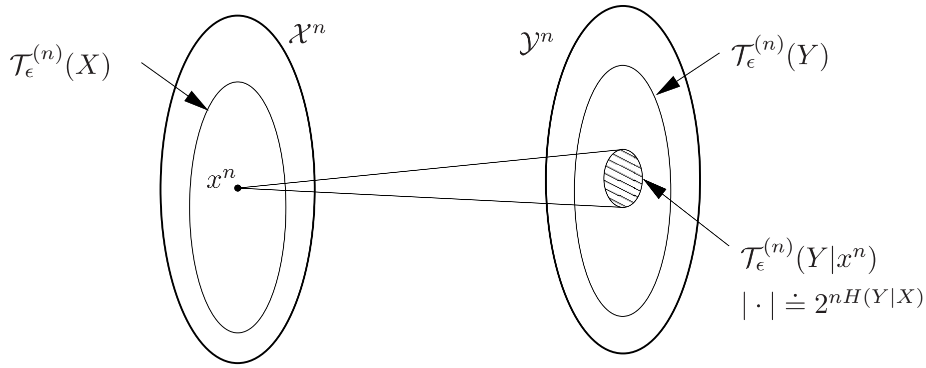

Figure 3.1 Conditional typicality lemma [2, Page 2-21]

3.2.2.2 Joint typicality lemma [2] [6] [38]

Given a sequence of random triples $(X, Y, Z) \sim p(x, y, z)$

Let $(X, Y, Z) \sim p(x, y, z)$. Then there exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that, given $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)$, let $\tilde{Z}^n$ be distributed according to $\prod_{i=1}^n p_{Z|X}(\tilde{z}_i|x_i)$ (instead of $p_{Z|X,Y}(\tilde{z}_i|x_i, y_i)$). Then

- $P\left\{(x^n, y^n, \tilde{Z}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\right\} \leq 2^{-n(I(Y;Z|X)-\delta(\epsilon))}$
- for sufficiently large n, $P\left\{(x^n, y^n, \tilde{Z}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\right\} \geq (1-\epsilon)2^{-n(I(Y;Z|X)+\delta(\epsilon))}$

The detailed proof for joint typicality lemma can be found at [2], [6]. The joint typicality lemma implies that given two jointly typical sequences (x^n, y^n), and another sequence z^n generated according to distribution $\prod_{i=1}^n p_{Z|X}(\tilde{z}_i|x_i)$. The probability of z^n to be jointly typical with the sequences (x^n, y^n) is close to $2^{-n(I(Y;Z|X)}$.

Next we look at Shannon's channel coding theorem. The theorem provides a useful result on the limits of information transfer between two points. The random coding argument uses conditional typicality and joint typicality lemmas. The proof provides the general idea and concepts used for achievability proofs in information theory.

3.3 Channel coding

The general discrete memoryless channel is shown in Fig. 3.2 and defined in [3]. It is represented by $(\mathcal{X}, p(y|x), \mathcal{Y})$, where $\mathcal{X}, \mathcal{Y}$ are finite sets and $p(.|x)$ is a collection of probability distributions on $\mathcal{Y}$, one for each $x \in \mathcal{X}$; x is the transmitted symbol at the source and y is the received symbol at the destination terminal.

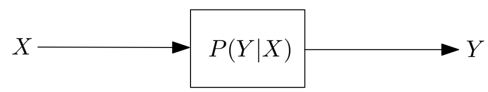

Figure 3.2 Point to point channel model

An (M, n) code for the channel consists of a set of integers $\mathcal{M} = \{1, 2, \ldots, M\}$, an encoding function $x_1 : \mathcal{M} \to \mathcal{X}_1^n$ and a decoding function $g : \mathcal{Y}_3^n \to \mathcal{M}$.
Define $\lambda(w) = p(g(Y) \neq w)$ as the probability of error of the decoding function of the channel and let λ_n be the maximal probability of error over all possible messages w. The rate $R = (1/n) \log M$ of an (M, n) code is said to be achievable by a channel if for any $\epsilon > 0$ and for sufficiently large n, there exists a code with $M \geq 2^{nR}$ such that $\lambda_n < \epsilon$. A similar definition of achievable rate can be made for relay channels.

The channel coding theorem [6, Theorem 8.7.1]: All rates below the capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$. Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$. The capacity C for a point to point channel is

$$C = \max_{p(x)} I(X; Y)$$

The brief sketch of the proof can be seen from the packing lemma which uses conditional typicality and joint typicality lemmas.

3.3.1 Packing lemma

The encoding and decoding process for transmitting information reliably over a point to point channel is described here. A detailed reference for the packing lemma is provided in [2].

Codebook generation: Generate 2^{nR} i.i.d. sequences x^n according to probability distribution $p(X)$. The codebook is revealed to the receiver so that it knows all the possible sequences that could be transmitted at the source.

With randomly generated symbols x^n transmitted at the source. We know that these sequences lie in the typical set by the way they are generated. At the receiver, we look for a unique codeword x^n that is jointly typical with the recieved symbol. There are $2^{n(H(Y|X))}$ possible sequences that are jointly typical and equally likely given x^n is transmitted. We don't want two X sequences producing the same sequence Y. Then we won't be able to decode the message X at the destination uniquely. The total number of possible typical Y sequences is close to $2^{nH(Y)}$. This set is divided into $2^{n(H(Y|X))}$ disjoint sets which are the possible sequences at the receiver given X is transmitted. The total number of such disjoint sets is given by $2^{nI(X;Y)}$ which is the number sequences that can be transmitted reliably. Fig. 3.3 provides a graphical representation of the packing lemma.

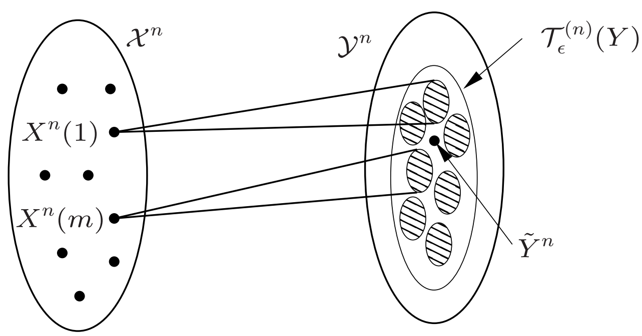

Figure 3.3 Packing lemma [2, Page 3-19]

We next look at another lemma in the context of lossy source coding. The idea is to send a message X and decode $\hat{X}$ which is the reproduction of the symbol X with some pre defined distortion level. The concept is also useful in sending compressed information over a rate limited channel.

3.3.2 Covering lemma

We look at more general lossy source coding problem where the source and destination have knowledge of some side information U. Let X^n represent the source sequence that is transmitted. Let $(U^n, X^n) \sim p(u^n, x^n)$ be jointly typical random sequences. Let $\hat{X}^n$ be conditionally independent of X^n given U^n and distributed according to $\prod_{i=1}^{n} p_{\hat{x}|u}(\hat{x}_i|u_i)$. Then,

$$\Pr\{(U^n, X^n, \hat{X}^n) \notin \mathcal{T}_\epsilon^{(n)}\} \to 0 \quad as \quad n \to \infty$$

if $R > I(X; \hat{X}|U)$.

Fig. 3.4 shows the graphical representation of the covering lemma [2] for the lossy source coding case with no side information. The lemma is to shows the existence of at least one reproduction sequence that is jointly typical with X^n provided $R > I(X; \hat{X}|U)$.

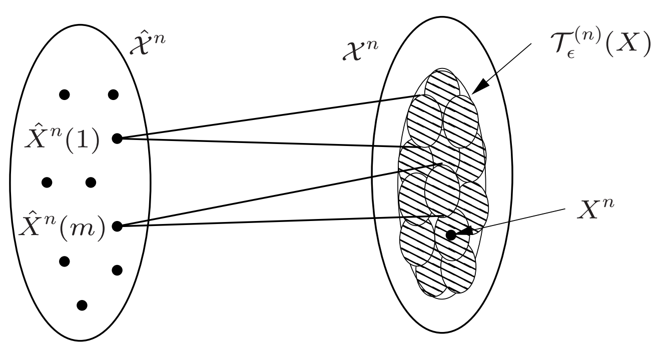

Figure 3.4 Covering lemma [2, Page 3-52]

3.3.3 Binning

Binning is a source encoding scheme [6]. Given a random variable X generated according to distribution $p(x)$, we know that there are $2^{nH(X)}$ typical sequences. To transmit the X symbols, the encoder simply bins all the sequences into 2^{nR} bins. The bin index is then transmitted for the symbols. The decoder decodes the bin index and looks for typical sequences in the bin index. If we make R large enough, then we can ensure that there is only one typical sequence in each bin. Thus, using the binning technique, we can recover the symbols X reliably if $R > H(X)$.

We next look at the distributed source coding problem. There are two sources sending correlated messages X and Y to the destination. The encodings are done separately at the sources, without the knowledge of what is being transmitted at the other source. By the binning scheme described before, a rate $R_1 > H(X)$ is sufficient to recover the messages X. A rate $R_2 > H(Y)$ is sufficient to recover the messages Y. A rate of $H(X) + H(Y)$ is sufficient to recover both messages when they are encoded separately. Slepian and Wolf [40] showed that a rate of $H(X, Y)$ is sufficient to recover correlated messages even if the sources are encoded separately. A brief outline of idea of the proof is provided here. Each source randomly bins its messages into 2^{nR_1} and 2^{nR_2} bins. The destination terminal receives both the indices and looks for sequences X and Y that are jointly typical. There are $2^{nH(X,Y)}$ such typical sequences.

How do we send the messages Y at the rate $H(Y|X)$ when the messages X are sent at rate $H(X)$? This is possible even without the Y encoder knowing the X messages transmitted at the other encoder. The Y encoder is required to bin the messages in the typical set of $Y|X$. Since the Y encoder does not know the messages X or the typical set $Y|X$, it randomly bins all the messages Y into 2^{nR_2} bins. If $R_2 > H(Y|X)$, then with high probability there would be only one typical Y sequence in the set of typical $Y|X$ sequences at the decoder.

3.3.4 Wyner-Ziv coding

We next look at the source coding problem with side information [41] shown in Fig. 3.5. Two messages X and Y are encoded separately. Only X is to be recovered at the destination. If R_2 is the rate allowed for Y, what is the rate required for transmitting X reliably. If the rate $R_2 > I(Y; U)$ for some auxiliary random variable U, then X can be reliably transmitted with rates $R_1 > I(X; U)$. The random variables X-Y-U form a Markov chain.

Now, how many bits are required to transmit the messages X within a distortion D given the side information Y at the decoder? The side information Y is correlated with the transmit messages X. In Wyner-Ziv [41] binning scheme, the source looks for a jointly typical codeword with the message and then sends the bin index of the typical codeword. The decoder looks for a typical codeword in the decoded bin and uses the side information to zero in the unique codeword. The key idea is to keep the typical codewords in each bin to be small enough such

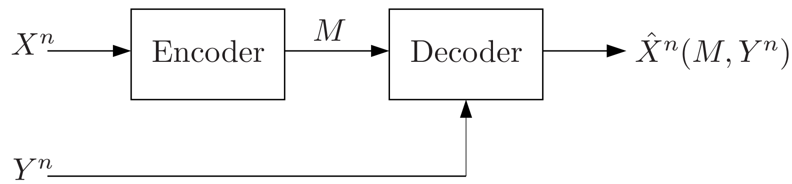

Figure 3.5 Source coding with side information [2, Page 12-16]

that side information can be used to identify the unique codeword within the allowed rate distortion. The following rates are achievable by using the Wyner-Ziv binning scheme for rate distortion with side information

$$R_Y(D) = \min_{p(w|x)} \min_{f} (I(X;W) - I(Y;W))$$

where f is the function mapping decoded codeword and side information to the output sequence $\hat{X}$.

3.4 Encoding

In this section, we provide a brief description of different encoding techniques used at the source.

3.4.1 Superposition encoding and block Markov encoding

In block Markov encoding scheme [2] [6], messages are transmitted over B blocks, each of n symbols. A sequence of $B-1$ messages, is sent over the channel in nB transmissions. Codebooks are generated randomly and independently for each block. All the messages transmitted in a given block carry new messages and are statistically dependent on messages transmitted in the previous block. Thus messages decoded in the previous block are also transmitted without interfering with the new messages being transmitted in the current block. This is possible by using superposition encoding [2] [6]. This scheme is specifically advantageous in the relay channel where the relay transmits the symbols it decoded in the previous block.

A general idea of superposition encoding is to superimpose the new messages m_1 over the message from the previous block m_2.

Fix $p(u)p(x|u)$. Generate 2^{nR_2} independent codewords $u^n(m_2)$ i.i.d. drawn according to distribution $p(u)$. These codewords form the cloud centers in the codebook. For each $u^n(m_2)$, generate 2^{nR_1} conditionally independent codewords $x^n(m_1, m_2)$ according to distribution $p(x|u)$. These would be the satellite codewords in the codebook. The codebook generation is explained graphically in Fig. 3.6.

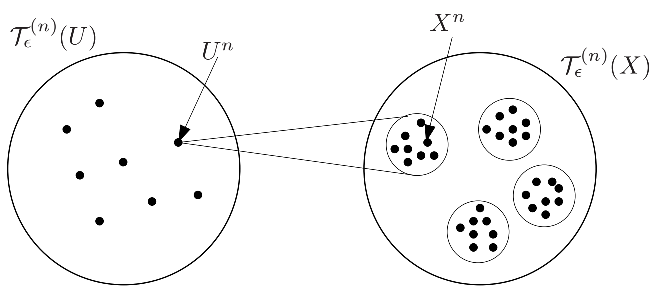

Figure 3.6 Superposition encoding [2, Page 5-14]

The destination decodes the messages in a successive way. It first decodes the cloud center $u^n(m_2)$ in one block. The destination also decodes the satellite codeword given the cloud center if the channel conditions are good.

3.4.1.1 Irregular encoding

The encoding scheme known as irregular encoding [5] uses superposition and is described below

Codebook generation: Fix $p(u)p(x|u)$.

- Randomly and conditionally independently generate 2^{nR_1} codewords $U(s)$ distributed as $\prod_{i=1}^n u_i$.
- For each $U(s)$, randomly and conditionally independently generate 2^{nR} codewords $X(m|s)$ according to $\prod_{i=1}^n p(x_i|u_i(s))$

This defines the codebook which is revealed to both transmitter and receiver. Randomly bin the codewords m into the 2^{nR_1} bins. The cell index for each message is now given as $s(m)$.

Encoding: If the message transmitted in block $b-1$ is m_{b-1}, find the bin index corresponding to the codeword, $s_b(m_{b-1})$. Then transmit $X(m_b|s_b)$.
The decoding can be done using successive or joint decoding techniques. The decoding techniques will be discussed in the next section.

3.4.1.2 Regular encoding

The regular encoding scheme [39] also uses superposition and block Markov encoding. The difference in the scheme is that it does not use binning. Instead, 2^{nR} U codewords are generated which would be the cloud center consisting of previous messages.
Codebook generation: Fix $p(u)p(x|u)$.

- Randomly and conditionally independently generate 2^{nR} codewords $U(s)$ distributed as $\prod_{i=1}^{n} u_i$.
- For each $U(s)$, randomly and conditionally independently generate 2^{nR} codewords $X(m|s)$ according to $\prod_{i=1}^{n} p(x_i|u_i(s))$

This defines the codebook which is revealed to both transmitter and receiver.
Encoding: If the message transmitted in block $b-1$ is m_{b-1}, Then transmit $X(m_b|m_{b-1})$ in block b.

The total codewords used in regular encoding is larger than the total number of codewords required in irregular encoding.

3.4.2 Message repetition encoding

Message repetition encoding transmits the same message over each of the B blocks. The message is now from the set of size 2^{nBR_1} and decoding is done after B blocks of transmission using joint decoding. This allows collaboration among all the blocks and each block provides a better estimate of the message to be decoded. The message repetition scheme is introduced and used in noisy network coding [36]. Noisy network coding and message repetition will be explained more deeply in chapter 5.

3.5 Decoding

Decoding is the process employed to find a unique transmitted message with the probability of error going to zero. In general we look for a jointly typical transmitted codeword with the received symbol. The decoding technique can vary depending on whether we chose to decode the message after every block or after b blocks of transmission. There are many possible decoding schemes which are explained in brief here.

3.5.1 List decoding

A $(2^{nR}, 2^{nL}, n)$ code for a discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of an encoder mapping messages $m \in 2^{nR}$ to a set of codewords X. The decoder after receiving y^n tries to find a list of codewords $\mathcal{L}(y^n) \subset 2^{nR}$ of size $|\mathcal{L}| \leq 2^{nL}$ that contains the transmit message. An error occurs if the list does not contain the transmit message.

List decoding is useful in the relay channel where the destination decodes a list of possible messages transmitted by the source in block $b-1$. The destination used the additional information sent by the relay in block b to look for a unique codeword within the list of codewords decoded.

3.5.2 Successive decoding

The succesive decoding for relay channels was introduced in [5]. For two messages transmitted at the source using superposition encoding, the decoding is performed in two steps

- The decoder decodes the cloud center of the transmitted message. $\hat{m}_2$ is declared to be the message transmitted if

$$(u(\hat{m}_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$$

- After decoding m_2, the decoder then finds a unique message m_1, the satellite codeword such that

$$(u(\hat{m}_2), x(\hat{m}_1|\hat{m}_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$$

If such a typical sequence is not uniquely found, an error is declared. The achieved rate can be different depending on which message is decoded first.

3.5.3 Joint decoding

A detailed description of joint decoding is provided in [2]. The decoder declares that the messages $(\hat{m}_1, \hat{m}_2)$ were transmitted if

$$(u(\hat{m}_2), x(\hat{m}_1|\hat{m}_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$$

The error events correspond to the following situations

- Both the original codewords are not typical, this probability of error goes to zero by conditional typicality lemma.
- Message m_1 is jointly typical while an error is made in finding the unique jointly typical m_2. This achieves one of the corner points of successive decoding.
- Message m_2 is jointly typical while an error is made in finding the unique jointly typical m_1. This achieves the other corner points of successive decoding.
- Error is made in decoding both the messages. This gives a bound on the sum rate of both the messages and this is achieved by timesharing in successive decoding.

Joint decoding is stronger than successive decoding since it does not need time sharing to achieve all the points in the achievable region.

3.5.4 Backward decoding

Decoding at the receiver is done backwards after all the b blocks have been decoded [13] [12]. The decoder uses the estimate of message transmitted in block b to decode the message transmitted in block $b-1$. Backward decoding has a delay of B blocks.

3.6 Conclusion

Now, we have a good understanding of the tools useful in information theory and especially in information theoretic analysis of relay channels. We will use these tools to derive the achievability region for the relay channel and develop new encoding and decoding schemes.

An analysis of superposition-forward is provided in the next chapter. This strategy has not been understood completely so far and is believed to provide better rates than decode-forward or compress-forward. The chapter is followed by introduction of superposition noisy network coding which is an extension of superposition-forward scheme with noisy network coding. The generalization of decode-forward and superposition noisy network coding to single source and multiple source multicast networks is provided.

CHAPTER 4. SUPERPOSITION-AND-FORWARD FOR AWGN RELAY CHANNEL

We analyze the achievable rate of the superposition of block Markov encoding (decode-forward) and side information encoding (compress-forward) for the three-node Gaussian relay channel. It is generally believed that the superposition can out perform decode-and-forward or compress-and-forward due to its generality. We prove that within the class of Gaussian distributions, the superposition scheme only achieves a rate that is equal to the maximum of the rates achieved by decode-forward or compress-forward individually. We also present a superposition scheme that combines broadcast with decode-forward, which even though does not achieve a higher rate than decode-forward, provides us the insight to the main result mentioned above.

4.1 Introduction

The relay channel, introduced by van der Meulen [4] is a fundamental building block in network information theory. It consists of a relay terminal assisting communication between a source terminal and a destination terminal, facilitating a higher data rate than a point to point channel. Cover and El Gamal [5] introduced two new coding strategies and a cut-set upper bound for the relay channel. They derived the capacity of the degraded and reversely degraded relay channels. Capacity results have been derived for special cases of the relay channel like the semi-deterministic case [42] but the capacity of the general relay channel is still unknown.

The main achievability strategies known for the relay channel are decode-forward and compress-forward [5]. The DF scheme is also known as the general block Markov encoding scheme. The relay decodes the transmitted message and jointly transmits the message from the source to the destination terminal. The DF strategy is optimal and achieves the cut-set bound when the source to relay channel is strong. The CF scheme is known as the side-information encoding scheme. The relay compresses the received signal without decoding and transmits to the destination terminal. The destination terminal treats the compressed information as side information and decodes the original message. The CF scheme is asymptotically optimum and achieves the cut-set bound when the relay to destination channel is strong. This allows the received signal at the relay to be conveyed faithfully to the destination. A combination of the two strategies that superimposes DF and CF was also proposed in [5, Theorem 7]. We refer to this scheme as superposition-forward. The SF scheme achieves the capacity for the special cases of degraded, reversely degraded and semi-deterministic relay channels. Due to the generality of the result in [5, Theorem 7], it is expected it can offer higher achievable rates than DF or CF alone.

In this chapter, we investigate the coding scheme for the general Gaussian relay channel. The initial motivation for the work was to develop new coding strategies with higher achievable rates. A new coding strategy was designed which superimposes decode-forward and broadcast, as presented in Section 4.3. The scheme unfortunately yields a rate that is inferior to DF. This attempt, though not successful, prompted us to investigate the general superiority of SF, especially for the Gaussian relay channel. It is found that for Gaussian relay channel, within the class of Gaussian distributions, the SF can achieve at most the larger rate achievable by DF or CF alone — there is no need to do superposition for Gaussian distributions (Section 4.4). We also provide a numerical example that verified the theoretical result in Section 4.5. Section 4.6 concludes the chapter.

Notation: For random variables X, Y, Z, we use $p(x, y, z)$ to denote the joint distribution, when there is no confusion, as a short cut to $p_{X,Y,Z}(x, y, z)$. When X and Z are conditionally independent given Y (i.e., X, Y, and Z form a Markov chain), we write $X - Y - Z$.

4.2 Preliminaries

We present the mathematical models for the discrete-memoryless and Gaussian relay channels in this section, and also include the known results on achievable rates that will be used later.

4.2.1 Discrete memoryless relay channel

The general discrete memoryless relay channel (DMRC) is the same as defined in [5]. A brief description is given here for completeness. The DMRC is denoted by $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3)$, where $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_2, \mathcal{Y}_3$ are finite sets and $p(.,.|x_1, x_2)$ is a collection of probability distributions on $\mathcal{Y}_2 \times \mathcal{Y}_3$, one for each $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$; x_1 and x_2 are the transmitted symbols at the source and the relay respectively; y_2 and y_3 are the received symbols at the relay and the destination terminal.

An (M, n) code for the relay channel consists of a set of integers $\mathcal{M} = \{1, 2, \ldots, M\}$, an encoding function $x_1 : \mathcal{M} \to \mathcal{X}_1^n$ a set of relay functions $\{f_i\}_{i=1}^n$ such that

$$x_{2i} = f_i\left(Y_{21}, Y_{22}, \ldots, Y_{2(i-1)}\right), \quad 1 \leq i \leq n,$$

and a decoding function $g : \mathcal{Y}_3^n \to \mathcal{M}$. The joint probability mass function on $\mathcal{M} \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}_2^n \times \mathcal{Y}_3^n$ is

$$p(w, x_1, x_2, y_2, y_3) = p(w) \prod_{i=1}^{n} p(x_{1i}|w) p(x_{2i}|y_{21}, y_{22}, \ldots, y_{2i-1}) p(y_{2i}, y_{3i}|x_{1i}, x_{2i}). \tag{4.1}$$

Define $\lambda(w) = p(g(Y) \neq w)$ as the probability of error of the decoding function of the relay channel and let λ_n be the maximal probability of error over all possible messages w. The rate $R = (1/n)\log M$ of an (M, n) code is said to be achievable by a relay channel if for any $\epsilon > 0$ and for sufficiently large n, there exists a code with $M \geq 2^{nR}$ such that $\lambda_n < \epsilon$.

The dependency graph of the discrete memoryless relay channel is shown in Fig. 4.1

4.2.2 Gaussian relay channel

Fig. 4.2 shows the Gaussian relay channel model that we will be using. The received symbols at the relay and the destination terminal are given respectively by

$$Y_2 = aX_1 + Z_1 \tag{4.2}$$

$$Y_3 = X_1 + bX_2 + Z_2 \tag{4.3}$$

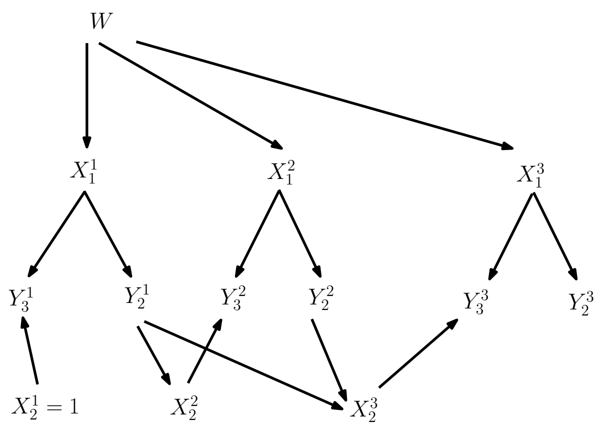

Figure 4.1 Dependency graph for the relay channel

where the noise terms Z_1 and Z_2 are uncorrelated zero mean Gaussian random variables with variances N_1 and N_2 respectively, and a and b are the channel gain constants. As a result, we have

$$p(y_2, y_3|x_1, x_2) = \frac{1}{2\pi\sqrt{N_1 N_2}} \exp\left[-\frac{(y_2 - ax_1)^2}{2N_1} - \frac{(y_3 - x_1 - bx_2)^2}{2N_2}\right], \tag{4.4}$$

which will be the channel assumed throughout the chapter.

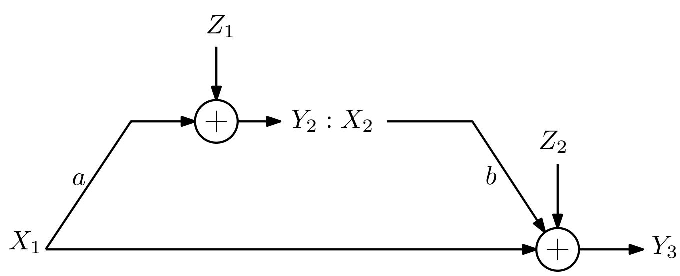

Figure 4.2 Gaussian relay channel

The average power constraints at the transmitters are

$$\frac{1}{n}\sum_{i=1}^{n} x_{1i}^2(k) \leq P_1, \quad \forall k \in \mathcal{M}, \tag{4.5}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} x_{2i}^2 \leq P_2, \quad \forall y_2^n \in \Re^n. \tag{4.6}$$

4.2.3 Known achievable rates

We briefly review the known results for DF, Partial-DF, CF, and the SF.

4.2.3.1 Decode-forward

For DMRC, the DF scheme achieves any rate less than [5, Theorem 1]

$$R_{DF} = \sup \min\{I(X_1; Y_2|X_2), I(X_1, X_2; Y_3)\} \tag{4.7}$$

where the supremum is taken over all possible $p(x_1, x_2)$. A brief description of the encoding and decoding algorithm along with the probability of error analysis is provided.

Codebook Generation: Consider p_{X_1,X_2}. We use regular encoding and superposition encoding. For each block $b = 1, \ldots, B+1$,

- generate 2^{nR} codewords $x_{2b}^n(v)$ i.i.d. using p_{X_2} for all $v = 1, \ldots, 2^{nR}$.
- generate 2^{nR} codewords $x_{1b}^n(v, m)$ i.i.d. using $p_{X_1|X_2}$ for all $m = 1, \ldots, 2^{nR}$.

Source Transmission: In block b, the source transmits the codeword $x_{1b}^n(m_{b-1}, m_b)$.

Relay Terminal: Checks for a typical $\tilde{m}_b$ using $\hat{m}_{b-1}$ and y^n

$$\mathbf{x}_{1b}^n(\hat{m}_{b-1}, \tilde{m}_b), \mathbf{x}_{2b}^n(\hat{m}_{b-1}), y_{2b^n} \in \mathcal{T}_\epsilon^{(n)}$$

An error is declared if no such message is found or more than one such message is jointly typical. Using conditional typicality, joint typicality and Markov lemmas, it can be shown that the probability of error is close to zero for sufficiently large n if

$$R < I(X_1; Y_2|X_2)$$

This can be interpreted as the rate at which messages can be reliably transmitted from the source to the relay.

Sink Terminal: We use sliding window decoding which is same as backward decoding but the delay is only 2 blocks now. The terminal uses the symbols received in two blocks to decode one message. The coding scheme was introduced by Carleial [13] in 1982. The sink terminal uses $y_{3,b-1}^n, y_{3b}^n, \hat{m}_{b-2}$ to find a jointly typical message $\tilde{m}_{b-1}$ such that

- $\{\mathbf{x}_{1,b-1}^n(\hat{m}_{b-2}, \tilde{m}_{b-1}), \mathbf{x}_{2,b-1}^n(\hat{m}_{b-2}), \mathbf{y}_{3,b-1}^n\} \in \mathcal{T}_\epsilon^{(n)}$
- $\{\mathbf{x}_{2,b}^n(\tilde{m}_{b-1}), \mathbf{y}_{3,b}^n\} \in \mathcal{T}_\epsilon^{(n)}$

Again using typicality lemmas, it can be shown that the probability of error is close to zero if

$$R < I(X_1, X_2; Y_3)$$

This rate can be interpreted as the rate at which both source and the relay coherently transmit information to the destination. Thus the rate achieved by decode forward scheme for discrete memoryless relay channel is derived. The rate can also be achieved using random partitioning/binning (irregular encoding) scheme described before. Backward decoding can also be employed at the sink terminal.

4.2.3.2 Partial decode-forward

Partial decode-forward [5] [42] is similar to decode-forward but the relay does not decode the message completely. This relaxes the constraint to decode the message at higher rate at the relay when the source relay link has low signal to noise ratio. In the next section, we explain encoding and decoding strategy for partial decode forward.

*Codebook generation:*Fix $p(u, x_1, x_2)$.

- Generate $2^{nR'}$ sequences $x_{2b}(v)$ according to probability $p(x_2)$ and $v \in [1, 2^{nR'}]$.
- For each $x_{2b}(v)$ randomly generate $2^{nR'}$ i.i.d. sequences $u_b^n(v, m)$ according to probability $p(u|x_2)$ where $m \in [1, 2^{nR'}]$.
- For every $x_{2b}(v)$ and $u_b^n(v, m)$, generate $2^{nR''}$ i.i.d. sequences $x_{1b}(v, m_b, t_b)$ according to $p(x_1|u, x_2)$ and $t \in [1, 2^{nR''}]$

Source terminal: The source terminal transmits codeword $x_{1b}(m_{b-1}, m_b, t_b)$

Relay terminal: The relay uses y_{2b}^n and $\hat{m}_{b-1}$ to find a unique message $\tilde{m}_b$ such that

$$\{\mathbf{u}_{1,b}^n(\hat{m}_{b-1}, \tilde{m}_b), \mathbf{x}_{2,b}^n(\hat{m}_{b-1}), \mathbf{y}_{2,b}^n\} \in \mathcal{T}_\epsilon^{(n)}$$

This would decode the correct message with high probability provided

$$R' < I(U; Y_2|X_2)$$

Sink Terminal: The sink uses $y^n_{3,b-1}$, y^n_{3b} and $\hat{m}_{b-2}$ to decode the messages $\tilde{m}_{b-1}$ and $\tilde{t}_{b-1}$ by the following joint typicality rule

$$\{\mathbf{u}^n_{1,b-1}(\hat{m}_{b-2}, \tilde{m}_{b-1}), \mathbf{x}_{1,b-1}(\hat{m}_{b-2}, \tilde{m}_{b-1}, \tilde{t}_{b-1}), \mathbf{x}^n_{2,b-1}(\hat{m}_{b-2}), \mathbf{y}^n_{3,b-1}\} \in \mathcal{T}^{(n)}_\epsilon \tag{4.8}$$

$$\{\mathbf{x}^n_{2,b-1}(\hat{m}_{b-2}), \mathbf{y}^n_{3,b-1}\} \in \mathcal{T}^{(n)}_\epsilon \tag{4.9}$$

The messages $\tilde{t}_{b-1}$ would be decoded correctly if

$$R'' < I(X_1; Y_3|X_2, U)$$

The message $\tilde{m}_{b-1}$ would be decoded correctly if

$$R' < I(U; Y_3|X_2) + I(X_2; Y_3)$$

The rates achieved by partial decode forward for the discrete memoryless relay channel is then

$$R' + R'' = \max_{p_{U,X_1,X_2}} \min\{I(U; Y_2|X_2) + I(X_1; Y_3|X_2, U), I(X_1, X_2; Y_3)\}$$

If the relay channel is semi-deterministic, that is $Y_2 = f(X_1, X_2)$. Then we can chose $U = Y_2$ without violating the Markov chain

$$U \leftrightarrow [X_1, X_2] \leftrightarrow [Y_2, Y_3]$$

This choice of U achieves the capacity of the semi deterministic relay channel [42] given by

$$C \leq \max_{p_{U,X_1,X_2}} \min\{H(Y_2|X_2) + I(X_1; Y_3|X_2, Y_2), I(X_1, X_2; Y_3)\}$$

4.2.3.3 Compress-forward

The relay does not decode the message in the compress-forward scheme [5] and uses Wyner-Ziv encoding to transmit side information to the destination.

Codebook generation:

- Generate 2^{nR} sequences $x_{1b}(m)$ according to probability $p(x_1)$ where $m \in [1, 2^{nR}]$.
- Generate 2^{nR_2} i.i.d. sequences $x^n_{2b}(v)$ according to probability $p(x_2)$ where $v \in [1, 2^{nR_2}]$.

- For every $x_{2b}(v)$, generate $2^{nR_2'}$ i.i.d. sequences $\hat{y}_{2b}(v,t,u)$ according to $p(\hat{y}_2|x_2)$ and $t \in [1, 2^{nR_2'}], u \in [1, 2^{nR_2}]$

Source Terminal: Transmit $x_{1b}(m)$.

Relay Terminal: The relay finds $(\tilde{t}_b, \tilde{u}_b)$ such that

$$\{\mathbf{y}_{2b}^n, \mathbf{x}_{2b}^n(v_b), \hat{y}_{2b}(v_b, \tilde{t}_b, \tilde{u}_b)\} \in \mathcal{T}_\epsilon^{(n)}$$

and sets $\tilde{u}_b = v_{b+1}$. If more than one such message is found, one of them is chosen. If no message is found to be typical, then set $v_{b+1} = 1$. We will find a jointly typical sequence $\tilde{u}_b$ if

$$R_2 + R_2' > I(\hat{Y}_2; Y_2|X_2)$$

Sink Terminal: The destination uses $y_{3b}^n, y_{3,b-1}^n$ to find a jointly typical $\tilde{m}_{b-1}$

$$\{\mathbf{x}_{2b}^n(v_b), \mathbf{y}_{3b}^n\} \in \mathcal{T}_\epsilon^{(n)}$$

which gives the rate bound

$$R_2 < I(X_2; Y_3)$$

The sink terminal finds a typical $\tilde{t}_{b-1}$

$$\{\mathbf{x}_{2,b-1}^n(\hat{v}_{b-1}), \mathbf{y}_{3,b-1}^n, \hat{\mathbf{y}}_{2,b-1}^n(\hat{v}_{b-1}, \tilde{t}_{b-1}, \hat{v}_b)\} \in \mathcal{T}_\epsilon^{(n)}$$

which gives the rate bound

$$R_2' < I(\hat{Y}_2; Y_3|X_2)$$

The sink terminal also looks for a jointly typical $\tilde{m}_{b-1}$ such that

$$\mathbf{x}_{1,b-1}^n(\tilde{m}_{b-1}), \mathbf{x}_{2,b-1}^n(\hat{v}_{b-1}), \mathbf{y}_{3,b-1}^n, \hat{\mathbf{y}}_{2,b-1}^n(\hat{v}_{b-1}, \tilde{t}_{b-1}, \hat{v}_b)$$

with a rate constraint

$$R < I(X_1; \hat{Y}_2, X_2, Y_3) = I(X_1; \hat{Y}_2, Y_3|X_2)$$

Combining the above rate bounds and using the Markov chain $Y_3 \leftrightarrow (X_2, Y_2) \leftrightarrow \hat{Y}_2$. The CF scheme achieves any rate less than [5, Theorem 6]

$$R_{CF} = \sup I(X_1; \hat{Y}_2, Y_3|X_2), \quad \text{such that } I(X_2; Y_3) \geq I(\hat{Y}_2; Y_2|X_2, Y_3) \tag{4.10}$$

where supremum is taken over all joint probability distributions of the form

$$p(x_1, x_2, y_2, y_3, \hat{y}_2) = p(x_1)p(x_2)p(y_2, y_3|x_1, x_2)p(\hat{y}_2|y_1, x_2). \tag{4.11}$$

El Gamal, Mohseni, and Zahedi [25] put forth an equivalent characterization of the CF scheme. That is, it achieves any rate less than

$$R_{CF} = \sup \min\{I(X_1; \hat{Y}_2, Y_3|X_2), I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_3)\} \tag{4.12}$$

where supremum is still taken over all joint probability distributions of the same form as in (4.11).

4.2.3.4 Superposition-forward

The supremum of rates achievable by superimposing DF and CF [5, Theorem 7] is

$$R_{SF} = \sup(\min\{I(X_1; Y_3, \hat{Y}_2'|X_2, U) + I(U; Y_2|X_2, V), \\ I(X_1, X_2; Y_3) - I(\hat{Y}_2'; Y_2|U, X_1, X_2, Y_3)\}) \tag{4.13}$$

where the supremum is over all joint probability distributions of the form

$$p(u, v, x_1, x_2, y_2', y_3, \hat{y}_2) = p(v)p(u|v)p(x_1|u)p(x_2|v)p(y_2, y_3|x_1, x_2)p(\hat{y}_2'|x_2, y_2, u) \tag{4.14}$$

subject to the constraint

$$I(X_2; Y_3|V) \geq I(\hat{Y}_2'; Y_2|X_2, Y_3, U). \tag{4.15}$$

Finally, the rate is upper bounded by the cut-set bound [5] [6]

$$R_{CS} = \sup \min\{I(X_1, X_2; Y_3), I(X_1; Y_2, Y_3)\}, \tag{4.16}$$

where the supremum is taken over all possible distributions $p(x_1, x_2)$.

4.3 Broadcast over decode-forward

Before investigating the coding scheme that superimposes CF and DF for the Gaussian relay channel, we will first look at a simpler coding scheme. In this scheme, partial information is decoded first at both the relay and the destination terminals like in a broadcast channel.

The remaining message is decoded and forwarded given the partial information available at the relay and destination terminal. The coding scheme is equivalent to superimposing broadcast over decode and forward.

We split the message M into two parts M' and M'' with respective rates R' and R''. We demand M' be decoded at both relay and destination. The relay also decodes the message M'' which the destination could not decode and sends this extra information to the destination in a block Markov encoding fashion. This strategy can be designed using an auxiliary random variable U and a block Markov superposition encoding explained below.

Theorem 1. *For any relay channel ($\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3$), the rate R is achievable where*

$$R < \sup_P \{\min\{I(U;Y_3), I(U;Y_2|X_2)\} + \min\{I(X_1;Y_2|X_2,U), I(X_1,X_2;Y_3|U)\}\} \qquad (4.17)$$

and the supremum is taken over all probability distribution functions of the form

$$p(u, x_1, x_2, y_2, y_3) = p(u)p(x_2)p(x_1|x_2, u)p(y_2, y_3|x_1, x_2).$$

Proof:

1. Codebook Generation: Encoding is performed in $K+1$ blocks. For each block k, generate $2^{nR'}$ codewords $u_k^n(s), s = 1, 2, \ldots, 2^{nR'}$ by choosing the $u_{ki}(s)$ independently using the distribution $P_U(\cdot)$. Generate $2^{nR''}$ codewords $x_{2k}^n(t), t = 1, 2, \ldots, 2^{nR''}$ by choosing $x_{2ki}(t)$ independently using the probability distribution $P_{X_2}(\cdot)$. Now use superposition coding and generate $2^{nR''}$ codewords $x_{1k}^n(r|s,t), r = 1, 2, \ldots, 2^{nR''}$ for every pair of $(u_k^n(s), x_{2k}^n(t))$, by choosing the $x_{1k,i}(r|s,t)$ independently using $P_{(X_1|X_2,U)}(.|u_{k,i}(s), x_{2k,i}(t))$.

2. Encoding: Let s_k be the message index of M' and t_k be the message index of M'' respectively to be sent in block k. The source encoder then transmits $x_{1k}^n(t_k|s_k, t_{k-1})$ where t_{k-1} is the index of M'' sent in the previous block. The relay in block k will send $x_{2k}^n(\hat{t}_{k-1})$, where $\hat{t}_{k-1}$ is the estimate of t_{k-1} at the relay.

3. Decoding at relay terminal: Assume that decoding of s_{k-1} and t_{k-1} in block $k-1$ has been successful. Upon receiving y_{2k}^n in block k, the relay looks for a unique $\hat{s}_k$ such that

$$\left(u_{1k}^n(\hat{s}_k), x_{2k}^n(\hat{t}_{k-1}), y_{2k}^n\right) \in T_\epsilon^n(P_{U,X_2,Y_2}).$$

Having decoded $\hat{s}_k$, the relay now looks for a unique $\hat{t}_k$ such that

$$(x_{1k}^n(\hat{t}_k|\hat{s}_k, \hat{t}_{k-1}), u_{1k}^n(\hat{s}_k), x_{2k}^n(\hat{t}_{k-1}), y_{2k}^n) \in T_\epsilon^n(P_{U,X_1,X_2,Y_2}).$$

4. Decoding at the sink terminal: Upon receiving y_{3k}^n, the destination terminal looks for a unique $\tilde{s}_k$ such that $(u_{1k}^n(\tilde{s}_k), y_{3k}^n) \in T_\epsilon^n(P_{U,Y_3})$. Now, the destination decodes the additional information that the source sends in a block Markov decoding fashion. The destination terminal tries to find a unique $\tilde{t}_{k-1}$ such that $\left(x_{2k}^n(\tilde{t}_{k-1}), u_{1k}^n(\tilde{s}_k), y_{3k}^n\right) \in T_\epsilon^n(P_{U,X_2,Y_3})$ and

$$\left(x_{1k}^n(\tilde{t}_{k-1}|\tilde{s}_{k-1}, \tilde{t}_{k-2}), u_{1k}^n(\tilde{s}_{k-1}), x_{2k}^n(\tilde{t}_{k-2}), y_{3(k-1)}^n\right) \in T_\epsilon^n(P_{U,X_1,X_2,Y_3}).$$

5. Rate analysis: At the relay, since we have a single user channel from U to Y_2, we will be able to decode the U codewords with low probability of error if $R' < I(U;Y_2|X_2)$. We can also decode the index t_k if

$$R'' < I(X_1;Y_2|U,X_2).$$

The destination first decodes the codeword U with a low probability of error provided $R' < I(U;Y_3)$, and then decodes the message t_k using successive interference cancellation on the messages from the relay and the source. The message would be decoded with low probability of error provided

$$R'' < I(X_2;Y_3|U) + I(X_1;Y_3|X_2,U).$$

Combining all the bounds, the desired result (4.17) follows.

In this scheme, the source message is split into two parts. The message M' is broadcast to both relay and destination. The message M'' is decoded by relay first and then cooperatively transmitted to the destination. Unfortunately, the above achievable rate does not outperform the DF strategy, as shown below:

$$R \leq \min\{I(U;Y_2|X_2) + I(X_1;Y_2|X_2,U), I(U;Y_3) + I(X_1X_2;Y_3|U)\} \tag{4.18}$$

$$= \min\{I(U,X_1;Y_2|X_2), I(X_1,X_2;Y_3)\} \tag{4.19}$$

$$= \min\{I(X_1;Y_2|X_2), I(X_1X_2;Y_3)\}. \tag{4.20}$$

where (4.20) follows from the Markov chains $U - X_1 - Y_2$ and $U - X_1 - Y_3$. Rate equation (4.20) is the rate achieved by the decode-forward strategy.

Although not providing a higher rate, the above proposed scheme of broadcast over decode and forward gives us a good insight on the superposition strategy. The cause of sub optimality arises due to the fact that the messages M' and M'' even though are generated from the same source, act as interference on each other. This limits the rate of decoding at the relay and destination terminals. This interference would also be present if we superimpose DF and CF. The rate achievable using the superposition strategy is investigated in the next section for the case of Gaussian relay channels.

4.4 Achievable rate of superposition-forward scheme

In this section, we focus on the Gaussian relay channel. We show that when considering only jointly Gaussian distribution for all the random variables involved in (4.13), superposition does not offer higher rate than DF or CF alone. To be more specific, we will show that when all the random variables involved are Gaussian, then $R_{SF} \leq \max(R_{DF}, R_{CF})$. Trivially, only one of two cases can be true

1. Case A: $R_{DF} \geq R_{CF}$;
2. Case B: $R_{CF} > R_{DF}$.

It is then enough to show that in Case A, $R_{SF} \leq R_{DF}$; and in Case B, $R_{SF} \leq R_{CF}$.

4.4.1 Gaussian distribution assumption

We assume that all random variables in (4.13) are zero mean and jointly Gaussian distributed. The distribution will then depend only on the variances and the cross-correlations of the random variables. For two generic random variables X and Y, let

$$\phi_{X,Y} := \frac{E\left\{(X - E[X])(Y - E[Y])\right\}}{\sqrt{E[X^2]E[Y^2]}}$$

denote the correlation coefficient between them. The following lemma is useful in deducing correlations from known ones.

Lemma 1. *Let $X - Y - Z$ be a Markov chain of jointly Gaussian random variables. Then $\phi_{X,Z} = \phi_{X,Y}\phi_{Y,Z}$.*

Proof: Assume without loss of generality that all three random variables are zero mean. We have

$$\begin{aligned}\phi_{X,Z} &= \frac{\mathsf{E}[XZ]}{\sqrt{\mathsf{E}[X^2]\,\mathsf{E}[Z^2]}} \\ &= \frac{\mathsf{E}\{\mathsf{E}[XZ|Y]\}}{\sqrt{\mathsf{E}[X^2]\,\mathsf{E}[Z^2]}} \\ &= \frac{\mathsf{E}\{\mathsf{E}[X|Y]\,\mathsf{E}[Z|Y]\}}{\sqrt{\mathsf{E}[X^2]\,\mathsf{E}[Z^2]}} \\ &= \frac{\mathsf{E}\{\sqrt{\mathsf{E}[X^2]/\,\mathsf{E}[Y^2]}\phi_{X,Y}Y \cdot \sqrt{\mathsf{E}[Z^2]/\,\mathsf{E}[Y^2]}\phi_{Y,Z}Y\}}{\sqrt{\mathsf{E}[X^2]\,\mathsf{E}[Z^2]}} \\ &= \phi_{X,Y}\phi_{Y,Z}\end{aligned} \tag{4.21}$$

Returning to the random variables involved in R_{SF}, we denote $\alpha = \phi_{U,V}$, $\beta = \phi_{V,X_2}$, and $\gamma = \phi_{U,X_1}$. Using Lemma 1, we obtain from the Markov chain $U - V - X_2$ that

$$\delta := \phi_{X_1,X_2} = \phi_{V,U} \cdot \phi_{V,X_2} = \alpha\beta, \tag{4.22}$$

and from the Markov chain $X_1 - U - X_2$ that

$$\rho := \phi_{X_1,X_2} = \phi_{X_1,U} \cdot \phi_{U,X_2} = \gamma\delta = \alpha\beta\gamma. \tag{4.23}$$

Fig. 4.3 shows the codebook generation and correlation between the random variables along with their dependencies on each other.

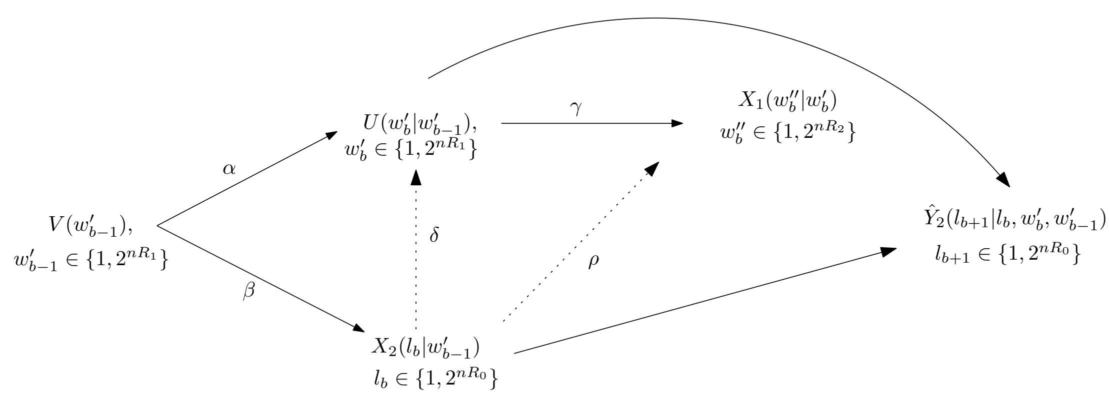

Figure 4.3 Dependency graph of random variables with correlation coefficients

4.4.2 Main result

The main result is stated in the following theorem. Two lemmas that are needed in the proof are stated and proved in the appendix.

Theorem 2. *Let* $(X_1, X_2, Y_2, Y_3, \hat{Y}_2', \hat{Y}_2, U, V,)$ *be a set of jointly Gaussian random variable whose joint distribution can be factorized in the following form:*

$$p(u, v, x_1, x_2, y_2, y_3, \hat{y}_2') = p(v)p(u|v)p(x_2|v)p(x_1|u)p(y_2, y_3|x_1x_2)p(\hat{y}_2'|y_2, u, x_2)p(\hat{y}_2|y_2, x_2), \tag{4.24}$$

where $p(y_2, y_3|x_1x_2)$ *is as given in* (4.4)*. Let* $\mathcal{P}$ *denote the class of distributions specified by* (4.24)*. Let* $\mathcal{P}'$ *denote a subset of* $\mathcal{P}$ *with distributions that also satisfy the constraint* (4.15)*. We have*

$$\sup_{\mathcal{P}'} \min\{I(X_1; Y_3, \hat{Y}_2'|X_2, U) + I(U; Y_2|X_2, V), I(X_1, X_2; Y_3) - I(\hat{Y}_2'; Y_2|X_1, X_2, U, Y_3)\} \tag{4.25}$$

$$= \max\{\sup_{\mathcal{P}} \min\{I(X_1; Y_2|X_2), I(X_1, X_2; Y_3)\}, \tag{4.26}$$

$$\sup_{\mathcal{P}} \min\{I(X_1; \hat{Y}_2, Y_3|X_2), I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_3)\}\}. \tag{4.27}$$

Proof: The rates appearing in (4.25)–(4.27) are R_{SF}, R_{DF}, and R_{CF}, respectively. Since through the judicious choice the random variables U and V, DF and CF can be cast as special cases of SF [5], we have $R_{SF} \geq R_{DF}$ and $R_{SF} \geq R_{CF}$. It is then sufficient to show that $R_{SF} \leq \max(R_{DF}, R_{CF})$.

Under the Gaussian assumption, the compressed version $\hat{Y}_2'$ of Y_2 in (4.14) can be written as

$$\hat{Y}_2' = c_1 Y_2 + c_2 U + c_3 X_2 + Z_w' \tag{4.28}$$

where c_1, c_2, c_3 are constant parameters, Z_w' is Gaussian and independent of Y_2, U, and X_2. Since in both (4.13) and (4.15), the three mutual information terms involving $\hat{Y}_2'$, namely,

$$I(X_1; Y_3, \hat{Y}_2'|X_2, U), \quad I(\hat{Y}_2'; Y_2|X_2, X_1, U, Y_3), \quad I(\hat{Y}_2'; Y_2|X_2, U, Y_3)$$

are all conditioned on U and X_2, the coefficients c_2 and c_3 do not affect the values of these terms. Therefore we can set $c_2 = c_3 = 0$. It is also true that scaling $\hat{Y}_2'$ by a constant does not change any of the terms. So unless $c_1 = 0$, we can assume $c_1 = 1$, as we do in the following. The case $c_1 = 0$ is known as the so called partial decoding and forward scheme, which is known to be inferior to the full DF scheme [25]. We denote the variance of Z'_w as Δ'. The amount of compression, which is controlled by the parameter Δ', depends on the constraint (4.15) imposed by the relay link channel and the encoding scheme at the relay. In summary, we can take without loss of generality

$$\hat{Y}'_2 = Y_2 + Z'_w, \tag{4.29}$$

The following is a broad outline of the proof. Given any rate achieved by the SF scheme, we can find a CF scheme or a DF scheme which can achieve a rate higher than or equal to SF. The $\hat{Y}_2$ for the CF scheme is set to be statistically equal to $\hat{Y}'_2$ of the SF scheme in (4.29):

$$\hat{Y}_2 = Y_2 + Z_w, \tag{4.30}$$

where Z_w is zero mean Gaussian with variance $\Delta = \Delta'$. Such $\hat{Y}_2$ would qualify as the compressed version of Y_2 in CF. This choice of $\hat{Y}_2$ is enough to achieve a higher rate than SF even though it can be suboptimal to the possible rates achievable by CF.

First, we have

$$\begin{aligned}
I(\hat{Y}'_2; Y_2|X_1, X_2, U, Y_3) &= h(Y_2|X_1, X_2, U, Y_3) - h(Y_2|X_1, X_2, U, Y_3, \hat{Y}'_2) && (4.31)\\
&= h(Y_2|X_1, X_2, Y_3) - h(Y_2|X_1, X_2, U, Y_3, \hat{Y}'_2) && (4.32)\\
&\geq h(Y_2|X_1, X_2, Y_3) - h(Y_2|X_1, X_2, Y_3, \hat{Y}'_2) && (4.33)\\
&\geq h(Y_2|X_1, X_2, Y_3) - h(Y_2|X_1, X_2, Y_3, \hat{Y}_2) && (4.34)\\
&= I(\hat{Y}_2; Y_2|X_1, X_2, Y_3) && (4.35)
\end{aligned}$$

where (4.32) is due to the Markov chain $U - (X_1, X_2, Y_3) - Y_2$; (4.33) uses the fact that conditioning does not increase entropy; and (4.34) is because given (X_2, U), $\hat{Y}'_2$ is statistically equivalent to $\hat{Y}_2$.

Thus, we have shown

$$I(X_1, X_2; Y_3) - I(\hat{Y}'_2; Y_2|X_1, X_2, U, Y_3) \leq I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_3). \tag{4.36}$$

It then remains to be shown that

$$I(X_1; Y_3, \hat{Y}_2'|X_2, U) + I(U; Y_2|X_2, V) \leq \max\{I(X_1; Y_2|X_2), I(X_1; \hat{Y}_2, Y_3|X_2)\}. \tag{4.37}$$

Depending on which one of the two terms on the right hand side is bigger, we have two cases. In the first case,

$$I(X_1; Y_2|X_2) \geq I(X_1; Y_3, \hat{Y}_2|X_2) \tag{4.38}$$

and we have

$$\begin{aligned} & I(U; Y_2|V, X_2) + I(X_1; Y_3, \hat{Y}_2'|X_2, U) && (4.39)\\ &= I(U; Y_2|V, X_2) + I(X_1; Y_3, \hat{Y}_2|X_2, U) && (4.40)\\ &= I(U; Y_2|X_2) - I(V; Y_2|X_2) + I(X_1; Y_3, \hat{Y}_2|X_2, U) && (4.41)\\ &= I(X_1; Y_2|X_2) - I(X_1; Y_2|X_2, U) - I(V; Y_2|X_2) + I(X_1; Y_3, \hat{Y}_2|X_2, U) && (4.42)\\ &\leq I(X_1; Y_2|X_2) - I(X_1; \hat{Y}_2, Y_3|X_2, U) - I(V; Y_2|X_2) + I(X_1; Y_3, \hat{Y}_2|X_2, U) && (4.43)\\ &= I(X_1; Y_2|X_2) - I(V; Y_2|X_2) && (4.44)\\ &\leq I(X_1; Y_2|X_2) && (4.45) \end{aligned}$$

where (4.40) follows by our choice of $\hat{Y}_2$ to be statistically the same as $\hat{Y}_2'$; (4.41) follows from the Markov chain $V - (U, X_2) - Y_2$; (4.42) follows from the Markov chain $U - (X_1, X_2) - Y_2$; (4.43) follows from (4.38) and Lemma 2, which is stated and proved in Appendix 4.A.1; and (4.45) follows from the fact that mutual information is nonnegative.

In the second case,

$$I(X_1; Y_2|X_2) < I(X_1; Y_3, \hat{Y}_2|X_2) \tag{4.46}$$

and we have

$$\begin{aligned} & I(X_1; Y_3, \hat{Y}_2'|X_2, U) + I(U; Y_2|V, X_2) \\ &= I(X_1; Y_3, \hat{Y}_2|X_2, U) + I(U; Y_2|V, X_2) && (4.47)\\ &= I(X_1; Y_3, \hat{Y}_2|X_2, U) + I(U; Y_2|X_2) - I(V; Y_2|X_2) && (4.48)\\ &\leq I(X_1; Y_3, \hat{Y}_2|X_2, U) + I(U; Y_3, \hat{Y}_2|X_2) - I(V; Y_2|X_2) && (4.49)\\ &= I(X_1; Y_3, \hat{Y}_2|X_2) - I(V; Y_2|X_2) && (4.50)\\ &\leq I(X_1; Y_3, \hat{Y}_2|X_2) && (4.51) \end{aligned}$$

where (4.47) follows by our choice of $\hat{Y}_2$ to be statistically the same as $\hat{Y}_2'$; (4.48) follows from the Markov chain $V - (U, X_2) - Y_2$; (4.49) follows from (4.46) and Lemma 3, which is stated and proved in Appendix 4.A.1; (4.50) follows from the Markov chain $U - (X_1, X_2) - \hat{Y}_2, Y_3$; and (4.51) follows from the fact that mutual information is nonnegative.

Thus we have shown (4.37) holds. And the whole proof is complete.

4.4.3 Discussion

We have shown that the SF does not outperform both DF and CF. We provide some intuitive explanation in the following.

Observe from (4.30) that $\hat{Y}_2$ is the quantized signal of Y_2 in the CF scheme. The variance of Z_w is Δ, which in general could be different from Δ', the variance of Z_w' in (4.28). From the constraint (4.10), we have $\Delta \geq \Delta_{CF}$, where

$$\Delta_{CF} = \frac{N_1 N_2 + \left(N_1 + a^2 N_2\right) P_1}{b^2 P_2}. \tag{4.52}$$

Although the constraint is not explicitly imposed in the formulation in (4.12), it can be shown that setting $\Delta = \Delta_{CF}$ actually maximizes the two terms on the right hand side of (4.12), and equalizes them:

$$I(X_1; \hat{Y}_2, Y_3 | X_2) = I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_3). \tag{4.53}$$

It can be verified that

1. $I((X_1; \hat{Y}_2, Y_3 | X_2)$ is a monotonically decreasing function of Δ (coarser compression reduces the useful information about X_1 in $\hat{Y}_2$);

2. $I(\hat{Y}_2; Y_2 | X_1, X_2, Y_3)$ is a monotonically increasing function of Δ.

Therefore the minimum of the two functions is maximized at their crossing point, which happens at $\Delta = \Delta_{CF}$. In other words, for CF, within the relay-destination link rate limit $I(X_2; Y_3)$, more compression yields higher rate over all. For the SF, however, the situation is different. The parameter Δ', which controls the amount of compression in (4.28) needs to be chosen to satisfy the constraint (4.15). In particular, we have $\Delta' \geq \Delta_{SF}$, where

$$\Delta_{SF} = \frac{(N_2 + P_1(1 - \alpha^2\gamma^2))(N_1 N_2 + (N_1 + a^2 N_2) P_1 (1 - \gamma^2))}{b^2 P_2 (1 - \beta^2)[N_2 + P_1(1 - \gamma^2)]} \tag{4.54}$$

In general Δ_{SF} can be less than Δ_{CF}; e.g., when $\gamma > 0$, $\alpha = 1$ and $\beta = 0$. In contrast to the CF case, it is not true for SF that more compression (smaller Δ') necessarily yields higher rate. The intuitive reason is that the relay has two messages to transmit to the destination: the partially decoded message carried by U and the compressed version of Y_2 carried by $\hat{Y}_2'$. Although reducing Δ' will provide to the destination a more faithful representation of Y_2, and enlarge the term $I(X_1; Y_3, \hat{Y}_2'|X_2, U) + I(U; Y_2|X_2, V)$, it will reduce the relay's ability to cooperate with the source through the message U, and hence enlarge the gap $I(\hat{Y}_2'; Y_2|X_1, X_2, U, Y_3)$ from the multiple-access cut-set bound $I(X_1, X_2; Y_3)$, which then becomes the rate limiting factor. The optimum amount compression turns out to the be same as in the CF case. And superposition of DF and CF does not help the rate, which agrees with the observation that we have made in Section 4.3.

4.5 Numerical result

Considering an example Gaussian relay channel such that the source and the destination are separated by a unit distance [24], and the relay is at distance d from the source and $1 - d$ from the destination. The channel model is shown in Fig. 4.4. The channel gain between any two nodes is inversely proportional to their distance. So $a = 1/d$ and $b = 1/(1 - d)$. The additive noises at the relay and the destination are independent but have the same variance $N_1 = N_2 = 1$. The transmit powers are set to $P_1 = P_2 = 5$ dB. The choice of power constraint may vary depending on the practical system analyzed. The performance as function of relay distance d is only scaled with varying power constraints. The relative performance of different schemes remain the same.

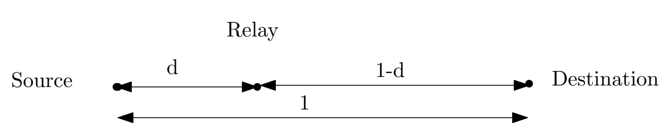

Figure 4.4 Example of a Gaussian relay channel

Fig. 4.5 shows the numerical rates achievable by DF, CF and the cut set bound (4.16) as a function of distance d of the relay from the source terminal. Depending on d, there are three cases:

1. When d is small (roughly $d < 0.2$), DF is optimal. The rate achieved by DF is equal to $I(X_1, X_2; Y_3)$ the multiple-access cut-set bound. The reason is that the source message can be fully decoded at the relay.

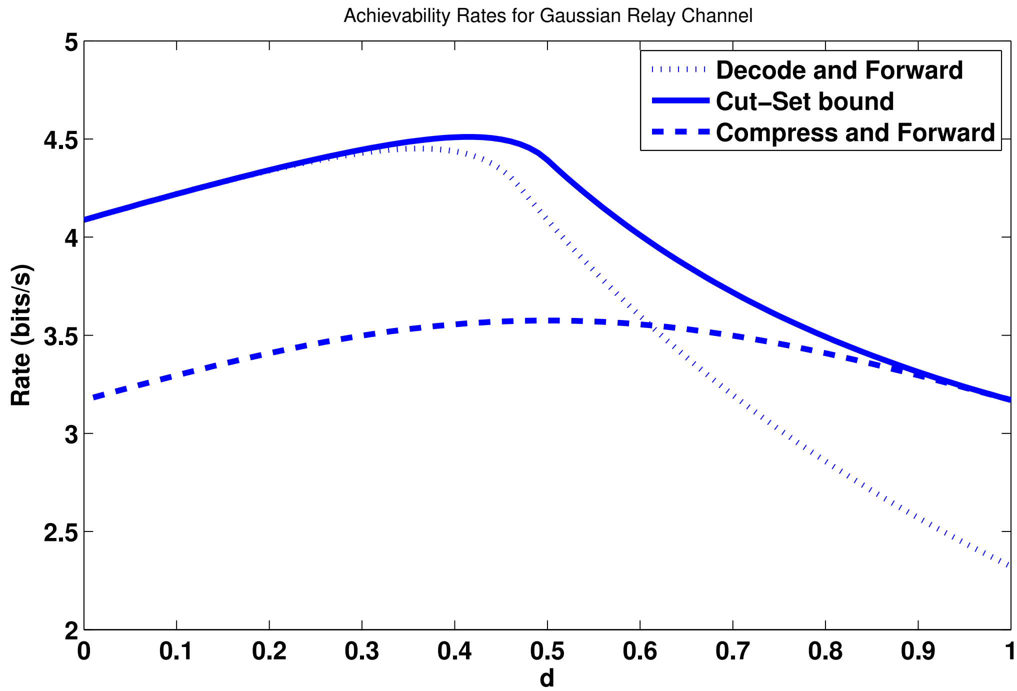

Figure 4.5 Achievable rates for the Gaussian relay channel

2. For medium d (roughly $0.2 < d < 0.6$), DF is not optimal, but still performs better than CF. In this case, the rate of DF is dominated by $I(X_1, Y_2|X_2)$, the amount information can be decoded at the relay, which dictates the amount of cooperation possible between source and relay. In this region, the relay-sink channel is "poor" so that sending "finely" compressed version of Y_2 is not possible.

3. For large d (roughly $0.6 < d \leq 1$), CF out performs DF. In this region, the ability of the relay to decode the source is weak, and it is more fruitful to send compressed version of the relay's observation. Only in the extreme case, $d = 1$, does CF actually achieve the cut-set bound.

The rate achievable by superimposing DF and CF given by (4.13) is numerically compared with the rates achieved by CF, DF and the cut-set bound. The mutual information terms of

(4.13) are evaluated for the choice of appropriate Gaussian Random variables, according to (4.61) and

$$I(U;Y_2|X_2,V) = C\left(\frac{\frac{P_1}{d^2}\gamma^2(1-\alpha^2)}{N_1+\frac{P_1}{d^2}(1-\gamma^2)}\right), \tag{4.55}$$

$$I(X_1X_2;Y_3) = C\left(\frac{P_1+\frac{P_2}{(1-d)^2}+\frac{2\rho\sqrt{P_1P_2}}{(1-d)}}{N_2}\right), \tag{4.56}$$

$$I(Y_2;\hat{Y}_2|U,X_1,X_2,Y_3) = C\left(\frac{N_1}{\Delta}\right). \tag{4.57}$$

The constraint $I(\hat{Y}_2;Y_2|U,X_2,Y_3) \leq I(X_2;Y_3|V)$ is evaluated to $\Delta' \geq \Delta_{SF}$, where Δ_{SF} is as given in (4.54). The correlation terms α, β, γ and the variance Δ' are optimizing parameters, which control the amount of information that is decoded and the amount that is compressed. When all the parameters have been optimized within the constraint posed by (4.54), the SF is found to achieve the maximum of R_{DF} and R_{CF}, as shown in Fig. 4.6.

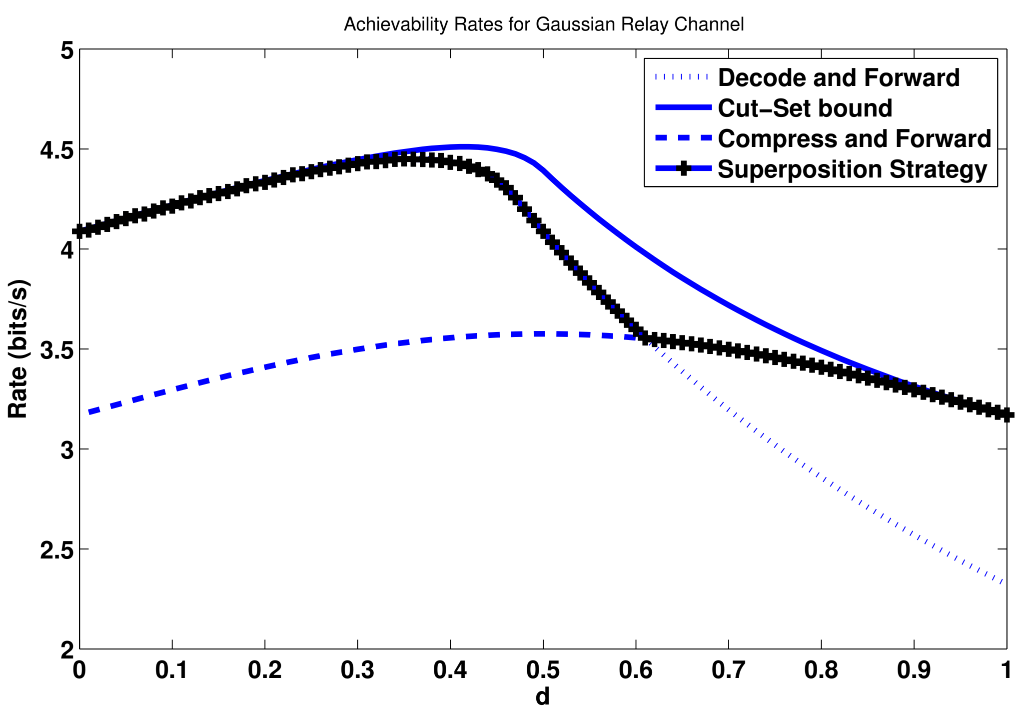

Figure 4.6 Achievable rates for Gaussian relay channel. The parameters of superposition-forward are optimized to maximize the achievable rate

4.6 Conclusion

We analyzed the coding strategy of superimposing CF and DF for the Gaussian relay channel. We note that superposition of CF and DF does not provide higher achievable rates than the individual DF and CF for the Gaussian case. Superposition scheme achieves the best of CF and DF rates for single relay channel. We conclude that we should look for new strategies or look for non-Gaussian distributions for the superposition scheme, or try to find tighter upper bounds than the cut-set bound.

We next look at the discrete memoryless network and propose a new coding scheme based on superposition and noisy network coding. This scheme is derived for single source and multiple source multicast networks.

4.A Appendix

4.A.1 Two lemmas needed in the proof of Theorem 2

We prove two lemmas in the following that will be useful in the proof of Theorem 2. Lemma 2 is used in the case $R_{DF} \geq R_{CF}$. Lemma 3 is used in the case $R_{DF} < R_{CF}$.

Lemma 2. *Let $(X_1, X_2, Y_2, Y_3, \hat{Y_2}, U, V)$ be jointly Gaussian random variables with joint distribution*

$$p(u, v, x_1, x_2, y_2, y_3, \hat{y_2}) = p(v)p(u|v)p(x_2|v)p(x_1|u)p(y_2, y_3|x_1x_2)p(\hat{y_2}|y_2, x_2),$$

where $p(y_2, y_3|x_1, x_2)$ is as given in (5.22).

If $I(X_1; Y_2|X_2) \geq I(X_1; Y_3, \hat{Y_2}|X_2)$ then $I(X_1; Y_2|X_2, U) \geq I(X_1; \hat{Y_2}, Y_3|X_2, U)$.

Proof: Under the Gaussian assumption, we have

$$I(X_1; Y_2|X_2) = \frac{1}{2} \log \left\{1 + \frac{a^2 P_1(1-\rho^2)}{N_1}\right\} \tag{4.58}$$

$$I(X_1; Y_2|X_2, U) = \frac{1}{2} \log \left\{1 + \frac{a^2 P_1(1-\gamma^2)}{N_1}\right\} \tag{4.59}$$

$$I(X_1; \hat{Y_2}, Y_3|X_2) = \frac{1}{2} \log \left\{1 + P_1(1-\rho^2)\frac{(N_1+\Delta) + a^2 N_2}{(N_1+\Delta)N_2}\right\} \tag{4.60}$$

$$I(X_1; \hat{Y_2}, Y_3|X_2, U) = \frac{1}{2} \log \left\{1 + P_1(1-\gamma^2)\frac{(N_1+\Delta) + a^2 N_2}{(N_1+\Delta)N_2}\right\} \tag{4.61}$$

Obviously when $\rho = 1$ and hence $\gamma = 1$ (because $\rho = \alpha\beta\gamma$), the lemma holds. We thus assume that $\rho < 1$. Since $I(X_1; Y_2|X_2) \geq I(X_1; \hat{Y}_2, Y_3|X_2)$, from (4.58) and (4.59) we have

$$\frac{a^2 P_1(1-\rho^2)}{N_1} \geq P_1(1-\rho^2)\frac{(N_1+\Delta)+a^2N_2}{(N_1+\Delta)N_2}. \tag{4.62}$$

Multiplying both sides with $(1-\gamma^2)/(1-\rho^2)$, we obtain

$$\frac{a^2 P_1(1-\gamma^2)}{N_1} \geq P_1(1-\gamma^2)\frac{(N_1+\Delta)+a^2N_2}{(N_1+\Delta)N_2}. \tag{4.63}$$

It then follows that $I(X_1; Y_2|X_2, U) \geq I(X_1; \hat{Y}_2, Y_3|X_2, U)$ from the monotonic property of the logarithmic function.

Lemma 3. *Let* $(X_1, X_2, Y_2, Y_3, \hat{Y}_2, U, V)$ *be jointly Gaussian random variables with distribution* $p(u, v, x_1, x_2, y_2, y_3, \hat{y}_2) = p(v)p(u|v)p(x_2|v)p(x_1|u)p(y_2, y_3|x_1, x_2)p(\hat{y}_2|y_2, x_2)$,

where $p(y_2, y_3|x_1, x_2)$ *is as given in* (5.22).

If $I(X_1; Y_2|X_2) \leq I(X_1; \hat{Y}_2, Y_3|X_2)$ *then* $I(U; Y_2|X_2) \leq I(U; \hat{Y}_2, Y_3|X_2)$.

Proof:

Under the Gaussian variable assumptions, we have

$$I(X_1; Y_2|X_2) = \frac{1}{2}\log\left\{1 + \frac{a^2P_1(1-\rho^2)}{N_1}\right\} \tag{4.64}$$

$$I(U; Y_2|X_2) = \frac{1}{2}\log\left\{1 + \frac{a^2P_1(\gamma^2-\rho^2)}{N_1 + a^2P_1(1-\gamma^2)}\right\} \tag{4.65}$$

$$I(X_1; \hat{Y}_2, Y_3|X_2) = \frac{1}{2}\log\left\{1 + P_1(1-\rho^2)\frac{(N_1+\Delta)+a^2N_2}{(N_1+\Delta)N_2}\right\} \tag{4.66}$$

$$I(U; \hat{Y}_2, Y_3|X_2) = \frac{1}{2}\log\left\{1 + \frac{P_1(\gamma^2-\rho^2)[(N_1+\Delta)+a^2N_2]}{(N_1+\Delta)N_2 + P_1(1-\gamma^2)[(N_1+\Delta)+a^2N_2]}\right\} \tag{4.67}$$

It can be verified that when $\gamma = 1$, $I(X_1; Y_2|X_2) = I(U; Y_2|X_2)$ and $I(X_1; \hat{Y}_2, Y_3|X_2) = I(U; \hat{Y}_2, Y_3|X_2)$, so that the desired result holds in this case. In the following, we assume that $\gamma < 1$, and therefore $\rho = \alpha\beta\gamma < 1$.

Since $I(X_1; Y_2|X_2) \leq I(X_1; \hat{Y}_2, Y_3|X_2)$, it follows from (4.64) and (4.65) that

$$\frac{a^2P_1(1-\rho^2)}{N_1} \leq \frac{P_1(1-\rho^2)[(N_1+\Delta)+a^2N_2]}{(N_1+\Delta)N_2}. \tag{4.68}$$

Multiplying both sides of (4.68) with $(1-\gamma^2)/(1-\rho^2)$ we obtain

$$\frac{a^2P_1(1-\gamma^2)}{N_1} \leq \frac{P_1(1-\gamma^2)[(N_1+\Delta)+a^2N_2]}{(N_1+\Delta)N_2}. \tag{4.69}$$

Adding the numerator to the denominator on both sides, we obtain

$$\frac{a^2P_1(1-\gamma^2)}{N_1+a^2P_1(1-\gamma^2)} \leq \frac{P_1(1-\gamma^2)[(N_1+\Delta)+a^2N_2]}{(N_1+\Delta)N_2+P_1(1-\gamma^2)[(N_1+\Delta)+a^2N_2]}. \tag{4.70}$$

Multiplying both sides of (4.70) by $(\gamma^2-\rho^2)/(1-\gamma^2)$, we obtain

$$\frac{a^2P_1(\gamma^2-\rho^2)}{N_1+a^2P_1(1-\gamma^2)} \leq \frac{P_1(\gamma^2-\rho^2)[(N_1+\Delta)+a^2N_2]}{(N_1+\Delta)N_2+P_1(1-\gamma^2)[(N_1+\Delta)+a^2N_2]} \tag{4.71}$$

It then follows that $I(U;Y_2|X_2) \leq I(U;\hat{Y}_2,Y_3|X_2)$ due to the monotonic property of the logarithmic function.

CHAPTER 5. NOISY NETWORK CODING WITH SUPERPOSITION

5.1 Introduction

An N node Discrete Memoryless Network (DMN) is a network model where each node transmits its message to a set of destination nodes. The nodes can also relay messages from other nodes. With increasing applications in Ad-Hoc networks, Wi-Fi and sensor networks, the DMN has gained significant importance. The DMN is the most general network model used in studying the characteristics of cooperative communication and in multi-user information theory. It includes many important class of channels like noiseless, erasure and deterministic networks as special cases [29], [27], [8]. The DMN also includes the relay, broadcast, interference and multiple access channels which are the fundamental building blocks for any multi-user communication network.

The primary research focus in multi-user information theory is to find the capacity of the general discrete memoryless network. The best known upper bound for the DMN is the cut-set bound [6]. Several coding schemes have been developed that are close to optimal for some important classes of the DMN. The achievable rates and the cut-set upper bound do not coincide and the capacity of the DMN is not known in general. In this work, a novel scheme is proposed which achieves a higher rate than the existing schemes in literature under specific conditions. This reduces the gap between the achievable rate and the upper bound and is a step towards finding the capacity of the DMN.

The existing schemes for the relay channel would be a good starting point to design novel coding schemes for the DMN. Cover and El Gamal [5] introduced the main coding schemes for the general discrete memoryless single relay channel. The schemes introduced in brief are Decode Forward (DF), Compress Forward (CF) and Superposition Forward (SF).

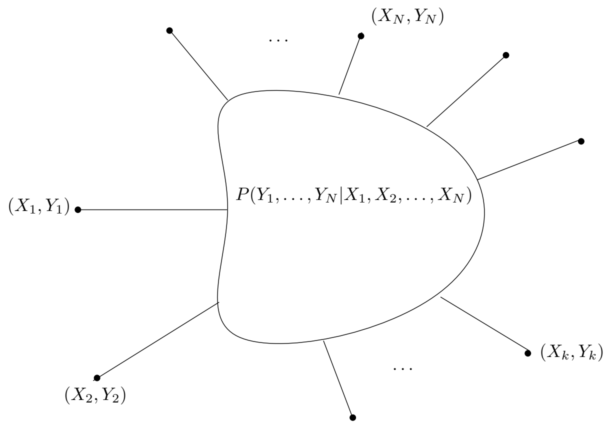

Figure 5.1 An N-node discrete memoryless network (DMN).

- In decode forward, the relay decodes the message transmitted from the source and coherently transmits to the destination.

- In compress forward, the relay compresses the received signal without decoding and helps the destination decode at a higher rate. The compress-forward scheme achieves any rate less than [5, Theorem 6]

$$R_{CF} = \sup I(X_1; \hat{Y}_2, Y_3 | X_2), \quad \text{such that } I(X_2; Y_3) \geq I(\hat{Y}_2; Y_2 | X_2, Y_3) \tag{5.1}$$

  where supremum is taken over all joint probability distributions of the form

$$p(x_1, x_2, y_2, y_3, \hat{y_2}) = p(x_1)p(x_2)p(y_2, y_3 | x_1, x_2)p(\hat{y_2} | y_1, x_2). \tag{5.2}$$

  The relay uses Wyner-Ziv [41] binning to send maximum information through the relay-destination link.

- Superposition-forward combines decode-forward and compress forwad using superposition. Decode-forward and compress-forward are special cases of superposition-forward. The superposition-forward scheme achieves the optimal rate for all the special cases where the capacity of the relay channel is known.

- El Gamal, Mohseni, and Zahedi [25] put forth an equivalent characterization of the CF scheme. That is, it achieves any rate less than

$$R_{CF} = \sup \min\{I(X_1; \hat{Y}_2, Y_3 | X_2), I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_3)\} \tag{5.3}$$

where supremum is still taken over all joint probability distributions of the same form as in (5.2). This equivalent representation naturally extends to the networks case.

Lim et al. [36] introduced a general lower bound for the discrete memoryless network using network coding [28] and the equivalent characterization of compress-forward. The new scheme is termed "Noisy network coding". The key ideas are message repetition encoding, no Wyner-Ziv [41] binning at the relay and joint decoding at the destination. The scheme achieves a higher rate than the better known compress-forward scheme for networks with multiple relays [43]. The noisy network coding scheme naturally extends to single and multiple source multicast networks.

We improve the achievable rates of the noisy network coding scheme by allowing the nodes to decode a part of message and use the message to make a better compressed signal to be relayed. The superposition noisy network coding scheme combines superposition forward with network coding. Modifications are made to the superposition-forward scheme to make it applicable to the network coding scenario. Specifically, the input distributions at each node are chosen to be independent.

In superposition noisy network coding, the message at each node is split in two parts. A part of the message is required to be decoded at each relay after every block. The other part of the message is transmitted over b blocks using repetition coding. The relay nodes use compress-forward to transmit this message. The destination nodes decode the messages after b blocks of transmission using joint decoding. Similar to noisy network coding, our scheme does not use Wyner-Ziv encoding at the relay, employs repetition encoding for a part of the message and uses joint decoding. These techniques have been shown to improve the achievable rates [36].

For simplicity and ease of understanding, the superposition noisy network coding scheme is first explained for a simple 3 node relay channel. In Section 5.2, the scheme is designed and achievable rates derived for a single relay channel. In Section 5.3, the scheme is further extended to single source multicast network where there is only a single source node transmitting information to a set of destination nodes. All other nodes act as relays. In Section 5.4, the superposition noisy network coding scheme is designed for multiple source multicast net-

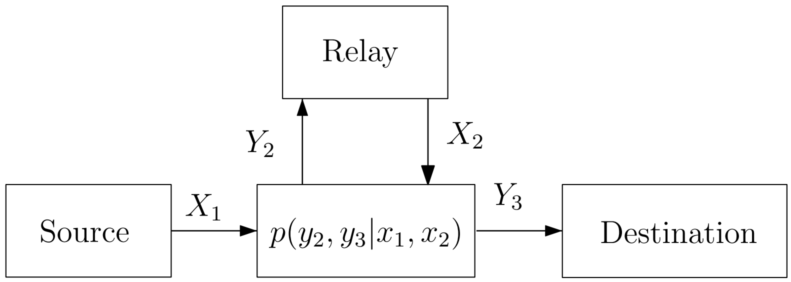

Figure 5.2 Discrete memoryless relay channel

works. This scheme is then derived for discrete memoryless single relay channel and two-way relay channel. The rates achieved are numerically compared to existing schemes to quantify performance.

5.2 Superposition noisy network coding for single relay channel

Consider the discrete memoryless relay channel $p(y_2, y_3|x_1, x_2)$ shown in Fig. 5.2. The source node is terminal 1, relay node is terminal 2, and destination is terminal 3. x_k and y_k denote the transmitted and received symbol at terminal k respectively.

The rate achieved by superposition-forward scheme [5, Theorem 7] for discrete memoryless relay channel is

$$R_{SF} = \sup(\min\{I(X_1; Y_3, \hat{Y}_2|X_2, U) + I(U; Y_2|X_2, V),$$
$$I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2|U, X_1, X_2, Y_3)\}) \quad (5.4)$$

where the supremum is over all joint probability distributions of the form

$$p(u, v, x_1, x_2, \hat{y}_2, y_3, y_2) = p(v)p(u|v)p(x_1|u)p(x_2|v)p(y_2, y_3|x_1, x_2)p(\hat{y}_2|x_2, y_2, u) \quad (5.5)$$

subject to the constraint

$$I(X_2; Y_3|V) \geq I(\hat{Y}_2; Y_2|X_2, Y_3, U). \quad (5.6)$$

Network coding requires the input distribution at each node to be independent [44]. Thus, we modify the superposition forward strategy such that the auxiliary random variables U and V are generated independent of each other. This may lead to a rate loss compared to original scheme under certain conditions. This happens since the maximization is now over all joint probability distributions of the form

$$p(u, v, x_1, x_2, y_2, y_3, \hat{y}_2) = p(v)p(u)p(x_1|u)p(x_2|v)p(y_2, y_3|x_1, x_2)p(\hat{y}_2|x_2, y_2, u) \quad (5.7)$$

which is a subset of (4.14).

For a single relay channel, we may choose the input distribution to be dependent and achieve the same rates as superposition-forward scheme. The rate achieved by superposition noisy network coding scheme for a single relay channel is stated in Theorem 3. The encoding and decoding process of the coding scheme is explained followed by the analysis of probability of error.

Theorem 3. *For any discrete memoryless relay channel, the rate* $\sup_P R' + R''$ *is achievable, where*

$$R' < \min\{I(U_1; Y_2|X_2),\ I(U_1, V_2; Y_3)\}. \quad (5.8)$$

$$R'' < \min\{I(X_1; \hat{Y}_2, Y_3|X_2, U_1),\ I(X_1, X_2; Y_3|U_1, V_2) - I(\hat{Y}_2; Y_2|U_1, X_1, X_2, Y_3)\}. \quad (5.9)$$

and the supremum is taken over all joint probability distributions of the form

$$p(u_1, v_2, x_1, x_2, y_2, y_3, \hat{y}_2) = p(u_1)p(v_2)p(x_1|u_1)p(x_2|v_2)p(y_2, y_3|x_1, x_2)p(\hat{y}_2|x_2, y_2, u_1) \quad (5.10)$$

Proof. The message m at the source is split in two parts m' and m''. m' is further split into b messages m'_j. The message $m'_j \in [1 : 2^{nR'}]$ is transmitted over every block j and the message $m'' \in [1 : 2^{nbR''}]$ is transmitted over b blocks of transmission. The source node transmits $\mathbf{x}_{1,j}(m''|m'_j)$ for each block $j \in [1 : b]$. $\mathbf{x}_{k,j}$ or $\mathbf{x}_{kj}$ is the message transmitted by node k in block j. After block j, the relay decodes the message $\hat{m}'_j$ and maps it to a $\mathbf{v}_{2,j+1}(\hat{m}'_j)$ codeword. It also finds a "compressed" version $\hat{\mathbf{y}}_{2j}(l_j|l_{j-1}, \hat{m}'_j, \hat{m}'_{j-1})$ of the relay output $\mathbf{y}_{2j}$ conditioned on $\mathbf{x}_{2j}(l_{j-1}|\hat{m}'_{j-1})$ and $\mathbf{u}_{1j}(\hat{m}'_j)$. The relay transmits a codeword $\mathbf{x}_{2,j+1}(l_j|\hat{m}'_j)$ in the next block. After b blocks of transmission, the decoder jointly decodes the message $m'' \in [1 : 2^{nbR}]$ using $(\mathbf{y}_{31}, \ldots, \mathbf{y}_{3b})$ received over b blocks simultaneously. The decoder has decoded all the messages m'_j by sliding window decoding for each block $j \in [1 : b]$. The details are as follows.

Block	1	2	3	$\cdots$	$b-1$	b
U_1	$\mathbf{u}_{11}(m'_1)$	$\mathbf{u}_{12}(m'_2)$	$\mathbf{u}_{13}(m'_3)$	$\cdots$	$\mathbf{u}_{1,b-1}(m'_{b-1})$	$\mathbf{u}_{1b}(m'_b)$
V_2	$\mathbf{v}_{21}(1)$	$\mathbf{v}_{22}(m'_1)$	$\mathbf{v}_{23}(m'_2)$	$\cdots$	$\mathbf{v}_{2,b-1}(m'_{b-2})$	$\mathbf{v}_{2b}(m'_{b-1})$
X_1	$\mathbf{x}_{11}(m''\|m'_1)$	$\mathbf{x}_{12}(m''\|m'_2)$	$\mathbf{x}_{13}(m''\|m'_3)$	$\cdots$	$\mathbf{x}_{1,b-1}(m''\|m'_{b-1})$	$\mathbf{x}_{1b}(m''\|m'_b)$
$\hat{Y}_2$	$\hat{\mathbf{y}}_{21}(l_1\|1,1,m'_1)$	$\hat{\mathbf{y}}_{22}(l_2\|l_1,m'_1,m'_2)$	$\hat{\mathbf{y}}_{23}(l_3\|l_2,m'_2,m'_3)$	$\cdots$	$\hat{\mathbf{y}}_{2,b-1}(l_{b-1}\|l_{b-2},m'_{b-2},m'_{b-1})$	$\hat{\mathbf{y}}_{2b}(l_b\|l_{b-1},m'_{b-1},m'_b)$
X_2	$\mathbf{x}_{21}(1\|1)$	$\mathbf{x}_{22}(l_1\|m'_1)$	$\mathbf{x}_{23}(l_2\|m'_2)$	$\cdots$	$\mathbf{x}_{2,b-1}(l_{b-2}\|m'_{b-2})$	$\mathbf{x}_{2b}(l_{b-1}\|m'_{b-1})$
Y_3	$\emptyset$	$\hat{m}'_1$	$\hat{m}'_2$	$\cdots$	$\hat{m}'_{b-2}$	$\hat{m}'_{b-1}$,$\hat{m}''$

Table 5.1 Superposition noisy network coding for the relay channel.

Codebook generation: Fix $p(u_1)p(x_1|u_1)p(v_2)p(x_2|v_2)p(\hat{y}_2|y_2,x_2,u_1)$. The codebooks are generated randomly and independently for each block

1. Generate $2^{nR'}$ sequences $\mathbf{u}_{1j}(m'_j)$, $m'_j \in [1:2^{nR'}], j \in [1:b]$ each with probability $\prod_{i=1}^{n} p_{U_1}(u_{1,(j-1)n+i})$.

2. For every $\mathbf{u}_{1j}(m'_j)$, generate $2^{nbR''}$ sequences $\mathbf{x}_{1j}(m''|m'_j)$, $m'' \in [1:2^{nbR''}]$, each with probability $\prod_{i=1}^{n} p_{X_1|U_1}(x_{1,(j-1)n+i}|u_{1,(j-1)n+i}(m'_j))$.

3. Generate $2^{nR'}$ sequences $\mathbf{v}_{2j}(m'_{j-1})$, $m'_{j-1} \in [1:2^{nR'}], j \in [1:b]$ each with probability $\prod_{i=1}^{n} p_{V_2}(v_{2,(j-1)n+i})$.

4. For every $\mathbf{v}_{2j}(m'_{j-1})$, generate $2^{n\hat{R}_2}$ sequences $\mathbf{x}_{2j}(l_{j-1}|m'_{j-1})$, $l_{j-1} \in [1:2^{n\hat{R}_2}], m'_{j-1} \in [1:2^{nR'}]$, each with probability
$\prod_{i=1}^{n} p_{X_2|V_2}(x_{2,(j-1)n+i}|v_{2,(j-1)n+i}(m'_{j-1}))$.

5. For every $\mathbf{x}_{2j}(l_{j-1}|m'_{j-1})$, and $\mathbf{u}_{1j}(m'_j)$, generate $2^{n\hat{R}_2}$ sequences $\hat{\mathbf{y}}_{2j}(l_j|l_{j-1},m'_{j-1},m'_j)$, $l_j, l_{j-1} \in [1:2^{n\hat{R}_2}]$, $m'_{j-1}, m'_j \in [1:2^{nR'}]$, each with probability
$\prod_{i=1}^{n} p_{\hat{Y}_2|X_2,U_1}(\hat{y}_{2,(j-1)n+i}|x_{2,(j-1)n+i}(l_{j-1},m'_{j-1}),u_{1,(j-1)n+i}(m'_j))$.

The codebook is

$$\mathcal{C}_j = \{\mathbf{u}_{1j}(m'_j),\mathbf{v}_{2j}(m'_{j-1}),\mathbf{x}_{1j}(m''|m'_j),\mathbf{x}_{2j}(l_{j-1}|m'_{j-1}),\hat{\mathbf{y}}_{2j}(l_j|l_{j-1},m'_{j-1},m'_j)$$
$$: m'_j, m'_{j-1} \in [1:2^{nR'}], m'' \in [1:2^{nbR''}], l_j, l_{j-1} \in [1:2^{n\hat{R}_2}]\} \quad (5.11)$$

for $j \in [1:b]$.

Encoding: The messages transmitted and decoded at each node in superposition noisy network coding is shown in table I. m'_j is the message transmitted in block j and m'' is the message transmitted over b blocks. The relay, upon receiving $\mathbf{y}_{2j}$ at the end of block $j \in [1:b]$, finds an index m'_j such that

$$(\mathbf{u}_{1j}(m'_j), \mathbf{y}_{2j}, \mathbf{x}_{2j}(l_{j-1}|\hat{m}'_{j-1})) \in \mathcal{T}_{\epsilon'}^{(n)},$$

where $\mathbf{x}_{2j}(l_{j-1}|\hat{m}'_{j-1})$ is the symbol transmitted by the relay in block j and $\mathcal{T}_{\epsilon'}^{(n)}$ is a set of ϵ'-typical sequences [2]. The relay then finds an index l_j such that

$$(\mathbf{u}_{1j}(\hat{m}'_j), \hat{\mathbf{y}}_{2j}(l_j|l_{j-1}, \hat{m}'_j, \hat{m}'_{j-1}), \mathbf{y}_{2j}, \mathbf{x}_{2j}(l_{j-1}|\hat{m}'_{j-1})) \in \mathcal{T}_{\epsilon'}^{(n)},$$

where $l_0 = 1$ by convention. If there is more than one such index, choose one of them at random. If there is no such index, choose an arbitrary index at random from $[1:2^{n\hat{R}_2}]$. The codeword pair $(\mathbf{x}_{1j}(m''|m'_j), \mathbf{x}_{2j}(l_{j-1}|m'_{j-1}))$ is transmitted in block $j \in [1:b]$.

Decoding: Let $\epsilon > \epsilon'$. After block j, the decoder uses $\mathbf{y}_{3(j-1)}$ and $\mathbf{y}_{3j}$ to find a unique message $\hat{m}'_{j-1} \in [1:2^{nR'}]$. The unique message satisfies the following two conditions simultaneously

$$(\mathbf{u}_{1j}(m'_j), \mathbf{v}_{2j}(\hat{m}'_{j-1}), \mathbf{y}_{3(j-1)}) \in \mathcal{T}_{\epsilon}^{(n)}$$
$$(\mathbf{v}_{2j}(m'_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_{\epsilon}^{(n)}$$

At the end of block b, after decoding the messages m'_j, $j \in [1:(b-1)]$ the decoder finds a unique message $\hat{m}'' \in [1:2^{nbR''}]$ such that

$$(\mathbf{u}_{1j}(\hat{m}'_j), \mathbf{v}_{2j}(\hat{m}'_{j-1}), \hat{\mathbf{y}}_{2j}(l_j|l_{j-1}, \hat{m}'_{j-1}, \hat{m}'_j), \mathbf{x}_{1j}(m''_j|\hat{m}'_{j-1}), \mathbf{x}_{2j}(l_{j-1}|\hat{m}'_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_{\epsilon}^{(n)} \quad \text{for all } j \in [1:b]$$

for some $l_1, l_2, \ldots, l_b$. If there is none or more than one such message, it declares an error.

Analysis of the probability of error: Let $\{M'_j\}$, denote the set of messages sent at the source node for all $j \in [1:(b-1)]$. To bound the probability of error in decoding the messages $\{M'_j\}$, assume without loss of generality that $\{M'_j = 1, j = 1, \ldots, b\}$. Define

$$\mathcal{E}_{2\{m'_j\}(0)} := \{(\mathbf{U}_{1j}(m'_j), \mathbf{Y}_{2j}, \mathbf{X}_{2j}(l_{j-1}|m'_{j-1})) \notin \mathcal{T}_{\epsilon'}^{(n)}\},$$
$$\mathcal{E}_{2\{m'_j\}(1)} := \{(\mathbf{U}_{1j}(m'_j), \mathbf{Y}_{2j}, \mathbf{X}_{2j}(l_{j-1}|m'_{j-1})) \in \mathcal{T}_{\epsilon'}^{(n)}, \forall m'_j \neq 1\}.$$

$\mathcal{E}_{2\{m'_j\}(0)}$ is the error event when the relay does not find any jointly typical message m'_j. $\mathcal{E}_{2\{m'_j\}(1)}$ is the error event where the relay finds a jointly typical message $m'_j \neq 1$ different from what was transmitted. The probability of error in decoding the message m'_j at the relay is upper bounded by

$$\mathsf{P}(\mathcal{E}_2) \leq \mathsf{P}\big(\bigcup_{j=1}^{b}\{\mathcal{E}_{2\{m'_j\}(1)} \cup \mathcal{E}_{\{m'_j\}(0)}\}\big)$$

The error event at each block can be analyzed independently assuming the messages have been decoded correctly till the previous block. The total probability of error would then be $(b+1)$ times the maximum block probability error.

By law of large numbers and conditional typicality lemma [2], the maximum probability of error at each block is

- $\mathsf{P}(\mathcal{E}_{2\{m'_j\}(0)}) \to 0$ as $n \to \infty$
- Assuming the messages have been decoded correctly till the previous block, $\mathsf{P}\big(\mathcal{E}_{2\{m'_j\}(1)}|\mathcal{E}^c_{2\{m'_j\}(0)}\big) \to 0$ as $n \to \infty$ if

$$R' \leq I(U_1; Y_2|X_2) \tag{5.12}$$

To bound the probability of error in decoding the message m'_j at the destination. Define the events

$$\mathcal{E}_{3\{m'_j\}(0)} := \big\{(\mathbf{U}_{1,j}(m'_j), \mathbf{V}_{2,j}(m'_{j-1}), \mathbf{Y}_{3j}) \notin \mathcal{T}_\epsilon^{(n)}\big\} \bigcup \big\{(\mathbf{V}_{2,(j+1)}(m'_j), \mathbf{Y}_{3,(j+1)}) \notin \mathcal{T}_\epsilon^{(n)},\ \big\}.$$

$$\mathcal{E}_{3\{m'_j\}(1)} := \big\{(\mathbf{U}_{1,j}(m'_j), \mathbf{V}_{2,j}(m'_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}\big\} \bigcup \big\{(\mathbf{V}_{2,(j+1)}(m'_j), \mathbf{Y}_{3,(j+1)}) \in \mathcal{T}_\epsilon^{(n)},\ \forall m'_j \neq 1\big\}.$$

$\mathcal{E}_{3\{m'_j\}(0)}$ is the error event when the destination does not find any jointly typical message m'_j. $\mathcal{E}_{3\{m'_j\}(1)}$ is the error event where the destination finds a jointly typical message $m'_j \neq 1$ different from what was transmitted. The probability of error in decoding the message m'_j at the destination is then upper bounded by

$$\mathsf{P}(\mathcal{E}_2) \leq \mathsf{P}\big(\bigcup_{j=1}^{b}\{\mathcal{E}_{2\{m'_j\}(1)} \cup \mathcal{E}_{\{m'_j\}(0)}\}\big)$$

Once again, we can analyze the error event for each block independently assuming all the messages till the previous block have been decoded correctly by the destination.

- By law of large numbers, $\mathsf{P}(\mathcal{E}_{3\{m'_j\}(0)}) \to 0$ as $n \to \infty$

- Since the codebooks are generated independently for each block, the two events of $\mathcal{E}_{3\{m'_j\}(1)}$ are independent. Thus, by conditional typicality lemma [2], $\mathsf{P}\big(\mathcal{E}_{3\{m'_j\}(1)}|\mathcal{E}^c_{3\{m'_j\}(0)}\big) \to 0$ as $n \to \infty$ if

$$R' \leq I(U_1;Y_3|V_2) + I(V_2;Y_3) = I(U_1,V_2;Y_3). \tag{5.13}$$

Combining (5.12) and (5.13), and for large b, the following rate is achievable

$$R' < \min\{I(U_1;Y_2|X_2), I(U_1,V_2;Y_3)\}$$

After decoding the messages m'_j for $j \in [1:(b-1)]$, the destination decodes the message M'' after b blocks. The probability of error analysis for message M'' is similar to the noisy network coding scheme [36], given the partial information of the messages m'_j. The probability of error analysis is explained here in brief.

Let M'' be the message sent at the source node over b blocks and L_j denote the indices chosen by the relay at block $j \in [1:b]$. To analyze the probability of error in decoding the unique message M'', define

$$\mathcal{E}_0 := \bigcup_{j=1}^{b}\{(\mathbf{U}_{1j}(\hat{m}'_j), \hat{\mathbf{Y}}_{2j}(l_j|L_{j-1},\hat{m}'_j,\hat{m}'_{j-1}), \mathbf{Y}_{2j}, \mathbf{X}_{2j}(L_{j-1}|\hat{m}'_{j-1})) \notin \mathcal{T}^{(n)}_{\epsilon'} \forall l_j\},$$

$$\begin{aligned}\mathcal{E}_{m''} := \big\{&(\mathbf{U}_{1j}(\hat{m}'_j), \mathbf{V}_{1j}(\hat{m}'_{j-1}), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1},\hat{m}'_{j-1},\hat{m}'_j), \mathbf{X}_{1j}(m''_j|\hat{m}'_{j-1}),\\ &\mathbf{X}_{2j}(l_{j-1}|\hat{m}'_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}^{(n)}_{\epsilon},\ j \in [1:b] \text{ for some } l_1, l_2, \ldots, l_b\big\}.\end{aligned}$$

To bound the probability of error, assume without loss of generality that $M'_j = 1$ for all j and $M'' = 1$. Assume all the M'_j's have been correctly decoded as 1 for sufficiently large n and $R' < \min\{I(U_1;Y_2|X_2), I(U_1,V_2;Y_3)\}$. Then the probability of error is upper bounded by

$$\mathsf{P}(\mathcal{E}) \leq \mathsf{P}(\mathcal{E}_0) + \mathsf{P}(\mathcal{E}^c_0 \cap \mathcal{E}^c_{1''}) + \mathsf{P}(\cup_{m''\neq 1}\mathcal{E}_{m''})$$

- By the covering lemma [2], $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$, if

$$\hat{R}_2 > I(\hat{Y}_2;Y_2|X_2,U). \tag{5.14}$$

- By conditional typicality lemma, $\mathsf{P}(\mathcal{E}^c_0 \cap \mathcal{E}^c_{1''}) \to 0$ as $n \to \infty$.

- To bound $\mathsf{P}(\cup_{m''\neq 1}\mathcal{E}_{m''})$, define the events

$$\mathcal{A}_j(m'', l_{j-1}, l_j) := \{(\mathbf{X}_{1j}(m''|1'), \mathbf{U}_{1j}(1'), \mathbf{V}_{1j}(1'), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}, 1', 1'), \mathbf{X}_{2j}(l_{j-1}|1'), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}\}.$$

for $j \in [1:b]$, $m'' \in [1:2^{nbR''}]$, and $l_{j-1}, l_j \in [1:2^{n\hat{R}_2}]$. Then, by the independence of codebooks in each block and the memoryless property of the channel, we upper bound the probability of error in decoding message m'' over all possible choices of l_j made at the relay and over all blocks j.

Following an analysis similar to [36], it can be shown that

$$\sum_{m''\neq 1}\sum_{l^b}\prod_{j=2}^{b}\mathsf{P}(\mathcal{A}_j(m'', l_{j-1}, l_j) \mid L_{j-1} = L_j = 1) \to 0$$

as $n \to \infty$, provided that

$$R'' < \frac{b-1}{b}(\min\{I_1,\ I_2 - \hat{R}_2\}) - \frac{1}{b}\hat{R}_2.$$

where

$$I_1 = I(X_1, \hat{Y}_2; Y_3|X_2, U_1, V_2)$$
$$I_2 = I(X_1, X_2; Y_3|U_1, V_2) + I(\hat{Y}_2; X_1, Y_3|X_2, U_1).$$

Finally, by eliminating $\hat{R}_2 > I(\hat{Y}_2; Y_2|X_2, U_1)$ and letting $b \to \infty$, we have

$$R'' < \min\{I(X_1; \hat{Y}_2, Y_3|X_2, U_1),\ I(X_1, X_2; Y_3|U_1, V_2) - I(\hat{Y}_2; Y_2|U_1, X_1, X_2, Y_3)\}.$$

□

The superposition noisy network coding scheme has been designed and analyzed for a single relay channel. It is in general advantageous to decode extra information available at the relay and use it to transmit more information to destination. The superposition noisy network coding scheme achieves the same rate as compress forward when the relay is ”close” to the destination. The rates achieved by superposition noisy network coding are higher than compress forward but less than those achieved by [5, Theorem 7] when the relay is ”close” to the source. The loss in performance arises due to the constraint that U_1 and V_2 are forced to be independent of each other. The superposition noisy network coding scheme is now extended to single source multicast networks.

5.3 Superposition noisy network coding for single source multicast networks

We now describe the superposition noisy network coding scheme for a single source discrete memoryless multicast networks (DMN-MC) $p(y_2, \ldots, y_N|x^N)$. Source terminal 1 splits the message in two parts m' and m'' and transmits using superposition forwarding. The message m' is transmitted in the same fashion decode-forward is extended to multicast relay networks [24], [23]. The scheme is modified to make the input distributions at each node independent of each other. After decoding the partial information m', the message m'' is decoded using noisy network coding [36] given the partial information.

Theorem 4. *For a discrete memoryless single source multicast network $p(y_2, \ldots, y_N|x^N)$, the rate $R' + R''$ is achievable if there exists some joint pmf $p(v_1)p(x_1|v_1)\prod_{k=2}^{N} p(v_k)p(x_k|v_k)p(\hat{y}_k|y_k, x_k, v_{k-1}^N)$ such that*

$$R' < \min_k I(V^{k-1}; Y_k|X_k, V_k^N)$$

$$R'' < \min_{\mathcal{S}} \Big(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_k|X(\mathcal{S}^c), V^N) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S})|X^N, \hat{Y}(\mathcal{S}^c), Y_k, V_{k-1}^N) \Big)$$

where $k \in \mathcal{D}$ the set of destination nodes. The source node is terminal 1 and the set of destination nodes $\mathcal{D} \subseteq [2, N]$. The minimum is over all possible cut-sets for node k.

Proof. The encoding and decoding process is similar to superposition noisy network coding for single relay channel. The relay nodes use an extension of decode-forward to multicast networks. The partial message is decoded at each of the nodes $\{2 : (k-1)\}$, and coherently transmitted to node k. The node k waits for $k-1$ transmissions to decode the partial information. After decoding the partial message, the remaining message is decoded using noisy network coding. The coding scheme is first explained for a DMN with a single destination node N. It is then extended to networks with a set of destination nodes, DMN-MC. An example of a four node discrete memoryless multicast network is shown in Fig. 5.3. This example would be used to explain the encoding and decoding process of superposition noisy network coding.

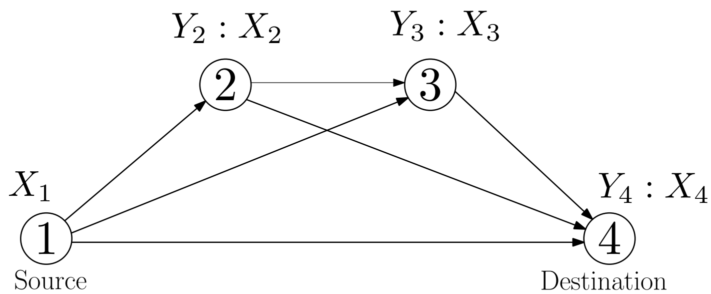

Figure 5.3 Four node discrete memoryless relay network

Codebook generation: Fix $p(v_1)p(x_1|v_1)p(v_k)p(x_k|v_k)p(\hat{y}_k|y_k, x_k, v_{k-1}^N)$ for $k \in (2:N)$. The codebooks are generated randomly and independently for each block $j \in [1:b]$.

1. Generate $2^{nR'}$ sequences $\mathbf{v}_{1j}(m'_j)$, $m' \in [1:2^{nR'}]$, each according to $\prod_{i=1}^n p_{V_1}(v_{1,(j-1)n+i})$.

2. For each $\mathbf{v}_{1j}(m'_j)$, generate $2^{nbR''}$ sequences $\mathbf{x}_{1j}(m''|m'_j)$, $m'' \in [1:2^{nbR''}]$, each according to $\prod_{i=1}^n p_{X_1|V_1}(x_{1,(j-1)n+i}|v_{1,(j-1)n+i}(m'_j))$.

3. Similarly, for each relay terminal $k \in [2:N-1]$, generate $2^{nR'}$ sequences $\mathbf{v}_{kj}(m'_{j-k+1})$, $m'_{j-k+1} \in [1:2^{nR'}]$, each according to $\prod_{i=1}^n p_{V_k}(v_{k,(j-1)n+i})$. We can assume $\mathbf{V}_N = \phi$ without loss of generality and to simplify notations.

4. For every $\mathbf{v}_{kj}(m'_{j-k+1}), k \in [2:N-1]$ generate $2^{n\hat{R}_2}$ sequences $\mathbf{x}_{kj}(l_{k,j-1}|m'_{j-k+1})$, $l_{k,j-1} \in [1:2^{n\hat{R}_2}], m'_{j-k+1} \in [1:2^{nR'}]$, each according to the distribution $\prod_{i=1}^n p_{X_k|V_k}(x_{k,(j-1)n+i}|v_{k,(j-1)n+i}(m'_{j-k+1}))$.

5. For every $\mathbf{x}_{kj}(l_{k,j-1}|m'_{j-k+1})$, $l_{k,j-1} \in [1:2^{n\hat{R}_2}]$ and $\mathbf{v}_{(k-1),j}(m'_{j-k+2}), \ldots, \mathbf{v}_{(N-1),j}(m'_{j-N+2})$, $m'_j \in [1:2^{nR'}]$, generate $2^{n\hat{R}_2}$ sequences $\hat{\mathbf{y}}_{kj}(l_{k,j}|l_{k,j-1}, m'^{(j-N+2)}_{j-k+2})$, $l_{k,j} \in [1:2^{n\hat{R}_2}]$, each according to $\prod_{i=1}^n p_{\hat{Y}_k|X_k,V_{k-1}^{N-1}}(\hat{y}_{k,(j-1)n+i}|x_{k,(j-1)n+i}(l_{k,j-1}, m'_{j-k+1}), v^{N-1}_{k-1,(j-1)n+i}(m'^{(j-N+2)}_{j-k+2}))$.

This defines the codebook

$$\mathcal{C}_j = \big\{\mathbf{v}_{1j}(m'_j), \mathbf{x}_{1j}(m''|m'_j), \mathbf{v}_{kj}(m'_{j-k+1}), \mathbf{x}_{kj}(l_{k,j-1}|m'_{j-k+1}), \hat{\mathbf{y}}_{kj}(l_{k,j}|l_{k,j-1}, m'^{(j-N+2)}_{j-k+2}) \\ : m'_j, m'_{j-k+1} \in [1:2^{nR'}], m'' \in [1:2^{nbR''}], l_{k,j}, l_{k,j-1} \in [1:2^{n\hat{R}_2}]\big\} \quad (5.15)$$

Block	1	2	3	$\cdots$	$j-1$
V_1	$\mathbf{v}_{11}(m'_1)$	$\mathbf{v}_{12}(m'_2)$	$\mathbf{v}_{13}(m'_3)$	$\cdots$	$\mathbf{v}_{1,j-1}(m'_{j-1})$
X_1	$\mathbf{x}_{11}(m''\|m'_1)$	$\mathbf{x}_{12}(m''\|m'_2)$	$\mathbf{x}_{13}(m''\|m'_3)$	$\cdots$	$\mathbf{x}_{1,j-1}(m''\|m'_{j-1})$
V_2	$\mathbf{v}_{21}(1)$	$\mathbf{v}_{22}(m'_1)$	$\mathbf{v}_{23}(m'_2)$	$\cdots$	$\mathbf{v}_{2,j-1}(m'_{j-2})$
X_2	$\mathbf{x}_{21}(1\|1)$	$\mathbf{x}_{22}(l_{21}\|m'_1)$	$\mathbf{x}_{23}(l_{22}\|m'_2)$	$\cdots$	$\mathbf{x}_{2,j-1}(l_{2,j-2}\|m'_{j-2})$
$\hat{Y}_2$	$\hat{\mathbf{y}}_{21}(l_{21}\|1,1,m'_1),l_{21}$	$\hat{\mathbf{y}}_{22}(l_{22}\|l_{21},m'_1,m'_2),l_{22}$	$\hat{\mathbf{y}}_{23}(l_{23}\|l_{22},m_1^{\prime 3}),l_{23}$	$\cdots$	$\hat{\mathbf{y}}_{2,j-1}(l_{2,j-1}\|l_{2,j-2},m_{j-1}^{\prime(j-3)}),l_{2,j-1}$
V_3	$\mathbf{v}_{31}(1)$	$\mathbf{v}_{32}(1)$	$\mathbf{v}_{33}(m'_1)$	$\cdots$	$\mathbf{v}_{3,j-1}(m'_{j-3})$
X_3	$\mathbf{x}_{31}(1\|1)$	$\mathbf{x}_{32}(l_{31}\|m'_1)$	$\mathbf{x}_{33}(l_{32}\|m'_1)$	$\cdots$	$\mathbf{x}_{3,j-1}(l_{3,j-2}\|m'_{j-3})$
$\hat{Y}_3$	$\hat{\mathbf{y}}_{31}(l_{31}\|1,1,m'_1),l_{31}$	$\hat{\mathbf{y}}_{32}(l_{32}\|l_{31},1,m'_1),l_{32}$	$\hat{\mathbf{y}}_{33}(l_{33}\|l_{32},m'_1,m'_2),l_{33}$	$\cdots$	$\hat{\mathbf{y}}_{3,j-1}(l_{3,j-1}\|l_{3,j-2},m'_{j-2},m'_{j-3}),l_{3,j-1}$
Y_4	$\emptyset$	$\emptyset$	$\hat{m}'_1$	$\cdots$	$\hat{m}'_{j-3}$

Table 5.2 Superposition noisy network coding for a four node DMN.

for $j \in [1:b]$ and $k \in [2:N-1]$.

Encoding: Let m'_j be the message to be sent in block j and m'' be the message to be sent over b blocks. The relay node k, upon receiving $\mathbf{y}_{k(j+k-2)}$ at the end of block $j+k-2 \in [1:b]$, finds an estimate m'_{kj} of the message m'_j. The relay also finds an estimate of the compressed signal $\hat{\mathbf{y}}_{k,j+k-2}(l_{k,j+k-2}|l_{k,j+k-3}, m_j^{\prime(j+k-N)})$. In block $j+k-1$, it transmits $\mathbf{x}_{k,j+k-1}(l_{k,j+k-2}|m'_{kj})$. An example of a four node discrete memoryless multicast network is shown in Fig. 5.3. The messages transmitted and decoded for this network is explained in table 5.2.

Decoding and probability of error analysis: At the end of block $j+k-2$, node k finds a unique message m'_j such that

$$(\mathbf{V}_{1j}(m'_j), \mathbf{V}_{2j}(\hat{m}'_{j-1}), \ldots, \mathbf{V}_{k-1,j}(\hat{m}'_{j-k+2}), \ldots, \mathbf{V}_{Nj}(\hat{m}'_{j-N+1}), \mathbf{X}_{kj}(l_{k,j-1}|m'_{j-k+1}), \mathbf{Y}_{kj}) \in \mathcal{T}_{\epsilon'}^{(n)},$$
$$(\mathbf{V}_{2,j+1}(m'_j), \ldots, \mathbf{V}_{k,j+1}(\hat{m}'_{j-k+2}), \ldots, \mathbf{V}_{N,j+1}(\hat{m}'_{j-N+2}), \mathbf{X}_{k,j+1}(l_{k,j}|m'_{j-k+2}), \mathbf{Y}_{k,j+1}) \in \mathcal{T}_{\epsilon'}^{(n)},$$
$$\vdots$$
$$(\mathbf{V}_{k-1,j+k-2}(m'_j), \ldots, \mathbf{V}_{N,j+k-2}(m'_{j+k-N-1}), \mathbf{X}_{k,j+k-2}(l_{k,j+k}|m'_{j-1}), \mathbf{Y}_{k,j+k-2}) \in \mathcal{T}_{\epsilon'}^{(n)}.$$

The symbols $\mathbf{V}_k^N$ transmitted by the nodes after k are already known to terminal k. In fact, node k instructs the terminals what to transmit in future blocks.

By the independence of codebooks, the Law of large numbers and joint typicality lemma, it can be shown that the probability of error in decoding the message m'_j at relay node k tends to 0 as $n \to \infty$, provided that

$$R' < \min_k I(V_1^{k-1}; Y_k | X_k, V_k^N) \tag{5.16}$$

The minimum is over all possible cutsets for the relay k.

Each relay node k also finds an index $l_{k,j}$ after block j such that

$$(\mathbf{V}_{k-1,j}(\hat{m}'_{j-k+2}),\ldots,\mathbf{V}_{N,j}(\hat{m}'_{j-N+1}),\ldots,\hat{\mathbf{Y}}_{k,j}(l_{k,j}|l_{k,j-1},\hat{m}'^{(j-N+2)}_{j-k+2}),$$
$$\mathbf{Y}_{kj},\mathbf{X}_{kj}(l_{k,j-1}|\hat{m}'_{j-k+1}))\in\mathcal{T}^{(n)}_{\epsilon'}$$

where $l_0 = 1$ by convention. If there is more than one such index, choose one of them at random. If there is no such index, choose an arbitrary index at random from $[1 : 2^{n\hat{R}_2}]$. Note that $V_{k-1,j}(m'_{j-k})$ is decoded at node k after block j.

By covering lemma, the probability of error goes to 0 as $n \to \infty$ if

$$\hat{R}_k > I(\hat{Y}_k; Y_k|X_k, V^N_{k-1})$$

The codeword $(\mathbf{x}_{1j}(m''|m'_j), \mathbf{x}_{kj}(l_{k,j-1}|m'_{j+k-1}))$ is transmitted in block $j \in [1 : b]$.

For decoding the message m'' at the end of b blocks, assume without loss of generality that $\mathbf{L}_1 = \cdots = \mathbf{L}_b = \mathbf{1}$, where $\mathbf{L}_j := (L_{1j},\ldots,L_{Nj})$. We also assume the messages m'_j transmitted were all 1's and were decoded correctly for all the blocks j. The probability of error analysis is similar to [36], [2] conditioned on the decoded messages m'.

To bound the probability of error in decoding the message m'', define the event

$$\mathcal{A}_j(\mathbf{m}'',\mathbf{l}_{j-1},\mathbf{l}_j,\vec{1}) := \{(\mathbf{X}_{1j}(m''_1|1),\ldots,\mathbf{X}_{Nj}(l_{N,j-1}|\vec{1}),\mathbf{V}_{1j}(1'),\ldots,\mathbf{V}_{Nj}(1'),$$
$$\hat{\mathbf{Y}}_{2j}(l_{2j}|l_{2,j-1},\vec{1}),\ldots,\hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1},\vec{1}),\mathbf{Y}_{Nj})\in\mathcal{T}^{(n)}_{\epsilon}\}$$

for $\mathbf{m}'' \neq \mathbf{1}$ and all $\mathbf{l}_j$.

Using the independence of codebooks for each block j and the memoryless property of the channel, the probability of error event can be upper bounded over all possible choices of indices $l_{k,j}$ at the relay for every block j.

Following an analysis similar to [36], it can be shown that

$$\sum_{\mathbf{m}''\neq\mathbf{1}}\sum_{\mathbf{l}^b}\prod_{j=2}^{b}\mathsf{P}(\mathcal{A}_j(\mathbf{m}'',\mathbf{l}_{j-1},\mathbf{l}_j))\to 0$$

which tends to zero as $n \to \infty$ if

$$R'' < \frac{b-1}{b}\left(\min_{\mathcal{S}}\left(I_1(\mathcal{S})+I_2(\mathcal{S})-\sum_{k\in\mathcal{S}}\hat{R}_k\right)\right)-\frac{1}{b}\left(\sum_{k=2}^{N-1}\hat{R}_k\right)$$

The minimum is over $\mathcal{S} \subseteq [1:N]$ such that $N \in \mathcal{S}^c$ and

$$I_1(\mathcal{S}) := I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_N | X(\mathcal{S}^c), V^N),$$
$$I_2(\mathcal{S}) := \sum_{k \in \mathcal{S}} I(\hat{Y}_k; \hat{Y}(\mathcal{S}^c \cup \{k' \in \mathcal{S} : k' < k\}), Y_N, X^N | X_k, V_{k-1}^N).$$

By eliminating $\hat{R}_k > I(\hat{Y}_k; Y_k | X_k, V_{k-1}^N)$ and letting $b \to \infty$, the probability of error tends to zero as $n \to \infty$ if

$$R'' < \min_{\mathcal{S}} \left(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} I(\hat{Y}_k; Y_k | X_k, V_{k-1}^N) \right)$$

Finally, note that

$$I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} I(\hat{Y}_k; Y_k | X_k, V_{k-1}^N) = -I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, V_{k-1}^N).$$

Therefore, the probability of error tends to zero as $n \to \infty$ if

$$R'' < \min_{\mathcal{S}} \left(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_N | X(\mathcal{S}^c), V^N) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, V_{k-1}^N) \right) \quad (5.17)$$

for all $\mathcal{S} \subseteq [1:N]$ such that $N \in \mathcal{S}^c$.

The rate achieved by the single destination node N using superposition noisy network coding is

$$R' < \min_k I(V^{k-1}; Y_k | X_k, V_k^N)$$
$$R'' < \min_{\mathcal{S}} \left(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_N | X(\mathcal{S}^c), V^N) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, V_{k-1}^N) \right)$$

For multicast network with a set of destination nodes, the relay function remains the same as the single destination case. The partial message is decoded at all the nodes. Each destination node also decode the superimposed message m'' after b blocks using the decoding scheme for the single terminal case. The probability of error $\to 0$ as $n \to 0$ if the rates for each destination node $d \in \mathcal{D}$ satisfies

$$R' < \min_k I(V^{k-1}; Y_k | X_k, V_k^N)$$
$$R'' < \min_{\mathcal{S}} \left(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c), V^N) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d, V_{k-1}^N) \right)$$

for all $d \in \mathcal{D}$, the set of destination nodes. The minimum is over all possible cut-sets for node d.

The achievable rate is also maximized over all possible permutations of the order in which the relay nodes decode the messages. Depending on the channel conditions, a relay node may be able to decode the source message at a higher rate and coherently transmit to subsequent relays. The order in which the relays decode the messages should be chosen such that the achievable rate of the channel is maximized.

This derives the achievable rate by superposition noisy network coding for single source discrete memoryless multicast networks. □

5.4 Superposition noisy network coding for multiple source multicast networks

The superposition noisy network coding scheme is generalized to an N node discrete memoryless multiple source multicast network $p(y^N|x^N)$, [36] shown in Fig. 5.4. In the general setup each node sends its independent message to a set of destination nodes while acting as relays for messages from other sources.

A general assumption is made on the multiple source multicast network to make the application of superposition noisy network coding easier. The source nodes are restricted not to act as relays. This is a reasonable assumption since the special cases that we are mainly interested do not require the source to act as relays. Two-way relay channel and interference relay channel are good examples. Let the nodes 1 to k_0 be the source nodes that do not relay the messages of other users. The nodes can act as destination and decode messages transmitted by other users.

The channel model is now similar to single source multicast network with a replacement of the source node with k_0 independent nodes. The partial information is transmitted the same way decode-forward is extended for the single source multicast network in Section 5.3. The relay decodes the message from all the sources using the decoding scheme of an m-user multiple access channel [6]. After decoding the partial information from the source nodes, the relay uses

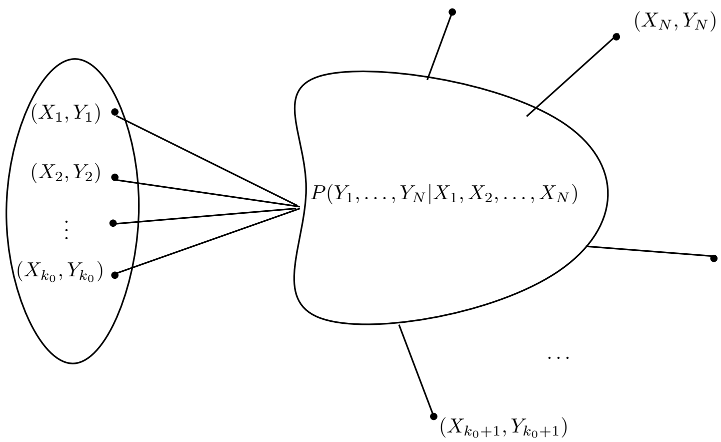

Figure 5.4 An N-node discrete memoryless network with k_0 sources.

binning to transmit the decoded information [30]. Further relays and destination nodes decode the message in the same multiple access fashion. The relays that have decoded the messages act as source nodes and coherently transmit the partial information. The remaining message is superimposed and decoded using noisy network coding.

Theorem 5. *For an N node discrete memoryless multiple source multicast network, the sum of the rates $R'+R''$ is achievable for any probability distribution $\prod_{k=1}^{k_0} p(v_k)p(x_k|v_k)\prod_{k=k_0+1}^{N} p(v_k)p(x_k|v_k)p(\hat{y}_k|y_k,$ The closure of the convex hull of the sum of rate vectors is achievable*

$$R'(\mathcal{S}) < \min_{\mathcal{S}} I(V(\mathcal{S}); Y_k | X_k, V(\mathcal{S}^c))$$

$$R''(\mathcal{S}) < \min_{\mathcal{S}} \Big(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_N | X(\mathcal{S}^c), V^N) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, V_{k-1}^N) \Big)$$

where k_0 is the number of source nodes, and the maximum is over all possible cut-sets $\mathcal{S} \subset$ $\{1, \ldots, k_0, \ldots, k-1\}$.

Proof. Codebook generation: Fix $\prod_{k=1}^{k_0} p(v_k)p(x_k|v_k)\prod_{k=k_0+1}^{N} p(v_k)p(x_k|v_k)p(\hat{y}_k|y_k, x_k, v_{k-1}^N)$. We randomly and independently generate a codebook for each block.

1. For each $j \in [1:b]$ and $k \in [1:k_0]$, generate $2^{nR'_k}$ sequences $\mathbf{v}_{k,j}(m'_k)$, $m'_{k,j} \in [1:2^{nR'_k}]$. each according to $\prod_{i=1}^{n} p_{V_k}(v_{k,(j-1)n+i})$

2. For each $\mathbf{v}_{k,j}(m'_k)$, $j \in [1:b]$ and $k \in [1:k_0]$, generate $2^{nbR''_k}$ sequences $\mathbf{x}_{k,j}(m''_k|m'_k)$, such that $m''_k \in [1:2^{nbR''_k}]$, $m'_k \in [1:2^{nbR'_k}]$. The sequences are generated independently

according to the distribution

$\prod_{i=1}^{n} p_{X_k|V_k}(x_{k,(j-1)n+i}|v_{k,(j-1)n+i}(m'_k))$

3. For all relay nodes $k \in [k_0+1:N]$ generate $2^{n\tilde{R}'_k}$ codewords $\mathbf{v}_{k,j}(\kappa_{j-k+1}(m_1^{'k_0}))$. The rate $\tilde{R}'_k$ is chosen such that

$$\tilde{R}'_k \geq \max_{d\in\mathcal{D}} I(V_k;Y_d|V_{k-1}^N)$$

The maximum is over $\mathcal{D}$ the set of all destination nodes. $\kappa_{j-k+1}(m_1^{'k_0})$ is the bin index of the messages $m_1^{'k_0}$ decoded at the relay k during block j. We drop the subscript $j+k-1$ for simplicity. The relay and block of each index would be evident from the subscripts of the symbol $\mathbf{v}$.

4. For each $\mathbf{v}_{k,j}(\kappa(m_1^{'k_0}))$ and $k\in[k_0+1:N]$, generate $2^{n\hat{R}_k}$ sequences $\mathbf{x}_{k,j}(l_{k,j-1}|\kappa(m_1^{'k_0}))$, such that $m'_k \in [1:2^{nR'_k}]$, $l_{k,j-1}\in[1:2^{n\hat{R}_k}]$, each according to the probability distribution $\prod_{i=1}^{n} p_{X_k|V_k}(x_{k,(j-1)n+i}|v_{k,(j-1)n+i}(\kappa(m_1^{'k_0})))$.

5. For each node $k\in[k_0+1:N]$ and each $\mathbf{x}_{k,j}(l_{k,j-1}|\kappa(m_1^{'k_0}))$, $\mathbf{v}_{k-1,j}(\kappa(m_1^{'k_0}))$, $\ldots,\mathbf{v}_{Nj}(\kappa(m_1^{'k_0}))$, such that $m'_k\in[1:2^{nR'_k}]$, $l_{k,j-1}\in[1:2^{n\hat{R}_k}]$, generate $2^{n\hat{R}_k}$ sequences $\hat{\mathbf{y}}_{kj}(l_{kj}|l_{k,j-1},\kappa_{j-k+2}^{j-N+1}(m_1^{'k_0}))$, $l_{kj}\in[1:2^{n\hat{R}_k}]$, each according to $\prod_{i=1}^{n} p_{\hat{Y}_k|X_k,V_k}(\hat{y}_{k,(j-1)n+i}|x_{k,(j-1)n+i}(m''_k|m'_k),v_{k-1,(j-1)n+i}^{N,(j-1)n+i}(\kappa(m_1^{'k_0})))$.

This defines the codebook

$$\begin{aligned}\mathcal{C}_j = \big\{&\mathbf{v}_{k,j}(m'_k),\mathbf{x}_{k,j}(m''_k|m'_k),k\in[1:k_0],\\ &\mathbf{v}_{k,j}(\kappa(m_1^{'k_0})),\mathbf{x}_{k,j}(l_{k,j-1}|\kappa(m_1^{'k_0})),\hat{\mathbf{y}}_{kj}(l_{kj}|l_{k,j-1},\kappa_{j-k+2}^{j-N+1}(m_1^{'k_0})),k\in[k_0+1:N]\\ &:m'_k\in[1:2^{nR'_k}],m''_k\in[1:2^{nbR''_k}],l_{kj},l_{k,j-1}\in[1:2^{n\hat{R}_k}]\big\}\end{aligned}$$

for $j\in[1:b]$.

Encoding: Let $(m'_1,\ldots,m'_{k_0},m''_1,\ldots,m''_{k_0})$ be the messages transmitted. Each relay node $k\in[k_0+1:N]$, upon receiving $\mathbf{y}_{kj}$ at the end of block $j\in[1:b]$, decode the messages $m_1^{'k_0}$ as shown in the decoding step. The relay node then finds a bin index $\kappa(m_1^{'k_0})$. After finding $m_1^{'k_0}$ and the bin index, the node finds an index l_{kj} such that

$$(\hat{\mathbf{Y}}_{kj}(l_{kj}|l_{k,j-1},\kappa(m_1^{'k_0})),\mathbf{Y}_{kj},\mathbf{X}_{k,j}(l_{k,j-1}|\kappa(m_1^{'k_0})),\mathbf{V}_{k-1,j}^{N,j}(\kappa(m_1^{'k_0})))\in\mathcal{T}_{\epsilon'}^{(n)},$$

where $l_{k,0} = 1$, $k \in [k_0 + 1 : N]$, by convention. If there is more than one such index, choose one of them at random. If there is no such index, choose an arbitrary index at random from $[1 : 2^{n\hat{R}_k}]$. Then each node $k \in [k_0 + 1 : N]$ transmits the codeword $\mathbf{x}_{k,j}(l_{k,j-1}|\kappa(m_1^{'k_0}))$ in block $j \in [1 : b]$.

Decoding: Let $\epsilon > \epsilon'$.After each block, the decoder $d \in \mathcal{D}$ decodes the messages $m_1^{'k_0}$. The messages are decoded as a k_0 user multiple access channel [6, Theorem 14.3.5]. The relays use joint typicality decoding and looks for messages $m_1^{'k_0}$ satisfying

$$(\mathbf{V}_{1,j}(m_1'), \ldots, \mathbf{V}_{k_0,j}(m_{k_0}'), \mathbf{V}_{k_0+1,j}(\kappa(m_1^{'k_0})), \ldots, \mathbf{V}_{N,j}(\kappa(m_1^{'k_0})),$$
$$\mathbf{X}_{k,j}(l_{k,j-1}|\kappa(m_1^{'k_0})), Y_{kj}) \in \mathcal{T}_\epsilon^{(n)}$$

for all $j \in [1 : b]$

After b blocks,the decoder $d \in \mathcal{D}$ finds a unique message set $(\hat{m}_{1d}'', \ldots, \hat{m}_{k_0 d}'')$, where $\hat{m}_{kd}'' \in [1 : 2^{nbR_k''}]$, such that there exist some $(l_{1j}, \ldots, l_{Nj})$, $l_{kj} \in [1 : 2^{n\hat{R}_k}]$, and $j \in [1 : b]$, satisfying

$$(\mathbf{V}_{1,j}(m_1'), \ldots, \mathbf{V}_{k_0,j}(m_{k_0}'), \mathbf{V}_{k_0+1,j}(\kappa(m_1^{'k_0})), \ldots, \mathbf{V}_{Nj}(\kappa(m_1^{'k_0})), \mathbf{X}_{1,j}(m_1''|m_1'), \ldots,$$
$$\mathbf{X}_{k_0,j}(m_{k_0}''|m_{k_0}'), \mathbf{X}_{(k_0+1),j}(l_{(k_0+1),j-1}|\kappa(m_1^{'k_0})), \ldots, \mathbf{X}_{N,j}(l_{N,j-1}|\kappa(m_1^{'k_0})),$$
$$\hat{\mathbf{Y}}_{(k_0+1),j}(l_{(k_0+1),j}|l_{k_0+1,j-1}, \kappa(m_1^{'k_0})), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1}, \kappa(m_1^{'k_0})), \mathbf{Y}_{Nj}) \in \mathcal{T}_\epsilon^{(n)}$$

for all $j \in [1 : b]$, given that the messages $m_1^{'k_0}$ have been decoded correctly.

Analysis of the probability of error: Relay node k can find the unique messages $m_1^{'k_0}$ by joint typical decoding. The messages are decoded in a manner similar to multiple access channels. The relays will find a set of unique messages $m_1^{'k_0}$ satisfying

$$(\mathbf{V}_{1,j}(m_1'), \ldots, \mathbf{V}_{k_0,j}(m_{k_0}'), \mathbf{V}_{k_0+1,j}(\kappa(m_1^{'k_0})), \ldots, \mathbf{V}_{N,j}(\kappa(m_1^{'k_0})),$$
$$\mathbf{X}_{k,j}(l_{k,j-1}|\kappa(m_1^{'k_0})), Y_{kj}) \in \mathcal{T}_\epsilon^{(n)}$$

for all $j \in [1 : b]$ if,

$$R'(\mathcal{S}) < I(V(\mathcal{S}); Y_k|V(\mathcal{S}^c), X_k)$$

for all cut-sets $\mathcal{S}$. By time sharing, the closure of the convex hull of the above rate vectors can be achieved.

Let M''_k denote the message sent at node $k \in [1:k_0]$ and L_{kj}, $j \in [1:b]$, denote the index chosen by node $k \in [k_0+1, N]$ for block j. To bound the probability of error for decoder $d \in \mathcal{D}$, assume without loss of generality that $(M'_1, \dots, M'_{k_0}) = (M''_1, \dots, M''_{k_0}) = (1, \dots, 1) =: \mathbf{1}$. Define the probability of errors given that $m_1^{\prime k_0}$ is decoded correctly.

$$\begin{aligned}
\mathcal{E}_0 :=& \bigcup_{j=1}^{b} \big(\mathbf{V}_{(k-1),j}(\kappa(m_1^{\prime k_0})), \dots, \mathbf{V}_{Nj}(\kappa(m_1^{\prime k_0})), \mathbf{X}_{k,j}(l_{k,j-1}|\kappa(m_1^{\prime k_0})), \\
&\hat{\mathbf{Y}}_{k,j}(l_{k,j}|l_{k,j-1}, \kappa(m_1^{\prime k_0})), \mathbf{Y}_{kj}\big) \in \mathcal{T}_\epsilon^{(n)} \\
\mathcal{E}_{\mathbf{m}''} :=& \big\{\big(\mathbf{V}_{1,j}(1), \dots, \mathbf{V}_{k_0,j}(1), \mathbf{V}_{k_0+1,j}(\vec{1}), \dots, \mathbf{V}_{Nj}(\vec{1}), \mathbf{X}_{1,j}(m''_1|1), \dots, \\
&\mathbf{X}_{k_0,j}(m''_{k_0}|1), \mathbf{X}_{(k_0+1),j}(l_{(k_0+1),j-1}|\vec{1}), \dots, \mathbf{X}_{N,j}(l_{N,j-1}|\vec{1}), \\
&\hat{\mathbf{Y}}_{(k_0+1),j}(l_{(k_0+1),j}|l_{k_0+1,j-1}, \vec{1}), \dots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1}, \vec{1}), \mathbf{Y}_{dj}\big) \in \mathcal{T}_\epsilon^{(n)}\big\}. \\
&\text{for some} \quad (\mathbf{l}_1, \dots, \mathbf{l}_b), j \in [1:b].
\end{aligned}$$

Here, $\mathbf{l}_j = (l_{1j}, \dots, l_{Nj})$ for $j \in [1:b]$. Then the probability of error is upper bounded as

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}(\mathcal{E}_0) + \mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{\mathbf{1}''}^c) + \mathsf{P}(\cup_{\mathbf{m}'' \ne \mathbf{1}} \mathcal{E}_{\mathbf{m}''}), \tag{5.18}$$

As in the single relay channel case, by the covering lemma, $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$, if $\hat{R}_k > I(\hat{Y}_k; Y_k | X_k, V_{k-1}^N)$, $k \in [k_0+1:N]$, and by the conditional typicality lemma $\mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{\mathbf{1}''}^c) \to 0$ as $n \to \infty$. For the third term, assume without loss of generality that $\mathbf{L}_1 = \cdots = \mathbf{L}_b = \mathbf{1}$, where $\mathbf{L}_j := (L_{1j}, \dots, L_{Nj})$. Define the events

$$\begin{aligned}
\mathcal{A}_j(\mathbf{m}'', \mathbf{l}_{j-1}, \mathbf{l}_j, \vec{1}) :=& \big\{\big(\mathbf{V}_{1,j}(1), \dots, \mathbf{V}_{k_0,j}(1), \mathbf{V}_{k_0+1,j}(\vec{1}), \dots, \mathbf{V}_{Nj}(\vec{1}), \mathbf{X}_{1,j}(m''_1|1), \dots, \\
&\mathbf{X}_{k_0,j}(m''_{k_0}|1), \mathbf{X}_{(k_0+1),j}(l_{(k_0+1),j-1}|\vec{1}), \dots, \mathbf{X}_{N,j}(l_{N,j-1}|\vec{1}), \\
&\hat{\mathbf{Y}}_{(k_0+1),j}(l_{(k_0+1),j}|l_{k_0+1,j-1}, \vec{1}), \dots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1}, \vec{1}), \mathbf{Y}_{Nj}\big) \in \mathcal{T}_\epsilon^{(n)}\big\}
\end{aligned}$$

for $\mathbf{m}'' \ne \mathbf{1}$ and all $\mathbf{l}_j$. Then,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_{\mathbf{m}''}) &= \mathsf{P}(\cup_{\mathbf{l}^b} \cap_{j=1}^b \mathcal{A}_j(\mathbf{m}'', \mathbf{l}_{j-1}, \mathbf{l}_j, \vec{1})) \\
&\le \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(\mathbf{m}'', \mathbf{l}_{j-1}, \mathbf{l}_j, \vec{1})),
\end{aligned}$$

For each $\mathbf{l}^b$ and $j \in [2:b]$, let $\mathcal{S}_j(\mathbf{l}^b) \subseteq [1:N]$ such that $\mathcal{S}_j(\mathbf{l}^b) = \{k : l_{k,j-1} \ne 1\}$.

Define $\mathbf{X}_j(\mathcal{S}_j(\mathbf{l}_{j-1}))$ to be the set of $\mathbf{X}_{kj}(l_{k,j-1}|\vec{1})$, $k \in \mathcal{S}_j(\mathbf{l}_{j-1})$, where $l_{k,j-1}$ are the corresponding elements in $\mathbf{l}^b$. Similarly define $\hat{\mathbf{Y}}_j(\mathcal{S}_j(\mathbf{l}_{j-1}))$ and $\mathbf{Y}_j(\mathcal{S}_j(\mathbf{l}_{j-1}))$. Then, by joint typicality lemma

$$\mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \le 2^{-n(I_1(\mathcal{S}(\mathbf{l}_{j-1}))+I_2(\mathcal{S}(\mathbf{l}_{j-1})))},$$

where

$$I_1(\mathcal{S}) := I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c), V^N),$$
$$I_2(\mathcal{S}) := \sum_{k\in\mathcal{S}} I(\hat{Y}_k; \hat{Y}(\mathcal{S}^c \cup \{k' \in \mathcal{S} : k' < k\}), Y_d, X^N | X_k, V_{k-1}^N)$$

Following an analysis similar to [36], it can be shown that

$$\sum_{\mathbf{m}''\neq \mathbf{1}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(\mathbf{m}'', \mathbf{l}_{j-1}, \mathbf{l}_j)) \to 0$$

as $n \to \infty$ if

$$R''(\mathcal{S}) < \frac{b-1}{b}\left(\min_{\mathcal{S}}\left(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k\in\mathcal{S}} \hat{R}_k\right)\right) - \frac{1}{b}\left(\sum_{k=2}^{N-1} \hat{R}_k\right)$$

where the minimum is over $\mathcal{S} \subseteq [1:N]$ such that $d \in \mathcal{S}^c$

By eliminating $\hat{R}_k > I(\hat{Y}_k; Y_k | X_k, V_{k-1}^N)$ and letting $b \to \infty$, the probability of error tends to zero as $n \to \infty$ if

$$R''(\mathcal{S}) < \min_{\mathcal{S}}\left(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k\in\mathcal{S}} I(\hat{Y}_k; Y_k | X_k, V_{k-1}^N)\right)$$

Finally, note that

$$\begin{aligned}
I_2(\mathcal{S}) - \sum_{k\in\mathcal{S}} I(\hat{Y}_k; Y_k | X_k, V_{k-1}^N) &= \sum_{k\in\mathcal{S}} -I(\hat{Y}_k; Y_k | X^N, \hat{Y}(\mathcal{S}^c), Y_N, \hat{Y}(\{k' \in \mathcal{S} : k' < k\}), V_{k-1}^N) \\
&= \sum_{k\in\mathcal{S}} -I(\hat{Y}_k; Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, \hat{Y}(\{k' \in \mathcal{S} : k' < k\}), V_{k-1}^N) \\
&= -I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, V_{k-1}^N).
\end{aligned}$$

Therefore, the probability of error tends to zero as $n \to \infty$ if

$$R''(\mathcal{S}) < \min_{\mathcal{S}}\left(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_N | X(\mathcal{S}^c), V^N) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, V_{k-1}^N)\right) \quad (5.19)$$

for all $\mathcal{S} \subseteq [1:N]$ such that $N \in \mathcal{S}^c$.

The rate achieved by superposition noisy network coding on a multiple source multicast network is

$$R'(\mathcal{S}) < \min_{\mathcal{S}} I(V(\mathcal{S}); Y_k | X_k, V(\mathcal{S}^c))$$
$$R''(\mathcal{S}) < \min_{\mathcal{S}} \left(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_N | X(\mathcal{S}^c), V^N) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_N, V_{k-1}^N) \right)$$

□

We next apply the superposition noisy network coding scheme to AWGN single and two-way Gaussian relay channels. We compare the achievable rate to the existing schemes of noisy network coding, compress-forward and the cut-set upper bound.

5.5 Numerical results

Consider a Gaussian relay channel model,

$$Y_2 = aX_1 + Z_1 \tag{5.20}$$

$$Y_3 = X_1 + bX_2 + Z_2 \tag{5.21}$$

where the noise terms Z_1 and Z_2 are uncorrelated zero mean Gaussian random variables with variances N_1 and N_2 respectively, and a and b are the channel gain constants. As a result, we have

$$p(y_2, y_3 | x_1, x_2) = \frac{1}{2\pi\sqrt{N_1 N_2}} \exp\left[-\frac{(y_2 - ax_1)^2}{2N_1} - \frac{(y_3 - x_1 - bx_2)^2}{2N_2} \right], \tag{5.22}$$

which will be the channel assumed throughout the section. The average power constraints at the transmitters are

$$\frac{1}{n}\sum_{i=1}^{n} x_{1i}^2(k) \le P_1, \quad \forall k \in \mathcal{M}, \tag{5.23}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} x_{2i}^2 \le P_2, \quad \forall y_2^n \in \Re^n. \tag{5.24}$$

The Gaussian relay channel model analyzed is such that all the terminals are aligned in a line [24]. The source and destination are at unit distance. d is the distance between the source and relay with $a = 1/d$. $1-d$ is the distance of relay from the destination with $b = 1/(1-d)$. Fig. 5.5 plots the rates achieved by superposition noisy network coding for $P1 = P2 = 5$. The rates achieved are compared to the schemes of noisy network coding, compress-forward and cut-set bound. Noisy network coding achieves the same rate as compress-forward scheme for a single relay channel. It is observed that the superposition noisy network coding scheme has

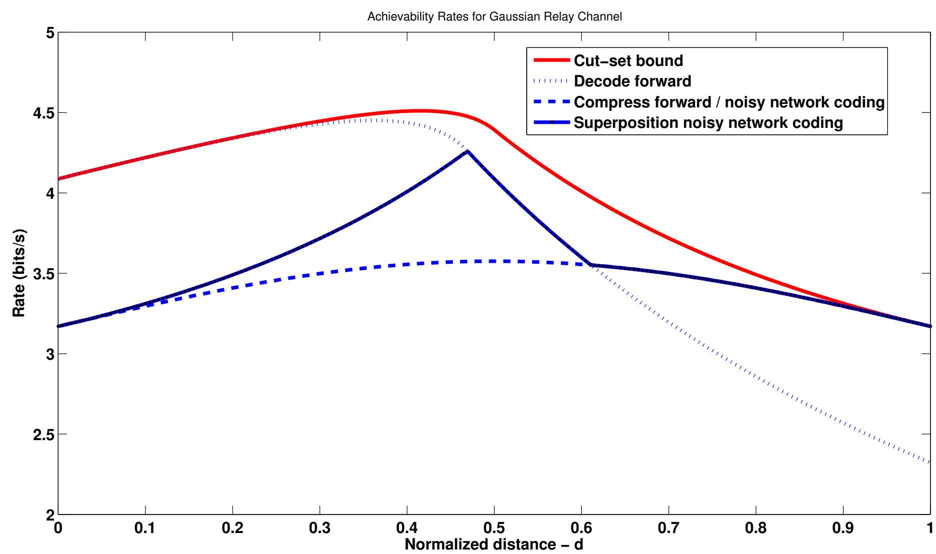

Figure 5.5 Achievable rates for an AWGN single relay channel using superposition noisy network coding

an advantage over the noisy network coding scheme when the relay is close to the source. This advantage arises due to a strong source-relay link. The relay can decode partial information depending on the strength of the link. The rate decreases as the relay gets closer to source.

For Gaussian relay channel, the rate loss caused by forcing U_1 and V_2 to be independent would be present only in the region where the relay is "close" to the source, specifically given by the condition

$$I(X_1; Y_2|X_2) > I(X_1X_2; Y_3).$$

The term $I(X_1; Y_2|X_2)$ represents the rate at which information can be transferred from source to the relay. The term $I(X_1X_2; Y_3)$ represents the rate at which source and relay coherently transmit information to the destination. The inequality states that the source relay link is not the bottleneck in capacity. For all other conditions, the source relay link $I(X_1; Y_2|X_2)$

is the capacity determining link, in which case mutual information is maximized by keeping U_1 and V_2 independent. As a result, we do not see any loss of rate in other regions by forcing U_1 and V_2 to be independent. Nevertheless, the rates achieved are still higher than the noisy network coding scheme in spite of the disadvantage in lack of beam-forming. Beam-forming can be incorporated by allowing U_1 and V_2 to be dependent for the single relay channel. The rates achieved would then be same as the superpostion-forward strategy. We would have extended the superposition-forward strategy using the ideas from noisy network coding.

The superposition noisy network coding scheme has been designed and analyzed for a single relay Gaussian channel. It is in general advantageous to decode extra information available at the relay and use it to transmit more information to destination. The superposition noisy network coding scheme is now extended to two-way Gaussian relay channel which is an example of single source multicast network.

5.5.1 Two-way relay channel

The two-way relay channel shown in Fig. 5.6 was first introduced by Shannon [45]. The two-way relay channel is a fundamental building block for multi-user information theory. Rankov et al. [46] derived the achievable rates for the two-way relay channel using the schemes decode-forward and compress-forward. The rates achieved by superposition noisy network coding is derived for the two way relay channel and compared to the existing rates.

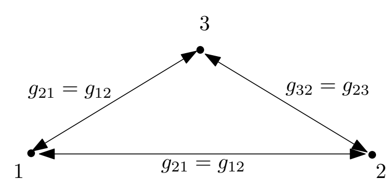

Figure 5.6 Gaussian two-way relay channel

The AWGN two-way relay channel is given by

$$\begin{aligned} Y_1 &= g_{21}X_2 + g_{31}X_3 + Z_1, \\ Y_2 &= g_{12}X_1 + g_{32}X_3 + Z_2, \\ Y_3 &= g_{13}X_1 + g_{23}X_2 + Z_3, \end{aligned} \tag{5.25}$$

where the channel gains are $g_{12} = g_{21} = 1$, $g_{13} = g_{31} = d^{-\gamma/2}$ and $g_{23} = g_{32} = (1-d)^{-\gamma/2}$, and $d \in [0,1]$ is the location of the relay node between nodes 1 and 2 (which are unit distance apart). Source nodes 1 and 2 wish to exchange messages reliably with the help of relay node 3. Specializing Theorem 5 to the two-way relay channel gives the inner bound that consists of all rate pairs (R_1, R_2) such that

$$R_1' \leq \min\{I(U_1; Y_2|U_2, V_3, X_3), I(U_1, V_3; Y_2|U_2, X_2)\} \tag{5.26}$$

$$R_2' \leq \min\{I(U_2; Y_1|U_1, V_3, X_3), I(U_2, V_3; Y_1|U_1, X_1)\} \tag{5.27}$$

$$R_1' + R_2' \leq I(U_1, U_2; Y_3|V_3, X_3) \tag{5.28}$$

$$R_1'' \leq \min\{I(X_1; Y_2, \hat{Y}_3|X_2, X_3, U_1, U_2), \tag{5.29}$$

$$I(X_1, X_3; Y_2|X_2, U_1, V_3) - I(Y_3; \hat{Y}_3|X_1, X_2, X_3, Y_2, U_1, U_2)\} \tag{5.30}$$

$$R_2'' \leq \min\{I(X_2; Y_1, \hat{Y}_3|X_1, X_3, U_1, U_2), \tag{5.31}$$

$$I(X_2, X_3; Y_1|X_1, U_2, V_3) - I(Y_3; \hat{Y}_3|X_1, X_2, X_3, Y_1, U_1, U_2)\}$$

for some $p(q)p(u_1)p(u_2)p(v_3)p(x_1|u_1,q)p(x_2|u_2,q)p(x_3|v_3,q)p(\hat{y}_3|y_3,x_3,q)$.

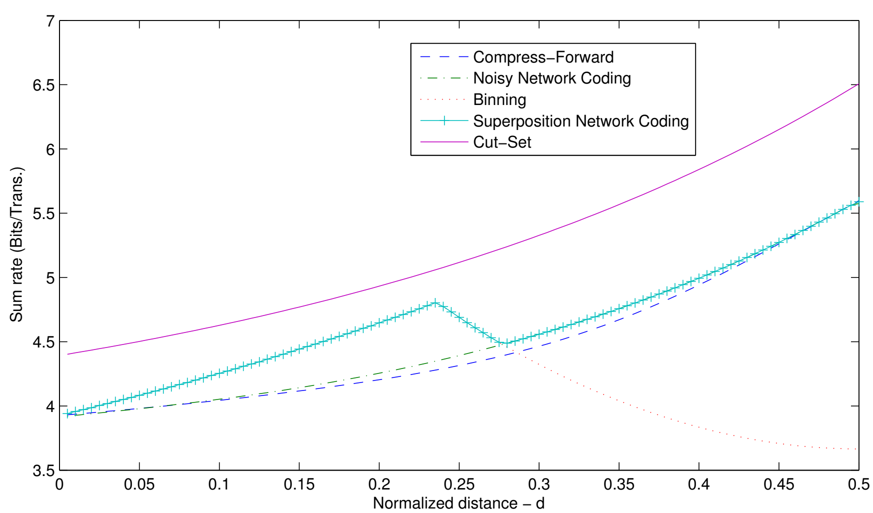

Figure 5.7 Achievable rates for an AWGN two-way relay channel

Fig. 5.7 compares the achievable rates of the schemes derived as a function of relay distance. The power constraint at the nodes are $P_1 = P_2 = P_3 = 10$. It is observed that superposition noisy network coding provides higher rates than compress-forward, noisy network coding and binning scheme. The binning scheme [30] is similar to decode and forward where the relay decodes the messages from both the sources. The decoded messages are binned randomly and the bin index is transmitted in the next block.

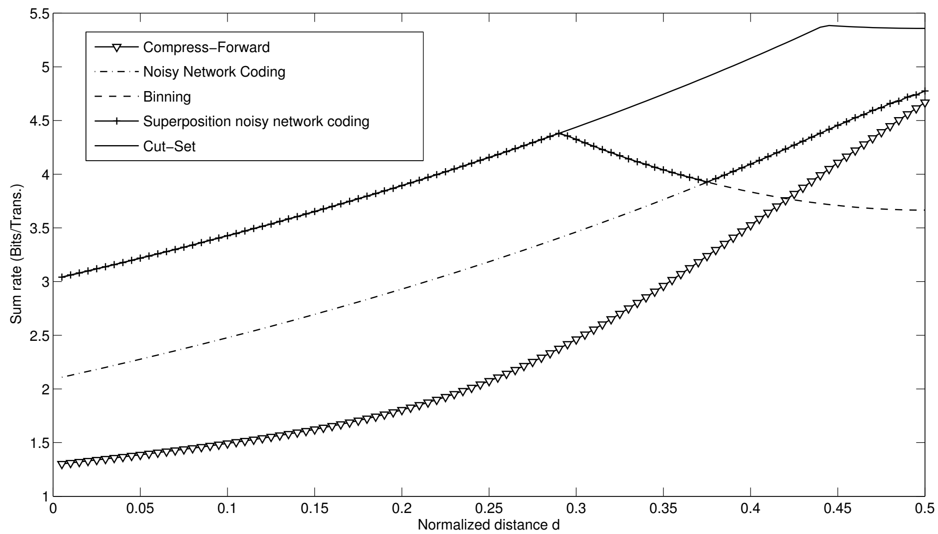

Figure 5.8 Achievable rates for an AWGN separated two-way relay channel

Fig. 5.8 compares the achievable rates for the separated two-way relay channel. We set $g_{12} = g_{21} = 0$. The power constraints are $P_1 = P_2 = 10$ dB and $P_3 = 5$ dB. It is observed that the superposition noisy network scheme achieves higher rates than compress-forward, noisy network coding and binning scheme.

Noisy network coding is a special case of superposition noisy network coding scheme. If we chose not to decode any message at the relay, the superposition noisy network coding scheme would achieve the same rate as noisy network coding scheme. As compared to noisy network coding, the superposition scheme performs better when the relay is close to either of the sources and decoding partial information is advantageous to the sum rate. In general, superposition network coding achieves the best of binning and noisy network coding scheme.

5.6 Conclusion

The noisy network coding for discrete memoryless channel is improved by superimposing partial decode and forward of the messages. The encoding and decoding strategies are derived for the superposition noisy network coding. Noisy network coding is a special case of superposition noisy network coding. The rates achieved by superposition noisy network coding is higher than the rates achieved by noisy network coding. It provides higher rates when the channel from the source to the relay nodes are stronger. This observation is confirmed by numerical results on single relay channel.

Superposition noisy network coding scheme is extended to multiple source multicast networks. The achievable rates were derived and generalized to two-way relay channel. It is numerically observed that the superposition noisy network coding scheme provides higher rates than noisy network coding, compress-forward and binning. The observation is consistent with AWGN single relay channel and AWGN two-way relay channel.

CHAPTER 6. SUMMARY

Many works in the literature of relay channel conclude with the proposition that superposition scheme can provide better achievable rate. The superposition-forward scheme is analyzed and it is shown that it can only achieve rates of at most decode-forward or compress-forward. Both schemes are special cases of superposition-forward. This shows that SF scheme is the best known achievable rate so far and it achieves capacity for all the special cases of relay channel where the capacity is known.

In another parallel work [36], it was shown that noisy network coding combines network coding with compress-forward to improve on the classic compress-forward achievable rate. This is shown to provide higher achievable rates for the case of relay networks with more than one source.

As a natural generalization, superposition noisy network coding is designed and analyzed. This scheme combines network coding with superposition-forward and promises the best achievable scheme so far. Further work is carried on to extend the superposition noisy network coding to discrete memoryless multiple source multicast networks. The multiple source network is constrained such that the source nodes do not act as relays. The decode forward scheme is extended to multicast networks using binning and m-user multiple access channel schemes. The achievable rate of the superposition noisy network coding scheme for multiple source multicast network is applied to two way relay channel and single source relay channel. It is observed that superposition scheme can be advantageous compared to individual existing schemes like compress forward or decode forward. The new coding scheme of superposition noisy network coding scheme achieves higher rates than the existing noisy network coding scheme for multicast networks.

6.1 Future Work

The superposition noisy network coding should be extended to a general multi source multicast network where the source node can also act as relays. This would provide the achievability results of superposition noisy network coding for the most general multiple source multicast network.

As observed in [37], message repetition encoding is not required to achieve the rates in noisy network coding for channels with single source. The noisy network coding rate can be achieved by using successive encoding and joint decoding at the destination. This can simplify the superposition noisy network coding scheme for the single source multicast network.

Another consideration is to look at the restriction on the input messages at the source and the relays. We force the input messages to be independent which results in a loss of rate as compared to superposition forward case. For a single relay case, it is straightforward to implement the dependency of the input random variables. This would provide the rates achieved by superposition forward. The superposition noisy network coding scheme can be extended with beam-forming to single source multicast networks.

A new direction of research would be to consider new schemes of hybrid decoding and hybrid digital-analog noisy network coding scheme. The hybrid coding scheme unifies amplify-forward and noisy network coding. We should look at new ideas to improve the superposition noisy network coding and work on finding better lower bounds on capacity.

The best known upper bound for the relay channel is still the cut set bound and it is known for some channels like the diamond relay channel, that the cut set is not a tight bound. Efforts can be made to find a tighter upper bound and make progress to find the capacity limits for the relay channel. The gap to cut set bound should also be quantized for the superposition noisy network scheme.

Lattice coding is another technique parallel to random coding and promises higher achievability rates. We should consider if superposition can be applied in conjunction with lattice coding. Existing results should be compared with techniques using lattice coding.

Bibliography

[1] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. of Allerton Conf.*, 2007, pp. 129–135.

[2] A. E. Gamal and Y.-H. Kim, "Lecture notes on network information theory," *CoRR*, vol. abs/1001.3404, 2010.

[3] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, pp. 379–423, 623–656, Jul., Oct. 1948.

[4] E. C. van der Meulen, "Three-terminal communication channels," *Advanced Applied Probability*, vol. 3, pp. 120–154, 1971.

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

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., 2006.

[7] B. Schein and R. Gallager, "The gaussian parallel relay network," in *Proc. IEEE Intl. Symp. on Info. Theory*, 2000, p. 22.

[8] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *CoRR*, vol. abs/0906.5394, 2009.

[9] W. Chang, S.-Y. Chung, and Y. H. Lee, "Gaussian relay channel capacity to within a fixed number of bits," 2010. [Online]. Available: http://arxiv.org/abs/1011.5065/

[10] A. El Gamal, "On information flow in relay networks," in *NTC '81; National Telecommunications Conference*, vol. 2, 1981.

[11] M. R. Aref, "Information flow in relay networks," Ph.D. dissertation, Department of Electrical Engg., Stanford University, 1980.

[12] F. Willems, "The feedback capacity of a class of discrete memoryless multiple access channels," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 93–95, 1982.

[13] A. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 841 – 850, Nov. 1982.

[14] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1106 –1110, Sep. 1988.

[15] H. Liao, "Multiple access channels," Ph.D. dissertation, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[16] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 2–14, 1972.

[17] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 197–207, 1973.

[18] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredaci Informaccii*, vol. 10, no. 3, pp. 3–14, 1974.

[19] T. Cover, "An achievable rate region for the broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-21, pp. 399–404, 1975.

[20] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 306–311, 1979.

[21] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. IT-23, pp. 60–64, 1977.

[22] P. Gupta and P. Kumar, "Towards an information theory of large networks: an achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877 – 1894, Aug. 2003.

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

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

[25] A. El Gamal, M. Mohseni, and S. Zahedi, “Bounds on capacity and minimum energy-per-bit for AWGN relay channels,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[26] T. M. Cover and Y. han Kim, “Capacity of a class of deterministic relay channels,” in *Proc. IEEE Int. Symp. Information Theory*, 2007.

[27] N. Ratnakar and G. Kramer, “The multicast capacity of deterministic relay networks with no interference,” *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2425 – 2432, Jun. 2006.

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

[29] A. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, “Capacity of wireless erasure networks,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789 –804, Mar. 2006.

[30] L.-L. Xie, “Network coding and random binning for multi-user channels,” in *10th Canadian Workshop on Information Theory, 2007. CWIT ’07*, Jun. 2007, pp. 85 –88.

[31] C. Nair and A. El Gamal, “An outer bound to the capacity region of the broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 350 –355, Jan. 2007.

[32] ——, “The capacity region of a class of three-receiver broadcast channels with degraded message sets,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4479 –4493, Oct. 2009.

[33] C. Nair, “Capacity regions of two new classes of two-receiver broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4207 –4214, Sep. 2010.

[34] V. Cadambe, S. Jafar, and S. Shamai, "Interference alignment on the deterministic channel and application to fully connected gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 269 –274, Jan. 2009.

[35] T. Kim, M. Skoglund, and G. Caire, "Quantifying the loss of compress-forward relaying without wyner-ziv coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1529 –1533, Apr. 2009.

[36] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," in *IEEE Information Theory Workshop (ITW), 2010*, Jan. 2010, pp. 1 –5.

[37] X. Wu, G. Fan, and L.-L. Xie, "An optimality-robustness tradeoff in the compress-and-forward relay scheme," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, Sep. 2010, pp. 1 –5.

[38] R. Yeung, *A First Course in Information Theory*. Springer, 2006.

[39] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, pp. 265–444, Apr. 2008. [Online]. Available: http://dl.acm.org/citation.cfm?id=1454713.1454714

[40] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 471–480, 1973.

[41] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1 – 10, Jan. 1976.

[42] A. E. Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 536–536, Mar. 1982. [Online]. Available: doi:10.1109/TIT.1982.1056502

[43] X. Wu and L.-L. Xie, "On the optimality of successive decoding in compress-and-forward relay schemes," in *Proc. of Allerton Conf.*, Oct. 2010, pp. 534 –541.

[44] J. Ponniah and L.-L. Xie, "An achievable rate region for the two-way two-relay channel," in *Proc. IEEE Intl. Symp. on Info. Theory*, Jul. 2008, pp. 489 –493.

[45] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, vol. 1. Univ. California Press, 1961.

[46] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE Intl. Symp. on Info. Theory*, Jul. 2006. [Online]. Available: http://www.nari.ee.ethz.ch/wireless/pubs/p/isit2006