AN INFORMATION-THEORETIC STUDY OF
COOPERATION IN NETWORKS

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Ron Dabora
May 2007
AN INFORMATION-THEORETIC STUDY OF COOPERATION IN NETWORKS

Ron Dabora, Ph.D.
Cornell University 2007

This thesis presents a study of cooperation in networks using the tools of information theory. We first review the basic network models, with an emphasis on the relay channel, as this is the most basic configuration of cooperative communication. We focus on the estimate-and-forward (EAF) relaying strategy, which is a scheme that does not require the relay to decode the source messages. We investigate EAF with assignments of the auxiliary random variable that satisfy the feasibility constraint and present an alternative characterization of the classic EAF result of [Cover & El-Gamal, 1979] without a feasibility constraint, thus simplifying the description of the rate.

Next, we combine the relay channel with the broadcast channel. This combination is used to study communication over the general discrete memoryless broadcast channel (BC) with partially cooperating receivers. In our setup, the receivers are able to exchange messages over noiseless conference links of finite capacities, prior to decoding the messages sent from the transmitter. We first find the capacity region of the physically degraded BC with cooperating receivers. Then, we derive an achievable rate region for the general BC with three independent messages – two private messages and a common message, where the receivers hold a $K$-cycle conference. Additionally, we consider a special case of the general setup, the case
of the general BC with just a single message. For this case we obtain explicit rate expressions. We also identify two scenarios in which these explicit rate expressions achieve capacity.

We then consider the discrete, memoryless, multiple-relay channel and derive an explicit achievable rate expression based on the EAF scheme. This expression is amenable to numerical evaluation. We demonstrate the benefits of this result via a discrete memoryless multiple-relay channel example, in which it is superior to multi-relay decode-and-forward. Finally, we consider the Gaussian relay channel with coded modulation at the transmitter and an orthogonal relay-destination link of finite capacity. Here we show that an EAF strategy implementing a three-level quantization outperforms the Gaussian quantization commonly used to evaluate the rates that the EAF scheme achieves in this scenario.
BIOGRAPHICAL SKETCH

Ron Dabora received his B.Sc. and M.Sc. degrees in electrical engineering in 1994 and 2000 respectively, from Tel-Aviv University, Tel-Aviv, Israel. From 1994 to 2000, he was with the Signal Corps of the Israel Defense Forces, where he served as an officer in the R&D department of the Electronic Warfare Unit. From 2000 to 2003, he was with the Algorithms Development Group at Millimetrix Broadband Networks, Israel. Since 2003 he is a Ph.D. student at Cornell University, Ithaca, New York.
ACKNOWLEDGEMENTS

It is quite common in the United States that when a student completes his academic studies, he has debts. My case is no different. However, in contrast to regular debts, that are normally paid in a certain number of years, in my case, I cannot really tell how long would it take me to pay my debts. If ever. The most I can do, I am afraid, is to list those to whom I am in debt.

To my advisor, Sergio, for his friendship in hard times, especially in ISIT 2006 and during ITA 2007. I will never forget his help. One of the most important lessons I learned in my life was from Sergio: never give up – always get back on your feet and keep going.

To my committee members: to Professor Berger for agreeing to be on my committee despite the fact that he was near retirement, and to Professor Fine who agreed to serve as his proxy for the B-Exam. To Professor Tong for serving on my committee and also for his support and reassurance in some hard times. To Professor Verdú for agreeing to serve on my committee despite the fact that he is not a Cornell Faculty.

To my friends back home, Ronen, Yaron and Osnat who, despite the large distance, stay close friends. Especially Osnat, whose friendship, support and advice are among my most valuable possessions.

To my friends in Ithaca – Geoff who had to put up with me for three years, and I suspect that one of the reasons he got married was to avoid being my roommate for the fourth year in a row. To Steve who had to put up with me for two long, untidy years, and to An-swol and Emilia who got off easy – only one semester. I am sure that if there is Heaven, then being my roommate is one of the things that gets you a free pass. To my officemates – An-swol who had three and a half years
of misery, and Mingbo and Yiorgos who got their (smaller) share.

Last but not least – to my family back home who had their share of struggles, but did not give up. I am humbled by their courage and spirit.

If I will be able to apply in the future the lessons I learned during my time here – then it was well worth it.
# TABLE OF CONTENTS

## 1 Introduction
1.1 Motivation ......................................................... 1  
1.2 Multi-User Information Theory ................................. 2  
  1.2.1 Discrete Memoryless Three-Node Networks ............... 2  
  1.2.2 Other Basic Models ........................................ 4  
1.3 The Relay Channel ................................................ 5  
  1.3.1 Relaying Strategies ........................................ 6  
  1.3.2 Related Work ............................................... 11  
  1.3.3 The Gaussian Relay Channel with Coded Modulation ..... 13  
1.4 The Broadcast Channel ........................................... 15  
  1.4.1 The Discrete Memoryless Broadcast Channel (DMBC) with Independent Decoders ........................................... 16  
  1.4.2 Marton’s Achievable Rate Region for the General DMBC .. 18  
  1.4.3 The Broadcast Channel with Cooperating Decoders: A Combination of Broadcasting and Relaying .................... 21  
1.5 The Multiple-Access Channel .................................... 25  
  1.5.1 MAC-BC Duality .............................................. 26  
1.6 Main Contributions and Organization .......................... 27  
1.7 Notations ......................................................... 32

## 2 Estimate-and-Forward Relaying with Time-Sharing Auxiliary Mapping
2.1 Definitions for the Relay Channel ............................... 34  
2.2 An Information Flow Interpretation of Estimate-and-Forward ... 35  
2.3 The Single Relay EAF with Time-Sharing Assignment ........... 40  
2.4 Joint-Decoding and Time-Sharing ................................ 43  
  2.4.1 Discussion .................................................. 47

## 3 The Broadcast Channel with Cooperating Decoders
3.1 Definitions for the Broadcast Channel with Cooperating Decoders . 49  
3.2 The Capacity Region of the PhysicallyDegraded Broadcast Channel with Cooperating Receivers .......................... 52  
  3.2.1 Achievability Proof ......................................... 53  
  3.2.2 Converse Proof .............................................. 57  
  3.2.3 Discussion .................................................. 60  
3.3 The Cooperative General Broadcast Channel with Two Private Messages and One Common Message ....................... 63  
  3.3.1 An Achievable Rate Region for the General Broadcast Channel with Cooperating Decoders Holding a K-Cycle Conference 63  
  3.3.2 An Upper Bound ............................................. 74  
  3.3.3 Special Cases ............................................... 75
3.4 The Cooperative General Broadcast Channel with a Single Common Message
3.4.1 A Single-Cycle Conference with TS Auxiliary Mapping  
3.4.2 Special Cases

4 Application of Time-Sharing to Relaying Scenarios
4.1 An Achievable Rate for the Relay Channel with Multiple Relays
4.1.1 Statement of the Main Theorem
4.1.2 Proof of Theorem 4.1
4.1.3 Discussion
4.2 The Gaussian Relay Channel with Coded Modulation
4.2.1 The Gaussian Relay Channel with a Gaussian Source Codebook
4.2.2 The Gaussian Relay Channel with Coded Modulation
4.2.3 Time-Sharing Deterministic Hard-Decision (TS-DHD)
4.2.4 When the SNR on the Source-Destination Link Approaches 0 (\(\sigma^2 \rightarrow \infty\))
4.2.5 Discussion

5 Conclusions and Future Work
5.1 Future Work

A Proof of Proposition 2.1
A.1 Codes Construction, Encoding and Decoding
A.2 Analysis of the Probability of Error
A.3 Bounding the Probability \(\Pr(E_{2,i}^c \cap E_{1,i}^c \cap E_{0,i}^c | F_{i-1}^c)\)
A.4 Bounding \(E_{y,x_2} \{||\mathcal{L}(i-1)|| | F_{i-1}^c\}\)

B Proof of Corollary 3.3

C The Expressions of [12, theorem 3] with Time-Sharing Assignments

D Expressions for Section 4.2
D.1 Hard-Decision Estimate-and-Forward
D.2 Evaluation of the Information Rate with DHD
D.2.1 DHD when \(T \rightarrow 0\)
D.3 Evaluating the Information Rate with TS-DHD
D.3.1 Evaluating \(I(X; Y, \hat{Y}_1)\)
D.3.2 Evaluating \(I(\hat{Y}_1; Y_1 | Y)\)
D.4 Gaussian-Quantization Estimate-and-Forward
D.5 Approximation of HD-EAF for \(\sigma^2 \rightarrow \infty\)

References
LIST OF TABLES

4.1 \( p(y, y_1, y_2|x, x_1, x_2) \) for the EAF example. .......................... 100
4.2 Optimal distribution for DAF ........................................ 101
4.3 Optimal distribution for EAF ........................................ 101
# LIST OF FIGURES

1.1 The relay channel. The encoder sends a message $W$ to the decoder.  
1.2 A schematic description of the MAC bound and the broadcast bound for the relay channel. .................................................. 6  
1.3 The broadcast channel: a single transmitter sends messages to $N$ receivers.  
1.4 The broadcast channel with independent receivers. The encoder sends three messages, a common message $W_0$, a private message to $R_{x1}$, $W_1$, and a private message to $R_{x2}$, $W_2$. $\hat{W}_0$ and $\hat{W}_0$ are the estimates of $W_0$ at $R_{x1}$ and $R_{x2}$ respectively. ........................................... 10  
1.5 The broadcast channel with cooperating receivers. The encoder sends three messages, a common message $W_0$, a private message to $R_{x1}$, $W_1$, and a private message to $R_{x2}$, $W_2$. $\hat{W}_0$ and $\hat{W}_0$ are the estimates of $W_0$ at $R_{x1}$ and $R_{x2}$ respectively. The receivers have noiseless links of finite capacities $C_{12}$ and $C_{21}$ between them.  
1.6 The multiple-access channel: $N$ transmitters send messages to a single receiver. ................................................................. 15  
2.1 The information flow budget for the general relay channel with compression at the relay. ................................................. 20  
3.1 The physically degraded BSBC. $p_U$, $p_1$ and $p_2$ are the transition probabilities at the left, middle and right segments respectively.  
3.2 The capacity region for the physically degraded BSBC. Top, middle and bottom lines correspond to maximum possible cooperation, partial cooperation and no-cooperation scenarios respectively.  
3.3 The single message broadcast channel with cooperating receivers. $\hat{W}$ and $\hat{\hat{W}}$ are the estimates of $W$ at $R_{x1}$ and $R_{x2}$ respectively.  
3.4 The achievable rate $R$ vs. conference links capacity $C$, for corollary 3.2 (dash-dot), proposition 3.2 (dash) and corollary 3.4 (solid), for the symmetric broadcast channel.  
4.1 The Gaussian relay channel with a finite capacity noiseless link between the relay and the destination.  
4.2 Information rate with BPSK for HD-EAF mapping at the relay vs. source-relay channel gain $g$, for different values of $C$.  
4.3 Information rate with BPSK, for DHD mapping at the relay vs. source-relay channel gain $g$, for different values of $C$.  
4.4 $I(\hat{Y}_1; Y_1 | Y)$ and $I(X; Y, \hat{Y}_1)$ vs. threshold $T$ for $(g, C) = (0.4, 0.8)$ (left) and $(g, C) = (1.4, 0.8)$ (right). The bold solid line represents $I(\hat{Y}_1; Y_1 | Y)$, the horizontal bold dashed line represents $C = 0.8$, $I(X; Y, \hat{Y}_1)$ is represented by the dash-dot line and the resulting information rate is depicted with the solid line.
4.5 Information rate with BPSK for HD-EAF, DHD and TS-DHD mappings at the relay vs. source-relay channel gain $g$, for different values of $C$. .......................... 114
4.6 Information rate with BPSK, for DAF, TS-DHD and GQ-EAF mappings at the relay vs. source-relay channel gain $g$, for different values of $C$. .......................... 115
4.7 The best relaying strategy (out of DAF, TS-DHD and GQ-EAF) for the Gaussian relay channel with BPSK modulation. .................. 116
4.8 Information rate with DAF, DHD, HD-EAF and GQ-EAF vs. source-relay channel gain $g$, for different values of $C$, at low SNR on the source-destination link. .......................... 122
Chapter 1

Introduction

1.1 Motivation

As the world becomes an increasingly connected environment, the demand for wireless network throughput increases rapidly. Broadly speaking, when focusing on the physical layer there are three paths to increase the throughput of a wireless network: the first path is to improve performance at the device level. This means the development of integrated circuits (ICs), such as amplifiers, mixers, digital-to-analog and analog-to-digital converters, that have better noise and linearity characteristics. A second way is to improve the design of the physical layer algorithms – develop more robust techniques for estimation, synchronization, modulation and coding that can operate at lower signal-to-noise ratios (SNR). A third approach would be to modify the philosophy of the network operation: instead of having each receiving node in the network decode its information independently of the other nodes, we can let other nodes, generally in the vicinity of the target node, help it decode its message by transmitting a version of their received signal to the target node. This way, by using cooperation between the nodes we can increase the throughput of the network even when the performance of the ICs and the physical layer algorithms remain the same. In this work we focus on the third approach. We conduct our study within the framework of multi-user information theory. We first study the most basic form of cooperation, the relay channel. Then, we study the application of relaying to the broadcast channel with cooperating decoders – the hybrid broadcast-relay channel. This problem comes up naturally in sensor networks, where a transmitter external to the sensor network wants to download
data into the network, e.g., to configure the sensor array. Here, the sensors are already connected via a network, so they can communicate with each other, helping other sensors to decode the messages broadcasted from the external transmitter.

1.2 Multi-User Information Theory

Multi-user information theory started in 1961 with a paper by Shannon which introduced the two-way channel (TWC) [1]. In this channel two transceivers communicate simultaneously with each other. In its most general formulation, each transceiver can adjust its transmission based on its received signal and, of course, can use its knowledge of its own transmitted signal to assist in decoding the messages sent from the other transceiver. In his work, Shannon gave inner and outer bounds on the rates for the general case and also showed that these bounds provide the capacity region of a special case of the general setup, in which the encoders do not use their knowledge of the received signal when encoding the transmitted messages. In [1] Shannon also suggested another multiple-user channel, a channel with two senders and one receiver, later known as the multiple-access channel. This was the first configuration of the most basic multiple-user setup – the three-node network.\(^1\)

1.2.1 Discrete Memoryless Three-Node Networks

The most basic form of the multi-user scenario is the three-node network. For this network there are three basic configurations:

\(^1\)Although one may argue that the TWC is a smaller multiple-user setup, we note that in the TWC, each node consists of an encoder and a decoder. Thus we treat the TWC conceptually as a four-node network.
1. The multiple-access channel (MAC): two senders communicate with one receiver. The senders operate independently.

2. The broadcast channel (BC): one sender communicates with two receivers. The receivers operate independently.

3. The relay channel: one sender communicates with one receiver. The third node is a relay whose purpose is to assist the communication between the other two nodes.\(^2\)

In the classic formulation of the multiple-access and the broadcast scenarios there is no active cooperation between the nodes in the network (i.e., nodes do not transmit signals intended to help other nodes). We note that cooperation is represented by the relay channel since this is the most basic scenario in which one node actively helps another node, hence the importance of this channel. In this work we focus on the broadcast and relay scenarios. In fact, we discuss here for the first time the combination of broadcast and relay. Furthermore, we shall dedicate most of the attention to multi-user scenarios governed by discrete memoryless channel models. In the following we give an overview of the results obtained for each of the three basic configurations. As we focus on the last two configurations, we give only a brief overview of the multiple-access channel, although this scenario has received most of the attention.

\(^2\)Note that the relay does not have any private information to encode, nor it is required to decode anything. Thus we do not treat the relay node as two nodes as we treated the transceivers in the TWC.
1.2.2 Other Basic Models

Another multi-user scenario that stems directly from Shannon’s TWC is the case in which the encoders and decoders do not cooperate. This channel was also suggested (but not analyzed) by Shannon in his work [1]. This scenario is commonly referred to as the interference channel (IFC). As the general IFC is a hard problem, several simplified models were considered in the following years. The capacity of the case where each decoder has to decode both messages was found by Ahlswede in [2]. In his paper, [2], Ahlswede noted the fact that the capacity region of this special case is equal to the intersection of the capacity regions of the two component MACs. The capacity of the case where both receivers have the same channel output (termed the twin two-user channel by Sato) was found in [2] and [3]. In [4] Carleial derived an achievable rate region using superposition coding for the general IFC. Carleial also considered in [4] a simplified channel model, the so-called statistically equivalent channel, in which all the marginal distributions are equal. The Gaussian IFC was also studied: Carleial found in [5] the capacity region for the case of very strong interference (i.e. when the rate region is equal to the case of two orthogonal point-to-point channels), and Sato [6] found the capacity region for the case of strong interference, where he showed it to be equivalent to the intersection of the capacity regions of the two component MACs. A recent outer bound on the capacity region of the Gaussian IFC can be found in [7] in which the author used a combination of the bounds of Costa in [8] and Sato [9]. The case of partial and unidirectional transmitter cooperation was studied in [10], [11].

In recent years additional models that extend three-node networks into four-node networks were studied. The two extended models considered were the multiple-access-relay channel (MARC) and the broadcast/relay channel (BRC). In these
models the classic MAC and BC were extended by adding a forth node which functions as a relay. Thus, each node in the network has a single function: a transmitter, a receiver or a relay. These scenarios were investigated in [12]. In our work we consider an extended cooperation model for the BC by letting each receiver have a dual function, i.e., each receiver also acts as a relay for the other receiver. Therefore, conceptually this model can be considered as a five-node network. However, there is a fundamental difference between the model considered in our work (and in [13]) and the other models just mentioned: we allow the receivers to hold a conference. This implies that the conference messages are determined in an interactive manner rather than be determined entirely at the beginning of a block. In contrast, in the classic scenarios, the relay picks a message to transmit at the beginning of the block, and then dedicates the block duration to the transmission of that message. We refer to such schemes as one-step conference schemes. The interactive conference generally supports higher broadcast rates than the one-step conference. There are instances, however, in which one conference step is enough to achieve capacity, and we shall discuss such cases as well.

1.3 The Relay Channel

The relay channel was introduced by van der Meulen in 1968 [14]. In this setup, a single transmitter with channel input $X^n$ communicates with a single receiver with channel output $Y^n$, where the superscript $n$ denotes the length of a vector. In addition to the transmitter and receiver, there is a transceiver, called a relay, that listens to the channel and can input signals to the channel. We denote the relay’s channel output with $Y_1^n$ and its channel input with $X_1^n$. The sole purpose of the relay is to assist the communication between the transmitter and the receiver.
This setup is depicted in figure 1.1. This model represents the most basic form of cooperation.

![Relay Channel Diagram](image)

Figure 1.1: The relay channel. The encoder sends a message $W$ to the decoder.

### 1.3.1 Relaying Strategies

In [15] Cover & El-Gamal introduced two relaying strategies commonly referred to as decode-and-forward (DAF) and estimate-and-forward (EAF). In DAF the relay decodes the message sent from the transmitter and then, at the next time interval, transmits a codeword based on the decoded message. The rate achievable with DAF is given in [15, theorem 1]:

**Theorem 1.1** *(achievability of [15, theorem 1])* For the general relay channel any rate $R$ satisfying

$$R \leq \min \{I(X;Y_1|X_1), I(X,X_1;Y)\}$$  \hspace{1cm} (1.1)

for some joint distribution $p(x,x_1,y,y_1) = p(x,x_1)p(y,y_1|x,x_1)$, is achievable.

Examining the rate expression in equation (1.1), we note that the first term in the minimum stems from the requirement that the relay will decode the message. The second term represents the rate to the receiver achieved when the relay and the transmitter coordinate their codewords.
The DAF strategy achieves the capacity of a special case of the relay channel – the physically degraded channel, for which $p(y, y_1|x, x_1) = p(y_1|x, x_1)p(y|y_1, x_1)$. We note that for DAF to be effective, the rate to the relay has to be greater than the point-to-point rate, i.e., the probability distribution $p(x, x_1)$ should satisfy

$$I(X; Y_1|X_1) > I(X; Y|X_1).$$  \hfill (1.2)

This is because when (1.2) does not hold, the DAF rate is $I(X; Y_1|X_1)$, but higher rates than that can be obtained by using the same $p(x|x_1)$ without using the relay at all, and simply fixing $X_1 = x_1$ that maximizes $I(X; Y|x_1)$. We note that now $I(X; Y_1|x_1)$ my be higher than $I(X; Y|x_1)$, but still, when fixing $X_1$ the rate to the destination is $I(X; Y|x_1)$ irrespective of the value of $I(X; Y_1|x_1)$. For relay channels where DAF is not useful or not optimal, [15] proposed the EAF strategy. In this strategy, the relay sends a function of its channel output to the destination, without decoding the source message at all. The rate achievable with EAF is given in [15, theorem 6]:

**Theorem 1.2 ([15, theorem 6])** For the general relay channel any rate $R$ satisfying

$$R \leq I(X; Y, \hat{Y}_1|X_1),$$  \hfill (1.3)

subject to $I(X_1; Y) \geq I(Y_1; \hat{Y}_1|X_1, Y)$, \hfill (1.4)

for some joint distribution $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1)$, where $||\hat{Y}_1|| < \infty$, is achievable.

The relay operation can be explained in terms of Wyner-Ziv (WZ) compression [16]: in the WZ setup, an encoder wants to send a RV $X$ to a decoder. The encoder generates information to the decoder based only on $X$. The decoder has
side information $Y$ correlated with $X$ according to $p(x,y)$. The decoder then uses the data received from the encoder and the side information $Y$ to decode $X$ up to a given distortion $D$. In EAF, the rate from the encoder (relay) to the decoder (receiver) is determined by the point-to-point channel between the two nodes and is equal to $I(X_1;Y)$. The relay now compresses its channel output $Y_1|X_1$ into $\hat{Y}_1|X_1$, which is the distorted version of $Y_1$ to be recovered at the destination. The receiver uses the information received from the relay via the point-to-point transmission and the side information $Y|X_1$, to decode $\hat{Y}_1$ according to the WZ scheme. The distortion is determined by the relay mapping $p(\hat{y}_1|x_1,y_1)$ used in the compression. All quantities are conditioned on $X_1$, as $X_1$ is known at the relay and is recovered by the destination prior to applying the WZ scheme. The EAF expression gives insight into the rate increase that can be obtained using WZ compression at the relay: it shows that an increase of $I(X;\hat{Y}_1|X_1,Y)$ over the point-to-point rate $I(X;Y|X_1)$ can be obtained, as long as $p(\hat{y}_1|y_1,x_1)$ satisfies (1.4). Equation (1.4) can be interpreted as requiring that the new information that $\hat{Y}_1$ contains on $Y_1$, represented by the expression $I(Y_1;\hat{Y}_1|X_1,Y)$, will be deliverable from the relay to the destination via a point-to-point transmission. We call this condition the \textit{feasibility condition} as it defines the region in which the rate $I(X;Y,\hat{Y}_1|X_1)$ is feasible.

At this point we make two important observations regarding the computability of the expressions (1.3)–(1.4): first, we note that the maximum rate of the EAF strategy is not computable since there is no cardinality bound on $\hat{Y}_1$. Second, even if a cardinality bound could be established for $\hat{Y}_1$, finding the maximum EAF rate requires a non-convex search. These two properties of the general EAF expression make it unhelpful in answering even the simplest questions. For example, assume
that we are given a binary-input discrete memoryless relay channel. Which method is better suited for this channel, EAF or DAF? While for DAF we can determine the maximum rate, even through a brute-force search over all possible binary input distributions $p(x, x_1)$, there is no way to compute the rate that EAF provides, while keeping the general expression. Part of this work will discuss ways to answer the above question.

Lastly, we note that one can combine the DAF and EAF schemes by performing partial decoding at the relay, thus obtaining higher rates as in [15, theorem 7]. There are special cases in which [15, theorem 7] achieves capacity. One such example is the semi-deterministic relay channel [17], in which $Y_1$ is a deterministic function of $X$ and $X_1$.

We note that the relay channel is a special case of the multiple-access channel with generalized feedback considered in [18]. In the generalized MAC setup, each transmitter receives a channel output through a feedback channel. The two feedback channels are different. Therefore, by eliminating the feedback to one transmitter, and setting the private rate of the other transmitter to zero (hence it essentially functions as a relay), we obtain the relay scenario [12]. We note that the scheme in [18], when specialized to the relay case, achieves the same rate as in [15, theorem 1] through a different coding scheme: in [18], both the transmitter and the relay have two codebooks of the same size, and they switch codebooks each message transmission. The destination uses a sliding-window decoding with simultaneous decoding over two consecutive blocks [12]. This scheme is commonly referred to as regular encoding/sliding windows decoding (RESWD). In contrast, the scheme in [15, theorem 1] uses a superposition codebook at the transmitter (also referred to as Markov encoding) with successive decoding at the destination.
In their paper [15], Cover & El-Gamal also proposed an upper bound, now referred to as the cut-set bound [19, theorem 14.10.1], [20]. This upper bound is the best known upper bound for the general discrete memoryless relay channel, and it states the following:

**Theorem 1.3 ([15, theorem 4])** For the general relay channel, any achievable rate $R$ has to satisfy

$$R \leq \sup_{p(x_1,x_2)} \min \{I(X,X_1;Y), I(X;Y,Y_1|X_1)\}$$

$$= \sup_{p(x_1,x_2)} \left\{I(X;Y|X_1) + \min \{I(X_1;Y), I(X;Y_1|X_1,Y)\}\right\}.$$  

The term $I(X,X_1;Y)$ is referred to as the “MAC bound” as this is the rate obtained when both source and relay cooperate in transmitting to the destination. The term $I(X;Y,Y_1|X_1)$ is referred to as the “broadcast bound” as this is the rate in which the source can transmit to both the relay and the destination. See schematic description in figure 1.2. We note that DAF can achieve the MAC bound, therefore, when the MAC condition is bad (which implies that for the relay-destination link, $I(X_1;Y)$ is low) and the BC condition is good ($I(X;Y_1|X_1)$

![Figure 1.2: A schematic description of the MAC bound and the broadcast bound for the relay channel.](image-url)
is high\(^3\), DAF achieves capacity. Another important point to note here is that in the upper bound, the source and relay are coordinating their codewords, as implied by the joint distribution \(p(x, x_1)\). This introduces an inherent gap between EAF and the upper bound. Hence, the capacity for the cases where DAF does not achieve it, is still unknown.

Recently, a third relaying method, particularly suitable for continuous relay channels, such as the Gaussian relay channel and the fading relay channel, was proposed in [21]. In this method, the relay transmits a scaled version of its channel output such that the power constraint at the relay is satisfied. Therefore the relay does not decode the message, nor does it use WZ compression to generate its signal. This method is usually referred to as amplify-and-forward (AAF). This method is simple to implement, has no significant delay and is more amenable to analysis than EAF, but it suffers from noise enhancement.

#### 1.3.2 Related Work

In recent years, the research in relaying has mainly focused on the multiple-relay channel and the multiple-input multiple-output (MIMO) relay channel. In the context of multiple-relay schemes based on DAF, several DAF variations were considered. In [20] Cover & El-Gamal’s block-Markov encoding/successive decoding (MESD) DAF scheme was extended to degraded relay networks. The capacity of deterministic relay networks without interference was also obtained in [20]. In [22] MESD was extended to the multiple-relay channel for both the discrete and the Gaussian models. Later work [23], [24] and [25] applied Carleial’s RESWD and

\[^3\text{Since } I(X; Y, Y_1|X_1) > I(X; Y_1|X_1), \text{ then if } I(X; Y_1|X_1) \text{ is high, so is } I(X; Y, Y_1|X_1).\]
the regular encoding/backward decoding (REBD) technique, originally developed by Willems in [26, ch. 7], to the multiple-relay scenario (in [25] REBD was applied only to the single relay case). We note that for multiple relays, both regular encoding schemes achieve the same rate, which is higher than the rate achieved by the multiple-relay extension of the block-Markov encoding developed in [22]. However, the delay in the RESWD scheme is smaller than in REBD, hence the RESWD scheme of [24] is the superior multiple-relay DAF scheme [12]. In [27] DAF was applied to the physically degraded Gaussian relay channel with multiple relays.

The EAF strategy was also extended to the multiple-relay scenario. The work in [12], for example, considered the EAF strategy for multiple-relay scenarios and the Gaussian relay channel, in addition to considering the DAF strategy. See also [28].

In [29] the DAF strategy was applied to the Gaussian MIMO relay channel and the Rayleigh MIMO relay channel. Another approach applied recently to the relay channel is that of iterative decoding. In [30] the three-node network in the half-duplex regime was considered. In the relay case, [30] uses a feedback scheme in which the receiver first uses EAF to send information to the relay and then the relay decodes and uses DAF at the next time interval to help the receiver decode. Combinations of EAF and DAF were also considered in [31], where conferencing schemes over orthogonal relay-receiver channels were analyzed and compared. Both [30] and [31] focus on the Gaussian relay channel. Finally we note that in [32] a relay that transmits the minimum mean-squared error (MMSE) estimate of its received symbol (based on symbol-by-symbol decisions) is proposed for relay channels with real inputs and outputs. This method generalizes AAF by letting
the relay process its received signal by a more general memoryless transformation, rather than simply multiply it by a constant to satisfy the output power requirement at the relay. The paper [32] considers only the scenario in which there is no direct link from the source to the destination. Another generalization of the AAF strategy is proposed in [33, section V]. In this scheme the relay transmits a weighted sum of all its previously received channel outputs. In [33] an example is provided in which the generalized AAF scheme outperforms both DAF and EAF with a Gaussian auxiliary RV and Gaussian codebooks.

A special case of the relay channel was considered in [34], in which the capacity of a class of independent relay channels with a noiseless, finite capacity relay-destination link was derived. Another class of relay channels that was studied is the permuting relay channel, considered in [35] and [36]. We also note a recent paper [37] in which the capacity of the relay channel with a noiseless relay-destination link, where also the channel output at the relay, $Y_1$, is a deterministic function of the transmitter’s channel input $X$ and the receiver’s channel output $Y$, was determined.

1.3.3 The Gaussian Relay Channel with Coded Modulation

One important instance of the relay channel that we consider in this work is the Gaussian relay channel with coded modulation. This scenario is important in evaluating the rates achievable with practical communication systems, since practically, components in the receive chain, for example the equalizer, require a uniformly distributed finite constellation for optimal operation. In Gaussian relay channels, most often three types for relaying techniques are encountered:
The first technique is DAF. This technique achieves capacity for the physically degraded Gaussian relay channel (see [15, section IV]), and also for more general Gaussian relay channels under certain conditions (see [31]). In [12, section VII-B] it is shown that for asymptotically high SNR on the source-relay link, DAF achieves capacity for the Gaussian relay channel.

The second technique is EAF, in which the auxiliary variable $\hat{Y}_1|Y_1$ is assigned a Gaussian distribution. For example, in [33, section IV] a Gaussian auxiliary random variable (RV) is used together with time-sharing at the transmitter. The Gaussian assignment achieves capacity for the Gaussian relay channel when the SNR on the relay-destination link approaches infinity [12, remark 31]. In [38] an achievable rate with full duplex relay transmission employing Gaussian EAF over the Rayleigh relay channel, is obtained for the high SNR regime (here $Y_1$ is not Gaussian, but compression is performed by adding a Gaussian RV to the received signal at the relay. Knowledge of the fading coefficients is assumed. High SNR is assumed for both the relay and the receiver).

The third technique is linear relaying, where the relay transmits a weighted sum of all its previously received channel outputs [33, section V]. AAF is an important subclass of this family of relaying functions. In [39] AAF was combined with DAF resulting in the decode-amplify-and-forward scheme. A related approach to AAF was proposed in [32], in which the relay finds an MMSE estimate of its received symbol on a symbol-by-symbol basis, and uses it to generate its transmitted symbol.
Several recent papers consider the Gaussian relay channel with coded modulation. In [40] the author considered the performance of half-duplex DAF relaying for different practical systems. In [21] DAF and amplify-and-forward were considered for coherent orthogonal binary phase-shift keying (BPSK) signalling and in [32] examples with BPSK modulation were considered as well.

As indicated by several authors (see [33] for example) it is not known if a Gaussian auxiliary RV is indeed optimal for EAF relaying over the Gaussian relay channel. In this work we show that for the case of coded modulation, there are situations in which non-Gaussian assignments of the auxiliary random variable in the EAF scheme result in higher rates than the commonly applied Gaussian assignment.

1.4 The Broadcast Channel

The broadcast channel is a one-to-many communication scenario. This scenario is illustrated in figure 1.3. In the classic broadcast scenario the receivers decode their messages independently of each other. However, as pointed out earlier, the
increasing demand for wireless network throughput motivates the consideration of broadcast scenarios in which each receiver, besides decoding its own information, tries to help other receivers in decoding. This prompted us to study the effect of receiver cooperation on the rates for the broadcast channel.

1.4.1 The Discrete Memoryless Broadcast Channel (DMBC) with Independent Decoders

The broadcast channel was introduced by Cover in [41]. In this scenario, one transmitter communicates with two receivers. The transmitter sends three messages simultaneously, two private messages – one to each receiver, and a common message to both. These three messages are encoded into a single channel codeword $X^n$. The receivers receive noisy versions of the transmitted codeword – $Y_1^n$ at $R_{x_1}$ and $Y_2^n$ at $R_{x_2}$. After reception, each receiver decodes its message pair based only on its channel output. This scenario is depicted in figure 1.4.

Figure 1.4: The broadcast channel with independent receivers. The encoder sends three messages, a common message $W_0$, a private message to $R_{x_1}, W_1$, and a private message to $R_{x_2}, W_2$. $\hat{W}_0$ and $\hat{W}_0$ are the estimates of $W_0$ at $R_{x_1}$ and $R_{x_2}$ respectively.

the BC is that despite the fact that the receivers’ channel outputs are jointly distributed according to $p(y_1, y_2|x)$, the fact that the receivers decode independently
implies that the probability of error depends only on the marginal distributions \( p(y_1|x) \) and \( p(y_2|x) \). This implies, in turn, that different channels with the same marginal distributions have the same rate region.

Following Cover’s initial work, Bergmans proved an achievability result for the degraded BC, [42], and also a partial converse that holds only for the Gaussian broadcast channel [43]. Bergmans’s converse used a conditional entropy power inequality, for the first time since Shannon in 1948. In [44] Gallager established a converse that holds for any discrete memoryless degraded broadcast channel. In [45] El-Gamal generalized the capacity result for the degraded broadcast channel to the “more capable” case, and in [46] and [47] he showed that feedback does not increase the capacity region of the physically degraded BC. Several other classes of broadcast channels were studied in the following years. For example, the sum and product of two degraded broadcast channels were considered in [48], and in [49], [50], [51] and [52] the deterministic broadcast channel was analyzed. We note here that the capacity region of the deterministic BC exhibits an interesting relation to the Slepian-Wolf (SW) region for the noiseless encoding of a pair of sources [53]: the capacity region of the deterministic BC and the SW region for encoding a pair of sources that have a joint distribution that is equal to the joint distribution of the received signals in the BC setup, share the same edge representing the sum-rate bound. The SW region is above that edge and the BC region is below that edge (see [54, figure 4]).

For the general broadcast channel, Cover derived an achievable rate region for the case of three independent messages in [55]. In [56] Körner and Marton derived the capacity region of the general broadcast channel with a degraded message set. The best achievable region and an upper bound for the two private messages case
were derived by Marton in [57], and a simple proof of Marton’s achievable region (for a simplified structure of the auxiliary RVs) appeared later in [58]. Marton’s region is the capacity region when the BC has one deterministic component [57, theorem 4]. The best upper bound for the two independent senders case was presented recently in [59]. Another upper bound for the general broadcast channel, the so-called degraded, same-marginals (DSM) bound, was presented in [60]. This bound is weaker than the upper bound in [59] but stronger than Sato’s upper bound previously presented in [61]. We note, however, that while the upper bound in [59] is the strongest, it is valid only for the two-receiver case, while Sato’s bound and the DSM bound can be extended to more than two receivers. The effect of feedback on the capacity of the Gaussian broadcast channel was studied in [62] and [63], and in [64] the case of correlated sources was considered. A survey on the topic, with extensive references to previous work (up to 1998), can be found in [54]. In recent years the MIMO Gaussian broadcast channel has attracted a lot of attention. Initially, the sum-rate capacity was characterized in [65], [66], [67], [68], and finally, in [69] the capacity region was obtained.

### 1.4.2 Marton’s Achievable Rate Region for the General DMBC

As Marton’s achievable region is the most important result for the DMBC we look into this result in more detail. In its original formulation, the achievability theorem of [57] considers only two private messages:

**Theorem 1.4 ([57, theorem 2])** For the general discrete memoryless broadcast
channel, any rate pair \((R_1, R_2)\) satisfying

\[
R_1 \leq I(W, U; Y_1)
\]

\[
R_2 \leq I(W, V; Y_2)
\]

\[
R_1 + R_1 \leq \min \{I(W, Y_1), I(W, Y_2)\} + I(U; Y_1|W) + I(V; Y_2|W) - I(U; V|W),
\]

for some joint distribution \(p(w, u, v, x, y_1, y_2) = p(w, u, v, x)p(y_1, y_2|x)\), is achievable.

The non-trivial rate constraint here is the sum-rate constraint. This constraint actually gives the insight into the mechanism of Marton’s coding scheme (the following explanation is based on the construction of [58]): \(W\) is a RV representing information that can be decoded by both receivers (although that information is intended only for one. \(W\) can also be split between the two users and then each “part of \(W\)” is intended only for one of the receivers). Therefore, in the sum-rate it appears inside a minimum, and “counted” only once. In fact, incorporating a common message into Marton’s coding scheme can be done easily by letting \(W\) represent the common information (actually, letting “part of \(W\)” represent the common information). The variables \(U\) and \(V\) represent private information that is decoded only by its intended receiver. The receivers use standard joint-typicality decoding and it is actually the code construction at the transmitter that gives rise to the sum-rate constraint: the two independently selected sequences \(U^n\) and \(V^n\) have to be jointly typical in order to facilitate joint-typicality decoding at the receivers. It is this requirement that decreases the number of potential sequences that the transmitter can use, and this is reflected in the term \(-I(U; V|W)\) that appears in the sum-rate constraint. We note that the fundamental difference between the code construction for the degraded scenarios (stochastically degraded,
“less noisy” and “more capable”) and Marton’s construction (and actually also Cover’s construction in [55]) is the fact that in the former cases superposition coding is used, while in the latter, the codewords representing each message are selected independently.

Marton’s code construction can also be interpreted as an implementation of dirty paper coding (DPC) [70]: examine the sum-rate constraint (setting \( W \) to be a constant for simplicity):

\[
R_1 + R_2 \leq I(U;Y_1) + I(V;Y_2) - I(U;V).
\]

We can interpret this as transmitting to \( R_{x1} \) at rate \( I(U;Y_1) \) and then, the message to \( R_{x2} \) is generated while taking the codeword to \( R_{x1} \), \( U^n \), as a vector of states known non-casually to the transmitter. Then, the marginal p.d.f of the signal received at \( R_{x2} \) is \( p(y_2|v,u) \). Thus, for \( R_{x2} \) we obtain the Gelfand-Pinsker rate of [71], \( R_2 = I(V;Y_2) - I(U;V) \). Since the Gelfand-Pinsker rate is the capacity for that scenario, this interpretation may indicate that this sum-rate is optimal (but, of course, this is not easy to prove). This interpretation also suggests that the cardinality of \( U \) and \( V \) should be related, as from the Gelfand-Pinsker result we have that \( ||V|| \leq ||X|| + ||U|| \).

Although theorem 1.4 gives a single letter characterization of an achievable rate region is has a significant weakness: the maximum achievable rates are, in general, not computable. The reason is that there are no bounds on the cardinalities of the auxiliary RVs \( U \) and \( V \), in contrast to the degraded case (the cardinality of \( W \) can be bounded as in [64, theorem 2]). This lack of cardinality bounds is because of the way the code is constructed at the transmitter (for the degraded message set case over the general BC there are cardinality bounds [56]). So far there was only one attempt to calculate the maximum achievable rate pair – for the binary BC.
without the common RV [72]. However, the technique in [72] requires the auxiliary random variables to be independent, and thus it applies to Cover’s construction [55] but not to Marton’s general formulation [57]. The method of Ahlswede and Körner [73], [74] cannot be applied to the general BC since the auxiliary RVs in the general BC need to preserve a dependence relationship between themselves, see [64, comment after theorem 2].

We also note here that there is a duality between Marton’s region (when $W$ is a constant) and the CEO rate region [75] obtained using the Berger-Tung code construction [76]. This duality states that if $U - Y_1 - Y_2 - V$ is a Markov chain, then both regions have the same sum-rate constraint [77].

### 1.4.3 The Broadcast Channel with Cooperating Decoders:

**A Combination of Broadcasting and Relaying**

None of the early work on the DMBC considered *direct* cooperation between the receivers (feedback can be considered as a form of indirect cooperation). In the

![Figure 1.5: The broadcast channel with cooperating receivers. The encoder sends three messages, a common message $W_0$, a private message to $R_{x1}$, $W_1$, and a private message to $R_{x2}$, $W_2$. $\hat{W}_0$ and $\hat{W}_0$ are the estimates of $W_0$ at $R_{x1}$ and $R_{x2}$ respectively. The receivers have noiseless links of finite capacities $C_{12}$ and $C_{21}$ between them.](image-url)
cooperative broadcast scenario, a single transmitter sends two private messages, one to each receiver, and a common message to both receivers. These messages are encoded into a single channel codeword $X^n$. Each of the receivers gets a noisy version of the codeword, $Y_1^n$ at $R_{x1}$ and $Y_2^n$ at $R_{x2}$. After reception, the receivers exchange messages over noiseless conference links of finite capacities $C_{12}$ and $C_{21}$, as depicted in figure 1.5. The conference messages are, in general, functions of $Y_1^n$ (at $R_{x1}$), $Y_2^n$ (at $R_{x2}$), and the previous conference messages received from the other decoder. After conferencing, each receiver decodes its own message. This scenario extends the single common message cooperative broadcast scenario studied in [13] for the independent BC, to the most general setting of three independent messages transmitted over the general BC.

The scenario in which one transceiver helps a second transceiver in decoding a message is clearly a relay scenario. Hence, cooperative broadcast can be viewed as a generalization of the broadcast and relay scenarios into a hybrid broadcast-relay system, which better describes future communication networks.

Scenarios of this type have attracted considerable attention recently both from the practical and the theoretical aspects. From the practical aspect, new protocols are proposed for the collaborative broadcast scenario. For example in [78] the authors present a protocol for collaborative decision making involving broadcasting and relaying. From the theoretical aspect, there is a considerable effort invested in characterizing the capacity of an entire network. This work started with [14] and recent results appear in [79] and the following work [80], [22] and [23]. This work focuses on the Gaussian case. Another approach for studying the performance of an entire network is the network coding approach sparked by the work of [81], which focuses on encoding at the nodes for maximizing the network throughput,
separately from the channel coding. A third approach for studying the performance of a network is to combine the basic building blocks of a network, namely multiple access, relaying and broadcasting and study the capacities of these combinations. This is the approach we follow in this work.

In part of this work we study the combination of broadcast and relay. The hybrid broadcast-relay channel was introduced in [13] in which the authors applied a combination of EAF and DAF to the independent broadcast channel (i.e. $p(y_1, y_2|x) = p(y_1|x)p(y_2|x)$) with a single common message. In [13] the authors first considered the case in which the receivers hold a single-cycle conference over orthogonal conferencing channels, and then extended their result to the multi-cycle conference. We note here that although for the classic BC, the rate region is completely characterized by the marginal distributions, for the cooperative BC this is not true. Therefore, a more comprehensive investigation of cooperation over the general BC is of interest. In [82] we used both a single-step and a two-step conference with orthogonal conferencing channels in the discrete memoryless framework. In [83] we extended the conference to $K$ cycles. An important result on multi-cycle conferencing in the context of rate-distortion theory was derived in [84]. Another relevant work in this context appeared in [85] where superposition encoding at the transmitter and DAF at the relays were used to derive achievable rate regions for the discrete memoryless broadcast-relay channel with multiple relays/receivers.

An investigation of the broadcast-relay channel was also carried out in [86] and [87], in which the authors applied the DAF strategy to the case where only one receiver is helping the other receiver, and also presented an upper bound for this case. The upper bound in [87] is shown to be non-strictly contained in the cut-set bound, rather than strictly contained. These bounds coincide for the physically
degraded BC. In [87] also the fully cooperative scenario was analyzed. In this case both receivers send their conference messages simultaneously, where one receiver uses DAF and the other uses EAF. The codebook at the transmitter was generated using superposition. Another related work is [88] which also focuses on the broadcast-relay channel. In [88] additional achievable regions based on the partially cooperative scheme (i.e., only one receiver is assisting the other receiver) with DAF at the relay were derived, together with new upper bounds. The work [88] also considers several special cases of the BC. Both [87] and [88] use the RESWD code construction. Finally, a study of the rates achievable with cooperating receivers over the Gaussian broadcast channel with a single common message was carried out in [89]. In this context we note that for the multi-cycle conference considered in our work, we let the auxiliary RVs follow a more general relationship than in [89] thus achieving higher rates.

As a final comment we note that for multi-node cooperative scenarios, conceptually, EAF is preferable to DAF. The reason is that the DAF scheme is intended to help a single target node as the relay has to decode the message prior to helping the target node (unless a more stringent requirement that the relay will decode the messages of several users is imposed). Thus, while DAF evidently helps the target node, it may increase the interference to other nodes. In EAF, since cooperation does not depend on the messages transmitted from the sources, all nodes benefit from the relay transmission. This emphasized the importance of EAF relaying in cooperative scenarios.
1.5 The Multiple-Access Channel

The multiple-access channel is a many-to-one communication scenario. This sce-
nario is illustrated in figure 1.6.

![Image of the multiple-access channel]

Figure 1.6: The multiple-access channel: $N$ transmitters send messages to a single
receiver.

The introduction of the MAC can be traced back to a paper by Shannon in
1961 [1]. The capacity region of the MAC was obtained in 1971 by Ahlswede [90].
Another characterization of the capacity region appeared later in [2] and [91]. We
present this characterization as this is the most common one:

**Theorem 1.5** ([2, theorem 1]) The capacity region for sending two independent
messages over the multiple-access channel $(X_1 \times X_2, p(y|x_1, x_2), \mathcal{Y})$ is the convex
hull of all the rate pairs $(R_1, R_2)$ that satisfy

\[
R_1 \leq I(X_1; Y|X_2) \\
R_2 \leq I(X_2; Y|X_1) \\
R_1 + R_2 \leq I(X_1, X_2; Y),
\]

for the joint distribution $p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2)$. 

25
This region lends itself to a simple interpretation: the maximum rate from user 1 is obtained if the decoder has full knowledge of user 2’s message, thus it can cancel out the interference from user 2’s codeword. Hence we obtain \( I(X_1; Y|X_2) \) as the maximum rate for user 1 and similarly \( I(X_2; Y|X_1) \) for user 2. The sum-rate is obtained from the point-to-point channel where the input distribution \( p(x_1, x_2) \) is constrained to be the product of the marginals, \( p(x_1, x_2) = p(x_1)p(x_2) \). This constraint is because each transmitter does not have knowledge of the other transmitter’s message, thus they cannot coordinate their codewords.

Several types of encoder cooperation were considered for the MAC: in [92] the case where there is one common message known at both encoders and a private message at only one encoder was considered (see also [93], [94]). The capacity region for the MAC with two private messages and one common message known at both encoders was derived in [95]. The MAC with feedback was investigated by Cover & Leung in [96] and recently an improved achievable region was found in [97]. We note that the MAC was the first channel for which it was demonstrated that feedback increases the capacity [98]. The case of conferencing encoders was investigated by Willems in [99]. In this thesis we apply Willems’ conference to the receivers in the broadcast channel. Finally, we note that the capacity region of the additive white Gaussian noise (AWGN) MAC was found by Wyner in [100].

### 1.5.1 MAC-BC Duality

In recent years the relationship between the MAC and the BC received a lot of attention. This research led to the establishment of duality relationships between the Gaussian BC and MAC. It was found that the Gaussian BC and the Gaussian MAC are closely related: the DPC achievable rate region for the Gaussian BC
with a power constraint $P$ is equal to the union of a capacity regions of all the dual MACs which have the same sum power constraint $P$ (i.e. the sum of the powers all of the transmitters is less than or equal to $P$). This relationship holds for both the scalar case and the MIMO case [101], [67]. The dual MAC channel has the same gain, but decoding in the MAC is done in a reversed order of the encoding in the BC, and the powers are scaled accordingly. This duality relationship was used to show that the DPC achievable region achieves the sum-rate capacity of the MIMO BC, by showing that the union of the regions of the dual MACs of the MIMO BC is at least as large as the Sato upper bound for the BC. A similar result appeared also in [66] for the case where each receiver has a single antenna, by using a more complex duality structure.

Unfortunately, it is not simple to extend this duality to the discrete memoryless case. The only duality example for the discrete memoryless case appeared in [102] where a relationship between all deterministic BCs and all deterministic MACs subject to a certain relationship of their cardinalities was shown. This problem is much harder than the Gaussian case since it is not clear how to construct the dual channel.

1.6 Main Contributions and Organization

In the following we summarize the main contributions of this work.

- We give an intuitive insight into the relay channel in terms of an information flow on a graph, and show how to obtain the EAF result of [15, theorem 6] from flow considerations. This flow interpretation highlights the underlying assumptions of the EAF strategy. Using flow considerations we also obtain the rate of the EAF strategy when the receiver uses joint-decoding. A similar
expression can be obtained by specializing the result of [103] to the case where the relay does not perform partial decoding. We then show that joint-decoding does not exceed the maximum rate of the EAF strategy achieved by the sequential decoding of [15, theorem 6]. Moreover, we find the assignment of the auxiliary RV that obtains the joint-decoding rate expression from the classic EAF expression. We refer to the class of assignments that this assignment belongs to as time-sharing (TS) assignments. We also present another time-sharing assignment that always exceeds the joint-decoding rate. This result has the same supremum rate as the classic EAF result of [15, theorem 6] but does not have a feasibility constraint. Therefore, it provides a simpler characterization of the EAF achievable rate.

- We introduce an achievable rate expression for the multiple-relay scenario based on EAF, that is also practically computable. As discussed in section 1.3.1, when the channel from the source to the relay is “noisy”, EAF may outperform DAF. However, for the multiple-relay scenario there is no explicit, computationally practical expression based on EAF that can be compared against the best DAF-based result derived in [24], so that the better strategy can be identified. As indicated in [12, remark 22, remark 23], applying the general EAF to a network with an arbitrary number of relays is computationally prohibitive due to the large number of constraints that characterize the feasible region (in addition to the cardinality and convexity issues discussed in section 1.3.1). Therefore, it is interesting to explore a computationally simple assignment that allows to derive a result that extends to an arbitrary number of relays in a simple manner. We also provide an explicit numerical example to demonstrate that indeed there are discrete memoryless scenarios
where multi-relay EAF outperforms both multi-relay DAF and point-to-point transmission.

- We consider the optimization of the EAF auxiliary random variable for the Gaussian relay channel with an orthogonal relay-destination link. We focus on the coded modulation scenario, and show that there are three regions: high SNR on the source-relay link, where DAF is the best strategy, low SNR on the source-relay link in which the common EAF with a Gaussian assignment is best, and an intermediate SNR region where EAF with “hard-decision per symbol” is better than both DAF and Gaussian EAF. For this intermediate SNR region we consider two kinds of hard-decisions: deterministic and probabilistic, and show that each one of them can be superior, depending on the channel conditions. We thus combine both schemes and analyze the resulting hybrid deterministic-probabilistic quantization. We give some insight on the information rates obtained with the different schemes.

- We study a special case of the general cooperative broadcast setup formulated in section 1.4.3: the case of the physically degraded broadcast channel. Although the physically degraded BC is of little practical interest, it is useful in developing the coding concept for the general BC with cooperating receivers. For the physically degraded BC, we present both an achievability result and a converse. Together, these two results give the capacity region for this setup. Furthermore, this new region is shown to be a strict enlargement of the classic region without cooperation.

- Lastly, we consider the general broadcast channel with receivers holding a multi-cycle conference. We derive an achievable rate region, extending the
Marton rate region of [57, theorem 2] to the case where the receivers hold a $K$-cycle conference prior to decoding the messages. We also derive an upper bound on the achievable rates for this scenario. We then specialize the achievability result to the single common message case and obtain explicit expressions (without auxiliary RVs) for the single-cycle conference. Here we provide expressions that explicitly relate the capacities of the conference links to the increase in the information rate. Finally, we show that for a special case of the general BC with a single common message, namely when one channel is distinctly better than the other, the upper and lower bounds coincide, resulting in the capacity for that case. Our results also provide, as a special case, the capacity region of the deterministic BC with cooperating receivers originally derived in [104].

These contributions are summarized in the following papers:

**Journal Papers**


**Conference Papers**

International Symposium on Information Theory (ISIT), June 2004, Chicago, IL.


The rest of this work is organized as follows: in chapter 2 we discuss the EAF strategy with time-sharing and its relationship to joint-decoding. In chapter 3 we study the cooperative broadcast scenario. In chapter 4 we apply time-sharing to the multiple-relay case and to the Gaussian relay channel with coded modulation. In chapter 5 we present concluding remarks and suggestions for future research.
1.7 Notations

First, a word about notation: in the following we use $H(\cdot)$ to denote the entropy of a discrete random variable, $h(\cdot)$ to denote the differential entropy of a continuous RV and $I(\cdot;\cdot)$ to denote the mutual information between two random variables, as defined in [19, ch. 2, ch. 9]. We denote the real numbers with $\mathbb{R}$. We denote random variables with upper case letters e.g. $X$, $Y$, and their realizations with lower case letters $x$, $y$. A random variable $X$ takes values in a set $\mathcal{X}$. We use $|\mathcal{X}|$ to denote the cardinality of a finite discrete set $\mathcal{X}$, $p_X(x)$ to denote the probability mass function (p.m.f.) of a discrete RV $X$ on $\mathcal{X}$ and $f_X(x)$ to denote the probability density function (p.d.f.) of a continuous RV $X$ on $\mathbb{R}$. For brevity we may omit the subscript $X$ when it is obvious from the context. We use $p_{X|Y}(x|y)$ to denote the conditional distribution of $X$ given $Y$. We denote vectors with boldface letters, e.g. $\mathbf{x}$, $\mathbf{y}$; the $i$'th element of a vector $\mathbf{x}$ is denoted by $x_i$ and we use $\mathbf{x}^j_i$ where $i < j$ to denote $(x_i, x_{i+1}, \ldots, x_{j-1}, x_j)$; $\mathbf{x}^j$ is a short form notation for $\mathbf{x}^j_1$, and unless specified otherwise $\mathbf{x} \triangleq x^n$. We use $A_{n}^{(n)}(X)$ to denote the set of $\epsilon$-strongly typical sequences w.r.t. distribution $p_X(x)$ on $\mathcal{X}$, as defined in [105, ch. 5.1] and $A_{c}^{(n)}(X)$ to denote the set of $\epsilon$-weakly typical sequences as defined in [19, ch. 3]. When referring to a typical set we may omit the random variables from the notation, when these variables are clear from the context. We use $[a]^*$ to denote $\min\{a, 1\}$, where $a \in \mathbb{R}$. 

32
Chapter 2

Estimate-and-Forward Relaying with

Time-Sharing Auxiliary Mapping

In this chapter we consider the relay channel with the EAF scheme. As discussed in section 1.3.1, EAF has a basic limitation – its maximum achievable rate is not computable. In order to compare EAF with other strategies such as AAF and DAF, we need to obtain a computable expression. This requires using an assignment of the auxiliary RV that results in a rate expression that depends only of the parameters of the problem $X, X_1, Y$ and $Y_1$. One explicit choice we consider in the following is the *time-sharing assignment*. In this assignment the relay transmits its channel output in a certain percentage of the block time, and in the remaining time transmits an erasure signal. This can also be viewed as an “on/off” signalling. This allows the relay to set the time-sharing ratio such that the feasibility constraint (1.4) is satisfied. We then study the implications of such an assignment: it allows us to evaluate the usefulness of introducing joint-decoding at the destination receiver instead of the classic sequential decoding. Another useful application which motivates this assignment is the multiple-relay channel considered in chapter 4: extending the general EAF strategy to the multiple-relay channel results in a large number of constraints that characterize the feasible region, while the TS assignment produces rate expressions without a feasibility condition. These rates are generally smaller than the general multi-relay EAF rates, however, they are computable. Most importantly, this assignment allows us to derive an alternative characterization of the classic EAF result of [15, theorem 6] without a feasibility condition.
Before going into the details of the TS assignment, we first give the formal definition of the relay channel. We then offer an intuitive way to interpret the EAF strategy in terms of an information flow on a graph. Following that, we analyze the TS assignment and compare its rate to the joint-decoding rate.

### 2.1 Definitions for the Relay Channel

**Definition 2.1** The discrete relay channel is defined by two discrete input alphabets $\mathcal{X}$ and $\mathcal{X}_1$, two discrete output alphabets $\mathcal{Y}$ and $\mathcal{Y}_1$ and a probability mass function $p(y, y_1|x, x_1)$ giving the probability distribution on $\mathcal{Y} \times \mathcal{Y}_1$ for each $(x, x_1) \in \mathcal{X} \times \mathcal{X}_1$. We denote this channel by $(\mathcal{X} \times \mathcal{X}_1, p(y, y_1|x, x_1), \mathcal{Y} \times \mathcal{Y}_1)$. This definition extends in a straightforward manner to the continuous case by replacing $\mathcal{Y} \times \mathcal{Y}_1$ with $\mathbb{R}^2$ and using p.d.fs instead of p.m.fs. The relay channel is called **memoryless** if the probability of a block of $n$ transmissions is given by

$$p(y^n, y^n_1|x^n, x^n_1) = \prod_{i=1}^n p(y_i, y_{1,i}|x_i, x_{1,i}).$$

In this work we consider only the memoryless relay channel.

**Definition 2.2** A $(2^nR, n)$ code for the relay channel consists of a source message set $\mathcal{W} = \{1, 2, ..., 2^nR\}$, a mapping function $f$ at the encoder,

$$f : \mathcal{W} \mapsto \mathcal{X}^n,$$

a set of $n$ relay functions

$$x_{1,i} = t_i(y_{1,1}, y_{1,2}, ..., y_{1,i-1}), \quad i = 2, 3, ..., n,$$

where the $i$'th relay function $t_i$ maps the first $i - 1$ channel outputs at the relay into a transmitted relay symbol at time $i$, and $x_{1,1} = c$ for some arbitrary $c \in \mathcal{X}_1$. 

34
Lastly we have a decoder

\[ g : \mathcal{Y}^n \mapsto \mathcal{W}. \]

**Definition 2.3** The average probability of error for a code of length \( n \) for the relay channel is defined as

\[ P_e^{(n)} = \Pr(g(Y^n) \neq W), \]

where \( W \) is selected uniformly over \( \mathcal{W} \).

**Definition 2.4** A rate \( R \) is called achievable for the relay channel if for every \( \epsilon, \delta > 0 \) there exists a block length \( n \), such that a \((2^{n(R-\delta)}, n)\) relay channel code with \( P_e^{(n)} \leq \epsilon \) can be constructed.

### 2.2 An Information Flow Interpretation of Estimate-and-Forward

Consider the rate bound and the feasible region of theorem 1.2 given in equations (1.3) and (1.4). We note that the following intuitive explanation does not constitute a proof, but it does provide insight into the achievable rate expression of the EAF scheme. We emphasize that the achievable rates obtained in this section can also be proved rigorously. In the following we provide insight into the expressions of (1.3) and (1.4) in terms of a flow on a graph.

In constructing the information flow representation for the relay channel, we first need to specify the underlaying assumptions and the operations performed at the source, the relay and the destination receiver:

- The source and the relay generate their codebooks without coordination.

This implies that the joint distribution of the channel inputs is in fact the
product of the marginal distributions: \( p(x, x_1) = p(x)p(x_1) \). We note that this is in contrast to DAF in which the source and the relay codebooks are related through superposition construction.

- The relay compresses its channel output \( y_1 \) into \( \hat{y}_1 \), which represents the information conveyed to the destination receiver to assist in decoding the source message. Therefore, \( \hat{y}_1 \) can depend only on \( x_1 \) and \( y_1 \).

Based on the above two assumptions we have the following distribution chain:

\[
p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|x_1, y_1).
\]

Now we note that:

- The relay channel input \( x_1 \) is based only on the compressed \( \hat{y}_1 \).

- The destination uses \( x_1, \hat{y}_1 \) and \( y \) to decode the source codeword \( x \).

We also use the following representation for transmission, reception and compression:

- We represent an information source as a source whose output flow is equal to its information rate.

- We represent the compression operation as a flow sink whose flow consumption is equal to the mutual information between the original and the compressed sequences. In this sense, the compression can be viewed as a device that removes flow from the graph, leaving a smaller flow going through the output edge. The amount of flow removed is exactly the new information that \( \hat{Y}_1 \) contains on \( Y_1 \). By removing this flow, we introduce the requirement that in order for the destination to obtain this information we need to have a positive flow through the relay path.

- The destination is represented as a flow sink.
• As in a standard flow on a graph, the flows are additive, following the chain rule of mutual information.

Now consider the flow diagram of figure 2.1. As can be observed from the

\[
i_T = I(X; Y, \hat{Y}_1, X_1) = I(X; Y, \hat{Y}_1 | X_1).
\]

This follows from the fact that the destination uses \( x_1, \hat{y}_1 \) and \( y \) to decode \( x \) and the fact that \( X \) and \( X_1 \) are independent. This total flow reaches the receiver through two branches, the direct branch (D) which carries a flow of \( i_D = I(X; Y | X_1) \) and the relay branch (ABCE). Now, the quantities in the relay branch are calculated given \( X_1 \) and \( Y \) to represent only the rate increase over the direct path. The relay branch consists of four components: an edge (A) which carries a flow of \( i_A = i_T - i_D = I(X; \hat{Y}_1 | X_1, Y) \), a sink (B) with consumption \( I(Y_1; \hat{Y}_1 | X_1, Y) \),

![Figure 2.1: The information flow budget for the general relay channel with compression at the relay.](image_url)
a relay source (C) with an output flow of $I(X_1; Y)$ and an edge (E) from the relay to the destination. Here, the relay source flow, $i_C$, has a fixed value of $I(X_1; Y)$, independent of the type of compression $p(\hat{y}_1|x_1)$ used at the relay, since we always transmit from the relay to the destination at the maximum possible rate in order to obtain the best performance. As explained previously, the rate loss due to compression is represented by $I(\hat{Y}_1; Y_1|X_1, Y)$. In fact, $I(\hat{Y}_1; Y_1|X_1, Y)$ can be interpreted as the information that $\hat{Y}_1$ contains on $Y_1$ beyond what $X_1$ and $Y$ already do.

Now, from the law of flow addition, the net flow from the source to the destination through the relay branch is $i_E = i_A - i_B + i_C$. To assist the direct link (D) we need the flows on edges (A) and (E) to be positive. Note that we always have $i_A \geq 0$. In theorem 1.2 the scheme considers only the last two elements in the sum, $-i_B + i_C$, and verifies that their net flow is positive, namely

$$-I(Y_1; \hat{Y}_1|X_1, Y) + I(X_1; Y) > 0.$$  \hfill (2.1)

This condition guarantees a net positive flow on (E) since $i_A \geq 0$. Now, the flow to the destination can be obtained as the minimum

$$R \leq \min \{i_D + i_E, i_T\} \quad \hfill (2.2)$$

$$= i_D + \min \{i_A, i_E\},$$

where the second term in the minimum in (2.2) is obtained from the transmitter, since trivially the information rate at the receiver cannot exceed $i_T$. We note that because $-i_B + i_C \geq 0$, the minimum in (2.2) is $i_T$. Therefore, the resulting achievable rate is

$$R \leq I(X; Y, \hat{Y}_1|X_1),$$

which combined with (2.1) gives the result of [15, theorem 6].
However, the condition in (2.1) is not tight since even when \(-i_B + i_C < 0\) the flow on (E) is still non-negative if the entire sum \(i_A - i_B + i_C\) is non-negative, i.e.

\[
I(X; \hat{Y}_1|X_1, Y) - I(Y_1; \hat{Y}_1|X_1, Y) + I(X_1; Y) \geq 0.
\]  

(2.3)

Then, the achievable rate to the destination is bounded by

\[
R \leq i_D + i_E = I(X; Y|X_1) + I(X_1; Y) - I(\hat{Y}_1; Y_1|X, X_1, Y).
\]  

(2.4)

Indeed, when the flow through edge (E) is zero we obtain the non-cooperative rate

\(I(X; Y|X_1)\). Plugging the expression (2.4) into (2.2) yields the following achievable rate:

\[
R \leq \min \{i_D + i_E, i_T\}
= \min \left\{ I(X; Y|X_1) + I(X_1; Y) - I(\hat{Y}_1; Y_1|X, X_1, Y), I(X; Y, \hat{Y}_1|X_1) \right\}
= I(X; Y|X_1) + \min \left\{ I(X_1; Y) - I(\hat{Y}_1; Y_1|X, X_1, Y), I(X; \hat{Y}_1|X_1) \right\}.
\]

Combining this with (2.3), (informally) proves the following proposition:

**Proposition 2.1** For the general relay channel, any rate \(R\) satisfying

\[
R \leq I(X; Y|X_1) + \min \left\{ I(X_1; Y) - I(\hat{Y}_1; Y_1|X, X_1, Y), I(X; \hat{Y}_1|X_1) \right\},
\]

subject to

\[
I(X_1; Y) \geq I(\hat{Y}_1; Y_1|X_1, Y) = I(Y_1; \hat{Y}_1|X_1, Y) - I(X; \hat{Y}_1|X_1, Y),
\]

for some joint distribution \(p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|x_1, y_1)\), is achievable.

The proof of proposition 2.1 can be made formal using joint-decoding at the destination receiver, see appendix A. In the next subsection we show that this expression is a special case of [15, theorem 6] obtained by time-sharing.
2.3 The Single Relay EAF with Time-Sharing Assignment

Consider the following time-sharing assignment for the auxiliary random variable of theorem 1.2:

\[ p(\hat{y}_1|y_1, x_1) = \begin{cases} 
q & , \hat{y}_1 = y_1 \\
1 - q & , \hat{y}_1 = E \notin Y_1,
\end{cases} \tag{2.5} \]

\( q \in [0, 1] \). This means that each \( \hat{y}_1 \) generated at the relay will have approximately \( qn \) symbols of a \( y_1 \) sequence with which it is also jointly typical. Under this assignment, the feasibility condition of (1.4) becomes

\[
I(X_1; Y) \geq I(Y_1; \hat{Y}_1|X_1, Y) \\
= H(Y_1|X_1, Y) - H(Y_1|X_1, Y, \hat{Y}_1) \\
= H(Y_1|X_1, Y) - (1 - q)H(Y_1|X_1, Y) - qH(Y_1|X_1, Y, Y_1) \\
= qH(Y_1|X_1, Y),
\]

and the rate expression (1.3) becomes

\[
R \leq I(X; Y, \hat{Y}_1|X_1) \\
= I(X; Y|X_1) + I(X; \hat{Y}_1|X_1, Y) \\
= I(X; Y|X_1) + H(X|X_1, Y) - H(X|X_1, Y, \hat{Y}_1) \\
= I(X; Y|X_1) + H(X|X_1, Y) - (1 - q)H(X|X_1, Y) - qH(X|X_1, Y, Y_1) \\
= I(X; Y|X_1) + qI(X; Y_1|X_1, Y).
\]

Clearly, maximizing the rate implies maximizing \( q \) subject to the constraint \( q \in [0, 1] \). This gives the following corollary to theorem 1.2:

**Corollary 2.1** For the general relay channel any rate \( R \) satisfying

\[
R \leq I(X; Y|X_1) + \left[ \frac{I(X_1; Y)}{H(Y_1|X_1, Y)} \right]^* I(X; Y_1|X_1, Y), \tag{2.6}
\]

40
for the joint distribution \(p(x, x_1, y, y_1) = p(x)p(x_1)p(y, y_1|x, x_1)\), with \([a]^{*} \triangleq \min\{a, 1\}\), is achievable.

Now, consider the following distribution chain:

\[
p(x, x_1, y, y_1, \hat{y}_1, \hat{\hat{y}}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|x_1, y_1)p(\hat{\hat{y}}_1|\hat{y}_1).
\]  

(2.7)

We note that this extended chain can be put into the standard form by letting \(p(\hat{\hat{y}}_1|x_1, y_1) = \sum_{\hat{Y}_1} p(\hat{y}_1, \hat{\hat{y}}_1|x_1, y_1) = \sum_{\hat{Y}_1} p(\hat{y}_1|x_1, y_1)p(\hat{\hat{y}}_1|\hat{y}_1)\). This distribution chain represents a situation in which after compression of \(Y_1\) into \(\hat{Y}_1\), there is a second compression operation, compressing \(\hat{Y}_1\) into \(\hat{\hat{Y}}_1\). The output of the second compression is used to facilitate cooperation between the relay and the destination. Therefore, the receiver decodes the message based on \(\hat{\hat{y}}_1\) and \(y\). Repeating exactly the same steps as in the standard relay decoding, with \(\hat{y}_1\) replacing \(\hat{\hat{y}}_1\), the expressions of theorem 1.2 become

\[
R \leq I(X; Y, \hat{\hat{Y}}_1|X_1),
\]  

subject to \(I(X_1; Y) \geq I(Y_1; \hat{\hat{Y}}_1|X_1, Y)\).  

(2.9)

Now, applying TS to \(\hat{\hat{Y}}_1\) with

\[
p(\hat{\hat{y}}_1|\hat{y}_1) = \begin{cases} 
q & \hat{\hat{y}}_1 = \hat{y}_1 \\
1 - q & \hat{\hat{y}}_1 \not\in \hat{\hat{Y}}_1
\end{cases},
\]  

(2.10)
the expressions in (2.8) and (2.9) become

\[ R \leq I(X;Y|X) + I(X;\hat{Y}_1|X,Y) \]

\[ = I(X;Y|X) + H(X|Y) - H(Y|X,Y) \]

\[ = I(X;Y|X) + q(H(X|X,Y) - H(X|\hat{Y}_1,X,Y)) \]

\[ = I(X;Y|X) + qI(X;\hat{Y}_1|X,Y), \quad (2.11) \]

\[ I(X;Y) \geq I(Y_1;\hat{Y}_1|X,Y) \]

\[ = H(Y_1|X,Y) - H(Y_1|\hat{Y}_1,X,Y) \]

\[ = H(Y_1|X,Y) - (1-q)H(Y_1|X,Y) - qH(Y_1|\hat{Y}_1,X,Y) \]

\[ = qI(Y_1;\hat{Y}_1|X,Y). \quad (2.12) \]

Combining this with the constraint \( q \in [0,1] \) we obtain the following corollary to theorem 1.2:

**Proposition 2.2** For the general relay channel, any rate \( R \) satisfying

\[ R \leq I(X;Y|X) + \left[ \frac{I(X_1;Y)}{I(Y_1;\hat{Y}_1|X,X)} \right] I(X;\hat{Y}_1|X,Y), \]

for some joint distribution \( p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1| x, x_1)p(\hat{y}_1|x_1, y_1) \), is achievable.

This proposition generalizes on corollary 2.1 by performing a general Wyner-Ziv compression followed by TS (which is a specific type of WZ compression), intended to guarantee feasibility of the first compression step. Note that the supremum of the rate of proposition 2.2 is equal to the supremum of the rate of [15, theorem 6]. Therefore, proposition 2.2 provides an alternative representation of the classic EAF result without a feasibility constraint. In section 4.2 we apply a similar idea to EAF relaying in the Gaussian relay channel scenario with coded modulation.

Next we discuss the relationship between joint-decoding and time-sharing.
2.4 Joint-Decoding and Time-Sharing

In the original EAF result of [15, theorem 6], the decoding procedure at the destination receiver for decoding the message \( w_{i-1} \) at time \( i \) consists of three steps (the notations below are identical to those used in [15, theorem 6]. The reader is referred to the proof of [15, theorem 6] to recall the definitions of the sets and variables used in the following description):

1. Decode the relay index \( s_i \) using \( y(i) \), the received signal at time \( i \).

2. Decode the relay message \( z_{i-1} \), using \( s_i \), the received signal \( y(i-1) \) and the previously decoded \( s_{i-1} \).

3. Decode the source message \( w_{i-1} \) using \( y(i-1) \), \( z_{i-1} \) and \( s_{i-1} \).

Evidently, when decoding the relay message \( z_{i-1} \) at the second step, the receiver does not make use of the statistical dependence between \( \hat{y}_1(z_{i-1}|s_{i-1}) \) – the relay sequence at time \( i - 1 \), and \( x(w_{i-1}) \) – the transmitted source codeword at time \( i - 1 \). The way to use this dependence is to jointly decode \( z_{i-1} \) and \( w_{i-1} \) after decoding \( s_i \) and \( s_{i-1} \). The joint-decoding procedure at time \( i \), then, consists of following steps:

1. From \( y(i) \), the received signal at time \( i \), the receiver decodes \( s_i \) by looking for a unique \( s \in S \), the set of indices used to select the codeword \( x_1 \), such that \( (x_1(s), y(i)) \in A_e^{s(n)} \). This is done exactly as in the first decoding step in [15, theorem 6].

2. The receiver now knows the set \( S_{s_i} \) into which \( z_{i-1} \) (the relay message at time \( i - 1 \)) belongs. Additionally, from decoding at time \( i - 1 \) the receiver knows \( s_{i-1} \), used to generate \( z_{i-1} \).
3. The receiver generates the set

\[ \mathcal{L}(i-1) = \left\{ w \in \mathcal{W} : (x(w), y(i-1), x_1(s_{i-1})) \in A_e^{(n)} \right\}. \]

4. The receiver now looks for a unique \( w \in \mathcal{L}(i-1) \) such that \((x(w), y(i-1), \hat{y}_1(z|s_{i-1}), x_1(s_{i-1})) \in A_e^{(n)}\) for some \( z \in S_{s_i} \). If such a unique \( w \) exists then it is the decoded \( \hat{w}_{i-1} \), otherwise the receiver declares an error.

As noted in section 2.2, the rate expression resulting from this decoding procedure is given by proposition 2.1. See proof in appendix A.

Let us now compare the rate obtained with joint-decoding (proposition 2.1) with the rate obtained with the sequential decoding of [15, theorem 6]. To that end we consider the joint-decoding result of proposition 2.1 with the extended probability chain of (2.7):

\[ p(x, x_1, y, y_1, \hat{y}_1, \hat{\hat{y}}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|x_1, y_1)p(\hat{\hat{y}}_1|\hat{y}_1), \]

where \( \hat{Y}_1 \) represents the information relayed to the destination. We now expand the expressions of proposition 2.1, derived by applying the extended chain, using the assignment (2.10) in a similar manner to equations (2.11) and (2.12). The resulting expressions are:

\[
R \leq I(X; Y|X_1) + \min \left\{ I(X_1; Y) - qI(\hat{Y}_1; Y_1|X, X_1, Y), \right. \\
\left. qI(X; \hat{Y}_1|X_1, Y) \right\} \quad (2.13)
\]

subject to

\[
I(X_1; Y) \geq qI(\hat{Y}_1; Y_1|X, X_1, Y) \\
= q \left( I(\hat{Y}_1; Y_1|X_1, Y) - I(X; \hat{Y}_1|X_1, Y) \right). \quad (2.14)
\]

We now can make the following observations:
1. Setting \( q = 1 \) we obtain proposition 2.1. Additionally, if

\[ I(X_1; Y) > I(\hat{Y}_1; Y_1|X_1, Y) \]

then both proposition 2.1 and [15, theorem 6] give identical expressions.

2. When \( q = 1 \) and

\[ I(\hat{Y}_1; Y_1|X_1, Y) - I(X_1 ; \hat{Y}_1|X_1, Y) < I(X_1; Y) < I(\hat{Y}_1; Y_1|X_1, Y), \]  
(2.15)

then for the same mapping \( p(\hat{y}_1|x_1, y_1) \) we obtain that proposition 2.1 provides rate but [15, theorem 6] does not. The rate expression of proposition 2.1 under these conditions is

\[ R \leq I(X; Y|X_1) + I(X_1; Y) - I(\hat{Y}_1; Y_1|X, X_1, Y). \]  
(2.16)

3. Now, fix the probability chain \( p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|x_1, y_1) \) and examine the expressions (2.13) and (2.14) when (2.15) holds: for any \( 0 \leq q < 1 \), (2.15) guarantees that condition (2.14) is still satisfied. If \( q \) is close enough to 1 such that we also have \( I(X_1; Y) \leq qI(\hat{Y}_1; Y_1|X_1, Y) \), the rate from (2.13), i.e.,

\[ R \leq I(X; Y|X_1) + I(X_1; Y) - qI(\hat{Y}_1; Y_1|X, X_1, Y), \]

is now greater than (2.16). Thus, using TS the rate of proposition 2.1 is increased. In this case we can keep decreasing \( q \) until

\[ I(X_1; Y) - qI(\hat{Y}_1; Y_1|X, X_1, Y) = qI(X_1; \hat{Y}_1|X_1, Y) \]  
(2.17)

at which point the rate becomes

\[ R \leq I(X; Y|X_1) + qI(X_1; \hat{Y}_1|X_1, Y). \]  
(2.18)

This rate can be obtained from [15, theorem 6] by applying the extended probability chain of (2.7), as long as \( I(X_1; Y) \geq qI(\hat{Y}_1, Y_1|X_1, Y) \).
We now show that all the rates that joint decoding allows can also be obtained or exceeded by the original EAF with an appropriate time sharing. Note that equality in (2.17) implies

\[ q_{opt} = \min \left\{ 1, \frac{I(X_1; Y)}{I(\hat{Y}_1; Y_1 | X, X_1, Y)} + I(X; \hat{Y}_1 | X_1, Y) \right\} = \min \left\{ 1, \frac{I(X_1; Y)}{I(\hat{Y}_1; Y_1 | X_1, Y)} \right\}, \]

hence \( q_{opt} \) is the maximum \( q \) that makes the mapping \( p(\hat{y}_1 | x_1, y_1) \) feasible for [15, theorem 6]. Plugging \( q_{opt} \) into (2.18), we obtain the rate expression of proposition 2.2. Thus, we see that at optimality, joint-decoding (when considered under distribution chains that satisfy (2.15)) produces proposition 2.2.

Finally, consider again the region where joint-decoding is useful (2.15):

\[ I(\hat{Y}_1; Y_1 | X, X_1, Y) \leq I(X_1; Y) \leq I(\hat{Y}_1; Y_1 | X_1, Y) \]

\[ \Rightarrow 0 \leq I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y) \leq I(\hat{Y}_1; Y_1 | X_1, Y) \]

\[ \Rightarrow 0 \leq I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y) \leq I(X_1; \hat{Y}_1 | X_1, Y) \]

\[ \Rightarrow 0 \leq \frac{I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y)}{I(X; \hat{Y}_1 | X_1, Y)} \leq 1, \]

if \( I(X; \hat{Y}_1 | X_1, Y) > 0 \). Then, using time-sharing on \( \hat{Y}_1 \) with

\[ q = \frac{I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y)}{I(X; \hat{Y}_1 | X_1, Y)} \] (2.19)

plugged into equations (2.11) and (2.12) yields:

\[ I(X; Y | X_1) + qI(X; \hat{Y}_1 | X_1, Y) = I(X; Y | X_1) + I(X_1; Y) - I(\hat{Y}_1; Y_1 | X_1, Y), \]

as long as \( I(X_1; Y) \geq qI(\hat{Y}_1; Y_1 | X_1, Y) \), or equivalently for \( I(\hat{Y}_1; Y_1 | X_1, Y) > 0 \),

\[ q \leq \frac{I(X_1; Y)}{I(\hat{Y}_1; Y_1 | X_1, Y)}. \] (2.20)
Plugging assignment (2.19) into (2.20) we obtain:

\[
\frac{I(X_1;Y) - I(\hat{Y}_1;Y_1|X_1,Y)}{I(X;\hat{Y}_1|X_1,Y)} \leq \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X_1,Y)}
\]

\[
\Rightarrow \quad \left(I(X_1;Y) - I(\hat{Y}_1;Y_1|X_1,Y)\right) I(\hat{Y}_1;Y_1|X_1,Y) \leq I(X_1;Y) I(X;\hat{Y}_1|X_1,Y)
\]

\[
\Rightarrow \quad I(X_1;Y) I(\hat{Y}_1;Y_1|X_1,Y) - I(X_1;Y) I(X;\hat{Y}_1|X_1,Y) \leq I(\hat{Y}_1;Y_1|X_1,Y) \times I(\hat{Y}_1;Y_1|X_1,Y)
\]

\[
\Rightarrow \quad I(X_1;Y) \leq I(\hat{Y}_1;Y_1|X_1,Y),
\]

as long as \(I(\hat{Y}_1;Y_1|X_1,Y) > 0\), which is the region where joint-decoding is supposed to be useful. Hence the joint-decoding rate of proposition 2.1 can be obtained by time sharing on the [15, theorem 6] expression. Therefore, joint-decoding does not improve on the rate of [15, theorem 6]. In fact, from (2.20) we see that the rate of proposition 2.2 is always at least as large as that of proposition 2.1, in the region where joint-decoding is supposed to be superior to the classic decoding.

### 2.4.1 Discussion

We make the following comments on the results presented in this chapter.

- Time-sharing is an assignment that results in rate expressions without a feasibility constraint following proposition 2.2. Since we removed the feasibility constraint, we can further set \(\hat{Y}_1 = Y_1\) in proposition 2.2 and obtain an explicit rate expression without an auxiliary RV. Although this rate is sub-optimal it can give insight into the problem by allowing a comparison with the DAF and AAF rates. We can also evaluate this rate numerically.
• Although the result of proposition 2.2 has the same supremum rate as the classic EAF result of [15, theorem 6], for fixed distributions $p(x)$, $p(x_1)$ and $p(\hat{y}_1|x_1, y_1)$, proposition 2.2 is at least as good as [15, theorem 6] and may exceed it. This result simplifies the optimization problem for finding the maximum rate, but we still have to deal with the major problems of the lack of a cardinality bound on $\hat{Y}_1$ and the non-convex nature of the search.

• In this chapter we showed that for the single relay EAF, joint-decoding does not improve upon sequential decoding. However, for the multiple-relay channel this may not necessarily be the case. Specifically, consider the multiple-relay DAF strategy. As discussed in section 1.3.2, the multi-relay DAF based on MESD is inferior to the multi-relay schemes based on RESWD and REBD. However, it may be that a joint-decoding version of MESD for the multiple-relay case will outperform the sequential version of MESD used in [22].
Chapter 3
The Broadcast Channel with Cooperating Decoders

In this chapter we first consider the capacity region of the physically degraded BC with cooperating receivers. We then derive lower and upper bounds on the capacity region of the general BC with cooperating receivers holding a $K$-cycle conference. Lastly, we consider the general BC with a single common message and cooperating receivers, for which we obtain explicit rate expressions without auxiliary RVs. We begin this study with the formal definition of the BC with cooperating decoders.

3.1 Definitions for the Broadcast Channel with Cooperating Decoders

Definition 3.1 A discrete broadcast channel is defined by a discrete input alphabet $\mathcal{X}$, two discrete output alphabets, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a probability mass function, $p(y_1, y_2|x)$, giving the probability distribution on $\mathcal{Y}_1 \times \mathcal{Y}_2$ for each $x \in \mathcal{X}$. We denote this channel by the triplet $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$. The broadcast channel is called memoryless if the probability mass function of a sequence of $n$ symbols is given by $p(y_1^n, y_2^n|x^n) = \prod_{i=1}^{n} p(y_{1,i}, y_{2,i}|x_i)$.

In the following we consider only discrete and memoryless broadcast channels.

Definition 3.2 The physically degraded broadcast channel is a broadcast channel in which the probability mass function can be decomposed as $p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1)$. Hence, for the physically degraded BC we have that $X - Y_1 - Y_2$ form a Markov chain.
Definition 3.3 A \((C_{12}, C_{21})\)-admissible \(K\)-cycle conference consists of the following elements:

1. \(K\) message sets from \(R_{x1}\) to \(R_{x2}\), denoted by \(W_{12}^{(1)}, W_{12}^{(2)}, \ldots, W_{12}^{(K)}\), and \(K\) message sets from \(R_{x2}\) to \(R_{x1}\), denoted by \(W_{21}^{(1)}, W_{21}^{(2)}, \ldots, W_{21}^{(K)}\). Message set \(W_{12}^{(k)}\) consists of \(2^{nR_{12}^{(k)}}\) messages and message set \(W_{21}^{(k)}\) consists of \(2^{nR_{21}^{(k)}}\) messages.

2. \(K\) mapping functions, one for each conference step from \(R_{x1}\) to \(R_{x2}\):

\[
h_{12}^{(k)} : \mathcal{Y}_1^n \times W_{21}^{(1)} \times W_{21}^{(2)} \times \ldots \times W_{21}^{(k-1)} \mapsto W_{12}^{(k)},
\]

and \(K\) mapping functions, one for each conference step from \(R_{x2}\) to \(R_{x1}\):

\[
h_{21}^{(k)} : \mathcal{Y}_2^n \times W_{12}^{(1)} \times W_{12}^{(2)} \times \ldots \times W_{12}^{(k-1)} \mapsto W_{21}^{(k)},
\]

where \(k = 1, 2, \ldots, K\).

The conference rates satisfy:

\[
C_{12} = \sum_{k=1}^{K} R_{12}^{(k)}, \quad C_{21} = \sum_{k=1}^{K} R_{21}^{(k)}.
\]

Definition 3.4 A \((2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, (C_{12}, C_{21}), K)\) code for the general broadcast channel with a common message and two private messages, consists of three sets of source messages, \(\mathcal{M}_0 = \{1, 2, \ldots, 2^{nR_0}\}\), \(\mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\}\) and \(\mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\}\), a mapping function at the transmitter,

\[
f : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \mapsto \mathcal{X}^n,
\]

A \((C_{12}, C_{21})\)-admissible \(K\)-cycle conference, and two decoders,

\[
g_1 : W_{21}^{(1)} \times W_{21}^{(2)} \times \ldots \times W_{21}^{(K)} \times \mathcal{Y}_1^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1, \quad (3.1)
\]

\[
g_2 : W_{12}^{(1)} \times W_{12}^{(2)} \times \ldots \times W_{12}^{(K)} \times \mathcal{Y}_2^n \mapsto \mathcal{M}_0 \times \mathcal{M}_2. \quad (3.2)
\]
Definition 3.5 The average probability of error for a code of length \( n \) for the broadcast channel is defined as the average probability that at least one of the receivers does not decode its message pair correctly:

\[
P^{(n)}_e = \Pr \left( g_1 \left( W_{21}^{(1)}, W_{21}^{(2)}, \ldots, W_{21}^{(K)}, Y_1^n \right) \neq (M_0, M_1) \right) \text{ or } g_2 \left( W_{12}^{(1)}, W_{12}^{(2)}, \ldots, W_{12}^{(K)}, Y_2^n \right) \neq (M_0, M_2) \right),
\]

where each message is selected uniformly and independently over its respective message set.

We also define the average probability of error for each receiver as:

\[
P^{(n)}_{e1} = \Pr \left( g_1 \left( W_{21}^{(1)}, W_{21}^{(2)}, \ldots, W_{21}^{(K)}, Y_1^n \right) \neq (M_0, M_1) \right), \tag{3.3}
\]

\[
P^{(n)}_{e2} = \Pr \left( g_2 \left( W_{12}^{(1)}, W_{12}^{(2)}, \ldots, W_{12}^{(K)}, Y_2^n \right) \neq (M_0, M_2) \right). \tag{3.4}
\]

By the union bound we have that \( \max \left\{ P^{(n)}_{e1}, P^{(n)}_{e2} \right\} \leq P^{(n)}_e \leq P^{(n)}_{e1} + P^{(n)}_{e2} \). Hence, \( P^{(n)}_e \to 0 \) implies that both \( P^{(n)}_{e1} \to 0 \) and \( P^{(n)}_{e2} \to 0 \), and when both individual error probabilities go to zero then \( P^{(n)}_e \) goes to zero as well.

Definition 3.6 A rate triplet \((R_0, R_1, R_2)\) is said to be achievable for the broadcast channel with a \( K \)-cycle conference, if for every \( \epsilon, \delta > 0 \) there exists a block length \( n \), such that a \((2^{n(R_0-\delta)}, 2^{n(R_1-\delta)}, 2^{n(R_2-\delta)}, n, (C_{12}, C_{21}), K)\) broadcast channel code with \( P^{(n)}_e \leq \epsilon \) can be constructed.

Definition 3.7 The capacity region of the discrete memoryless broadcast channel with cooperating receivers holding a \( K \)-cycle conference is the convex hull of all achievable rate triplets.
3.2 The Capacity Region of the Physically Degraded Broadcast Channel with Cooperating Receivers

We now consider the physically degraded broadcast channel with three independent messages: a private message to each receiver and a common message to both. We note that for the physically degraded channel, following the argument in [19, theorem 14.6.4], we can incorporate a common rate to both receivers by replacing $R_2$, the private rate to the bad receiver, obtained for the two private messages case with $R_0 + R_2$, where $R_0$ denotes the rate of the common information. Without cooperation, the capacity region of the physically degraded BC $X - Y_1 - Y_2$ given in [19, theorem 14.6.4], is the convex hull of all the rate triplets $(R_0, R_1, R_2)$ that satisfy

$$R_1 \leq I(X; Y_1 | U),$$  \hspace{1cm} (3.5)

$$R_0 + R_2 \leq I(U; Y_2),$$  \hspace{1cm} (3.6)

for some joint distribution $p(u)p(x|u)p(y_1|x)p(y_2|y_1)$, where

$$||U|| \leq \min \{||X||, ||Y_1||, ||Y_2||\}.$$  \hspace{1cm} (3.7)

Next, consider cooperation between the receivers over the physically degraded BC. First note that for this case, the link from $R_{x2}$ to $R_{x1}$ does not contribute to increasing the rates due to cooperation, and that only the link from $R_{x1}$ to $R_{x2}$ does. This is due to the data processing inequality (see [19, theorem 2.8.1]): since $X - Y_1 - Y_2$ form a Markov chain, any information about $X$ contained in $Y_2$ will also be contained in $Y_1$, and thus conferencing cannot help:

$$I(X; Y_1, Y_2) = I(X; Y_1) + I(X; Y_2 | Y_1) = I(X; Y_1).$$
This also implies that a single conference step from $R_{x1}$ to $R_{x2}$ is sufficient to obtain all the benefits of cooperation. For the rest of this section then, we shall consider only a conference link from the good receiver $R_{x1}$, to the bad receiver $R_{x2}$ (i.e. we set $C_{21} = 0$). This implies that $W_{21}$ is a constant and we can thus omit it from the analysis. We begin with a statement of the theorem:

**Theorem 3.1** The capacity region for sending independent information over the discrete memoryless physically degraded broadcast channel $X - Y_1 - Y_2$, with cooperating receivers having a noiseless conference link of capacity $C_{12}$, as defined in section 3.1, is the convex hull of all rate triplets $(R_0, R_1, R_2)$ that satisfy

$$R_1 \leq I(X; Y_1 | U),$$  \hfill (3.8)

$$R_0 + R_2 \leq \min\left( I(U; Y_1), I(U; Y_2) + C_{12} \right),$$  \hfill (3.9)

for some joint distribution $p(u)p(x|u)p(y_1|x)p(y_2|y_1)$, where the auxiliary random variable $U$ has cardinality bounded by $||U|| \leq \min\{||X||, ||Y_1||\}$.

We note that this result, presented in [106], was simultaneously derived in [86] for the case of a non-orthogonal relay-destination link.

### 3.2.1 Achievability Proof

In this subsection, we show that the rate triplets of theorem 3.1 are indeed achievable. We shall show that the region defined by (3.8) and (3.9) with $R_0 = 0$ is achievable. Incorporating $R_0 > 0$ easily follows as explained earlier.

**Overview of Coding Strategy**

The coding strategy is a combination of a broadcast code as an “outer” code used to split the rate between $R_{x1}$ and $R_{x2}$, and an “inner” relay code for $R_{x2}$, using
the code construction for the physically degraded relay channel, described in [15, theorem 1]. We first generate codewords $U^n$ for $R_{x2}$, according to the relay channel code construction. Then, the codewords for $R_{x2}$ are used as “cloud centers” for the codewords transmitted to $R_{x1}$ (which are also the input to the channel). Upon reception, $R_{x1}$ decodes both its own message and the message for $R_{x2}$, and then uses the relay code selection to select the message relayed to $R_{x2}$. $R_{x2}$ uses its received signal, $Y^n_{2}$, to generate a list of possible $U^n$ sequences, and then uses the information from $R_{x1}$ to resolve for the correct codeword.

Details of Coding Strategy

Code Generation

1. Consider first the set of $M_R = 2^{nC_{12}}$ relay messages. These are the messages that the relay $R_{x1}$ transmits to $R_{x2}$ through the noiseless finite capacity conference link between the two receivers. Index these messages by $s$, where $s \in \{1, 2, ..., M_R\}$.

Next, fix $p_U(u)$ and $p_{X|U}(x|u)$.

2. For each index $s \in [1, M_R]$, generate $2^{nR_2}$ conditionally independent codewords $u(w_2|s) \sim \prod_{i=1}^{n} p_U(u_i)$, where $w_2 \in \{1, 2, ..., 2^{nR_2}\}$.

3. For each codeword $u(w_2|s)$ generate $2^{nR_1}$ conditionally independent codewords $x(w_1, w_2|s) \triangleq x(w_1|u(w_2|s)) \sim \prod_{i=1}^{n} p_{X|U}(x_i|u_i(w_2|s))$, where $w_1 \in \{1, 2, ..., 2^{nR_1}\}$.

4. Randomly partition the message set for $R_{x2}$, $\{1, 2, ..., 2^{nR_2}\}$, into $M_R$ sets $\{S_1, S_2, ..., S_{M_R}\}$, by independently and uniformly assigning to each message an index in $[1, M_R]$. 

54
We note here that since the relay transmission does not affect the received signals \( y_1 \) and \( y_2 \), it is actually not necessary to generate \( M_R \) codebooks for \( u \) and this result can be proved using a single relay codebook. However, since we completely rely on the relay result of [15, theorem 1] to prove the achievable rate to \( R_{x2} \), we keep the construction similar.

**Encoding Procedure**

Consider the transmission of \( B \) blocks, each block transmitted using \( n \) channel symbols. Here we use \( nB \) symbol transmissions to transmit \( B - 1 \) message pairs \((w_1,i, w_2,i) \in [1, 2^{nR_1}] \times [1, 2^{nR_2}], i = 1, 2, \ldots, B - 1\). As \( B \to \infty \) we have that the rates \((R_1, R_2) \frac{B-1}{B} \to (R_1, R_2)\). Hence, any rate pair achievable without blocking can be approached arbitrarily close with blocking as well. Let \( w_{1,i} \) and \( w_{2,i} \) be the messages intended for \( R_{x1} \) and \( R_{x2} \) respectively, at the \( i \)'th block, and also assume that \( w_{2,i-1} \in S_{s_i} \). \( R_{x1} \) has an estimate \( \hat{w}_{2,i-1} \) of the message sent to \( R_{x2} \) at block \( i - 1 \). Let \( \hat{w}_{2,i-1} \in S_{\hat{s}_i} \). At the \( i \)'th block the transmitter outputs the codeword \( x(w_{1,i}, w_{2,i}|s_i) \), and \( R_{x1} \) sends the index \( \hat{s}_i \) to \( R_{x2} \) through the noiseless conference link.

**Decoding Procedure**

Assume first that up to the end of the \((i - 1)\)'th block there was no decoding error. Hence, at the end of the \((i - 1)\)'th block, \( R_{x1} \) knows \((w_{1,1}, w_{1,2}, \ldots, w_{1,i-1}), (w_{2,1}, w_{2,2}, \ldots, w_{2,i-1})\) and \((s_1, s_2, \ldots, s_i)\), and \( R_{x2} \) knows \((w_{2,1}, w_{2,2}, \ldots, w_{2,i-2})\) and \((s_1, s_2, \ldots, s_{i-1})\). The decoding at block \( i \) proceeds as follows:

1. \( R_{x1} \) knows \( s_i \) from \( w_{2,i-1} \). Hence, \( R_{x1} \) determines uniquely \((\hat{w}_{1,i}, \hat{w}_{2,i})\) s.t. \((u(\hat{w}_{2,i}|s_i), x(\hat{w}_{1,i}, \hat{w}_{2,i}|s_i), y_1(i)) \in A^{(n)}_i\). If there is none or there is more
than one, an error is declared.

2. $R_{x2}$ receives $s_i$ from $R_{x1}$. From knowledge of $s_{i-1}$ and $y_2(i-1)$, $R_{x2}$ forms a list of possible messages, $L(i-1) = \{ w_2 \in [1, 2^{nR_2}] : (y_2(i-1), u(w_2|s_{i-1})) \in A_e(n) \}$. Now, $R_{x2}$ uses $s_i$ to find a unique $\hat{w}_{2,i-1} \in S_{s_i} \cap L(i-1)$. If there is none or there is more than one, an error is declared.

**Analysis of the Probability of Error**

The achievable rate to $R_{x2}$ can be proved using the same technique as in [15, theorem 1]. For the ease of description assume that $R_{x1}$ transmits through an orthogonal channel to $R_{x2}$ and let $X'$ denote the channel input from $R_{x1}$ and $Y'$ the corresponding channel output to $R_{x2}$. Thus, $R_{x2}$ has a combined channel output $(Y_2, Y')$. The overall channel is given by

$$p(y_1, y_2, y'|x, x') = p(y_1, y_2|x)p(y'|x).$$

(3.10)

Additionally, we select the transition matrix $p(y'|x')$ and the input and output alphabets $X'$, $Y'$ such that the capacity of the orthogonal channel $X' - Y'$ is $C_{12}$. An example for such a selection is letting $X' = Y' = \{ 0, 1, \ldots, 2^{[C_{12}]} - 1 \}$, where $[\cdot]$ is denotes the ceil function. Letting $[a]$ denotes the integer part of the real number $a$, we set the channel transition function to be

$$p(y'|x') = \begin{cases} 1 - \alpha, & y' = x' \\ \alpha, & y' = \text{mod} \left( x' + 2^{[C_{12}]}, 2^{[C_{12}]} \right), \end{cases}$$

with $\alpha$ selected such that $H(Y'|X') = [C_{12}] - C_{12}$ (note that the probabilities of $Y'$ given $X'$ are the same, independent of the value of $X'$. Thus $H(Y'|X')$ has the same value independent of the distribution of $X'$). The capacity of this channel is

56
$C_{12}$ and is achieved by letting $p(x') = \frac{1}{2^{C_{12}}}, \forall x' \in X'$. As $n \to \infty$ we get that this setup is equivalent to the original setup described in section 1.4.3.

Now consider the rate to $R_{x2}$. The Markov chain $U \rightarrow X \rightarrow (Y_1, Y_2)$ combined with the transition function of (3.10) induces the following probability mass function

$$p(u, y_1, y_2, y', x') = p(y_1, y_2|u)p(y'|x')p(u, x').$$

Now, applying [15, theorem 1], with $p(u, x') = p(u)p(x')$, we have that

$$R_2 \leq \min \{I(U, X'; Y_2, Y'), I(U; Y_1|X')\}$$

$$= \min \{I(U, X'; Y') + I(U, X'; Y_2|Y'), I(U; Y_1)\}$$

$$= \min \{I(X'; Y') + I(U; Y'|X') + I(U; Y_2|Y') + I(X'; Y_2|Y', U), I(U; Y_1)\}$$

$$= \min \{C_{12} + I(U; Y_2), I(U; Y_1)\}.$$

Next, consider the rate to $R_{x1}$. From the proof of [15, theorem 1] we have that $R_{x1}$ decodes $W_2$. Therefore, $R_{x1}$ can now use successive decoding similar to the decoding at $R_{x1}$ in [19, ch. 14.6.2], which implies that the achievable rate to $R_{x1}$ is given by $R_1 \leq I(X; Y_1|U)$. Combining both bounds we get the rate constraints of theorem 3.1.

### 3.2.2 Converse Proof

In this section we prove that for a code with $P_e^{(n)} \to 0$, the rates must satisfy the constraints in theorem 3.1. First, note that for the case of the physically degraded broadcast channel with cooperating receivers we have the following Markov chain:

$$X^n - Y_1^n - (W_{12}(Y_1^n), Y_2^n). \quad (3.11)$$

Considering the definitions of the decoders in equations (3.1) and (3.2), and the definitions of the probability of error for each of the receivers in equations (3.3)
and (3.4), we have from Fano’s lemma ([19, ch. 2.11]) that

\[
H(W_1|Y_1^n) \leq P_{e_1}^{(n)} \log_2 \left( 2^{nR_1} - 1 \right) + H_b(P_{e_1}^{(n)}) \tag{3.12}
\]

\[\triangleq n\delta(P_{e_1}^{(n)}),\]

\[
H(W_2|Y_2^n, W_{12}(Y_1^n)) \leq P_{e_2}^{(n)} \log_2 \left( 2^{nR_2} - 1 \right) + H_b(P_{e_2}^{(n)}) \tag{3.13}
\]

\[\triangleq n\delta(P_{e_2}^{(n)}),\]

where \(H_b(P)\) is the entropy of a Bernoulli RV with parameter \(P\). Note that when \(P_{e_1}^{(n)} \to 0\) then \(\delta(P_{e_1}^{(n)}) \to 0\) and when \(P_{e_2}^{(n)} \to 0\) then \(\delta(P_{e_2}^{(n)}) \to 0\).

Now, for \(R_{x1}\) we have that

\[nR_1 = H(W_1) = I(W_1; Y_1^n) + H(W_1|Y_1^n).\]

Applying inequality (3.12), and then proceeding as in [44] we get the bound on \(R_1\) as

\[nR_1 \leq \sum_{k=1}^{n} I(X_k; Y_{1,k}|U_k) + n\delta(P_{e_1}^{(n)}),\]

where \(U_k \triangleq (Y_{1,1}, Y_{1,2}, \ldots, Y_{1,k-1}, W_2)\).

For \(R_{x2}\) we can write

\[nR_2 = H(W_2)\]

\[\leq I(W_2; Y_2^n, W_{12}(Y_1^n)) + n\delta(P_{e_2}^{(n)}) \tag{3.14}\]

\[= I(W_2; Y_2^n) + I(W_2; W_{12}(Y_1^n)|Y_2^n) + n\delta(P_{e_2}^{(n)}),\]

where the inequality in (a) is due to (3.13). Proceeding as in [44], we bound

\[I(W_2; Y_2^n) \leq \sum_{k=1}^{n} I(U_k; Y_{2,k}).\]

Next, we bound \(I(W_2; W_{12}(Y_1^n)|Y_2^n)\) as follows:

\[I(W_{12}(Y_1^n); W_2|Y_2^n) \leq H(W_{12}(Y_1^n)|Y_2^n)\]

\[\leq H(W_{12}(Y_1^n))\]

\[\leq nC_{12},\tag{3.15}\]

58
where the first inequality follows from the definition of mutual information, the second is due to removing the conditioning and the third is due to the admissibility of the conference. Combining both bounds we get that

\[ nR_2 \leq \sum_{k=1}^{n} I(U_k; Y_{2,k}) + nC_1 + n\delta(P_{e2}^{(n)}). \]  

(3.16)

The bound on \( R_2 \) can be developed in an alternative way. Begin with (3.14):

\[ nR_2 \leq I(W_2; Y_2^n, W_1(Y_1^n)) + n\delta(P_{e2}^{(n)}) \]

\[ \leq I(W_2; Y_2^n, Y_1^n) + n\delta(P_{e2}^{(n)}) \]

\[ = \sum_{k=1}^{n} I(W_2; Y_{1,k}, Y_{2,k}|Y_1^{k-1}, Y_2^{k-1}) + n\delta(P_{e2}^{(n)}), \]  

(3.17)

where (a) follows from the fact that \((W_1, W_2) - (Y_1^n, Y_2^n) - (W_12, Y_2^n)\) is a Markov relation combined with the data processing inequality. Next, we can write

\[ I(W_2; Y_{1,k}, Y_{2,k}|Y_1^{k-1}, Y_2^{k-1}) \]

\[ \overset{(a)}{=} I(W_2; Y_{1,k}|Y_1^{k-1}, Y_2^{k-1}) \]

\[ = H(Y_{1,k}|Y_1^{k-1}, Y_2^{k-1}) - H(Y_{1,k}|Y_1^{k-1}, Y_2^{k-1}, W_2) \]

\[ \overset{(b)}{\leq} H(Y_{1,k}) - H(Y_{1,k}|Y_1^{k-1}, Y_2^{k-1}, W_2) \]

\[ \overset{(c)}{=} H(Y_{1,k}) - H(Y_{1,k}|Y_1^{k-1}, W_2) \]

\[ = I(Y_{1,k}; Y_1^{k-1}, W_2) \]

\[ = I(Y_{1,k}; U_k), \]  

(3.18)

where the equality in (a) is due to the physical degradedness and memorylessness of the channel, (b) is due to removing the conditioning, and (c) is because the Markov chain makes \( Y_{1,k} \) independent of \( Y_2^{k-1} \) given \( Y_1^{k-1} \). Plugging this into (3.17), we obtain a second bound on \( R_2 \):

\[ nR_2 \leq \sum_{k=1}^{n} I(U_k; Y_{1,k}) + n\delta(P_{e2}^{(n)}). \]
Collecting the three bounds we have:

\begin{align*}
R_1 & \leq \frac{1}{n} \sum_{k=1}^{n} I(X_k; Y_{1,k}|U_k) + \delta(P^{(n)}_{e1}), \\
R_2 & \leq \frac{1}{n} \sum_{k=1}^{n} I(U_k; Y_{2,k}) + C_{12} + \delta(P^{(n)}_{e2}), \\
R_2 & \leq \frac{1}{n} \sum_{k=1}^{n} I(U_k; Y_{1,k}) + \delta(P^{(n)}_{e2}).
\end{align*}

(3.19) \hspace{1cm} (3.20) \hspace{1cm} (3.21)

Using the standard time-sharing argument as in [19, ch. 14.3], we can write the averages in (3.19)–(3.21) by introducing an appropriate time sharing variable, with cardinality upper bounded by 4. Therefore, if $P^{(n)}_{e1} \to 0$ and $P^{(n)}_{e2} \to 0$ as $n \to \infty$, the convex hull of this region can be shown to be equivalent to the convex hull of the region defined by

\begin{align*}
R_1 & \leq I(X; Y_1|U), \\
R_2 & \leq I(U; Y_2) + C_{12}, \\
R_2 & \leq I(U; Y_1).
\end{align*}

(3.22) \hspace{1cm} (3.23) \hspace{1cm} (3.24)

Finally, the bound on the cardinality of $U$ follows from the same arguments as in the converse for the non-cooperative case in [44]. Note however, that $||Y_2||$ is not included among the cardinalities in the minimum (cf. equation (3.7) for the non-cooperative case). The reason is that even when $||Y_2|| = 1$, information to $R_{x2}$ (represented by the random variable $U$), can be sent through the conference link between the two receivers.

\subsection{3.2.3 Discussion}

To illustrate the implications of theorem 3.1, consider the physically degraded binary symmetric broadcast channel (BSBC) depicted in figure 3.1. For this chan-
Figure 3.1: The physically degraded BSBC. $p_U$, $p_1$ and $p_2$ are the transition probabilities at the left, middle and right segments respectively.

Theorem 3.1 implies that $||U|| = 2$. Due to the symmetry of the channel, the probability distribution of $U$ which maximizes the rates, is a symmetric binary distribution, $\Pr(U = 0) = \Pr(U = 1) = \frac{1}{2}$. The resulting capacity region for this case is depicted in figure 3.2 for the case where $R_0 = 0$. In the figure, the bottom line (dash) is the non-cooperative capacity region, and the top line (dash-dot) is the maximum possible sum-rate, which requires that $C_{12} \geq H_b(p_{12}) - H_b(p_1)$, where

$$H_b(p) = -p \log_2(p) - (1 - p) \log_2(1 - p),$$

$$p_{12} = p_1(1 - p_2) + p_2(1 - p_1).$$

This maximum sum-rate of $I(X; Y_1)$ is obtained by summing the rate to $R_{x1}$ given by (3.22) and the maximum possible rate for $R_{x2}$ given by (3.24), and using the Markov chain relation $U - X - Y_1$. The middle line (solid) is the capacity region for the partial cooperation case where $0 < C_{12} < H_b(p_{12}) - H_b(p_1)$.

As can be seen from this example, the capacity region derived in this section is strictly larger than the capacity region for the non-cooperation case. Indeed, summing the constraints on $R_0$, $R_1$ and $R_2$ without cooperation (equations (3.5), (3.6)), results in a maximum achievable sum-rate of

$$R_0 + R_1 + R_2 \leq I(X; Y_1) - (I(U; Y_1) - I(U; Y_2)),$$

(3.25)

where the second term is always positive due to the Markov chain $U - X - Y_1 - Y_2$.
Figure 3.2: The capacity region for the physically degraded BSBC. Top, middle and bottom lines correspond to maximum possible cooperation, partial cooperation and no-cooperation scenarios respectively.

(assuming the degrading channel is non-invertible\textsuperscript{1}). In this setup, the maximum possible sum-rate, $I(X; Y_1)$, is achieved only when $U$ is a constant, and thus no information is sent to $R_{x2}$. When $R_0 + R_2 > 0$, because of the relationship $R_0 + R_2 \leq I(U; Y_2) < I(U; Y_1)$, we cannot achieve the maximum sum-rate of $I(X; Y_1)$ to $R_{x1}$. However, summing (3.23) or (3.24) with (3.22), results in a maximum achievable sum-rate with cooperating receivers of

$$R_0 + R_1 + R_2 \leq I(X; Y_1) + \min \{0, C_{12} - (I(U; Y_1) - I(U; Y_2))\}. \quad (3.26)$$

Comparing this to the non-cooperative sum-rate given by (3.25), it is clear that cooperation allows a net increase in the sum-rate of $C_{12}$, up to a maximum sum-rate of $I(X; Y_1)$.

\textsuperscript{1}It can be shown that $I(U; Y_1) - I(U; Y_2) = 0$ for the degraded channel setup implies that if $R_0 + R_2 > 0$ then $H(Y_1|Y_2) = 0$, i.e. the channel from $R_{x1}$ to $R_{x2}$ is invertible. Under these circumstances, this setup can be replaced by an equivalent setup in which both receivers get $Y_1$, but such a degraded setup is not interesting.
3.3 The Cooperative General Broadcast Channel with Two Private Messages and One Common Message

In this section we consider the cooperative broadcast scenario for the general broadcast channel depicted in figure 1.5. Previous work considered the single message case over the independent BC (i.e. \( p(y_1,y_2|x) = \prod_{i=1}^{n} p(y_{1,i}|x_i)p(y_{2,i}|x_i) \)) studied in [13], and the general BC with two private messages and a single cycle of conferencing studied in [82]. In the following we study the most general setup.

3.3.1 An Achievable Rate Region for the General Broadcast Channel with Cooperating Decoders Holding a \( K \)-Cycle Conference

For the classic general BC scenario, the best achievability result was derived by Marton in [57, theorem 2]. This result states that for the general BC, any rate triplet \((R_0, R_1, R_2)\) satisfying

\[
R_0 \leq \min \{I(W;Y_1), I(W;Y_2)\} \tag{3.27a}
\]

\[
R_0 + R_1 \leq I(W,U;Y_1) \tag{3.27b}
\]

\[
R_0 + R_2 \leq I(W,V;Y_2) \tag{3.27c}
\]

\[
R_0 + R_1 + R_2 \leq \min \{I(W;Y_1), I(W;Y_2)\} + I(U;Y_1|W) + I(V;Y_2|W) - I(U;V|W), \tag{3.27d}
\]

for some joint distribution \( p(w,u,v,x,y_1,y_2) = p(w,u,v)p(y_1,y_2|x) \), is achievable\(^2\).

\(^2\)Marton’s original region considers only the two private rates \( R_1 \) and \( R_2 \), but it can be easily extended to incorporate a common rate \( R_0 \).
We now consider cooperation between the receivers. We first present a general result for the cooperative broadcast scenario with a $K$-cycle conference, and then consider special cases. Denote with $\hat{Y}_1 = (\hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)})$ and $\hat{Y}_2 = (\hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K)})$. Let $R_1$ and $R_2$ be the private rates to $R_{x_1}$ and $R_{x_2}$ respectively, and let $R_0$ denote the rate of the common information. Then, the following rate triplets are achievable:

**Theorem 3.2** Consider the general broadcast channel $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ with cooperating receivers having noiseless conference links of finite capacities $C_{12}$ and $C_{21}$ between them. Let the receivers hold a conference that consists of $K$ cycles. Then, any rate triplet $(R_0, R_1, R_2)$ satisfying

\[
R_0 \leq \min \left\{ I \left( W; Y_1, \hat{Y}_2 \right), I \left( W; \hat{Y}_1, Y_2 \right) \right\} \tag{3.28a}
\]

\[
R_0 + R_1 \leq I (W, U; Y_1, \hat{Y}_2) \tag{3.28b}
\]

\[
R_0 + R_2 \leq I (W, V; \hat{Y}_1, Y_2) \tag{3.28c}
\]

\[
R_0 + R_1 + R_2 \leq \min \left\{ I \left( W; Y_1, \hat{Y}_2 \right), I \left( W; \hat{Y}_1, Y_2 \right) \right\} + I(U; Y_1, \hat{Y}_2|W) + I(V; \hat{Y}_1, Y_2|W) - I(U; V|W) \tag{3.28d}
\]

subject to,

\[
C_{12} \geq I(Y_1; \hat{Y}_1, \hat{Y}_2|Y_2) \tag{3.29a}
\]

\[
C_{21} \geq I(Y_2; \hat{Y}_2, \hat{Y}_1|Y_1) \tag{3.29b}
\]
for some joint distribution

\[
p(w, u, v, x, y_1, y_2, \hat{y}_1^{(1)}, \ldots, \hat{y}_1^{(K)}, \hat{y}_2^{(1)}, \ldots, \hat{y}_2^{(K)}) = \\
p(w, u, v, x)p(y_1, y_2|x)p(\hat{y}_1^{(1)}|y_1) p(\hat{y}_2^{(1)}|y_2, \hat{y}_1^{(1)}) \cdots \\
\times p(\hat{y}_1^{(K)}|y_1, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, \ldots, \hat{y}_1^{(K-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, \ldots, \hat{y}_2^{(K-1)}) \\
\times p(\hat{y}_2^{(K)}|y_2, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, \ldots, \hat{y}_1^{(K-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, \ldots, \hat{y}_2^{(K-1)}) \cdots \\
\times p(\hat{y}_2^{(1)}|y_2, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, \ldots, \hat{y}_1^{(K)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, \ldots, \hat{y}_2^{(K-1)}) \cdot \cdot \cdot
\]

is achievable. The cardinalities of the \(k\)'th auxiliary random variables are bounded by:

\[
||\hat{Y}_1^{(k)}|| \leq ||Y_1|| \times \prod_{l=1}^{k-1} ||\hat{Y}_1^{(l)}|| \times \prod_{l=1}^{k-1} ||\hat{Y}_2^{(l)}|| + 1, \quad k = 1, 2, \ldots, K
\]

\[
||\hat{Y}_2^{(k)}|| \leq ||Y_2|| \times \prod_{l=1}^{k} ||\hat{Y}_1^{(l)}|| \times \prod_{l=1}^{k-1} ||\hat{Y}_2^{(l)}|| + 1, \quad k = 1, 2, \ldots, K,
\]

and

\[
||W|| \leq \min \{||X||, ||Y_1|| \cdot ||Y_2||\}.
\]

Proof

Overview of Coding Strategy

The coding strategy is based on combining the BC code construction of [58], after incorporating the common information into the construction, with the \(K\)-cycle conference of [84]. The transmitter constructs a broadcast code that splits the rate between the three message sets. This is done independently of the relaying scheme. Each receiver generates its conference messages according to the construction of [84]. After \(K\) cycles of conferencing each receiver decodes its information
based on its channel output and the conference messages received from the other receiver. The fact that the channel encoding and the relay operation are performed independently, allows to easily create a hybrid coding scheme.

In the proof we let $W$ represent only the common information (i.e. $R_0 = \min \{ I(W; Y_1, \hat{Y}_2), I(W; \tilde{Y}_1, Y_2) \}$). The case where $W$ contains also private information (i.e. information that can be decoded by both receivers but is intended only for one of them) follows in a straightforward manner. Thus the individual rate bounds are $I(U; Y_1, \hat{Y}_2 | W)$ for $R_1$ and $I(V; \tilde{Y}_1, Y_2 | W)$ for $R_2$.

**Code Construction at The Transmitter**

- Fix all the distributions in (3.30). Fix $\epsilon > 0$ and let $\delta > 0$ be a positive number whose value is determined in the following steps. Let $S_{W|\delta}^{(n)}$ denote the set of all sequences $w \in W^n$ such that $w \in A_{\delta}^{*}(W)$ and $A_{\delta}^{*}(U, V|w)$ is non-empty, as defined in [105, corollary 5.11]. From [105, corollary 5.11] we have that $||S_{W|\delta}^{(n)}|| \geq 2^{n(H(W) - \phi)}$, where $\phi \to 0$ as $\delta \to 0$ and $n \to \infty$.

- Pick $2^{nR_0}$ sequences from $S_{W|\delta}^{(n)}$ in a uniform and independent manner according to

$$\Pr(w) = \begin{cases} \frac{1}{||S_{W|\delta}^{(n)}||}, & w \in S_{W|\delta}^{(n)} \\ 0, & \text{otherwise.} \end{cases}$$

Label these sequences with $l \in M_0 \triangleq \{1, 2, ..., 2^{nR_0}\}$.

- For each sequence $w(l)$, $l \in M_0$, consider the set $A_{\delta'}^{*}(U|w(l))$, $\delta' = \delta \max \{||U||, ||V||\}$. Since the sequences $w(l) \in S_{W|\delta}^{(n)}$ are such that $A_{\delta}^{*}(U, V|w(l))$ is non-empty and since $(u, v) \in A_{\delta}^{*}(U, V|w(l))$ implies $u \in A_{\delta'}^{*}(U|w(l))$, then also $A_{\delta'}^{*}(U|w(l))$ in non-empty, and by [105, theorem 5.9], $||A_{\delta'}^{*}(U|w(l))|| \geq 2^{n(H(U|W) - \psi)}$, $\psi \to 0$ as $\delta' \to 0$ and $n \to \infty$. 

66
For each $l \in \mathcal{M}_0$ pick $2^{n(I(U; Y_1, Y_2|W) - \epsilon)}$ sequences in a uniform and independent manner from $A_{\delta'}^{(n)}(U|w(l))$ according to

$$\Pr(u|l) = \begin{cases} \frac{1}{||A_{\delta'}^{(n)}(U|w(l))||}, & u \in A_{\delta'}^{(n)}(U|w(l)) \\ 0, & \text{otherwise.} \end{cases}$$

Label these sequences with $u(m|l)$, $m \in \mathcal{Z}_1 \triangleq \{1, 2, \ldots, 2^{n(I(U; Y_1, Y_2|W) - \epsilon)}\}$. Similarly, pick $2^{n(I(V; Y_1, Y_2|W) - \epsilon)}$ sequences in a uniform and independent manner from $A_{\delta'}^{(n)}(V|w(l))$ according to

$$\Pr(v|l) = \begin{cases} \frac{1}{||A_{\delta'}^{(n)}(V|w(l))||}, & v \in A_{\delta'}^{(n)}(V|w(l)) \\ 0, & \text{otherwise.} \end{cases}$$

Label these sequences with $v(j|l)$, $j \in \mathcal{Z}_2 \triangleq \{1, 2, \ldots, 2^{n(I(V; Y_1, Y_2|W) - \epsilon)}\}$. $\delta$ is selected such that $\forall l \in \mathcal{M}_0$ we have that $||A_{\delta'}^{(n)}(U|w(l))|| \geq 2^{n(I(U; Y_1, Y_2|W) - \epsilon)}$ and $||A_{\delta'}^{(n)}(V|w(l))|| \geq 2^{n(I(V; Y_1, Y_2|W) - \epsilon)}$.

- Partition the set $\mathcal{Z}_1$ into $2^{nR_1}$ subsets $B_{w_1}$, $w_1 \in \mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\}$ and let $B_{w_1} = (w_1 - 1)2^{n(I(U; Y_1, Y_2|W) - R_1 - \epsilon)} + 1, w_12^{n(I(U; Y_1, Y_2|W) - R_1 - \epsilon)}$. Similarly partition the set $\mathcal{Z}_2$ into $2^{nR_2}$ subsets $C_{w_2}$, $w_2 \in \mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\}$ and $C_{w_2} = (w_2 - 1)2^{n(I(V; Y_1, Y_2|W) - R_2 - \epsilon)} + 1, w_22^{n(I(V; Y_1, Y_2|W) - R_2 - \epsilon)}$.

- For each triplet $(l, w_1, w_2)$ consider the set

$$\mathcal{D}(w_1, w_2|l) \triangleq \{(m_1, m_2) : m_1 \in B_{w_1}, m_2 \in C_{w_2}, (u(m_1|l), v(m_2|l)) \in A_{\delta'}^{(n)}(U, V|w(l))\}.$$ 

By [58, lemma on pg. 121], we have that taking $n$ large enough we can make
For each \( l \in M_0 \), we pick a unique pair of \((m_1(w_1, w_2, l), m_2(w_1, w_2, l))\) \(\in D(w_1, w_2|l)\), \((w_1, w_2)\) \(\in M_1 \times M_2\). The transmitter generates the codeword \(x(l, w_1, w_2)\) according to
\[
p(x(l, w_1, w_2)) = \prod_{i=1}^{n} p_{x|U,V,W}(x_i|u_i(m_1(w_1, w_2, l)), v_i(m_2(w_1, w_2, l)), w_i(l)).
\]

When transmitting the triplet \((l, w_1, w_2)\) the transmitter outputs \(x(l, w_1, w_2)\).

### Codebook Generation at the Receivers

- For the first conference step from \(R_{x1}\) to \(R_{x2}\), \(R_{x1}\) generates a codebook with \(2^{nR_{12}^{(1)}}\) codewords indexed by \(z_{12}^{(1)} \in Z_{12}^{(1)} = \{1, 2, ..., 2^{nR_{12}^{(1)}}\}\) according to the distribution \(p_{\hat{y}_1^{(1)}}(\hat{y}_1^{(1)}): p(\hat{y}_1^{(1)}(z_{12}^{(1)})) = \prod_{i=1}^{n} p_{\hat{y}_1^{(1)}}(\hat{y}_1^{(1)}(z_{12}^{(1)})).\) \(R_{x1}\) uniformly and independently partitions the message set \(Z_{12}^{(1)}\) into \(2^{nR_{12}^{(1)}}\) subsets indexed by \(w_{12}^{(1)} \in W_{12}^{(1)} = \{1, 2, ..., 2^{nR_{12}^{(1)}}\}\). Denote these subsets with \(S_{12,w_{12}^{(1)}}^{(1)}\).

- For the first conference step from \(R_{x2}\) to \(R_{x1}\), \(R_{x2}\) generates a codebook with \(2^{nR_{21}^{(1)}}\) codewords indexed by \(z_{21}^{(1)} \in Z_{21}^{(1)} = \{1, 2, ..., 2^{nR_{21}^{(1)}}\}\), for each codeword \(\hat{y}_1^{(1)}(z_{12}^{(1)}), z_{12}^{(1)} \in Z_{12}^{(1)}\). Each codebook is generated in an i.i.d. manner according to \(p(\hat{y}_2^{(1)}(z_{21}^{(1)}|z_{12}^{(1)}) = \prod_{i=1}^{n} p(\hat{y}_{2,i}^{(1)}(z_{21}^{(1)}|z_{12}^{(1)})|\hat{y}_{1,i}^{(1)}(z_{12}^{(1)}))\). \(R_{x2}\) uniformly and independently partitions the message set \(Z_{21}^{(1)}\) into \(2^{nR_{21}^{(1)}}\).
subsets indexed by $w_{21}^{(1)} \in \mathcal{W}_{21}^{(1)} = \{1, 2, \ldots, 2^{nR_{21}^{(1)}}\}$. Denote these subsets with $S_{21,v_{21}^{(1)}}^{(1)}$.

- For the $k$'th conference step from $R_{x2}$ to $R_{x1}$ considers each combination of $z_{12}^{(1)}, z_{12}^{(2)}, \ldots, z_{12}^{(k-1)}, z_{21}^{(1)}, z_{21}^{(2)}, \ldots, z_{21}^{(k-1)}$. For each combination, $R_{x1}$ generates a codebook with $2^{nR_{12}^{(k)}}$ codewords indexed by $z_{12}^{(k)} \in \mathcal{Z}_{12}^{(k)} = \{1, 2, \ldots, 2^{nR_{12}^{(k)}}\}$, according to the distribution $p\left(\hat{y}_{1}^{(k)} | y_{1}^{(1)}, \hat{y}_{1}^{(2)}, \ldots, \hat{y}_{1}^{(k-1)}, y_{2}^{(1)}, \hat{y}_{2}^{(2)}, \ldots, \hat{y}_{2}^{(k-1)}\right)$. $R_{x1}$ uniformly and independently partitions the message set $\mathcal{Z}_{12}^{(k)}$ into $2^{nR_{12}^{(k)}}$ subsets indexed by $w_{12}^{(k)} \in \mathcal{W}_{12}^{(k)} = \{1, 2, \ldots, 2^{nR_{12}^{(k)}}\}$. Denote these subsets with $S_{12,v_{12}^{(k)}}^{(k)}$.

- The codebook for the $k$'th conference step from $R_{x2}$ to $R_{x1}$ is generated in a parallel manner for each combination of $\hat{z}_{12}^{(1)}, \hat{z}_{12}^{(2)}, \ldots, \hat{z}_{12}^{(k)}, \hat{z}_{21}^{(1)}, \hat{z}_{21}^{(2)}, \ldots, \hat{z}_{21}^{(k-1)}$.

Decoding and Encoding at $R_{x1}$ at the $k$'th Conference Cycle ($k \leq K$) at Transmission Block $i$

$R_{x1}$ needs first to decode the message $z_{21}^{(k-1)}$ sent from $R_{x2}$ at the $(k - 1)$'th cycle. To that end, $R_{x1}$ uses $w_{21}^{(k-1)}$, the index received from $R_{x2}$ at the $(k - 1)$'th conference step. In decoding $z_{21}^{(k-1)}$ we assume that all the previous $\hat{z}_{21}^{(1)}, \hat{z}_{21}^{(2)}, \ldots, \hat{z}_{21}^{(k-2)}$ were correctly decoded at $R_{x1}$. We denote the $\hat{y}_{2}^{(k)}$ sequences corresponding to $z_{21}^{(1)}, z_{21}^{(2)}, \ldots, z_{21}^{(k-2)}$ by $\hat{y}_{2}(1), \hat{y}_{2}(2), \ldots, \hat{y}_{2}(k-2)$, and similarly define $\hat{y}_{1}(1), \hat{y}_{1}(2), \ldots, \hat{y}_{1}(k-1)$.

- $R_{x1}$ first generates the set $\mathcal{L}_{1}(k-1)$ defined by:

$$
\mathcal{L}_{1}(k-1) = \left\{ z_{21}^{(k-1)} \in \mathcal{Z}_{21}^{(k-1)} : \left( \hat{y}_{2}^{(k-1)}(z_{21}^{(k-1)} | \hat{z}_{21}^{(1)}, \hat{z}_{12}^{(2)}, \ldots, \hat{z}_{12}^{(k-1)}, \hat{z}_{12}^{(1)}, \hat{z}_{21}^{(2)}, \ldots, \hat{z}_{21}^{(k-2)}), \hat{y}_{1}(1), \hat{y}_{1}(2), \ldots, \hat{y}_{1}(k-1), \hat{y}_{2}(1), \hat{y}_{2}(2), \ldots, \hat{y}_{2}(k-2), y_{1}(i) \right) \in A_{n}^{(n)} \right\}.
$$
• $R_{x1}$ then looks for a unique $z_{21}^{(k-1)} \in Z_{21}^{(k-1)}$ such that $z_{21}^{(k-1)} \in L_1(k-1) \cap S_{21,w_{21}^{(k-1)}}$. If there is none or there is more than one, an error is declared.

• From an argument similar to [84], the probability of error can be made arbitrarily small by taking $n$ large enough as long as

$$R_{21}^{(k-1)} < I \left( \hat{y}_2^{(k-1)} ; Y_1 | \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(k-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(k-2)} \right) + R_{21}^{(k-1)} - \epsilon.$$  

Here, $k > 1$, since for the first conference message from $R_{x1}$ to $R_{x2}$ no decoding takes place.

In generating the $k$’th conference message to $R_{x2}$, it is assumed that all the previous $k-1$ messages from $R_{x2}$ were decoded correctly.

• $R_{x1}$ looks for a message $z_{12}^{(k)} \in Z_{12}^{(k)}$ such that

$$R_{12}^{(k)} > I \left( \hat{y}_1^{(k)} ; Y_1 | \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(k-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(k-1)} \right) + \epsilon.$$  

• $R_{x1}$ looks for the partition of $Z_{12}^{(k)}$ into which $z_{12}^{(k)}$ belongs. Denote the index of this partition with $w_{12}^{(k)}$.

• $R_{x1}$ transmits $w_{12}^{(k)}$ to $R_{x2}$ through the conference link.
Decoding and Encoding at \( R_{x2} \) at the \( k \)'th Conference Cycle \((k \leq K)\) at Transmission Block \(i\)

Using similar arguments to the previous section, we obtain the following rate constraints:

- Decoding \( z_{12}^{(k)} \) at \( R_{x2} \) can be done with an arbitrarily small probability of error by taking \( n \) large enough as long as

\[
R_{12}^{(k)} < I \left( \hat{Y}_1^{(k)}; Y_2 \left| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) + R_{12}^{(k)} - \epsilon.
\]

- Encoding \( z_{21}^{(k)} \) can be done with an arbitrarily small probability of error by taking \( n \) large enough as long as

\[
R_{21}^{(k)} > I \left( \hat{Y}_2^{(k)}; Y_2 \left| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) + \epsilon.
\]

Combining All Bounds on Conference Transmission Rates

First consider the bounds on \( R_{12}^{(k)} \), \( k = 1, 2, ..., K \):

\[
I \left( \hat{Y}_1^{(k)}; Y_1 \left| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) + \epsilon < R_{12}^{(k)} < \\
I \left( \hat{Y}_1^{(k)}; Y_2 \left| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) + \epsilon
\]

This can be satisfied only if

\[
I \left( \hat{Y}_1^{(k)}; Y_2 \left| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) + R_{12}^{(k)} - \epsilon > \\
I \left( \hat{Y}_1^{(k)}; Y_1 \left| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) + \epsilon
\]

\[
\Rightarrow R_{12}^{(k)} > H \left( \hat{Y}_1^{(k)} \left| Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) \\
- H \left( \hat{Y}_1^{(k)} \left| Y_1, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)} \right. \right) + 2\epsilon
\]
Hence

\[ C_{12} = \sum_{k=1}^{K} R_{12}^{(k)} \]
\[ \geq \sum_{k=1}^{K} \left( I \left( \hat{Y}_{1}^{(k)}; Y_{1}, \hat{Y}_{1}^{(1)}, \hat{Y}_{1}^{(2)}, \ldots, \hat{Y}_{1}^{(k-1)}, \hat{Y}_{2}^{(1)}, \hat{Y}_{2}^{(2)}, \ldots, \hat{Y}_{2}^{(k-1)} \right) + 2\epsilon \right) \]
\[ = \sum_{k=1}^{K} \left[ I \left( \hat{Y}_{1}^{(k)}; Y_{1}, \hat{Y}_{1}^{(1)}, \hat{Y}_{1}^{(2)}, \ldots, \hat{Y}_{1}^{(k-1)}, \hat{Y}_{2}^{(1)}, \hat{Y}_{2}^{(2)}, \ldots, \hat{Y}_{2}^{(k-1)} \right) \right. \]
\[ + I \left( \hat{Y}_{2}^{(k)}; Y_{1}, \hat{Y}_{1}^{(1)}, \hat{Y}_{1}^{(2)}, \ldots, \hat{Y}_{1}^{(k-1)}, \hat{Y}_{2}^{(1)}, \hat{Y}_{2}^{(2)}, \ldots, \hat{Y}_{2}^{(k-1)} \right) \] \[ + 2K\epsilon \]
\[ = \sum_{k=1}^{K} I \left( \hat{Y}_{1}^{(k)}, \hat{Y}_{2}^{(k)}; Y_{1}, \hat{Y}_{1}^{(1)}, \hat{Y}_{1}^{(2)}, \ldots, \hat{Y}_{1}^{(k-1)}, \hat{Y}_{2}^{(1)}, \hat{Y}_{2}^{(2)}, \ldots, \hat{Y}_{2}^{(k-1)} \right) + 2K\epsilon \]
\[ = I \left( \hat{Y}_{1}^{(1)}, \hat{Y}_{2}^{(1)}, \ldots, \hat{Y}_{1}^{(K)}, \hat{Y}_{2}^{(1)}, \hat{Y}_{2}^{(2)}, \ldots, \hat{Y}_{2}^{(K)}; Y_{1}, Y_{2} \right) + 2K\epsilon, \quad (3.32) \]

and similarly

\[ C_{21} \geq I \left( \hat{Y}_{1}^{(1)}, \hat{Y}_{2}^{(1)}, \ldots, \hat{Y}_{1}^{(K)}, \hat{Y}_{2}^{(1)}, \hat{Y}_{2}^{(2)}, \ldots, \hat{Y}_{2}^{(K)}; Y_{1}, Y_{2} \right) + 2K\epsilon. \quad (3.33) \]

This provides the rate constraints on the conference auxiliary variables of (3.29a) and (3.29b).

**Decoding at** $R_{x1}$

$R_{x1}$ uses $y_{1}(i)$ and $\hat{y}_{2}^{(1)}, \hat{y}_{2}^{(2)}, \ldots, \hat{y}_{2}^{(K)}$ received from $R_{x2}$ at block $i$, to decode $(l_{i}, w_{1,i})$ as follows:

- $R_{x1}$ first looks for a unique message $l \in M_{0}$ such that

\[ \left( w(l), y_{1}(i), \hat{y}_{2}^{(1)}, \hat{y}_{2}^{(2)}, \ldots, \hat{y}_{2}^{(K)} \right) \in A_{\epsilon(n)}. \]

From the point-to-point channel capacity theorem (see [58]), this can be done with an arbitrarily small probability of error by taking $n$ large enough as long as

\[ R_{0} < I(W; Y_{1}, \hat{Y}_{2}) - \epsilon. \quad (3.34) \]
Denote the decoded message $\hat{l}_i$.

- Next, $R_{x_1}$ decodes $w_{1,i}$ by looking for a unique $m \in \mathbb{Z}_1$ such that
  \[
  (u(m|\hat{l}_i), w(\hat{l}_i), y_1(i), \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(K)}) \in A^{*}_{\epsilon(n)}.
  \]
  If a unique such $m$ exists, denote the decoded index with $\hat{m}$. Now $R_{x_1}$ looks for the partition of $\mathbb{Z}_1$ into which $\hat{m}$ belongs and sets $\hat{w}_{1,i}$ to be the index of that partition: $\hat{m} \in B_{\hat{w}_{1,i}}$. Similarly to the proof in [19, ch. 14.6.2], assuming successful decoding of $l_i$, the probability of error for decoding $w_{1,i}$ can be made arbitrarily small by taking $n$ large enough as long as
  \[
  \frac{1}{n} \log_2 ||\mathbb{Z}_1|| \leq I(U; Y_1, \hat{Y}_2|W),
  \]
  which is satisfied by construction. Finally, combining (3.31a) with (3.34), gives (3.28b).

**Decoding at $R_{x_2}$**

Repeating similar steps for decoding at $R_{x_2}$ we get that decoding $l_i$ can be done with an arbitrarily small probability of error by taking $n$ large enough as long as
  \[
  R_0 < I(W; \hat{Y}_1, Y_2) - \epsilon,
  \]
  and assuming successful decoding of $l_i$, decoding $w_{2,i}$ with an arbitrarily small probability of error requires that
  \[
  \frac{1}{n} \log_2 ||\mathbb{Z}_2|| \leq I(V; \hat{Y}_1, Y_2|W),
  \]
  which again is satisfied by construction. Combining (3.31b) and (3.35) gives (3.28c).
Finally, combining (3.34), (3.35) and (3.31c) gives the common rate constraint of (3.28a) and the sum-rate constraint of (3.28d). Equations (3.32) and (3.33) give the conference rate constraints of the theorem.

3.3.2 An Upper Bound

Proposition 3.1 Assume the broadcast channel setup of theorem 3.2. Then, for sending three independent messages, any achievable rate triplet \((R_0, R_1, R_2)\) must satisfy

\[
R_0 + R_1 \leq I(X; Y_1) + C_{21},
\]

\[
R_0 + R_2 \leq I(X; Y_2) + C_{12},
\]

\[
R_0 + R_1 + R_2 \leq I(X; Y_1, Y_2),
\]

for some distribution \(p(x)\) on \(X\).

Proof

The proof uses the cut-set bound [19, theorem 14.10.1]. First we define an equivalent system by introducing two orthogonal channels, \(X'_2 - Y'_1\) from \(R_{x2}\) to \(R_{x1}\) and \(X'_1 - Y'_2\) from \(R_{x1}\) to \(R_{x2}\). The joint probability distribution function then becomes

\[
p((y_1, y'_1); (y_2, y'_2)|x, x'_1, x'_2) = p(y_1, y_2|x)p(y'_1|x'_2)p(y'_2|x'_1),
\]

where the signal received at \(R_{x1}\) is \((Y_1, Y'_1)\) and the signal received at \(R_{x2}\) is \((Y_2, Y'_2)\). As in the proof in section 3.2.1, we select \(X'_1, X'_2, Y'_1, Y'_2, p(x'_1), p(x'_2), p(y'_1|x'_2)\) and \(p(y'_2|x'_1)\) such that the capacities of the channels \(X'_2 - Y'_1\) and \(X'_1 - Y'_2\) are \(C_{21}\) and \(C_{12}\) respectively. Additionally, the codewords for the conference channels are generated independently of each other and of the source codebook, hence we set
\( p(x, x'_1, x'_2) = p(x)p(x'_1)p(x'_2) \). Now, from the cut-set bound, letting the transmitter and \( R_{x_2} \) form one group and \( R_{x_1} \) the second group, we have

\[
R_0 + R_1 \leq I(X, X'_2; Y_1, Y'_1 | X'_1)
\]

\[
= I(X'_2; Y_1, Y'_1 | X'_1) + I(X; Y_1, Y'_1 | X'_1, X'_2)
\]

\[
= I(X'_2; Y'_1 | X'_1) + I(X'_2; Y_1 | X'_1, Y'_1) + I(X; Y'_1 | X'_1, X'_2)
\]

\[
+ I(X; Y_1 | X'_1, X'_2, Y'_1)
\]

\[
= I(X'_2; Y'_1) + I(X; Y_1)
\]

\[
= C_{21} + I(X; Y_1),
\]

where \( I(X'_2; Y_1 | X'_1, Y'_1) = I(X; Y'_1 | X'_1, X'_2) = 0 \) follows from direct application of the probability chain. Similarly we obtain the rate constraint on \( R_0 + R_2 \). Lastly, for the sum-rate let the transmitter form one group and the receivers form the second. Then, the cut-set bound results in

\[
R_0 + R_1 + R_2 \leq I(X; Y_1, Y_2, Y'_1, Y'_2 | X'_1, X'_2)
\]

\[
= I(X; Y_1, Y_2 | X'_1, X'_2) + I(X; Y'_1, Y'_2 | X'_1, X'_2, Y_1, Y_2)
\]

\[
= I(X; Y_1, Y_2),
\]

yielding the last constraint in the proposition. \( \blacksquare \)

### 3.3.3 Special Cases

Consider the case of a single cycle \((K = 1)\). Setting \( W \) to be a constant in theorem 3.2 we get the following corollary:

**Corollary 3.1** Let \((X, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)\) be a discrete memoryless broadcast channel, with cooperating receivers having noiseless conference links of finite capacities \( C_{12} \) and \( C_{21} \), as defined in section 3.1. Then, for sending independent
information, any rate pair \((R_1, R_2)\) satisfying

\[
R_1 \leq I(U; Y_1, \hat{Y}_2), \tag{3.36a}
\]

\[
R_2 \leq I(V; Y_2, \hat{Y}_1), \tag{3.36b}
\]

\[
R_1 + R_2 \leq I(U; Y_1, \hat{Y}_2) + I(V; Y_2, \hat{Y}_1) - I(U; V), \tag{3.36c}
\]

subject to,

\[
C_{21} \geq I(\hat{Y}_2; Y_2) - I(\hat{Y}_2; Y_1), \tag{3.37a}
\]

\[
C_{12} \geq I(\hat{Y}_1; Y_1) - I(\hat{Y}_1; Y_2), \tag{3.37b}
\]

for some joint distribution \(p(u, v, x, y_1, y_2, \hat{u}, \hat{v}) = p(u, v, x)p(y_1, y_2|x)p(\hat{y}_2|y_2)p(\hat{y}_1|y_1)\), is achievable, with \(u \in U, v \in V, \hat{y}_2 \in \hat{Y}_2, \hat{y}_1 \in \hat{Y}_1, \|\hat{Y}_2\| \leq \|Y_2\| + 1\) and \(\|\hat{Y}_1\| \leq \|Y_1\| + 1\).

We now examine the achievable rates for different values of \(C_{12}\) and \(C_{21}\).

**No Cooperation:** \(C_{12} = C_{21} = 0\)

Consider first cooperation from \(R_{x2}\) to \(R_{x1}\). Setting \(C_{21} = 0\) in corollary 3.1 implies that

\[
H(\hat{Y}_2|Y_1) = H(\hat{Y}_2|Y_2). \tag{3.38}
\]

From equation (3.36a), the constraint on \(R_1\) can be written in the form

\[
R_1 \leq I(U; Y_1) + I(U; \hat{Y}_2|Y_1). \tag{3.39}
\]

Now we find \(I(U; \hat{Y}_2|Y_1)\):

\[
I(U; \hat{Y}_2|Y_1) = H(\hat{Y}_2|Y_1) - H(\hat{Y}_2|Y_1, U)
\]

\[
\overset{(a)}{=} H(\hat{Y}_2|Y_2) - H(\hat{Y}_2|Y_1, U)
\]

\[
\overset{(b)}{=} H(\hat{Y}_2|Y_2, Y_1, U) - H(\hat{Y}_2|Y_1, U)
\]

\[
= -I(\hat{Y}_2; Y_2|Y_1, U).
\]

76
where (a) is due to (3.38), and (b) is due to the Markov chain $U - (U, V) - X - (Y_1, Y_2) - Y_2 - \hat{Y}_2$, which implies that given $Y_2$, $\hat{Y}_2$ is independent of $Y_1$ and $U$. Now, since mutual information is non-negative, we conclude that $I(U; \hat{Y}_2 | Y_1) = 0$. Hence, the rate constraint on $R_1$ becomes

$$ R_1 \leq I(U; Y_1). $$

Similarly, the maximum rate $R_2$ is given by $I(V; Y_2)$, and in conclusion when $C_{12} = C_{21} = 0$ we resort back to the rate region without cooperation derived in [57, theorem 2] (with a constant $W$).

**Full Cooperation**: $C_{12} = H(Y_1 | Y_2), C_{21} = H(Y_2 | Y_1)$

When $C_{12} = H(Y_1 | Y_2)$, we get from (3.37b) that

$$ H(Y_1 | Y_2) = C_{12} \geq I(\hat{Y}_1; Y_1) - I(\hat{Y}_1; Y_2) $$

$$ = H(\hat{Y}_1 | Y_2) - H(\hat{Y}_1 | Y_1), $$

which is satisfied when $\hat{Y}_1 = Y_1$. Plugging this into (3.36b), we get that when full cooperation from $R_{x1}$ to $R_{x2}$ is available, the rate constraint for $R_{x2}$ becomes

$$ R_2 \leq I(V; Y_2, Y_1). $$

Using the same reasoning we conclude that when full cooperation from $R_{x2}$ to $R_{x1}$ is available, the rate constraint for $R_{x1}$ becomes $R_1 \leq I(U; Y_1, Y_2)$.

**Partial Cooperation**

When $0 < C_{12} < H(Y_1 | Y_2)$ and $0 < C_{21} < H(Y_2 | Y_1)$, we get that

$$ C_{21} \geq H(\hat{Y}_2 | Y_1) - H(\hat{Y}_2 | Y_2) $$

$$ \Rightarrow H(\hat{Y}_2 | Y_1) \leq C_{21} + H(\hat{Y}_2 | Y_2). \quad (3.40) $$

77
Hence, the achievable rate to $R_{x1}$ is upper bounded by

$$R_1 \leq I(U; Y_1, \hat{Y}_2)$$

$$= I(U; Y_1) + I(U; \hat{Y}_2|Y_1)$$

$$= I(U; Y_1) + H(\hat{Y}_2|Y_1) - H(\hat{Y}_2|U, Y_1)$$

$$(a) \leq I(U; Y_1) + H(\hat{Y}_2) - H(\hat{Y}_2|U, Y_1) + C_{21}$$

$$(b) \equiv I(U; Y_1) + H(\hat{Y}_2|Y_2, Y_1, U) - H(\hat{Y}_2|U, Y_1) + C_{21}$$

$$R_1 \leq I(U; Y_1) + C_{21} - I(\hat{Y}_2; Y_2|U, Y_1). \quad (3.41)$$

where (a) is due to (3.40) and (b) follows from the same reasoning leading to equation (3.39). Similarly, $R_2 \leq I(V; Y_2) + C_{12} - I(\hat{Y}_1; Y_1|V, Y_2)$.

Note that there exist negative terms $-I(\hat{Y}_2; Y_2|U, Y_1)$ and $-I(\hat{Y}_1; Y_1|V, Y_2)$ in the upper bounds on the achievable rates. This can be explained as follows: the mutual information $I(\hat{Y}_2; Y_2|U, Y_1)$ can be considered as a type of “ancillary” information that $\hat{Y}_2$ contains, since this information is contained in $\hat{Y}_2$ while $U$ and $Y_1$ are already known – therefore, this information is a “noise” part of $Y_2$ which does not include any helpful information for decoding $U$ at $R_{x1}$. Thus, for cooperating in the optimal way, $\hat{Y}_2$ has to be a type of “sufficient and complete” cooperation information.

### 3.4 The Cooperative General Broadcast Channel with a Single Common Message

We now consider the case where only a single message is transmitted to both receivers. The main motivation for considering this case is that in the three independent messages case it is impossible to obtain explicit rate expressions without
auxiliary RVs, even if we explicitly specify the conditional distributions for the conference auxiliary RVs. Hence, we cannot identify directly the gain from cooperation, except in the case of full cooperation, and we also cannot numerically evaluate the achievable region. For the single message case, we are able to derive results for partial cooperation without auxiliary random variables, which make this rate explicitly computable. This scenario is depicted in figure 3.3.

Figure 3.3: The single message broadcast channel with cooperating receivers. \( \hat{W} \) and \( \hat{\hat{W}} \) are the estimates of \( W \) at \( R_{x1} \) and \( R_{x2} \) respectively.

For this scenario we need to specialize the definitions of a code and the average probability of error as follows:

- A \( (2^{nR}, n, (C_{12}, C_{21}), K) \) code for sending a single message over the broadcast channel with cooperating receivers having conference links of capacities \( C_{12} \) and \( C_{21} \) between them, is defined in a similar manner to definition 3.4 with \( ||\mathcal{M}_1|| = ||\mathcal{M}_2|| = 1 \).

- The average probability of error is defined similarly to definition 3.5 with \( M_1 \) and \( M_2 \) omitted.

The capacity for the non-cooperative single message BC scenario is given in
The upper bound on the achievable rate for the single message BC with cooperating receivers can be obtained from the bound for the three independent messages case in proposition 3.1:

**Corollary 3.2** Let \((\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)\) be a discrete memoryless broadcast channel, with cooperating receivers having noiseless conference links of finite capacities \(C_{12}\) and \(C_{21}\), as defined in section 3.1. Then, for sending a single message to both receivers, any achievable rate \(R\) must satisfy

\[
R \leq \sup_{p(x)} \left\{ \min \left\{ I(X; Y_1) + C_{21}, I(X; Y_2) + C_{12}, I(X; Y_1, Y_2) \right\} \right\}.
\]

**Proof**

Follows directly from proposition 3.1 by setting \(R_1 = R_2 = 0\).

Specializing the three independent messages achievable rate region to the single message case we obtain the following achievable rate with a \(K\)-cycle conference:

**Corollary 3.3** Consider the general broadcast channel \((\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)\) with cooperating receivers having noiseless conference links of finite capacities \(C_{12}\) and \(C_{21}\) between them. Let the receivers hold a conference that consists of \(K\) cycles. Then, any rate \(R\) satisfying

\[
R = \max \{R_{12}, R_{21}\},
\]

is achievable.

Here, \(R_{12}\) is defined as follows:

\[
R_{12} = \sup_{p_X(x), \alpha \in [0,1]} \min \{R_1, R_2\},
\]
with

\[ R_1 = I \left( X; Y_1, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)} \right) + \alpha C_{21}, \]  

(3.45a)

\[ R_2 = I \left( X; Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)} \right), \]  

(3.45b)

subject to

\[ C_{12} \geq I \left( Y_1; \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)} \mid Y_2 \right), \]  

(3.46a)

\[ (1 - \alpha)C_{21} \geq I \left( Y_2; \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)} \mid Y_1 \right), \]  

(3.46b)

for the joint distribution

\[
p \left( x, y_1, y_2, \tilde{y}_1^{(1)}, \tilde{y}_1^{(2)}, ..., \tilde{y}_1^{(K)}, \tilde{y}_2^{(1)}, \tilde{y}_2^{(2)}, ..., \tilde{y}_2^{(K-1)} \right) = \]

\[
p(x)p(y_1, y_2|x)p \left( \tilde{y}_1^{(1)} \mid y_1 \right) p \left( \tilde{y}_2^{(1)} \mid y_2, \tilde{y}_1^{(1)} \right) \cdots \]

\[
\times p \left( \tilde{y}_1^{(k)} \mid y_1, \tilde{y}_1^{(1)}, \tilde{y}_1^{(2)}, ..., \tilde{y}_1^{(k-1)}, \tilde{y}_2^{(1)}, \tilde{y}_2^{(2)}, ..., \tilde{y}_2^{(k-1)} \right) \]

\[
\times p \left( \tilde{y}_2^{(k)} \mid y_2, \tilde{y}_2^{(1)}, \tilde{y}_2^{(2)}, ..., \tilde{y}_2^{(k-1)} \right) \]

The cardinalities of the \( k \)'th auxiliary random variables are bounded by:

\[
||\hat{\mathcal{Y}}_1^{(k)}|| \leq ||\mathcal{Y}_1|| \times \prod_{l=1}^{k-1} ||\hat{\mathcal{Y}}_1^{(l)}|| \times \prod_{l=1}^{k-1} ||\hat{\mathcal{Y}}_2^{(l)}|| + 1, \quad k = 1, 2, ..., K
\]

\[
||\hat{\mathcal{Y}}_2^{(k)}|| \leq ||\mathcal{Y}_2|| \times \prod_{l=1}^{k} ||\hat{\mathcal{Y}}_1^{(l)}|| \times \prod_{l=1}^{k-1} ||\hat{\mathcal{Y}}_2^{(l)}|| + 1, \quad k = 1, 2, ..., K - 1.
\]

\( R_{21} \) is defined in a symmetric manner to \( R_{12} \), with \( R_{x2} \) performing the first conference step, and the appropriate changes in the probability chain.

The proof of corollary 3.3 is provided in appendix B. We note that this is not a direct specialization of theorem 3.2 as in the last conference step from \( R_{x2} \) to \( R_{x1} \).
in $R_{12}$, we use DAF and not EAF. This is because after the $K$’th conference step from $R_{x1}$ to $R_{x2}$, $R_{x2}$ decodes the message. Thus, its last conference message to $R_{x1}$ can be generated using DAF which is more efficient than EAF (in the sense that it does not transmit noise, see section 3.3.3).

We note that [13, theorem 2] presents a similar result for this scenario, under the constraint that the memoryless broadcast channel can be decomposed as $p(y_1, y_2|x) = \prod_{i=1}^{n} p(y_{1,i}|x_i)p(y_{2,i}|x_i)$, and considering only the sum-rate of the conference. Here we show that the same achievable rate expressions hold for the general memoryless broadcast channel. Another recent result appears in [89], in which the single common message case over the Gaussian BC is considered. In the multi-cycle conference considered in this section, we let the auxiliary RVs follow a more general chain than that of [89] – which results in a higher achievable rate.

### 3.4.1 A Single-Cycle Conference with TS Auxiliary Mapping

Consider the case where the receivers hold a single cycle of conferencing ($K = 1$). Specializing $R_{12}$ in corollary 3.3 to the single-cycle case we obtain

$$R_1 = I(X; Y_1) + C_{21} \tag{3.47a}$$

$$R_2 = I(X; Y_2, \hat{Y}_1^{(1)}) \tag{3.47b}$$

$$C_{12} \geq I(Y_1; \hat{Y}_1^{(1)}|Y_2). \tag{3.47c}$$

Now consider the TS assignment:

$$p(\hat{y}_1^{(1)}|y_1) = \begin{cases} 
q_1, & \hat{y}_1^{(1)} = y_1 \\
1 - q_1, & \hat{y}_1^{(1)} = E \notin \mathcal{Y}_1.
\end{cases}$$
Applying the TS assignment to (3.47c) and (3.47b) we obtain

\[ C_{12} \geq I(Y_1; \hat{Y}_1^{(1)}|Y_2) \]

\[ = H(Y_1|Y_2) - H(Y_1|Y_2, \hat{Y}_1^{(1)}) \]

\[ = H(Y_1|Y_2) - q_1 H(Y_1|Y_2, Y_1) - (1 - q_1) H(Y_1|Y_2) \]

\[ = q_1 H(Y_1|Y_2) \]

\[ R_2 = I(X; Y_2, \hat{Y}_1^{(1)}) \]

\[ = I(X; Y_2) + H(X|Y_2) - H(X|Y_2, \hat{Y}_1^{(1)}) \]

\[ = I(X; Y_2) + H(X|Y_2) - (1 - q_1) H(X|Y_2) - q_1 H(X|Y_2, Y_1) \]

\[ = I(X; Y_2) + q_1 I(X; Y_1|Y_2). \]  \hfill (3.48)

Maximizing \( R_2 \) requires maximizing \( q_1 \in [0, 1] \). Therefore setting \( q_1 = \left[ \frac{C_{12}}{H(Y_1|Y_2)} \right]^* \), we obtain \( R_2 = I(X; Y_2) + \left[ \frac{C_{12}}{H(Y_1|Y_2)} \right]^* I(X; Y_1|Y_2) \). Combining with \( R_1 \) we obtain that the rate when \( R_{x2} \) decodes first is given by

\[ R_{12} = \min \left\{ I(X; Y_1) + C_{21}, I(X; Y_2) + \left[ \frac{C_{12}}{H(Y_1|Y_2)} \right]^* I(X; Y_1|Y_2) \right\}, \]

for a given \( p(x) \). By a symmetric argument we can obtain \( R_{21} \). Therefore, the rate for the single-cycle conference with TS is given by the following corollary:

**Corollary 3.4** *For the single message setup of corollary 3.3 with a single-cycle conference (\( K = 1 \)), any rate \( R \) satisfying*

\[ R = \sup_{p(x)} \min \{ R_{12}, R_{21} \}, \]

\[ R_{12} = \min \left\{ I(X; Y_1) + C_{21}, I(X; Y_2) + \left[ \frac{C_{12}}{H(Y_1|Y_2)} \right]^* I(X; Y_1|Y_2) \right\}, \]

\[ R_{21} = \min \left\{ I(X; Y_1) + \left[ \frac{C_{21}}{H(Y_2|Y_1)} \right]^* I(X; Y_2|Y_1), I(X; Y_2) + C_{12} \right\}, \]

*is achievable.*
When the receiver that decodes first uses joint-decoding, the rate expression can be obtained from equation (3.48) with assignment (2.19), by setting \( \hat{Y}_1 = Y_1 \), \( X_1 = \emptyset \), \( Y = Y_2 \) and replacing \( I(X_1; Y) \) with \( C_{12} \). The resulting TS proportion for \( \hat{Y}^{(1)} \) is

\[
q_1 = C_{12} - \frac{H(Y_1|Y_2, X)}{I(X; Y_1|Y_2)}.
\]

Plugging (3.49) into (3.47) we obtain the joint-decoding rate. This assignment has first to satisfy the non-negativity constraint on \( q_1 \) hence

\[
C_{12} \geq H(Y_1|Y_2, X).
\]

Next, the feasibility constraint has to be satisfied: \( q_1 \leq \frac{C_{12}}{H(Y_1|Y_2)} \). We now show that this is indeed the case:

\[
\frac{C_{12}}{H(Y_1|Y_2)} \geq \frac{C_{12} - H(Y_1|Y_2, X)}{I(X; Y_1|Y_2)}
\]

\[
C_{12} I(X; Y_1|Y_2) \geq (C_{12} - H(Y_1|Y_2, X)) H(Y_1|Y_2)
\]

\[
C_{12} (I(X; Y_1|Y_2) - H(Y_1|Y_2)) \geq -H(Y_1|Y_2, X) H(Y_1|Y_2)
\]

\[
-C_{12} H(Y_1|Y_2, X) \geq -H(Y_1|Y_2, X) H(Y_1|Y_2)
\]

\[
C_{12} \leq H(Y_1|Y_2),
\]

as long as \( I(X; Y_1|Y_2) > 0 \) (otherwise there is no need for cooperation from \( R_{x1} \) to \( R_{x2} \)). We note that if \( H(Y_1|Y_2, X) = 0 \) than (a) is satisfied with equality, and then both quantities are equal. As discussed in section 2.4, the assignment (3.49) produces the rate obtained with joint-decoding. The resulting achievable rate is given in the following proposition:

**Proposition 3.2** Assume the broadcast channel setup of corollary 3.3. Then, for sending a single message to both receivers with a single-cycle conference \( (K = 1) \),
any rate $R$ satisfying

$$R \leq \sup_{p(x)} \left[ \max \left\{ R^{12}(p(x)), R^{21}(p(x)) \right\} \right]$$

$$R^{12}(p(x)) \triangleq \min \left\{ I(X; Y_1) + C_{21}, I(X; Y_2) - H(Y_1|Y_2, X) + \min \{ C_{12}, H(Y_1|Y_2) \} \right\},$$

$$R^{21}(p(x)) \triangleq \min \left\{ I(X; Y_2) + C_{12}, I(X; Y_1) - H(Y_2|Y_1, X) + \min \{ C_{21}, H(Y_2|Y_1) \} \right\},$$

with the appropriate $C_{12} > H(Y_1|Y_2, X)$ or $C_{21} > H(Y_2|Y_1, X)$ (the one used for the first conference step), is achievable.

This gives another partial cooperation result without auxiliary random variables.

We note that the rate of corollary 3.4 is always better than the point-to-point rate and also better than the joint-decoding rate of proposition 3.2 (whenever cooperation can provide a rate increase). However, for both corollary 3.4 and proposition 3.2, at least one conference link has to satisfy the Slepian-Wolf condition (i.e. $C_{12} \geq H(Y_1|Y_2)$ or $C_{21} \geq H(Y_2|Y_1)$) for the full cooperation rate to be achieved. We also note that using TS-EAF with more than a single cycle does not improve upon this result.

Finally, we demonstrate the results of proposition 3.2 and corollary 3.4 through a symmetric BC example: consider the symmetric broadcast channel in which $Y_1 = Y_2 = Y$ and

$$p_{Y_1|Y_2, X}(a|b, x) = p_{Y_2|Y_1, X}(a|b, x),$$

for any $(a, b) \in Y \times Y$ and $x \in \mathcal{X}$. Let $C_{21} = C_{12} = C$. For this scenario we have that $R_{12} = R_{21}$ in corollary 3.4 and also $R^{12}(p_X(x)) = R^{21}(p_X(x))$ in proposition 3.2. The resulting rates are depicted in figure 3.4 for a fixed probability $p(x)$. We can see that for this case, time-sharing exceeds joint-decoding for all values of $C$. Both methods meet the upper bound at $C = H(Y_1|Y_2)$. We note that this is a corrected version of the figure in [108].
3.4.2 Special Cases

The “Distinctly Better” Case

There are special cases in which the lower bound of corollary 3.4 coincides with the upper bound of corollary 3.2, yielding the capacity for these cases. For example, assume a strong version of the “more capable” condition of [45]: \( I(X; Y_1) >> I(X; Y_2) \)\(^3\) for all input distributions \( p(x) \) on \( \mathcal{X} \). Assume also that \( C_{21} < H(Y_2|Y_1) \) and \( C_{12} < H(Y_1|Y_2) \). Under these conditions, we have that

\[
I(X; Y_1) + C_{21} > I(X; Y_2) + \frac{C_{12}}{H(Y_1|Y_2)} I(X; Y_1|Y_2).
\]

Thus, if \( R_{x_1} \) is helping \( R_{x_2} \) first, the achievable rate is

\[
I(X; Y_2) + \frac{C_{12}}{H(Y_1|Y_2)} I(X; Y_1|Y_2).
\]

\(^3\)The precise condition requires that \( I(X; Y_1) > I(X; Y_2) + C_{12} - \frac{C_{21}}{H(Y_2|Y_1)} I(X; Y_2|Y_1) \) for all input distributions \( p(x) \).
If $R_{x2}$ is helping $R_{x1}$ first, then the achievable rate is $I(X; Y_2) + C_{12}$. Since 
$C_{12} \frac{I(X; Y_1 | Y_2)}{H(Y_1 | Y_2)} \leq C_{12}$, this cooperation scheme achieves the upper bound

$$R = \sup_{p(x)} \{ I(X; Y_2) + C_{12} \}.$$ 

### The Deterministic BC with Cooperating Decoders

The capacity region of the deterministic broadcast channel with cooperating receivers follows from proposition 3.2 and corollary 3.2. This region was derived in [104]. For this case we have that $H(Y_1 | X) = H(Y_2 | X) = 0$, hence $I(X; Y_i) = H(Y_i), i = 1, 2$. The achievable rate (from proposition 3.2) is given by

$$R \leq \min \{ H(Y_2) + C_{12}, H(Y_1) + \min (C_{21}, H(Y_2 | Y_1)) \}$$

$$= \min \{ H(Y_2) + C_{12}, H(Y_1) + C_{21}, H(Y_1, Y_2) \},$$

and the same is obtained from corollary 3.2.
Chapter 4

Application of Time-Sharing to Relaying Scenarios

In this chapter we apply the EAF strategy with the TS assignment to two relay scenarios: first we consider the discrete, memoryless, multiple-relay channel. For this scenario, extending the general EAF results in an achievable rate that has a complicated characterization and is not amenable to numerical evaluation. The second scenario is the Gaussian relay channel with coded modulation. Here we consider quantization with a finite alphabet at the relay based on TS-EAF. This approach is better suited to intermediate source-relay SNRs than both the DAF method and the Gaussian EAF.

4.1 An Achievable Rate for the Relay Channel with Multiple Relays

When the source-relay channel is very noisy then, as discussed in section 1.3.1, it may be better not to use the relay at all than to employ the decode-and-forward strategy. Alternatively, when decode-and-forward is not useful, the relay could employ estimate-and-forward. One result for multiple relays based on EAF can be found in [28] which considered the two-relay case. In [12, theorem 3] the EAF strategy, with partial decoding at the relays was applied to the multiple-relay channel, and in [12, theorem 4] a mixed EAF and DAF strategy was applied. However, as stated in [12, remark 22, remark 23] applying the general estimate-and-forward to a network with an arbitrary number of relays is computationally
prohibitive due to the large number of constraints that characterize the feasible region (for two relays we need to satisfy 9 constraints, when the relays employ partial decoding). Moreover, searching for the maximum rate is also a difficult task since it requires solving a non-convex optimization problem. In conclusion, for the discrete memoryless multiple-relay channel, an alternative achievable rate to that based on decode-and-forward, which can also be evaluated with a reasonable computational effort, has not been presented to date. In this section we derive an explicit achievable rate based on estimate-and-forward. The strategy we use is to pick the auxiliary random variables such that the feasibility constraints are satisfied. This is not a trivial choice since setting the auxiliary random variable in theorem 1.2 to be the relay channel output (i.e. $\hat{Y}_1 = Y_1$) does not remove this constraint, and we therefore need to incorporate time-sharing as discussed in the following.

4.1.1 Statement of the Main Theorem

We extend the idea of section 2.3 to the relay channel with $N$ relays. This channel consists of a transmitter with channel input $X$, $N$ relays where for relay $i$, $X_i$ denotes the channel input and $Y_i$ denotes the channel output, and a destination with channel output $Y$. This channel is denoted by $(\mathcal{X} \times \prod_{i=1}^{N} \mathcal{X}_i, p(y, y_1, ..., y_N|x, x_1, ..., x_N), \mathcal{Y} \times \prod_{i=1}^{N} \mathcal{Y}_i)$. Let $X = (X_1, X_2, ..., X_N)$ and $Y = (Y_1, Y_2, ..., Y_N)$. We begin with a statement of the theorem:

**Theorem 4.1** For the general multiple-relay channel with $N$ relays, $(\mathcal{X} \times \prod_{i=1}^{N} \mathcal{X}_i, p(y, y_1, ..., y_N|x, x_1, ..., x_N), \mathcal{Y} \times \prod_{i=1}^{N} \mathcal{Y}_i)$, any rate $R$ satisfying

$$R \leq I(X; Y|X) + \sum_{\theta=1}^{2^{N-1}} P(Bin_N(\theta)) I(X; Y_{Bin_N(\theta)}|X, Y),$$

89
where $Bin_N(\theta)$ is the $N$-bit binary representation of the integer $\theta$,

$$
P(Bin_N(\theta)) = \prod_{i : Bin_N(\theta)_i = 0} (1 - q_i) \prod_{i : Bin_N(\theta)_i = 1} q_i,
$$

$Bin_N(\theta)_i$ is the $i$'th bit of $Bin_N(\theta)$, $Y_{Bin_N(\theta)} = (Y_{i_1}, Y_{i_2}, ..., Y_{i_M(\theta)})$, where $i_1, i_2, ..., i_M(\theta)$ are the locations of the '1's in $Bin_N(\theta)$, and the indices refer to $Y$, and

$$
q_i = \left[ \frac{I(X_i; Y|\tilde{Z}_i)}{H(Y_i|X, Y) - \sum_{j=1}^{2^{L'_i-1}} P_r(Bin_{L'_i}(j))I(Y_i; \tilde{Y}_{r', Bin_{L'_i}(j)}|T_i)|X, Y) \right]^*, \quad (4.1)
$$

for the joint distribution $p(x, x_1, x_2, ..., x_N, y, y_1, y_2, ..., y_N) = p(x)p(x_1)...p(x_N) \times p(y, y_1, ..., y_N|x, x_1, ..., x_N)$, is achievable.

In (4.1) $\tilde{Z}_i = (X_{m_1}, X_{m_2}, ..., X_{m_{M_i}})$ is a vector containing all the variables $X_j$ decoded prior to decoding $X_i$ at the destination ($\tilde{Z}_{i-1} \subseteq \tilde{Z}_i \subseteq X$). $\tilde{T}_i = (Y_{i_1}', Y_{i_2}', ..., Y_{i_{L'_i}}')$ is a vector that contains all the variables $Y_p$ decoded prior to decoding $Y_i$ at the destination ($\tilde{T}_{i-1} \subseteq \tilde{T}_i \subseteq Y$), and $\tilde{Y}_{r', Bin_{L'_i}(j)}(\tilde{T}_i)$ contains all the $Y_{i_r'}$, such that $Y_{i_r'} \in \tilde{T}_i$, and $r$ is a location of '1' in $Bin_{L'_i}(j)$, the $L'_i$-bit binary representation of $j$. $L'_i$ is the number of elements in $\tilde{T}_i$. Note that if $Y_p \in \tilde{T}_i$ then we must have $X_p \in \tilde{Z}_i$.

To facilitate the understanding of the expressions in theorem 4.1, we first look at a simplified case in which the destination decodes each relay message independently of the messages of the other relays. This can be obtained from theorem 4.1 by setting $\tilde{Z}_i = \emptyset$ and $\tilde{T}_i = \emptyset$, $i = 1, 2, ..., N$. The result is summarized in the following corollary:

**Corollary 4.1** For the general multiple-relay channel with $N$ relays, $(X \times_{i=1}^N X_i$, $p(y, y_1, ..., y_N|x, x_1, ..., x_N), Y \times_{i=1}^N Y_i)$, any rate $R$ satisfying

$$
R \leq I(X; Y|X) + \sum_{\theta=1}^{2^{N-1}} P(Bin_N(\theta))I(X; Y_{Bin_N(\theta)}|X, Y), \quad (4.2)
$$

90
is achievable, where

\[ q_i = \left[ \frac{I(X_i; Y)}{H(Y_i|X,Y)} \right]^* , \quad (4.3) \]

for the joint distribution \( p(x, x_1, x_2, \ldots, x_N, y, y_1, y_2, \ldots, y_N) = p(x)p(x_1)\ldots p(x_N) \times p(y, y_1, \ldots, y_N|x, x_1, \ldots, x_N). \)

In the multi-relay strategy we employ in this section, each relay transmits its channel output \( Y_i \) with probability \( q_i \) and an erasure symbol \( E \) with probability \( 1 - q_i \), independent of the other relays and the transmitter. Therefore, considering the group of \( N \) relays, the probability that any subgroup of relays will transmit their channel outputs simultaneously is simply the product of all transmission probabilities \( q_i \) at each relay in the subgroup, multiplied by the product of erasure probabilities \( (1 - q_i) \) of each relay in the complementing subgroup. Now, considering the rate expression of (4.2) we observe that the rate is obtained by taking all possible subgroups of relays. For each subgroup the resulting rate is derived by using the channel outputs of all the relays in that subgroup to assist in decoding. This is the role of the term \( Y_{\text{Bin}_{N}(\theta)} \). This rate has to be weighted by the probability of such an overlap occurring, which is given by \( P(\text{Bin}_{N}(\theta)) \). We then sum over all possible subgroups to obtain the achievable rate. The parameter \( q_i \) for each relay, which is determined by (4.3), can be interpreted by considering the terms in the denominator and numerator: the denominator \( H(Y_i|X,Y) \) is the (exponent of the) size of the uncertainty at the destination receiver about relay \( i \)'s output \( Y_i^n \). The numerator is the (exponent of the) size of the message set that can be transmitted from relay \( i \) to the destination receiver. Therefore, the fraction \( \frac{I(X_i; Y)}{H(Y_i|X,Y)} \) can be interpreted as the maximal fraction of the uncertainty at the destination about relay \( i \)'s channel output \( Y_i^n \), that can be compensated by
the relay transmission. Of course, this fraction has to be upper bounded by one.

In the more general setup of theorem 4.1, decoding the information from relay $i$ is done by using the messages from the relays which were decoded prior to relay $i$ to assist in decoding. This results in the conditioning at the numerator and the negative terms in the denominator, in equation (4.1), both contribute to increasing the value of $q_i$.

We note here that the result of [12, theorem 3] with partial decoding at the relays does not specialize to theorem 4.1 by applying the TS assignment of (2.5). The rate expressions are the same but the feasibility conditions of [12, theorem 3] produce a non-linear set of equations in the parameters $\{q_i\}_{i=1}^N$ which is still complicated to solve. Here we actually also simplified the decoding procedure at the destination receiver, to allow us to obtain the $q_i$’s in a recursive manner. In appendix C we give the resulting expressions of [12, theorem 3] for the two-relay channel, when the auxiliary assignments for relay 1 and relay 2 are TS assignments, to allow comparison with theorem 4.1 (cf. $R^{TS-EAF}$ in section 4.1.3).

### 4.1.2 Proof of Theorem 4.1

**Overview of Coding Strategy**

The transmitter generates its codebook independent of the relays. Next, each relay generates its own codebook independent of the other relays and the transmitter, following the construction of [15, theorem 6], with the mapping $p(\hat{y}_i|x_i,y_i)$ at each relay set to the time-sharing mapping of (2.5) with parameter $q_i$. The destination receiver first decodes all the relay codewords $\{X^n_{i}\}_{i=1}^N$ and then uses this information to decode the relay messages $\{\hat{Y}_i^n\}_{i=1}^N$. To this end, the receiver decides on a decoding order for the $X_{i}^n$ sequences and a decoding order for the $\hat{Y}_i^n$ sequences.
These decoding orders determine the maximum value of $q_i$ that can be selected for each relay, thereby allowing us to determine the auxiliary variables’ mappings and obtain an explicit rate expression. Finally, the receiver uses all the decoded sequences, together with its channel output $Y^n$ to decode the source codeword $X^n$.

We now give the details of the construction: fix the distributions $p(x)$, $p(x_1)$, $p(x_2), \ldots, p(x_N)$, and $p(\hat{y}_i|x_i, y_i) = \begin{cases} q_i, & \hat{y}_i = y_i \\ 1 - q_i, & \hat{y}_i \notin \mathcal{Y}_i \end{cases}$, $i = 1, 2, \ldots, N$. Let $\mathcal{W} = \{1, 2, \ldots, 2^{nR_i}\}$ be the source message set.

**Code Construction at the Transmitter and the Relays**

- Code construction and transmission at the transmitter are the same as in [15, theorem 6].
- Code construction at the relays is done by repeating the relay code construction of [15, theorem 6] for each relay, where relay $i$ uses the distributions $p(\hat{y}_i|x_i, y_i)$ and $p(x_i)$. We denote the relay message, the transmitted message and the partition set at relay $i$ at time $k$ with $z_{i,k}$, $s_{i,k}$ and $S^{(i)}_{s_{i,k}}$ respectively. The message set for $s_i$ is denoted $\mathcal{W}_i = \{1, 2, \ldots, 2^{nR_i}\}$. The message set for $z_i$ is denoted $\mathcal{W}_i' = \{1, 2, \ldots, 2^{nR_i'}\}$. The relay codewords at relay $i$ are denoted $\hat{y}_i(z_i|s_i)$, and the transmitted codewords at relay $i$ are denoted $x_i(s_i)$, $s_i \in \mathcal{W}_i$, $z_i \in \mathcal{W}_i'$.

**Decoding and Encoding at the Relays**

Consider relay $i$ at time $k - 1$:
• From the relay transmission at time $k - 1$, the relay knows $s_{i,k-1}$. Now the relay looks for a message $z_i \in W'_i$, such that

$$(\hat{y}_i(z_i|s_{i,k-1}), y_i(k-1), x_i(s_{i,k-1})) \in A^*(n)(\hat{Y}_i, Y_i, X_i).$$

Following the argument in [15, theorem 6], for $n$ large enough there is such a message $z_i$ with a probability that is arbitrarily close to 1, as long as

$$R'_i > I(\hat{Y}_i; Y_i|X_i) + \epsilon = q_i H(Y_i|X_i) + \epsilon. \quad (4.5)$$

Denote this message with $z_{i,k-1}$.

• Let $s_{i,k}$ be the index of the partition of $W'_i$ into which $z_{i,k-1}$ belongs, i.e., $z_{i,k-1} \in S^{(i)}_{s_{i,k}}$.

At time $k$ relay $i$ transmits $x_i(s_{i,k})$.

**Decoding at the Destination**

• Consider the decoding of $w_{k-1}$ at time $k$, for a fixed decoding order: let $\tilde{Z}_i$ contain all the $X_j$’s whose $s_{j,k}$’s were decoded prior to decoding $s_{i,k}$. Therefore, decoding $s_{i,k}$ is done by looking for a unique message $s_i \in W_i$ such that

$$(x_i(s_i), x_{m_1}(s_{m_1,k}), x_{m_2}(s_{m_2,k}), ..., x_{m_{M_i}}(s_{m_{M_i},k}), y(k)) \in A^*(n)(X_i, \tilde{Z}_i, Y),$$

where $m_1, m_2, ..., m_{M_i}$ enumerate all the $X_j$’s in $\tilde{Z}_i = (X_{m_1}, X_{m_2}, ..., X_{m_{M_i}})$. Assuming correct decoding of $\{s_{m_{j,k}}\}_{j=1}^{M_i}$ at the previous steps, then by the point-to-point channel achievability proof we obtain that the probability of error for decoding $s_{i,k}$ can be made arbitrarily small by taking $n$ large enough as long as

$$R_i < I(X_i; Y, \tilde{Z}_i) - \epsilon = I(X_i; Y|\tilde{Z}_i) - \epsilon. \quad (4.6)$$
With a slight abuse of notation, let $\tilde{T}_i$ contain all the $\hat{Y}_l$’s whose $z_{l',k-1}$’s are decoded prior to decoding $z_{i,k-1}$. Note that all the $\{s_{i,k-1}\}_{i=1}^N$ were already decoded at the previous time interval (time $k-1$) when $w_{k-2}$ was decoded.

- The destination generates the set

$$\mathcal{L}_i(k-1) = \left\{ z_i \in \mathcal{W}_i' : (y(k-1), \hat{y}_i(z_i|s_{i,k-1}), \hat{y}_{l'_i}(z_{l'_i,k-1}|s_{l'_i,k-1}), \ldots, \hat{y}_{l'_i}(z_{l'_i,k-1}|s_{l'_i,k-1}), x_1(s_{1,k-1}), x_2(s_{2,k-1}), \ldots, x_N(s_{N,k-1})) \in A_{i}^{\ast(n)}(Y, \hat{Y}_i, \tilde{T}_i, X) \right\}, \quad (4.7)$$

where $l'_1, l'_2, \ldots, l'_{L'_i}$ enumerate all the $\hat{Y}_l$’s in $\tilde{T}_i$. The average size of $\mathcal{L}_i(k-1)$ can be bounded using the standard technique of [15, equation (36)] and the fact that when $z_i \neq z_{i,k-1}$, then the corresponding $\hat{y}_i(z_i|s_{i,k-1})$ is independent of all the variables in (4.7) except $x_i(s_{i,k-1})$. Assuming correct decoding of all $\{z_{l'_j,k-1}\}_{j=1}^{L'_i}$, the resulting bound is

$$E \{||\mathcal{L}_i(k-1)||\} \leq 1 + 2^{n(R'_i - I(\hat{Y}_i; Y, X_{-i}, \tilde{T}_i|X_i)) + 3\epsilon},$$

where $X_{-i}$ is an $N - 1$ element vector that contains all the elements of $X$ except $X_i$.

- Now, the destination looks for a unique $z_i \in \mathcal{L}_i(k-1) \cap S_{s_{i,k}}^{(i)}$. Assuming correct decoding of $s_{i,k}$, the probability of error can be made arbitrarily small by taking $n$ large enough as long as

$$R'_i < I(\hat{Y}_i; Y, X_{-i}, \tilde{T}_i|X_i) + I(X_i; Y|\tilde{Z}_i) - 4\epsilon. \quad (4.8)$$
Using the assignment (4.4) we can write

\[ I(\tilde{Y}_i; Y, X_{-i}, \tilde{T}_i|X_i) = H(Y, X_{-i}, \tilde{T}_i|X_i) - H(Y, X_{-i}, \tilde{T}_i|X_i, \tilde{Y}_i) \]

\[ = H(Y, X_{-i}, \tilde{T}_i|X_i) - (1 - q_i)H(Y, X_{-i}, \tilde{T}_i|X_i) \]

\[ -q_iH(Y, X_{-i}, \tilde{T}_i|X_i, Y_i) \]

\[ = q_i H(Y, X_{-i}, \tilde{T}_i|X_i) - q_i H(Y, X_{-i}, \tilde{T}_i|X_i, Y_i) \]

\[ = q_i \left( H(Y_i|X_i) - H(Y_i|Y, X_{-i}, \tilde{T}_i, \tilde{Y}_i) \right) \]

\[ = q_i \left( q_{i1} H(Y_i|X_i) + (1 - q_{i1})H(Y_i|X_i) \right) \]

\[ -q_{i1}H(Y_i|Y, X_{-i}, \tilde{T}_i, \tilde{Y}_i) \]

\[ - (1 - q_{i1}) H(Y_i|Y, X_{-i}, \tilde{T}_i, \tilde{Y}_i) \]

\[ = q_i \left( q_{i1} I(Y_i; Y, X_{-i}, \tilde{T}_i, \tilde{Y}_i) + (1 - q_{i1})I(Y_i; Y, X_{-i}, \tilde{T}_i) \right) \]

\[ \ldots \]

\[ = q_i \sum_{j=0}^{2^{L'_M} - 1} P_{H}(\text{Bin}_{L'_M}(j)) I(Y_i; Y, X_{-i}, \tilde{Y}_{j}; \text{Bin}_{L'_M}(j))(\tilde{T}_i)|X_i, \]

where \( P_{H}(\text{Bin}_{L'_M}(j)) = \prod_{i=1}^{r} \text{Bin}_{L'_M}(j)_{i} = 1 q_{tr} \times \prod_{i=0}^{r} \text{Bin}_{L'_M}(j)_{i} = 1 - q_{tr} \), \( \text{Bin}_{L'_M}(j)_{r} \) is the \( r \)-th bit of the \( L'_M \)-bit binary representation of \( j \), and \( \tilde{Y}_{j}; \text{Bin}_{L'_M}(j)(\tilde{T}_i) = (Y_{l'_n1}, Y_{l'_n2}, \ldots, Y_{l'_nM}) \), \( n_1, n_2, \ldots, n_M \) are the locations of ‘1’ in the \( L'_M \)-bit binary representation of \( j \), and \( l'_n1, l'_n2, \ldots, l'_nM \) are the indices of the \( \tilde{Y}_i \)'s in locations \( n_1, n_2, \ldots, n_M \) in \( \tilde{T}_i \). For example, if \( L'_M = 3 \) and \( j = 5 \) then Bin3(5) = (1, 0, 1) and \( M = 2, n_1 = 1, n_2 = 3 \).

Letting \( \tilde{T}_i = (\tilde{Y}_3, \tilde{Y}_1, \tilde{Y}_2) \) then \( l'_1 = 3, l'_2 = 1 \) and \( l'_3 = 2 \), and

\[ P_{H}(\text{Bin}_{L'_M}(5)) = q_{i1} (1 - q_{i2})q_{i3} = q_3(1 - q_1)q_2, \]

\[ \tilde{Y}_{j}; \text{Bin}_{L'_M}(5)(\tilde{T}_i) = (Y_{l'_3}, Y_{l'_2}) = (Y_3, Y_2), \]

96
Combining the Bounds on $R'_i$

Applying the above scheme requires that $R'_i$ satisfies (4.5) and (4.8):

$$q_i H(Y_i|X_i) + \epsilon < R'_i < q_i \sum_{j=0}^{2^{L'_i}-1} P_t(Bin_{L'_i}(j)) I(Y_i; Y, X_{-i}, \tilde{Y}_{j}, Bin_{L'_i}(j)(\tilde{T}_i)|X_i)$$

$$+ I(X_i; Y|\tilde{Z}_i) - 4\epsilon,$$

which is satisfied if

$$q_i < \frac{I(X_i; Y|\tilde{Z}_i) - 5\epsilon}{H(Y_i|X_i) - \sum_{j=0}^{2^{L'_i}-1} P_t(Bin_{L'_i}(j)) I(Y_i; Y, X_{-i}, \tilde{Y}_{j}, Bin_{L'_i}(j)(\tilde{T}_i)|X_i)}$$

$$= \frac{I(X_i; Y|\tilde{Z}_i) - 5\epsilon}{H(Y_i|X, Y) - \sum_{j=1}^{2^{L'_i}-1} P_t(Bin_{L'_i}(j)) I(Y_i; \tilde{Y}_{j}, Bin_{L'_i}(j)(\tilde{T}_i)|X, Y)}.$$

Combining with the constraint $0 \leq q_i \leq 1$ gives the condition in (4.1).

Finally, the achievable rate is obtained as follows: using the decoded $\{x_i(s_{i,k-1})\}_{i=1}^N$ and $\{\tilde{y}_i(z_{i,k-1}|s_{i,k-1})\}_{i=1}^N$ (assuming correct decoding of all $\{z_{i,k-1}\}_{i=1}^N$ and $\{s_{i,k-1}\}_{i=1}^N$), the receiver decodes the source message $w_{k-1}$ by looking for a unique message $w \in \mathcal{W}$ such that

$$\left( x(w), \tilde{y}_1(z_1,k-1|s_{1,k-1}), \tilde{y}_2(z_2,k-1|s_{2,k-1}), \ldots, \tilde{y}_N(z_{N,k-1}|s_{N,k-1}), x_1(s_{1,k-1}), x_2(s_{2,k-1}), \ldots, x_N(s_{N,k-1}), y(k-1) \right) \in A^*_n(X, \hat{Y}, X, Y),$$

where $\hat{Y} = (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N)$. This results in an achievable rate of

$$R \leq I(X; Y, \hat{Y}, X) = I(X; Y, \hat{Y}|X).$$

Plugging the assignments for all the $\hat{Y}_i$’s given by (4.4), we get the following explicit
rate expression:

\[ I(X; Y, \hat{Y}|X) = I(X; Y|X) + I(X; \hat{Y}|X, Y) \]

\[ = I(X; Y|X) + H(X|X, Y) - H(X|X, Y, \hat{Y}) \]

\[ = I(X; Y|X) + H(X|X, Y) - (1 - q_1)H(X|X, Y, \hat{Y}_2^N) \]

\[ - q_1 H(X|X, Y, \hat{Y}_2^N, Y_1) \]

\[ = I(X; Y|X) + (1 - q_1)I(X; \hat{Y}_2^N|X, Y) + q_1 I(X; \hat{Y}_2^N, Y_1|X, Y) \]

\[ = I(X; Y|X) + 2^{N-1} \sum_{\theta=1} P(\text{Bin}_N(\theta)) I(X; Y_{\text{Bin}_N(\theta)}|X, Y). \]

4.1.3 Discussion

To demonstrate the benefits of the explicit EAF-based achievable rate expression of theorem 4.1 we compare it with the DAF-based rate of [24, theorem 3.1] for the two-relay channel. This DAF rate is the highest achievable rate based on DAF for the multiple-relay channel. For this scenario there are five possible DAF setups,
and the maximum of the five resulting rates is taken as the DAF-based rate:

\[
R^{DAF} = \sup_{p(x, x_1, x_2)} \max \{R_1, R_2, R_{12}, R_{21}, R_G\}
\]

\[
R_1 = \max_{x_2 \in \mathcal{X}_2} \min_{x_1 \in \mathcal{X}_1} \{I(X; Y_1|X_1, x_2), I(X; Y|X_1, x_2) + I(X_1; Y|x_2)\}
\]

\[
R_2 = \max_{x_1 \in \mathcal{X}_1} \min_{x_2 \in \mathcal{X}_2} \{I(X; Y_2|X_2, x_1), I(X; Y|X_2, x_1) + I(X_2; Y|x_1)\}
\]

\[
R_{12} = \min \{I(X; Y_1|X_1, X_2), I(X; Y_2|X_1, X_2) + I(X_1; Y_2|X_2),
I(X; Y|X_1, X_2) + I(X_1; Y|X_2) + I(X_2; Y)\}
\]

\[
R_{21} = \min \{I(X; Y_2|X_1, X_2), I(X; Y_1|X_1, X_2) + I(X_2; Y_1|X_1),
I(X; Y|X_1, X_2) + I(X_2; Y|X_1) + I(X_1; Y)\}
\]

\[
R_G = \min \{I(X; Y_1|X_1, X_2), I(X; Y_2|X_1, X_2), I(X, X_1, X_2; Y)\}
\]

where \(R_1\) is the rate obtained when only relay 1 is active, \(R_2\) is the rate obtained when only relay 2 is active, \(R_{12}\) is the rate obtained when relay 1 decodes first and relay 2 decodes second and \(R_{21}\) is the rate obtained when this order is reversed. \(R_G\) is the rate obtained when both relays form one group and decode simultaneously\(^1\).

We also define the maximum point-to-point (PTP) rate as

\[
R^{PTP} = \sup_{p(x|x_1, x_2), (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2} I(X; Y|x_1, x_2).
\]

This rate is obtained when both relays fix their channel inputs.

Now, as in the single-relay case, DAF is limited by the worst source-relay link. Therefore, if

\[
R^{PTP} > \sup_{p(x|x_1, x_2), (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2} \max \{I(X; Y_1|x_1, x_2), I(X; Y_2|x_1, x_2)\}, \tag{4.9}
\]

then it is better not to use [24, theorem 3.1] at all, but rather set the relays to transmit the symbol pair \((x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2\) that maximizes the point-to-point rate.

\(^1\)\(R_G\) is given only for completeness. Clearly \(R_G\) does not exceed \(\max \{R_{12}, R_{21}\}\).
rate. However, the rate obtained using corollary 4.1 for the two-relay case is given by

\[ R_{TS-EAF} = \sup_{p(x)p(x_1)p(x_2)} I(X; Y) = q_1 (1 - q_2) I(X; Y_1 | X_1, X_2, Y) \\
+ (1 - q_1) q_2 I(X; Y_2 | X_1, X_2, Y) + q_1 q_2 I(X; Y_1, Y_2 | X_1, X_2, Y), \]

where \( q_1 \) and \( q_2 \) are non-negative and determined according to (4.3). This expression can, in general, be greater than \( R_{PTP} \) even when (4.9) holds, for channels where the relay-destination links are very good. Hence, this explicit achievable expression provides an easy way to improve upon the DAF-based achievable rates when the source-relay links are very noisy.

To demonstrate this, consider the channel given in table 4.1 over binary RVs \( X, X_1, X_2, Y, Y_1 \) and \( Y_2 \). The channel distribution was constructed under the independence constraint

\[ p(y, y_1, y_2 | x, x_1, x_2) = p(y_1 | x, x_1, x_2) p(y_2 | x, x_1, x_2) p(y | x, x_1, x_2, y_1, y_2), \]

i.e., given the channel inputs, the outputs of the two relays are independent. This channel is characterized by noisy source-relay links, while the link from relay 1 to the destination has low noise. Therefore, DAF is inferior to point-to-point
Table 4.2: Optimal distribution for DAF

<table>
<thead>
<tr>
<th>(x, x₁, x₂)</th>
<th>p(x, x₁, x₂)</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>5.6982e-009</td>
</tr>
<tr>
<td>001</td>
<td>5.2591e-017</td>
</tr>
<tr>
<td>010</td>
<td>6.7921e-009</td>
</tr>
<tr>
<td>011</td>
<td>4.4242e-001</td>
</tr>
<tr>
<td>100</td>
<td>4.3018e-009</td>
</tr>
<tr>
<td>101</td>
<td>4.7400e-017</td>
</tr>
<tr>
<td>110</td>
<td>3.2079e-009</td>
</tr>
<tr>
<td>111</td>
<td>5.5758e-001</td>
</tr>
</tbody>
</table>

transmission but EAF is able to exceed $R^{PTP}$, by giving up a small amount of rate on the source-destination link (compared to the point-to-point transmission) and gaining more rate through the relays. The numerical evaluation of the rates for this channel produces\(^2\)

\[
R^{PTP} = 0.2860, \\
R^{DAF} = 0.2409, \\
R^{TS-EAF} = 0.2925,
\]

where the optimal distributions that achieve these rates are summarized in tables 4.2 and 4.3. The optimal DAF distribution fixes both $X_1$ and $X_2$ to '1' and sets the probability of $X$ to be $\Pr(X = 0) \cong 0.4424$, as expected in the case where the

\(^2\)The resulting rates were obtained by optimizing for the rates with random initial input distributions. The optimization was repeated 50 times for each strategy and the maximum resulting rate was recorded. The results were also verified by a coarse grid search. The m-files used for this evaluation are available at http://cn.ece.cornell.edu.
relays limit the achievable rate. For TS-EAF, the useless relay 2 is fixed to ‘1’, to facilitate transmission with the useful relay 1. In accordance, we obtain time sharing proportions of \( q_1 \approx 0.1570 \) and \( q_2 \approx 0 \) for relay 1 and relay 2 respectively. We note that in this scenario, we actually have that even the single-relay TS-EAF outperforms the two-relay DAF.

Note that the maximum multiple-relay TS-EAF rate will always be at least as high as the PTP rate, since in the worst case, we can set the time-sharing proportions to zero, thus fixing the relays’ channel inputs. We also point out that if more than one message is transmitted from the source, e.g. two messages are transmitted simultaneously to two different destinations, DAF may reduce the rate also due to interference that relays that try to help only one receiver will create to the other receiver, who is interested in the other message.

4.2 The Gaussian Relay Channel with Coded Modulation

In this section we investigate the application of estimate-and-forward with time-sharing to the Gaussian relay channel. When applying EAF to this channel, the common practice is to use Gaussian codebooks and Gaussian quantization at the relay. The rate in Gaussian scenarios where coded modulation is employed, is usually analyzed by applying DAF or AAF at the relay. In this section we show that when considering coded modulation, one should select the relay strategy according to the channel condition: EAF with a Gaussian auxiliary RV seems a good choice when the SNR of the source-relay link is low and DAF appears to be superior when the source-relay link enjoys high SNR conditions. However, for intermediate SNR there is much room for optimizing the auxiliary mapping at the relay for the EAF strategy.
In the following we first recall the Gaussian relay channel with a Gaussian codebook, and then consider the Gaussian relay channel under BPSK modulation constraint. Since we focus on the mapping at the relay we consider here the Gaussian relay channel with an orthogonal relay-destination link of finite capacity \( C \), also considered in [31]. This scenario is depicted in figure 4.1.

Figure 4.1: The Gaussian relay channel with a finite capacity noiseless link between the relay and the destination.

Here, \( X \) is the channel input from the transmitter, \( Y_1 = g \cdot X + N_1 \) with \( g \in \mathbb{R} \), is the channel output at the relay and \( Y = X + N \) is the channel output at the receiver, which decodes the message based on \( Y^n \) and the information received from the relay through the orthogonal relay-destination link. Let \( \mathcal{N}(a, b^2) \), \( a \in \mathbb{R} \), \( b \in \mathbb{R}^+ \) denote a Gaussian RV with mean \( a \) and variance \( b^2 \). \( N_1 \) and \( N \) are independent Gaussian RVs:

\[
N_1 \sim \mathcal{N}(0, \sigma_1^2),
\]

\[
N \sim \mathcal{N}(0, \sigma^2).
\]

Hence \( X \), \( N_1 \) and \( N \) are real-valued, independent RVs. Let \( \mathcal{W} = \{1, 2, \ldots, 2^{nR}\} \) denote the source message set, and let the encoder satisfy an average power constraint
\( P: \)
\[
\frac{1}{n} \sum_{i=1}^{n} x_i^2(w) \leq P, \quad \forall w \in \mathcal{W}.
\]

The relay compresses its channel output \( Y_1^n \) into \( \hat{Y}_1^n \) using the EAF scheme. The relay then sends to the destination information through a noiseless link of finite capacity \( C \) to facilitate decoding of \( \hat{Y}_1^n \) at the destination. For this scenario the expressions of [15, theorem 6] specialize to

\[
R \leq I(X; Y, \hat{Y}_1) \tag{4.10a}
\]
\[
\text{subject to } C \geq I(\hat{Y}_1; Y_1|Y), \tag{4.10b}
\]

with the Markov chain \((X,Y) - Y_1 - \hat{Y}_1\).

We also consider in this section the DAF method whose achievable rate is given by (see [15, theorem 1])

\[
R_{DAF} = \min \{ I(X; Y_1), I(X; Y) + C \},
\]

and the upper bound of [15, theorem 3]:

\[
R_{upper} = \sup_{p(x)} \left\{ \min \{ I(X; Y) + C, I(X; Y, Y_1) \} \right\}.
\]

We note that although these expressions were originally derived for the finite and discrete alphabets case, following the method of Wyner in [109], they hold also for arbitrary sources and in particular for the Gaussian and the Gaussian-mixture cases.
4.2.1 The Gaussian Relay Channel with a Gaussian Source

Codebook

When $X \sim \mathcal{N}(0, P)$, i.i.d., then the channel outputs at the relay and the receiver are jointly Normal RVs:

$$\begin{pmatrix} Y \\ Y_1 \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P + \sigma^2 & gP \\ gP & g^2 P + \sigma_1^2 \end{pmatrix} \right).$$

The EAF compression is implemented by adding to $Y_1$ a zero mean independent Gaussian RV, $N_Q$:

$$\hat{Y}_1 = Y_1 + N_Q, \quad N_Q \sim \mathcal{N}(0, \sigma_Q^2). \tag{4.11}$$

We refer to the assignment (4.11) as Gaussian-quantization estimate-and-forward (GQ-EAF). Evaluating the mutual information expressions in (4.10a) and (4.10b) with assignment (4.11) results in (see also [31]):

$$I(X; Y, \hat{Y}_1) = \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma^2} + \frac{g^2 P}{\sigma_1^2 + \sigma_Q^2} \right) \tag{4.12a}$$

$$I(Y_1; \hat{Y}_1 | Y) = \frac{1}{2} \log_2 \left( 1 + \frac{\sigma_1^2 (\sigma^2 + P) + g^2 \sigma^2 P}{\sigma_Q^2 (P + \sigma^2)} \right). \tag{4.12b}$$

The feasibility condition (4.10b), combined with (4.12b), yields

$$\sigma_Q^2 \geq \frac{\sigma_1^2 (\sigma^2 + P) + g^2 \sigma^2 P}{(2^2C - 1)(P + \sigma^2)},$$

and because maximizing the rate (4.12a) requires minimizing $\sigma_Q^2$, the resulting GQ-EAF rate expression is

$$R \leq \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma^2} + \frac{g^2 P}{\sigma_1^2 + \sigma_Q^2} \left( \frac{\sigma_1^2 (\sigma^2 + P) + g^2 \sigma^2 P}{(2^2C - 1)(P + \sigma^2)} \right) \right).$$

Now, when using Gaussian quantization at the relay it is obvious that time-sharing does not help: the minimum $\sigma_Q^2$ is required in order to maximize the rate (4.12a).
This minimum is achieved only when the entire capacity of the relay-destination link is dedicated to the transmission of the (minimally) quantized $Y_1$. However, when we consider the Gaussian relay channel with coded modulation, the situation is quite different, as we show in the remaining of this section.

### 4.2.2 The Gaussian Relay Channel with Coded Modulation

Consider the Gaussian relay channel in which $X$ is an equiprobable BPSK signal of amplitude $\sqrt{P}$:

$$\Pr(X = \sqrt{P}) = \Pr(X = -\sqrt{P}) = \frac{1}{2}. \quad (4.13)$$

For this choice of $X$, the received symbols $(Y, Y_1)$ are no longer jointly Gaussian, but follow a Gaussian-mixture distribution:

$$f(y, y_1) = \Pr(X = \sqrt{P}) f(y, y_1 | x = \sqrt{P}) + \Pr(X = -\sqrt{P}) f(y, y_1 | x = -\sqrt{P})$$

$$= \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) G_{y_1}(g \sqrt{P}, \sigma_1^2) + G_y(-\sqrt{P}, \sigma^2) G_{y_1}(-g \sqrt{P}, \sigma_1^2) \right),$$

where

$$G_x(a, b) \triangleq \frac{1}{\sqrt{2\pi b}} e^{-\frac{(x-a)^2}{2b}}. \quad (4.14)$$

Contrary to the Gaussian codebook case, where it is hard to identify a mapping $p(\hat{y}_1|y_1)$ that will outperform Gaussian quantization, in this case it is a natural question to compare the Gaussian mapping of (4.11), which induces a Gaussian-mixture distribution on $\hat{Y}_1$ with other mappings. In the case of binary inputs it is natural to consider ternary mappings for $\hat{Y}_1$. We can predict that such mappings will perform well at high SNR on the source-relay link, when the probability of error for symbol-by-symbol detection at the relay is small, with a much smaller
complexity than Gaussian quantization. We start by considering two types of hard-decision (HD) mappings:

1. The first mapping is HD-EAF: The relay first makes a hard decision about every received $Y_1$ symbol, determining whether it is positive or negative, and then randomly decides whether to transmit this decision or transmit an erasure symbol, $E$, instead (note: this is only a qualitative explanation. In practice joint-typicality over blocks of $n$ symbols is used). The probability of transmitting an erasure, $1 - P_{\text{no erase}}$, is used to adjust the conference rate such that the feasibility constraint is satisfied. Therefore, the conditional distribution $p(\hat{y}_1|y_1)$ is given by:

$$
p(\hat{y}_1|y_1 > 0) = \begin{cases} P_{\text{no erase}}, & 1 \\ 1 - P_{\text{no erase}}, & E \end{cases} (4.15a)$$

$$
p(\hat{y}_1|y_1 \leq 0) = \begin{cases} P_{\text{no erase}}, & -1 \\ 1 - P_{\text{no erase}}, & E \end{cases} . (4.15b)$$

This choice is motivated by the time-sharing method considered in chapter 2: after making a hard decision on the received symbol’s sign – positive or negative, the relay applies TS to that decision so that the rate required to transmit the resulting random variable to the destination is less than $C$. This facilitates transmission to the destination through the conference link. Since the entropy of the sign decision is 1, then when $C \geq 1$ we can transmit the sign decisions directly without using an erasure. Therefore, we expect that for values of $C$ in the range $C > 1$, this mapping will not produce a rate increase over the rate obtained for $C = 1$. The focus is, therefore, on values of $C$ that are less than 1. The expressions for this assignment are given in appendix D.1.
2. The second mapping is deterministic hard-decision. In this approach, we select a threshold $T$ such that the range of $Y_1$ is divided into three regions: $Y_1 < -T$, $-T \leq Y_1 \leq T$ and $Y_1 > T$. Then, according to the value of each received $Y_1$ symbol, the corresponding $\hat{Y}_1$ is deterministically selected:

$$\hat{Y}_1 = \begin{cases} 
1, & Y_1 > T \\
E, & -T \leq Y_1 \leq T \\
-1, & Y_1 < -T 
\end{cases} \tag{4.16}$$

The threshold $T$ is selected such that the achievable rate is maximized subject to satisfying the feasibility constraint. We refer to this method as deterministic HD (DHD). Therefore, this is another type of TS in which the erasure probability is determined by the fraction of the time the relay input is between $-T$ and $T$. The expressions for evaluating the rate of the DHD assignment are given in appendix D.2.

The DHD method should be better than HD-EAF at high source-relay SNR since for HD-EAF, erasure is selected without any regard to the quality of the decision – both high quality sign decisions and low quality sign decisions are erased with the same probability. In contrast, in DHD, the erased area is the area where the decisions have low quality in the first place and all high quality decisions are sent. However, at low source-relay SNR and small capacity for the relay-destination link, HD-EAF may perform better than DHD since the erased area (i.e. the region between $-T$ and $+T$) for the DHD mapping has to be very large in order to facilitate transmission of the estimate through the relay-destination link, while HD-EAF may require less compression of the HD output. This is because the erasure symbol in DHD carries information while in HD-EAF it does not. Therefore, DHD requires more bandwidth for transmission of this information to
the destination.

We note here that there is related work that considers transmission based on symbol-by-symbol decisions at the relay. One such paper is [32] which compares the performance obtained by soft and hard symbol-by-symbol decisions at the relay with AAF. Another work we note is [110] in which the bit-error rate (BER) is compared for DAF and AAF where the focus is on the detector structure at the destination receiver.

We now examine the performance of each mapping using numerical evaluation: first, we examine the information rates with HD-EAF. The expressions are evaluated for $\sigma_1^2 = \sigma^2 = 1$ and $P = 1$. For every pair of values $(g, C)$ considered, the $P_{\text{no erase}}$ that maximizes the rate was selected. Figure 4.2 depicts the information rate vs. $g$ for $0.4 \leq C \leq 2$, together with the upper bound and the decode-and-forward rate. As can be observed from figure 4.2, the information rate of HD-EAF increases with $C$ until $C = 1$, after which it stops increasing with $C$. It is also seen that for small values of $g$, HD-EAF is better than DAF. This region of $g$ increases with $C$, and for $C \geq 1$ the crossover value of $g$ is approximately 1.71.

Next, examine DHD: as can be seen from figure 4.3, for small values of $C$, DAF exceeds the information rate of DHD for values of $g$ greater than 1, but for $C \geq 0.8$, DHD is superior to DAF, and in fact DAF approaches DHD from below. Another phenomenon obvious from the figure (examine $C = 0.8$ for example), is the existence of a threshold: for low values of $C$ there is some $g$ at which the DHD rate exhibits a jump. This can be explained by looking at figure 4.4, which depicts the values of $I(X;Y,\hat{Y}_1)$ and $I(\hat{Y}_1;Y_1|Y)$ vs. the threshold $T$: the bold-solid line of $I(\hat{Y}_1;Y_1|Y)$ can intersect the horizontal bold dashed line representing $C = 0.8$ at two values of $T$ at the most. We also note that for small $T$ the value of $I(X;Y,\hat{Y}_1)$
Figure 4.2: Information rate with BPSK for HD-EAF mapping at the relay vs. source-relay channel gain $g$, for different values of $C$.

is generally larger than for large $T$. Now, the jump can be explained as follows: as shown in appendix D.2.1, for small $T$ and $g$, $I(\hat{Y}_1; Y_1 | Y)$ is bounded from below. If this bound value is greater than $C$ then the intersection will occur only at a large value of $T$ (see left-hand side of figure 4.4), hence the low information rate. When $g$ increases, the value of $I(\hat{Y}_1; Y_1 | Y)$ for small $T$ decreases accordingly, until at some $g$ it intersects $C$ for a small $T$ as well as for a large $T$, as indicated by the arrow in the right-hand part of figure 4.4. This allows us to obtain the rates in the region of small $T$ which are in general higher than the rates for large $T$ and this is the source of the jump in the achievable rate.
Figure 4.3: Information rate with BPSK, for DHD mapping at the relay vs. source-relay channel gain $g$, for different values of $C$. 
Figure 4.4: $I(\hat{Y}_1; Y_1|Y)$ and $I(X; Y, \hat{Y}_1)$ vs. threshold $T$ for $(g, C) = (0.4, 0.8)$ (left) and $(g, C) = (1.4, 0.8)$ (right). The bold solid line represents $I(\hat{Y}_1, Y_1|Y)$, the horizontal bold dashed line represents $C = 0.8$, $I(X; Y, \hat{Y}_1)$ is represented by the dash-dot line and the resulting information rate is depicted with the solid line.
4.2.3 Time-Sharing Deterministic Hard-Decision (TS-DHD)

It is clearly evident from the above numerical evaluation that none of the two mappings, HD-EAF and DHD, is universally better than the other: when \( g \) is small and \( C \) is less than 1, then HD-EAF performs better than DHD, since the erased region in DHD is too large, and when \( g \) increases, DHD performs better than HD-EAF since it erases only the low quality decisions. It is therefore natural to consider a third mapping which combines both aspects of ternary mapping at the relay, namely deterministically erasing low quality decisions and then randomly gating (i.e. TS) the resulting discrete variable (in the region \( |Y_1| > T \)) in order to facilitate its transmission over the relay-destination link. This hybrid mapping is given in the following equations:

\[
p(\hat{y}_1|y_1 > T) = \begin{cases} 
P_{\text{no erase}}, & 1 \\ 1 - P_{\text{no erase}}, & E 
\end{cases} \tag{4.17a}
\]

\[
p(\hat{y}_1 = E | |y_1| \leq T) = 1 \tag{4.17b}
\]

\[
p(\hat{y}_1|y_1 < -T) = \begin{cases} 
P_{\text{no erase}}, & -1 \\ 1 - P_{\text{no erase}}, & E 
\end{cases} \tag{4.17c}
\]

In this mapping, the region \( |Y_1| \leq T \) is always erased, and the complement region is erased with probability \( 1 - P_{\text{no erase}} \). Of course, now both \( T \) and \( P_{\text{no erase}} \) have to be optimized. The expressions for TS-DHD can be found in appendix D.3. Figure 4.5 compares the performance of DHD, HD-EAF and TS-DHD. As can be seen, the hybrid method enjoys the benefits of both mappings and is the superior method.

Next, figure 4.6 compares the performance of TS-DHD, GQ-EAF and DAF (the expressions for GQ-EAF can be found in appendix D.4). As can be seen from the figure, Gaussian quantization is not always the optimal choice: for \( C = 0.6 \)
Figure 4.5: Information rate with BPSK for HD-EAF, DHD and TS-DHD mappings at the relay vs. source-relay channel gain $g$, for different values of $C$.

The lines with diamond-shaped markers) we see that GQ-EAF is the best method for $g < 1.05$, for $1.05 < g < 1.55$ TS-DHD is the best method and for $g > 1.55$ DAF achieves the highest rate. For $C = 1$ (x-shaped markers) TS-DHD is superior to both GQ-EAF and DAF for $g > 0.9$ and for $C = 2$, GQ-EAF is the superior method for all $g \leq 2$ (it intersects with DHD for some $g > 2$). This suggests that for the practical Gaussian relay scenario, where the modulation constraint is taken into account, there is room to optimize the mapping at the relay since the choice of Gaussian quantization is not always optimal.

Lastly, figure 4.7 depicts the regions in the $g-C$ plane in which each of the methods considered here is superior, in a similar manner to [31, figure 2]. As 

\[\text{The block shapes are due to the step-size of 0.025 in the values of } g \text{ and } C \text{ at which the rates were evaluated.}\]
can be observed from the figure, in the noisy region of small $g$ and also in the region of very large $C$ (high $C/g$) GQ-EAF is superior, and in the strong relay region of medium-to-high $g$ and medium-to-high $C$ (medium $C/g$) TS-DHD is the superior method. DAF is superior for small $C$ and high $g$ (low $C/g$). We note that the region where DAF achieves capacity is obtained by numerically evaluating the upper and lower bounds on the rate.

In some sense, the TS-DHD method is a hybrid method between the DAF which makes a hard-decision on the entire block and GQ-EAF which (can be thought of as if it) makes a soft decision every symbol (although it actually works in blocks of $n$ symbols), therefore it is superior in the transition region between the region where DAF is distinctly better, and the region where GQ-EAF is distinctly superior.
Figure 4.7: The best relaying strategy (out of DAF, TS-DHD and GQ-EAF) for the Gaussian relay channel with BPSK modulation.

4.2.4 When the SNR on the Source-Destination Link Approaches $0$ ($\sigma^2 \to \infty$)

In this subsection we analyze the relaying strategies discussed in this section as the SNR on the direct link $X - Y$ approaches zero. Because TS-DHD is a hybrid method combining both DHD and HD-EAF, we analyze the behavior of the components rather than the hybrid, to gain more insight. This analysis is particularly useful when trying to numerically evaluate the rates, since as the direct-link SNR goes to zero, the computer’s numerical accuracy does not allow to numerically evaluate the general expressions.
First we note that when the SNR of the direct link \( X - Y \) approaches 0 we have that \( I(X;Y) \to 0 \) as well. To see this we write

\[
I(X;Y) = h(Y) - h(Y|X)
\]

\[
= h(Y) - h(X + N|X)
\]

\[
= h(Y) - h(N),
\]

with \( h(Y) = -\int_{-\infty}^{\infty} f(y) \log_2(f(y))dy \), and from equation (D.3)

\[
f_Y(y) = \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) + G_y(-\sqrt{P}, \sigma^2) \right)
\]

\[
= \frac{1}{2} \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\sqrt{P})^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+\sqrt{P})^2}{2\sigma^2}} \right)
\]

\[
= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \left( \frac{1}{2} e^{y\sqrt{P}/\sigma^2} + \frac{1}{2} e^{-y\sqrt{P}/\sigma^2} \right) e^{-\frac{P}{2\sigma^2}}
\]

\[
\sigma^2 \rightarrow \infty \implies \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \approx G_y(0, \sigma^2),
\]

where the approximation is in the sense that for small \(|y|\) we have \( \cosh(|y|) \approx 1 \) and for large \(|y|\), \( e^{-\frac{y^2}{2\sigma^2}} \) drives the entire expression to zero as \( e^{-\frac{y^2}{2\sigma^2}} \), for \( \sigma^2 \to \infty \). This approximation reflects the intuitive notion that as the variance increases to infinity, the two-component, symmetric Gaussian mixture resembles more and more a zero-mean Gaussian RV with the same variance. Therefore, for low SNR, the output is very close to a zero-mean Normal RV with variance \( \sigma^2 \), and \( h(Y) \approx h(N) \),\(^4\) hence

\[
I(X;Y) \xrightarrow{\sigma^2 \to \infty} 0.
\]

Note that the upper bound and the decode-and-forward rate in this case are both

\(^4\)For \( \sigma = 20 \) we have that \( \int_{-\infty}^{\infty} |f_Y(y) - G_y(0, \sigma^2)|dy < 0.001 \), for \( \sigma = 55 \), \( h(Y) - h(N) \approx 0.001 \) and for \( \sigma = 200 \), \( h(Y) - h(N) < 0.0001 \).
\[ R_{DAF} = R_{upper} = \min \{C, I(X; Y_1)\}. \]

Now, let us evaluate the rate for HD-EAF as the SNR goes to zero. From (4.10a):
\[ R \leq I(X; Y, \hat{Y}_1) = I(X; \hat{Y}_1) + I(X; Y|\hat{Y}_1), \]
and
\[ I(X; Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(Y|X, \hat{Y}_1) \]
\[ = \Pr(\hat{Y}_1 = 1)h(Y|\hat{Y}_1 = 1) + \Pr(\hat{Y}_1 = E)h(Y|\hat{Y}_1 = E) \]
\[ + \Pr(\hat{Y}_1 = -1)h(Y|\hat{Y}_1 = -1) - h(N). \]

Using appendix D, equations (D.5) – (D.7), we have
\[ h(Y|\hat{Y}_1 = 1) = -\int_{y=\infty}^{\infty} f_{Y|\hat{Y}_1}(y|\hat{y}_1 = 1) \log_2 \left( f_{Y|\hat{Y}_1}(y|\hat{y}_1 = 1) \right) dy, \]
\[ f_{Y|\hat{Y}_1}(y|\hat{y}_1 = 1) = \frac{f_{Y,Y_1}(y, y_1 > 0) Pr(\text{no erase})}{Pr(Y_1 > 0) Pr(\text{no erase})}, \]
\[ f_{Y,Y_1}(y, y_1 > 0) = \frac{1}{2} \left( f_{Y,Y_1|X}(y, y_1 > 0|x = \sqrt{P}) + f_{Y,Y_1|X}(y, y_1 > 0|x = -\sqrt{P}) \right) \]
\[ = \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) Pr(Y_1 > 0|X = \sqrt{P}) \right. \]
\[ + G_y(-\sqrt{P}, \sigma^2) (1 - Pr(Y_1 > 0|X = \sqrt{P})) \right) \]
\[ = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \left( \frac{1}{2} e^{\frac{y\sqrt{P}}{\sigma^2}} Pr(Y_1 > 0|X = \sqrt{P}) \right. \]
\[ \left. + \frac{1}{2} e^{-\frac{y\sqrt{P}}{\sigma^2}} (1 - Pr(Y_1 > 0|X = \sqrt{P})) \right) e^{-\frac{P}{2\sigma^2}} \]
\[ = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \left( \frac{1}{2} - \delta \right) e^{\frac{y\sqrt{P}}{\sigma^2}} + \left( \frac{1}{2} + \delta \right) e^{-\frac{y\sqrt{P}}{\sigma^2}} \right) e^{-\frac{P}{2\sigma^2}} \]
\[ = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \left( \frac{1}{2} \cosh \left( \frac{y\sqrt{P}}{\sigma^2} \right) - \delta \sinh \left( \frac{y\sqrt{P}}{\sigma^2} \right) \right) e^{-\frac{P}{2\sigma^2}} \]
\[ \approx \frac{1}{2} G_y(0, \sigma^2), \]
when $\sigma^2 \to \infty$ and $\delta \in [-\frac{1}{2}, \frac{1}{2}]$ is selected such that $\Pr(Y_1 > 0 | X = \sqrt{P}) = \frac{1}{2} - \delta$. The approximation in (a) is because for small $|y|$, $\sinh \left( \frac{y \sqrt{P}}{\sigma^2} \right) \approx 0$ and $\cosh \left( \frac{y \sqrt{P}}{\sigma^2} \right) \approx 1$, and for large $|y|$, both $e^{-\frac{y^2}{2\sigma^2}} \sinh \left( \frac{y \sqrt{P}}{\sigma^2} \right)$ and $e^{-\frac{y^2}{2\sigma^2}} \cosh \left( \frac{y \sqrt{P}}{\sigma^2} \right)$ behave as $e^{-\frac{y^2}{2\sigma^2}}$. Hence

$$h(Y|\hat{Y}_1 = 1) \approx -\int_{y=-\infty}^{\infty} \frac{G_y(0, \sigma^2)}{2 \Pr(Y_1 > 0)} \log_2 \left( \frac{G_y(0, \sigma^2)}{2 \Pr(Y_1 > 0)} \right) dy$$

$$= -\frac{1}{2 \Pr(Y_1 > 0)} \int_{y=-\infty}^{\infty} G_y(0, \sigma^2) \left[ \log_2 \left( G_y(0, \sigma^2) \right) - \log_2 \left( 2 \Pr(Y_1 > 0) \right) \right] dy$$

$$= \frac{1}{2 \Pr(Y_1 > 0)} \left[ h(N) + \log_2 \left( 2 \Pr(Y_1 > 0) \right) \right],$$

and using $\Pr(Y_1 > 0) = \Pr(Y_1 \leq 0) = \frac{1}{2}$ and $h(Y|\hat{Y}_1 = 1) = h(Y|\hat{Y}_1 = -1)$, we obtain

$$h(Y|\hat{Y}_1) \approx \frac{1}{2} P_{\text{no erase}} h(N) + (1 - P_{\text{no erase}}) h(N) + \frac{1}{2} P_{\text{no erase}} h(N)$$

$$= h(N).$$

Therefore, as the SNR approaches 0, $Y$ and $\hat{Y}_1$ become independent. Then, $I(X; Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(N) \approx 0$ and the information rate becomes (see appendix D.5)

$$R \leq I(X; \hat{Y}_1) = H(\hat{Y}_1) - H(\hat{Y}_1|X)$$

$$= P_{\text{no erase}} (1 - H(P_1, 1 - P_1)),$$

where $H(p)$ is the discrete entropy for the specified discrete distribution $p$ and $P_1 = \Pr(Y_1 > 0 | X = \sqrt{P})$. Now, consider the feasibility condition $C \geq I(Y_1; \hat{Y}_1|Y)$:

$$I(Y_1; \hat{Y}_1|Y) = H(\hat{Y}_1|Y) - H(\hat{Y}_1|Y_1, Y)$$

$$\overset{(a)}{=} H(\hat{Y}_1) - H(\hat{Y}_1|Y_1)$$

$$= P_{\text{no erase}},$$

119
where (a) follows from the independence of $Y$ and $\hat{Y}_1$ at low SNR, see appendix D.5. Therefore, at low SNR on the source-destination link, we set $P_{\text{no erase}} = \min \{ C, 1 \}$ and the rate becomes

$$R_{HD-EAF} = \min \{ C, 1 \} \left( 1 - H(P_1, 1 - P_1) \right).$$

For the GQ-EAF scheme we first approximate $f_{Y_1,\hat{Y}_1}(y, \hat{y}_1)$ at low SNR starting with (D.8):

$$f_{Y_1,\hat{Y}_1}(y, \hat{y}_1) = \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) G_{\hat{y}_1}(g \sqrt{P}, \sigma_1^2 + \sigma_2^2) + G_y(-\sqrt{P}, \sigma^2) G_{\hat{y}_1}(-g \sqrt{P}, \sigma_1^2 + \sigma_2^2) \right)$$

$$= \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{y^2}{2\sigma^2}} \left( \frac{1}{2} G_{\hat{y}_1}(g \sqrt{P}, \sigma_1^2 + \sigma_2^2) e^{\frac{g\sqrt{P}}{\sigma^2}} + \frac{1}{2} G_{\hat{y}_1}(-g \sqrt{P}, \sigma_1^2 + \sigma_2^2) e^{-\frac{g\sqrt{P}}{\sigma^2}} \right) e^{-\frac{y^2}{2\sigma^2}}$$

$$\approx \frac{1}{\sqrt{2\pi} \sigma^2} G_y(0, \sigma^2) f_{\hat{Y}_1}(\hat{y}_1),$$

since the behavior of this expression vs. $y$ is largely determined by $G_y(0, \sigma^2)$, and $e^{\pm \frac{g\sqrt{P}}{\sigma^2}}$ have only a negligible effect as $\sigma^2 \to \infty$. We conclude that as the direct SNR approaches 0, $Y$ and $\hat{Y}_1$ become independent also in the Gaussian mapping case. Now, the rate is given by:

$$R \leq I(X; \hat{Y}_1, Y)$$

$$= h(Y, \hat{Y}_1) - h(Y, \hat{Y}_1 | X)$$

$$= h(Y) + h(\hat{Y}_1) - h(X + N_1, gX + N_1 + N_1, gX + N_1 + N_Q | X)$$

$$= h(Y) + h(\hat{Y}_1) - h(N_1, N_1 + N_Q | X)$$

$$= h(Y) - h(N | X) + h(\hat{Y}_1) - h(N_1 + N_Q | X)$$

$$= I(X; Y) + I(X; \hat{Y}_1)$$

$$\approx I(X; \hat{Y}_1)$$

$$= h(\hat{Y}_1) - h(N_1 + N_Q).$$

(4.18)
The feasibility condition becomes:

\[ C \geq I(\hat{Y}_1; Y_1 | Y) \]
\[ = h(\hat{Y}_1 | Y) - h(\hat{Y}_1 | Y_1) \]
\[ \approx h(\hat{Y}_1) - h(N_Q), \]  \hspace{1cm} (4.19)

with

\[ f_{\hat{Y}_1}(\hat{y}_1) = \frac{1}{2} \left[ G_{\hat{g}_1}(g\sqrt{P}, \sigma^2_1 + \sigma^2_Q) + G_{\hat{g}_1}(-g\sqrt{P}, \sigma^2_1 + \sigma^2_Q) \right]. \]

Finally, for DHD, as \( \sigma^2 \to \infty \) we have

\[ I(X; \hat{Y}_1, Y) = I(X; Y) + I(X; \hat{Y}_1 | Y) \]
\[ \approx I(X; \hat{Y}_1 | Y) \]
\[ = H(\hat{Y}_1 | Y) - H(\hat{Y}_1 | Y, X) \]
\[ \overset{(a)}{=} H(\hat{Y}_1) - H(\hat{Y}_1 | X) \]
\[ = I(X; \hat{Y}_1) \]

where (a) follows from the independence of \( Y \) and \( Y_1 \) as \( \sigma^2 \to \infty \) and the fact that \( \hat{Y}_1 \) is a deterministic function of \( Y_1 \), combined with the fact that given \( X, Y_1 \) and \( Y \) are independent. The feasibility condition becomes

\[ C \geq H(\hat{Y}_1 | Y) \approx H(\hat{Y}_1). \]

Because \( I(X; \hat{Y}_1) \) is not a monotone function of \( T \) we have to optimize over \( T \) to find the actual rate.

As can be seen from the expression for HD-EAF, when the SNR on the source-destination link decreases, the capacity of the conference link acts as a scaling factor on the rate of the binary channel from the source to the relay. This is due to the TS. In figure 4.8 we plotted the information rates for DHD, HD-EAF,
Figure 4.8: Information rate with DAF, DHD, HD-EAF and GQ-EAF vs. source-relay channel gain $g$, for different values of $C$, at low SNR on the source-destination link.

GQ-EAF and DAF (which coincides with the upper bound for asymptotically low SNR on the source-destination link). Comparing the three EAF strategies we note that DHD, which at intermediate SNR on the source-relay channel performs well for $C \geq 0.8$, has the worst performance at low SNR up to $C = 1.2$. At $C = 1.2$, DHD becomes the best scheme out of the three. The jump is again explained as in figure 4.4: at low SNR on the source-relay link we get that $I(\hat{Y}_1; Y_1|Y) = H(\hat{Y}_1)$ intersects with $C$ only for $C > 1$ (see also appendix D.2.1). For $C < 1.2$ and high SNR on the source-relay link, HD-EAF outperforms both DHD and GQ-EAF. For low SNR on the source-relay link, GQ-EAF is again superior.
4.2.5 Discussion

We make the following observations:

- As noted at the beginning of this section, for low SNR on the source-relay link, GQ-EAF outperforms TS-DHD. To see why, consider the distribution of $Y_1$:

$$f_{Y_1}(y_1) = G_{y_1}(0, \sigma_1^2) \cosh \left( \frac{\sqrt{P}y_1}{\sigma_1^2} \right) e^{\frac{g^2 P}{2\sigma_1^2}}$$

$$\xrightarrow{g \to 0} G_{y_1}(0, \sigma_1^2) \left(1 - \frac{g^2 P}{2\sigma_1^2}\right),$$

where the approximation is obtained using the first order Taylor expansion, and the fact that for large values of $Y_1$, $G_{y_1}(0, \sigma_1^2)$ determines the behavior of the expression. Therefore, as $g \to 0$, $Y_1$ approaches a zero-mean Gaussian RV: $Y_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2)$. As discussed in [19, ch. 13.3.2], the rate-distortion function for quantizing a Gaussian RV is minimized by Gaussian quantization (for squared error distortion). Therefore, it should be natural to guess that GQ will perform better at low SNR on the source-relay link.

We also note that in [12, section VII-B], EAF with a Gaussian auxiliary RV and Gaussian codebooks was evaluated for the general Gaussian relay channel. It was shown that at asymptotically high relay-destination SNR, this assignment of the codebooks and the auxiliary RV achieves capacity.

- At the other extreme, as $g \to \infty$, consider the DAF strategy: as $g \to \infty$, we
have that

\[ h(Y_1) = -\frac{1}{2} \left[ G_{y_1}(g\sqrt{P}, \sigma_1^2) + G_{y_1}(-g\sqrt{P}, \sigma_1^2) \right] \times \]

\[ \log_2 \left( \frac{1}{2} \left[ G_{y_1}(g\sqrt{P}, \sigma_1^2) + G_{y_1}(-g\sqrt{P}, \sigma_1^2) \right] \right) dy_1 \]

\[ \approx 1 - \int_{y_1 = -\infty}^{\infty} \frac{1}{2} G_{y_1}(g\sqrt{P}, \sigma_1^2) \log_2 G_{y_1}(g\sqrt{P}, \sigma_1^2) dy_1 \]

\[ - \int_{y_1 = -\infty}^{\infty} \frac{1}{2} G_{y_1}(-g\sqrt{P}, \sigma_1^2) \log_2 G_{y_1}(-g\sqrt{P}, \sigma_1^2) dy_1 \]

\[ = 1 + h(N_1), \]

where the approximation is due to the fact that as \( g \to \infty \), the two Gaussian peaks in the Gaussian-mixture distribution are so far from one another that the effect of the overlap can be neglected. Therefore,

\[ I(X; Y_1) = h(Y_1) - h(Y_1|X) \approx 1 + h(N_1) - h(N_1) = 1 = H(X). \]

Hence,

\[ R_{DAF} = \min \{ I(X; Y_1), I(X; Y) + C \} = \min \{ 1, I(X; Y) + C \}, \]

which is the maximal rate. Therefore, as \( g \to \infty \) DAF provides the maximum rate.

This conclusion is in accordance with [12, section VII-B], where it was shown that for the general Gaussian relay channel, DAF achieves capacity as the source-relay SNR goes to infinity. We note that for the general Gaussian relay channel, capacity is achieved with Gaussian codebooks. Here we showed that this holds also for BPSK modulation.

- We can expect that at intermediate SNR, methods that combine elements of the “soft-decision per symbol” of GQ-EAF and the hard-decision on the
entire codeword of DAF, will be superior to both. As discussed earlier, TS-DHD is such a method. Furthermore, we believe that as the SNR decreases, increasing the cardinality of $\hat{Y}_1$ accordingly will improve the performance.

- We note that we did not make a comparison with the AAF scheme. The reason is that AAF generates an output variable $X_1$ which is a Gaussian RV. However, such a RV cannot be transmitted through a finite capacity link, therefore AAF is not applicable to this scenario.
Chapter 5

Conclusions and Future Work

As communication networks evolve, it can be expected that in future networks, nodes that are close enough to be able to communicate directly, will use this ability to help each other in reception. This motivated us to investigate node cooperation strategies from the information theoretic point of view.

We first considered the EAF scheme. We showed that performing joint-decoding at the destination, instead of the sequential decoding used in the classic EAF scheme, results in a special case of the classic EAF of [15, theorem 6]. We also derived a simpler characterization of the classic EAF result without a feasibility constraint. We then considered an application of the EAF strategy to the general broadcast channel with cooperating receivers. The receivers use noiseless links of finite capacities for conferencing. We analyzed two scenarios: first we considered the physically degraded BC – for which we derived the capacity region using a single-step conference. Next, we considered the general broadcast channel with a multi-cycle conference between the receivers. We derived an achievable rate region for the most general scenario of three independent messages, and then specialized it to single message case with a single-cycle conference. For this setup we obtained an explicit achievable rate by applying TS-EAF. This rate is superior to the rate that can be obtained using joint-decoding at the receivers. For the single message scenario we also identified two special cases where capacity is achieved: the “distinctly better” case and the deterministic BC with cooperating receivers.

Next, we applied the time-sharing assignment to the auxiliary RVs of the EAF scheme for the discrete, memoryless, multiple-relay channel. This enabled us to
obtain a practically computable achievable rate expression based on the EAF strategy. This rate can be compared against the DAF-based results, so that the strategy that allows the higher rate for a given scenario can be identified. Note that although other mappings for the auxiliary RVs of the EAF scheme may provide higher rates, our choice has the advantage of satisfying the feasibility constraints, and is thus easier to evaluate. Lastly, we showed that for the Gaussian relay channel with coded modulation, the Gaussian auxiliary RV assignment is not always optimal, and a TS-EAF which behaves like a per-symbol hard decision performs better under certain channel conditions.

5.1 Future Work

The work reported here can be extended in several directions:

1. Relaying

   (a) The relay channel under a peak power constraint: The problem of relaying in fading channel scenarios received much attention recently. Focusing on the amplify-and-forward method, we note that the results were obtained under average power constraints at the relay and the transmitter. It is well known, however, that a peak power constraint much more accurately characterizes communication systems (as energy cannot be stored in significant amounts by practical transmitters) and in particular a combined peak-and-average power constraint is required to capture all the restrictions of practical communication systems. The investigation of AAF under peak and peak-and-average power constraints is therefore an interesting research problem.
(b) **Optimizing the relay mapping for fading channels:** When analyzing the relay channel in fading scenarios we find that usually DAF, Gaussian EAF or AAF is applied. However, as [83] shows, if the received signal is non-Gaussian then, at least in the intermediate SNR region, there is much to gain by selecting a mapping of the auxiliary RV that is better “matched” to the received statistics (actually it should capture the way the relay information interacts with the signal at the destination receiver). The objective of this research is to identify the mapping that will result in the highest possible rate for these relay scenarios.

2. The Interference Channel with Relay: The interference channel received a lot of attention recently: new upper bounds have been introduced and the capacities of some special cases were found. In contrast to the multiple access channel and the broadcast channel, for the interference channel, an analysis of the impact that adding a relay node can have on the achievable rates has not been performed. Introducing such a relay node raises many interesting questions and in particular how the relay should select which receiver to help the most. This is important since in a practical network, letting all transmitters obtain the channel state information seems impractical and in fact providing such information to a central station, so that it can decide which receiver should be helped, may be a more reasonable approach.

3. The Broadcast Channel with Cooperating Decoders: We can consider a different cooperation strategy for the three messages case: we note that the common information represented by $W$ is decoded by both receivers. Now, considering a $K$-cycle conference, we can let one receiver decode $W$ after the
$K_1$’th cycle, with $K_1 \leq K$ and then use DAF to help the second receiver decode $W$. The receivers can then continue the conference conditioned on the decoded $W$ – hence they can focus the remaining conference steps on helping each other decode their private messages.

4. The Discrete Memoryless Broadcast Channel: The capacity of the DMBC has been an open problem for over 30 years. Recently, a new upper bound that improves upon the Marton upper bound form 1979 has been introduced [59]. However, The attempts to develop an upper bound that captures the essential characteristics of the BC (as are evident in the Marton rate region) based only on manipulating the Fano inequality, have not been successful so far. Therefore another approach that begins with identifying the basic limitations inherent to this scenario may provide a way to improve the upper bound.
Appendix A

Proof of Proposition 2.1

A.1 Codes Construction, Encoding and Decoding

Overview of Coding Strategy

The code construction is essentially the same as the one devised in [15, theorem 6]. We also use the same procedure for decoding at the relay. The main change is in the decoding at the receiver: instead of first decoding the relay message and then using that message in the decoding of the source message, we jointly decode both the relay and source messages. In the following we specify precisely the receiver decoding and analyze the corresponding probability of error. We transmit $B - 1$ messages using $B$ blocks, each block contains $n$ symbols. In the analysis we use the approach of [15, theorem 1] and show that for a fixed $B$ the decoding of the $i$’th message can be done with an arbitrarily small probability of error assuming no decoding error at block $i - 1$. In the following we use the notation of Cover & El-Gamal in [15] for the relay channel: $(X_1 \times X_2, p(y, y_1|x_1, x_2), Y \times Y_1)$, where $X_2$ and $Y_1$ are the channel input and output, respectively, at the relay.

Details of Coding Strategy

Fix $p(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$ and $p(\hat{y}_1|x_2, y_1)$. 

Code Construction and Transmission at the Transmitter

The transmitter generates $2^{nR}$ codewords in an i.i.d. manner according to $p(x_1(w)) = \prod_{i=1}^{n} p_{X_1}(x_{1,i}), w \in W$. For transmission of the message $w_i$ at time $i$, the trans-
mitter outputs $x_1(w_i)$.  

\section*{Code Construction at the Relay}

1. The relay generates $2^{nR_0}$ codewords in an i.i.d. manner according to $p(x_2(s)) = \prod_{i=1}^{n} p_{X_2}(x_{2,i}), s \in S, S = \{1, 2, ..., 2^{nR_0}\}$.

2. For each $s \in S$ the relay generates a codebook with $2^{nR'}$ codewords according to

$$p(\hat{y}_1(m|s)) = \prod_{i=1}^{n} p_{Y_1|X_2}(\hat{y}_{1,i}|x_{2,i}(s)), m \in M, s \in S \text{ and } M = \{1, 2, ..., 2^{nR'}\}.$$  

The p.d.f. $p(\hat{y}_1|x_2)$ is given by

$$p(\hat{y}_1|x_2) = \sum_{x_1, y_1} p(x_1|x_2)p(y, y_1|x_1, x_2)p(\hat{y}_1|x_2, y_1),$$

and by construction $p(x_1|x_2) = p(x_1)$.  

3. The relay partitions its message set $M$ into $2^{nR_0}$ subsets in a uniform and independent manner. Denote these sets with $S_s, s \in S$.

\section*{Transmission at the Relay at Time $i$}

At time $i$ the relay knows $m_{i-1}$, the relay message received at time $i - 1$. Denote with $s_i$ the partition index into which $m_{i-1}$ belongs: $m_{i-1} \in S_{s_i}$. Then, at time $i$ the relay outputs $x_2(s_i)$ to the channel.

\section*{Decoding at the Relay at Time $i$}

Since at time $i$ the relay knows $s_i$, the relay decodes $m_i$, its received message at time $i$, by looking for $m \in M$ such that $(\hat{y}_1(m|s_i), y_1(i), x_2(s_i)) \in A_c^{(n)}$, where $y_1(i)$ is the received signal at the relay at time $i$. Noting that the relay knows $s_i$,
then as in [15, theorem 6], we can make the probability that such a message \( m_i \) exists arbitrarily close to 1 by taking \( n \) large enough, as long as

\[
R' \geq I(\hat{Y}_1; Y_1 | X_2).
\]  
(A.1)

Decoding at the Destination at Time \( i \)

At time \( i \) the receiver decodes \( w_{i-1} \).

1. From \( y(i) \), the received signal at time \( i \), the receiver decodes \( s_i \) by looking for a unique \( s \in S \) such that \((x_2(s), y(i)) \in A_e^{(n)}\). From the single channel capacity theorem, see [105, Ch. 8.4], the correct \( s_i \) can be decoded with an arbitrarily small probability of error by taking \( n \) large enough as long as

\[
R_0 \leq I(X_2; Y).
\]  
(A.2)

2. The receiver now knows the set \( S_{s_i} \) into which \( m_{i-1} \) belongs. Additionally, from decoding at time \( i - 1 \) the receiver knows \( s_{i-1} \), used to generate \( m_{i-1} \).

3. The receiver generates the set \( \mathcal{L}(i-1) = \{ w : (x_1(w), y(i-1), x_2(s_{i-1})) \in A_e^{(n)} \} \).

4. The receiver now looks for a unique \( w \in \mathcal{L}(i-1) \) such that \((x_1(w), y(i-1), \hat{y}_1(m | s_{i-1}), x_2(s_{i-1})) \in A_e^{(n)} \) for some \( m \in S_{s_i} \). If such a unique \( w \) exists then it is the decoded \( \hat{w}_{i-1} \), otherwise the receiver declares an error.

Note that the decoding of \( m_{i-1} \) at the receiver is not explicitly considered here, however, as indicated in the next section, it is an outcome of the new decoding procedure. Actually, we are not interested in the correct \( m_{i-1} \) as long as \( w_{i-1} \) can be decoded correctly.
A.2 Analysis of the Probability of Error

We begin by stating the error events for decoding at time \( i \). In the following we enumerate only the error events concerned with final decoding at the receiver. The error events for decoding at the relay can be accounted for in the same way as in [15, theorem 6]. Therefore, the error events we need to account for are

- \( E_{0,i} = E_{0,i}^{'} \cup E_{0,i}^{''} \), where
  \[ E_{0,i}^' = \left\{ (x_1(w_i), x_2(s_i), y_1(i), y(i) \} \notin A_{\epsilon}^{s(n)} \right\}, \]
  \[ E_{0,i}^{''} = \left\{ (x_1(w_{i-1}), x_2(s_{i-1}), y(i-1), \hat{y}_1(m_{i-1}|s_{i-1})) \notin A_{\epsilon}^{s(n)} \right\}. \]

- \( E_{1,i} = E_{1,i}^{'} \cup E_{1,i}^{''} \), where
  \[ E_{1,i}^' = \left\{ (x_2(s_i), y(i) \} \notin A_{\epsilon}^{s(n)} \right\}, \]
  \[ E_{1,i}^{''} = \left\{ \exists s \in S, s \neq s_i, (x_2(s), y(i)) \in A_{\epsilon}^{s(n)} \right\}. \]

This error event accounts for decoding error of \( s_i \) at the receiver.

- Finally, we consider the error events in decoding \( w_{i-1} \) at the receiver: \( E_{2,i} = E_{2,i}^{'} \cup E_{2,i}^{''} \), where
  \[ E_{2,i}^' = \left\{ \exists m \in S_{s_i} \text{ s.t. } (x_1(w_{i-1}), y(i-1), \hat{y}_1(m|s_{i-1}), x_2(s_{i-1})) \in A_{\epsilon}^{s(n)} \right\}, \]
  \[ E_{2,i}^{''} = \left\{ \exists m \in S_{s_i}, \exists w \in L(i-1), w \neq w_{i-1}, (x_1(w), y(i-1), \hat{y}_1(m|s_{i-1}), x_2(s_{i-1})) \in A_{\epsilon}^{s(n)} \right\}. \]

Now, as in the analysis in [15], we analyze the probability of error at block \( i \) assuming that at block \( i - 1 \) there was no decoding error. We denote with \( F_{i-1} \) the error event at time \( i - 1 \), defined as \( F_{i-1} = \left\{ \hat{s}_{i-1} \neq s_{i-1} \right\} \cup \{ \hat{w}_{i-2} \neq w_{i-2} \} \) at the receiver, or \( \left\{ \exists m \in M \text{ such that } (\hat{y}_1(m|s_{i-1}), y_1(i-1), x_2(s_{i-1})) \in A_{\epsilon}^{s(n)} \right\} \) at the relay. Note that for the general block Markov scheme we actually do not need successful decoding of \( w_{i-2} \) or \( m_{i-2} \) at the receiver to facilitate the analysis.
• From the properties of strongly typical sequences, see [105, theorem 5.8], we can make \( \Pr(E_{0,i}') \leq \epsilon \) by taking \( n \) large enough. Next, note that by construction we have that \( X_1, X_2, Y - X_2, Y_1 - \hat{Y}_1 \) is a Markov chain. Also, from decoding the relay message at section A.1 we have that \( F_{c,i-1} \) implies \( (\hat{y}_1(m_{i-1}|s_{i-1}), y_1(i-1), x_2(s_{i-1})) \in A_{*}^{(n)} \). Therefore, by the Markov lemma [107, lemma 4.2], we have that by taking \( n \) large enough we can make \( \Pr(E_{0,i}'|F_{c,i-1}) \leq \epsilon \) for arbitrary \( \epsilon > 0 \). Hence, we can make \( \Pr(E_{0,i}|F_{c,i-1}) \leq 2 \epsilon \) by taking \( n \) large enough.

• As explained in section A.1, equation (A.2) guarantees that taking \( n \) large enough we can make \( \Pr(E_{1,i} \cap E_{0,i}'|F_{c,i-1}) \leq \epsilon \) for arbitrary \( \epsilon > 0 \). By \( E_{1,i} \cap E_{0,i}' \) we have that the receiver has the correct \( s_i \); in addition, by construction \( m_{i-1} \in S_{s_i} \), thus \( \Pr(E_{2,i} \cap E_{1,i}' \cap E_{0,i}'|F_{c,i-1}) = 0 \). Lastly, in appendix A.3 we find a set of conditions (given by equations (A.14), (A.16) and (A.17)), that when satisfied, allow making \( \Pr(E_{2,i} \cap E_{1,i}' \cap E_{0,i}'|F_{c,i-1}) \leq \epsilon \) for arbitrary \( \epsilon > 0 \) by taking \( n \) large enough. Therefore, taking \( n \) large enough we can make \( \Pr(E_{2,i} \cap E_{1,i}' \cap E_{0,i}'|F_{c,i-1}) \leq \epsilon \) for any \( \epsilon > 0 \).

Now, examine (A.2), (A.14), (A.16) and (A.17), repeated here for convenience:

\[
R_0 \leq I(X_2;Y) \quad (A.3)
\]
\[
R < I(X_1;Y,\hat{Y}_1|X_2) - 9\eta \quad (A.4)
\]
\[
R < I(X_1;Y|X_2) + I(\hat{Y}_1;Y,X_1|X_2) - R' + R_0 - 9\eta \quad (A.5)
\]
\[
I(\hat{Y}_1;Y_1|X_2) \leq R' < R_0 + I(\hat{Y}_1;Y,X_1|X_2) - 4\eta. \quad (A.6)
\]

First we note that the maximum value of \( R_0 \) allowed by (A.3) also achieves the maximum rate in (A.5) and the maximum region for \( R' \) in (A.6). Therefore, we
can substitute $R_0$ in equations (A.5) and (A.6), with its upper bound $I(X_2;Y)$ maximizing both the rate and the feasible region. Thus, we get the following constraints:

$$R < I(X_1; Y, \hat{Y}_1 | X_2) - 9\eta$$  \hspace{1cm} (A.7)

$$R < I(X_1; Y|X_2) + I(\hat{Y}_1; Y, X_1|X_2) - R' + I(X_2; Y) - 9\eta$$ \hspace{1cm} (A.8)

$$I(\hat{Y}_1; Y_1 | X_2) \leq R' < I(X_2; Y) + I(\hat{Y}_1; Y, X_1 | X_2) - 4\eta.$$ \hspace{1cm} (A.9)

Now, as long as (A.9) is satisfied, we can maximize the rate (A.8) by using the lowest possible $R'$. From (A.9) we obtain the feasible region:

$$I(X_2; Y) \geq I(\hat{Y}_1; Y_1 | X_2) - I(\hat{Y}_1; Y, X_1 | X_2)$$ \hspace{1cm} (A.10)

$$= I(\hat{Y}_1; Y_1 | X_2) - I(\hat{Y}_1; Y | X_2) - I(\hat{Y}_1; X_1 | X_2, Y)$$

$$= H(\hat{Y}_1 | X_2) - H(\hat{Y}_1 | X_2, Y_1) - H(\hat{Y}_1 | X_2) + H(\hat{Y}_1 | Y_1, X_2)$$

$$- I(\hat{Y}_1; X_1 | X_2, Y)$$

$$= I(\hat{Y}_1; Y_1 | Y, X_2) - I(\hat{Y}_1; X_1 | X_2, Y)$$ \hspace{1cm} (A.11)

$$= H(\hat{Y}_1 | Y, X_2) - H(\hat{Y}_1 | Y_1, Y, X_2) - H(\hat{Y}_1 | X_2, Y) + H(\hat{Y}_1 | X_1, X_2, Y)$$

$$= H(\hat{Y}_1 | X_1, X_2, Y) - H(\hat{Y}_1 | Y_1, Y, X_2, X_1)$$

$$= I(\hat{Y}_1; Y_1 | X_1, X_2, Y).$$ \hspace{1cm} (A.12)
Next, combining (A.7), (A.8) and the lower bound in (A.9) we obtain

\[
R \leq \min \left( I(X_1; Y|X_2) + I(\hat{Y}_1; Y, X_1|X_2) - R' + I(X_2; Y), I(X_1; Y, \hat{Y}_1|X_2) \right)
\]

\[
\leq I(X_1; Y|X_2) + \min \left( I(\hat{Y}_1; Y, X_1|X_2) - I(\hat{Y}_1; Y_1|X_2) + I(X_2; Y), I(X_1; \hat{Y}_1|Y, X_2) \right)
\]

\[
= I(X_1; Y|X_2) + \min \left( -H(\hat{Y}_1|Y, X_1, X_2) + H(\hat{Y}_1|Y_1, X_2) + I(X_2; Y), I(X_1; \hat{Y}_1|Y, X_2) \right)
\]

\[
= I(X_1; Y|X_2) + \min \left( I(X_2; Y) - I(\hat{Y}_1; Y_1|Y, X_1, X_2), I(X_1; \hat{Y}_1|Y, X_2) \right). \tag{A.13}
\]

Finally, combining (A.11) and (A.13) yields the constraints of proposition 2.1.

Therefore, we proved that the under the conditions of proposition 2.1, the probability of error for decoding at block \(i\) can be made arbitrarily small by taking \(n\) large enough, assuming no decoding error at block \(i - 1\). Hence, for any fixed \(B\) the probability of error over the entire \(B\) blocks can be made arbitrarily small by taking \(n\) large enough. \qed

### A.3 Bounding the Probability \(\text{Pr}(E''_{2,i} \cap E^c_{1,i} \cap E^c_{0,i} | F^c_{i-1})\)

We find the conditions for making \(\text{Pr}(E''_{2,i} \cap E^c_{1,i} \cap E^c_{0,i} | F^c_{i-1})\) arbitrarily small. As explained in section A.2, \(E^c_{1,i} \cap E^c_{0,i}\) and \(F^c_{i-1}\) imply that the correct \(s_i\) (for which
\(m_{i-1} \in S_{s_i}\) and the correct \(s_{i-1}\) are known at the receiver. We can therefore write

\[
\Pr(E''_{2,i} \bigcap E_{1,i}^c \bigcap E_{0,i}^c | F_{i-1}) = \\
\Pr \left( \{ \exists w \in \mathcal{L}(i-1), w \neq w_{i-1}, \right.
\]
\[
(x_1(w), y(i-1), \hat{y}_1(m_{i-1}|s_{i-1}), x_2(s_{i-1})) \in A_{e}^{(n)} \} \bigcap E_{1,i}^c \bigcap E_{0,i}^c \bigg\vert F_{i-1} \bigg) \\
+ \Pr \left( \{ \exists m \in S_{s_i}, m \neq m_{i-1}, \exists w \in \mathcal{L}(i-1), w \neq w_{i-1}, \right.
\]
\[
(x_1(w), y(i-1), \hat{y}_1(m|s_{i-1}), x_2(s_{i-1})) \in A_{e}^{(n)} \} \bigcap E_{1,i}^c \bigcap E_{0,i}^c \bigg\vert F_{i-1} \bigg)
\]
\[\triangleq \Pr(E''_{2,a,i}) + \Pr(E''_{2,b,i}).\]
We first bound $\Pr(E''_{2,a,i})$:

$$\Pr(E''_{2,a,i}) =$$

$$\Pr\left( \{ \exists w \in \mathcal{L}(i-1), w \neq w_{i-1}, (x_1(w), y(i-1), \hat{y}_1(m_{i-1}|s_{i-1}), x_2(s_{i-1})) \in A_x^{(n)} \} \cap E_{1,i}^c \cap E_{0,i}^c \middle| F_{i-1}^c \right)$$

$$= \sum_{x_2(s_{i-1}), y(i-1)} \Pr\left( \{ \exists w \in \mathcal{L}(i-1), w \neq w_{i-1},
(x_1(w), y(i-1), \hat{y}_1(m_{i-1}|s_{i-1}), x_2(s_{i-1})) \in A_x^{(n)} \} \cap E_{1,i}^c \cap E_{0,i}^c \middle| x_2(s_{i-1}), y(i-1), F_{i-1}^c \right) \times \Pr (x_2(s_{i-1}), y(i-1) \middle| x_1(w_{i-1}) \text{ transmitted}, F_{i-1}^c)$$

$$\triangleq E_{y,x_2} \left\{ \sum_{w \in \mathcal{L}(i-1), w \neq w_{i-1}} \Pr\left( \{ (x_1(w), y(i-1), \hat{y}_1(m_{i-1}|s_{i-1}), x_2(s_{i-1})) \in A_x^{(n)} \} \cap E_{1,i}^c \cap E_{0,i}^c \middle| x_2(s_{i-1}), y(i-1), F_{i-1}^c \right) \right\}$$

$$\leq E_{y,x_2} \left\{ \sum_{w \in \mathcal{L}(i-1), w \neq w_{i-1}} \Pr (\hat{y}_1 \middle| y(i-1), x_2(s_{i-1})) \middle| F_{i-1}^c \right\}$$

$$\leq E_{y,x_2} \left\{ \sum_{w \in \mathcal{L}(i-1), w \neq w_{i-1}} 2^{n(H(\hat{y}_1|x_1,x_2,y)+2\eta)} \cdot 2^{-n(H(\hat{y}_1|x_2,y)-2\eta)} \middle| F_{i-1}^c \right\}$$

$$\leq E_{y,x_2} \left\{ \left| \mathcal{L}(i-1) \right| \mid F_{i-1}^c \right\} \cdot 2^{-n(I(X_1;\hat{Y}_1|X_2,Y)-4\eta)}$$

$$\leq (1 + 2^n(R-I(X_1;Y_1|X_2)+5\eta)) \cdot 2^{-n(I(X_1;\hat{Y}_1|X_2,Y)-4\eta)}$$

$$= 2^{-n(I(X_1;\hat{Y}_1|X_2,Y)-4\eta)} + 2^n(R-I(X_1;Y_1|X_2)+9\eta),$$

138
where (a) is due to the union bound, (b) follows from the definition of the conditionally typical set in [105, Ch. 5.3] and from the fact that for \( w \in \mathcal{L}(i - 1), \ x_2(s_{i-1}), y(i - 1) \) and \( x_1(w) \) are jointly typical, and (c) follows from the properties of strongly typical sequences, see [105, Ch. 5]. Also note that \( \eta \to 0 \) as \( \epsilon \to 0 \).

Lastly, the bound on \( E_{y,x_2} \{ ||\mathcal{L}(i - 1)|| \mid F_{i-1}^c \} \) in (d) is derived in appendix A.4, and follows from (A.18). We conclude that \( \Pr (E_{2,a,i}^n) \) can be made arbitrarily small by taking \( n \) large enough if

\[
R < I(X_1; Y, \hat{Y}_1|X_2) - 9\eta. \tag{A.14}
\]
Now bound $\Pr(E''_{2,b,i})$:

$$
\Pr(E''_{2,b,i}) = \Pr\left( \left\{ x_1(w), y(i-1), \hat{y}_1(m|s_{i-1}), x_2(s_{i-1}) \in A^*_t \right\} \cap E^c_{1,i} \cap E^c_{0,i} \right| F_{i-1}^c)
$$

$$
= E_{x_2,y,\hat{y}_1} \left\{ \Pr\left( \left\{ x_1(w), y(i-1), \hat{y}_1(m|s_{i-1}), x_2(s_{i-1}) \in A^*_t \right\} \right| F_{i-1}^c \right) \right)
$$

$$
\leq E_{x_2,y,\hat{y}_1} \left\{ \sum_{w \in \mathcal{L}(i-1)} \sum_{m \in S_{s_i}} \Pr(\hat{y}_1 | x_2(s_{i-1})) \right) \right)
$$

$$
= E_{x_2,y,\hat{y}_1} \left\{ \left| \mathcal{L}(i-1) \right| \cdot \left| S_{s_i} \right| \cdot 2^{-n(H(\hat{y}_1|X_1,X_2,Y)+2\eta)} \cdot 2^{-n(H(\hat{y}_1|X_2)-2\eta)} \right| F_{i-1}^c \right) \right)
$$

$$
\leq 2^{-n(I(\hat{y}_1;Y_1|X_1|X_2)-4\eta)} \left( 1 + 2^{n(R-I(X_1;Y)|X_2)+5\eta} \right) \left[ \left| S_{s_i} \right| \right] \left| F_{i-1}^c \right)
$$

where in (a) the expectation is over the values of the actual vectors $x_2(s_{i-1})$, $y(i-1)$ and $\hat{y}_1(m_{i-1}|s_{i-1})$ received when $x_1(w_{i-1})$ is transmitted. Thus, $\hat{y}_1(m|s_{i-1})$, $m \neq m_{i-1}$ is a random vector despite the conditioning. This conditioning is used to fix $S_{s_i}$ in the expectation. Next, in (b) $\eta \to 0$ as $\epsilon \to 0$ and in (c) we used the
fact that \( \hat{y}_1(m_{i-1}|s_{i-1}) \) uniquely determines the set \( S_{s_i} \). Lastly, (d) follows along
same lines used to derive (A.18). To make \( \Pr(E''_{2,b,i}) \) arbitrarily small by taking
\( n \) large enough we need to verify that each of the exponents goes to zero as \( n \) increases:

- By construction we have that \( I(\hat{Y}_1; Y, X_1|X_2) > 0 \) since \( X_2 \) carries only partial
information on \( \hat{Y}_1 \) (namely the set into which \( m \) belongs).

- \( R - I(X_1; Y|X_2) - I(\hat{Y}_1; Y, X_1|X_2) + 9\eta < 0 \) implies

\[
R \leq I(X_1; Y|X_2) + I(\hat{Y}_1; Y, X_1|X_2) - 9\eta
= I(X_1; Y|X_2) + I(\hat{Y}_1; Y|X_2) + I(\hat{Y}_1; X_1|Y, X_2) - 9\eta
= I(X_1; Y, \hat{Y}_1|X_2) + I(\hat{Y}_1; Y|X_2) - 9\eta,
\]

which is satisfied by constraint (A.14).

- \( R' - R_0 - I(\hat{Y}_1; Y, X_1|X_2) + 4\eta < 0 \) implies \( R' < R_0 + I(\hat{Y}_1; Y, X_1|X_2) - 4\eta \).

Combining with the constraint on \( R' \) from (A.1) we obtain

\[
I(\hat{Y}_1; Y_1|X_2) \leq R' < R_0 + I(\hat{Y}_1; Y, X_1|X_2) - 4\eta. \tag{A.16}
\]

- \( R + R' - R_0 - I(X_1; Y|X_2) - I(\hat{Y}_1; Y, X_1|X_2) + 9\eta < 0 \) implies

\[
R < I(X_1; Y|X_2) + I(\hat{Y}_1; Y, X_1|X_2) - R' + R_0 - 9\eta. \tag{A.17}
\]

Therefore, when conditions (A.14), (A.16) and (A.17) are satisfied, taking \( n \) large
enough we can make \( \Pr(E''_{2,i} \cap E^c_{1,i} \cap E^c_{0,i}|F^c_{i-1}) \) arbitrarily small.

### A.4 Bounding \( E_{y,x_2} \{ \| \mathcal{L}(i - 1) \| \big| F^c_{i-1} \} \)

We now bound \( E_{y,x_2} \{ \| \mathcal{L}(i - 1) \| \big| F^c_{i-1} \} \). This derivation is very similar to [15,
equation (36)], but because there is a slight difference from that derivation, we
provide the details here for completeness. Define first the function $\varphi_i(w)$,

$$
\varphi_i(w) = \begin{cases} 
1 , & (x_1(w), y(i), x_2(s_i)) \in A_\epsilon^{(n)} \\
0 , & \text{otherwise}
\end{cases}
$$

Now we can write $||L(i-1)|| = \sum_{w=1}^{2^nR} \varphi_{i-1}(w)$. Hence

$$
E_{y,x_2} \{||L(i-1)|| \mid F^c_{i-1} \} = \sum_{w=1}^{2^nR} E_{y,x_2} \{\varphi_{i-1}(w) \mid F^c_{i-1} \}
$$

$$
= 1 + \sum_{w=1, w \neq w_{i-1}}^{2^nR} E_{y,x_2} \{\varphi_{i-1}(w) \mid F^c_{i-1} \}
$$

$$
= 1 + \sum_{w=1, w \neq w_{i-1}}^{2^nR} \Pr((x_1(w), y(i-1), x_2(s_{i-1})) \in A_\epsilon^{(n)}))
$$

$$
\leq 1 + 2^n R \cdot 2^{n(H(X_1; X_2) + \eta)} 2^{-n(H(X_1) - \eta)} 2^{-n(H(X_2) - \eta)} 2^{-n(H(Y|X_2) - \eta)}
$$

$$
= 1 + 2^{n(R - I(X_1; Y|X_2) + 5\eta)},
$$

where the probability bounds in (a) follow from the properties of strongly typical sequences, see [105, Ch. 5], and $\eta \to 0$ as $\epsilon \to 0$. Therefore, we have that

$$
E_{y,x_2} \{||L(i-1)|| \mid F^c_{i-1} \} \leq 1 + 2^{n(R - I(X_1; Y|X_2) + 5\eta)}. \quad (A.18)
$$
Appendix B

Proof of Corollary 3.3

In the following we highlight only the modifications from the general broadcast result due to the application of DAF to the last conference step from \( R_{x1} \) to \( R_{x2} \), and the fact that we transmit a single message.

Codebook Generation and Encoding at the Transmitter

The transmitter generates \( 2^{nR} \) codewords \( x \) in an i.i.d. manner according to

\[
p(x(w)) = \prod_{i=1}^{n} p(x_i(w)), \quad w \in \mathcal{W} = \{1, 2, ..., 2^{nR}\}.
\]

For transmission of the message \( w_i \) at time \( i \) the transmitter outputs \( x(w_i) \).

Codebook Generation at the \( R_{x1} \)

The \( K \) conference steps from \( R_{x1} \) to \( R_{x2} \) are carried out exactly as in section 3.3.1. The first \( K - 1 \) steps from \( R_{x2} \) to \( R_{x1} \) are carried out as in section 3.3.1. The \( K' \)th conference step from \( R_{x2} \) to \( R_{x1} \), is different from that of theorem 3.2, as after the \( K' \)th step from \( R_{x1} \) to \( R_{x2} \), \( R_{x2} \) may decode the message since \( R_{x2} \) received all the \( K \) conference messages from \( R_{x1} \). Then, \( R_{x2} \) uses decode-and-forward for its \( K' \)th conference transmission to \( R_{x1} \). Therefore, \( R_{x2} \) simply partitions \( \mathcal{W} \) into \( 2^{n\alpha C_{21}} \) subsets in a uniform and independent manner.

Encoding and Decoding at the \( K' \)th Conference Step from \( R_{x2} \) to \( R_{x1} \)

- Before the \( K' \)th conference step, \( R_{x2} \) decodes its message using his channel input and all the \( K \) conference messages received from \( R_{x1} \). This can be done with an arbitrarily small probability of error as long as (3.45b) is satisfied.
• Having decoded its message, $R_{x2}$ uses the decode-and-forward strategy to select the $K$'th conference message to $R_{x1}$. The conference capacity allocated to this step is $R_{21}^{(K)} = \alpha C_{21}$.

• Having received the $K$'th conference message from $R_{x2}$, $R_{x1}$ can now decode its message using the information received at the first $K - 1$ steps, and combining it with the information from the last step using the decode-and-forward decoding rule. This gives rise to (3.45a).

Combining All the Conference Rate Bounds

The bounds on $R_{12}^{(k)}$, $k = 1, 2, \ldots, K$ can be obtained as in section 3.3.1:

$$C_{12} = \sum_{k=1}^{K} R_{12}^{(k)} \geq I \left( \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, \ldots, \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, \ldots, \hat{Y}_2^{(K-1)}; Y_1 | Y_2 \right) + 2K \epsilon,$$

and similarly

$$(1 - \alpha)C_{21} \geq I \left( \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, \ldots, \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, \ldots, \hat{Y}_2^{(K-1)}; Y_2 | Y_1 \right) + 2K \epsilon,$$

where $(1 - \alpha)C_{21}$ is the total capacity allocated to the first $K - 1$ conference steps from $R_{x2}$ to $R_{x1}$. This provides the rate constraints on the conference auxiliary variables.
Appendix C

The Expressions of [12, theorem 3] with Time-Sharing Assignments

We give the resulting expressions of [12, theorem 3] for the two-relay channel, when the auxiliary assignments for relay 1 and relay 2 are TS assignments, to allow comparison with theorem 4.1. The purpose is to show that simply using the TS assignment in the general expression does not yield theorem 4.1.

The two-relay channel without partial decoding at the relays is obtained by setting $U_1 = \emptyset$ and $U_2 = \emptyset$ in [12, theorem 3]. Then, the achievable rate is given by (here we use our notation for numbering the relays):

$$R = I(X; \hat{Y}_1, \hat{Y}_2, Y|X_1, X_2),$$

subject to

$$I(X_1; Y|X_2) \geq I(\hat{Y}_1; Y_1|X_1, X_2, \hat{Y}_2, Y) + I(\hat{Y}_1; X_2|X_1)$$

$$I(X_2; Y|X_1) \geq I(\hat{Y}_2; Y_2|X_1, X_2, \hat{Y}_1, Y) + I(\hat{Y}_2; X_1|X_2)$$

$$I(X_1, X_2; Y) \geq I(\hat{Y}_1, \hat{Y}_2; Y_1, Y_2|X_1, X_2, Y) + I(\hat{Y}_1; X_2|X_1) + I(\hat{Y}_2; X_1|X_2).$$

for the distribution $p(x)p(x_1)p(x_2)p(y, y_1, y_2|x, x_1, x_2)p(\hat{y}_1|x_1, y_1)p(\hat{y}_2|x_2, y_2)$.

Now, using the TS assignments,

$$p(\hat{y}_1|x_1, y_1) = \begin{cases} q_1, & \hat{y}_1 = y_1 \\ 1 - q_1, & \hat{y}_1 \in \mathcal{Y}_1 \end{cases},$$

$$p(\hat{y}_2|x_2, y_2) = \begin{cases} q_2, & \hat{y}_2 = y_2 \\ 1 - q_2, & \hat{y}_2 \in \mathcal{Y}_2 \end{cases},$$
we obtain the following result: let

\[ A_1 \triangleq H(Y_1|X_1, X_2, Y) + I(Y_1; X_2|X_1) \]
\[ A_2 \triangleq H(Y_2|X_1, X_2, Y) + I(Y_2; X_1|X_2) \]
\[ A_{12} \triangleq I(Y_1; Y_2|X_1, X_2, Y). \]

Then, [12, theorem 3] with TS has the form:

\[
R = I(X; Y|X_1, X_2) + q_1(1 - q_2)I(X; Y_1|X_1, X_2, Y)
+ (1 - q_1)q_2 I(X; Y_2|X_1, X_2, Y) + q_1q_2 I(X; Y_1, Y_2|X_1, X_2, Y),
\]

subject to

\[
I(X_1; Y|X_2) \geq q_1A_1 - q_1q_2A_{12}
\]
\[
I(X_2; Y|X_1) \geq q_2A_2 - q_1q_2A_{12}
\]
\[
I(X_1, X_2; Y) \geq q_1A_1 + q_2A_2 - q_1q_2A_{12}
\]
\[
0 \leq q_1 \leq 1, \quad 0 \leq q_2 \leq 1.
\]

As can be seen, obtaining \( q_1 \) and \( q_2 \) requires solving a non-linear system of equations. This is not needed in theorem 4.1, due to fixing the decoding order.
Appendix D

Expressions for Section 4.2

D.1 Hard-Decision Estimate-and-Forward

We evaluate $I(X; \hat{Y}_1, Y)$, with $p(\hat{Y}_1|Y)$ given by (4.15a) and (4.15b) using:

$$I(X; \hat{Y}_1, Y) = I(X; \hat{Y}_1) + I(X; Y|\hat{Y}_1).$$

1. Evaluating $I(X; \hat{Y}_1)$: Note that both $X$ and $\hat{Y}_1$ are discrete RVs, therefore $I(X; \hat{Y}_1)$ can be evaluated using the discrete entropies. The conditional distribution of $\hat{Y}_1$ given $X$ is given by:

$$p(\hat{Y}_1|X = \sqrt{P}) = \begin{cases} 
    P_1 \cdot P_{\text{no erase}}, & 1 \\
    1 - P_{\text{no erase}}, & E \\
    (1 - P_1)P_{\text{no erase}}, & -1 \end{cases} \quad (D.1)$$

where

$$P_1 = \Pr(Y_1 > 0|X = \sqrt{P}).$$

$p(\hat{Y}_1|X = -\sqrt{P})$ can be obtained from $p(\hat{Y}_1|X = \sqrt{P})$ by switching 1 and $-1$ in (D.1).

2. Evaluating $I(X; Y|\hat{Y}_1)$: write first

$$I(X; Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(Y|\hat{Y}_1, X),$$

and we note that

$$h(Y|\hat{Y}_1, X) = h(X + N|\hat{Y}_1, X) = h(N|\hat{Y}_1, X) = h(N) = \frac{1}{2} \log_2(2\pi e\sigma^2).$$

Using the chain rule we write

$$h(Y|\hat{Y}_1) = p(\hat{Y}_1 = 1)h(Y|\hat{Y}_1 = 1) + p(\hat{Y}_1 = E)h(Y|\hat{Y}_1 = E) + p(\hat{Y}_1 = -1)h(Y|\hat{Y}_1 = -1),$$
\( p(\hat{Y}_1) \) can be obtained by combining (4.13) and (D.1) which results in

\[
p(\hat{Y}_1) = \begin{cases} 
\frac{1}{2} P_{\text{no erase}}, & 1 \\
1 - P_{\text{no erase}}, & E \\
\frac{1}{2} P_{\text{no erase}}, & -1 
\end{cases}
\]  
(D.2)

and we note that \( h(Y|\hat{Y}_1 = E) = h(Y) \), since erasure is equivalent to no prior information. Finally we note that by definition

\[
h(Y) = -\int_{y=-\infty}^{\infty} f(y) \log_2(f(y)) dy,
\]

\[
f(Y) = \Pr(X = \sqrt{P}) f(Y|X = \sqrt{P}) + \Pr(X = -\sqrt{P}) f(Y|X = -\sqrt{P})
\]

\[= \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) + G_y(-\sqrt{P}, \sigma^2) \right), \quad \text{(D.3)}\]

where

\[
G_x(a, b) = \frac{1}{\sqrt{2\pi b}} e^{-\frac{(x-a)^2}{2b}}. \quad \text{(D.4)}
\]

Next, we have

\[
h(Y|\hat{Y}_1 = 1) = -\int_{y=-\infty}^{\infty} f(y|\hat{Y}_1 = 1) \log_2(f(y|\hat{Y}_1 = 1)) dy \quad \text{(D.5)}
\]

\[
f(Y|\hat{Y}_1 = 1) = \frac{f(Y, \hat{Y}_1 = 1)}{\Pr(\hat{Y}_1 = 1)}
\]

\[= \frac{f(Y, Y_1 > 0) P_{\text{no erase}}}{\Pr(Y_1 > 0) P_{\text{no erase}}}
\]

\[= \frac{f(Y, Y_1 > 0)}{\Pr(Y_1 > 0)}, \quad \text{(D.6)}\]

\[
f(Y, Y_1 > 0) = \Pr(X = \sqrt{P}) f(Y, Y_1 > 0|X = \sqrt{P})
\]

\[+ \Pr(X = -\sqrt{P}) f(Y, Y_1 > 0|X = -\sqrt{P})
\]

\[= \frac{1}{2} \left( f(Y, Y_1 > 0|X = \sqrt{P}) + f(Y, Y_1 > 0|X = -\sqrt{P}) \right) \quad \text{(D.7)}
\]

Using

\[
f_{Y,Y_1}(y, y_1|x) = \mathcal{N} \left( \begin{pmatrix} x \\ g \cdot x \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix} \right) = G_y(x, \sigma^2)G_{y_1}(g \cdot x, \sigma_1^2),
\]
we obtain
\[ f(Y, Y_1 > 0|X) = \int_{y_1=0}^{\infty} f(y, y_1|x)dy_1 = G_y(x, \sigma^2) \int_{y_1=0}^{\infty} G_{y_1}(g \cdot x, \sigma_1^2)dy_1. \]

Next we need to evaluate \( I(\hat{Y}_1; Y_1|Y) = h(Y_1|Y) - h(Y_1|Y, \hat{Y}_1) \):

1. \( h(Y_1|Y) = h(Y, Y_1) - h(Y) \). Here

\[
    h(Y, Y_1) = -\int_{y = -\infty}^{\infty} \int_{y_1 = -\infty}^{\infty} f(y, y_1) \log_2(f(y, y_1))dy dy_1,
\]
\[
    f(Y, Y_1) = \frac{1}{2} \left( f(Y, Y_1|X = \sqrt{P}) + f(Y, Y_1|X = -\sqrt{P}) \right),
\]
\[
    f(Y, Y_1|X) = G_y(x, \sigma^2)G_{y_1}(g \cdot x, \sigma_1^2).
\]

2. By the definition of conditional entropy we have

\[
    h(Y_1|Y, \hat{Y}_1) = p(\hat{Y}_1 = 1)h(Y_1|Y, \hat{Y}_1 = 1) + p(\hat{Y}_1 = E)h(Y_1|Y, \hat{Y}_1 = E)
    + p(\hat{Y}_1 = -1)h(Y_1|Y, \hat{Y}_1 = -1),
\]

where \( h(Y_1|Y, \hat{Y}_1 = E) = h(Y_1|Y) \), and for \( \hat{Y}_1 = 1 \), for example, we have

\[
    h(Y_1|Y, \hat{Y}_1 = 1) = -\int_{y = -\infty}^{\infty} \int_{y_1 = -\infty}^{\infty} f(y, y_1|\hat{y}_1 = 1) \log_2(f(y_1|y, \hat{y}_1 = 1))dy dy_1.
\]

Finally, we need to derive the distributions \( f(y, y_1|\hat{y}_1 = 1) \) and \( f(y_1|y, \hat{y}_1 = 1) \). Begin with

\[
    f_{Y, Y_1|\hat{y}_1}(y, y_1|\hat{y}_1 = 1) = \frac{f_{Y, Y_1, \hat{y}_1}(y, y_1, \hat{y}_1 = 1)}{\Pr(\hat{y}_1 = 1)} = \frac{f_{Y, Y_1, \hat{y}_1}(y, y_1, y_1 > 0)P_{\text{no erase}}}{\Pr(y_1 > 0)P_{\text{no erase}}} = f(y, y_1|y_1 > 0) = \begin{cases} f_{Y, Y_1|y_1 > 0}, & y_1 > 0 \\ 0, & y_1 \leq 0 \end{cases}
\]
and due to the symmetry, \( \Pr(Y_1 > 0) = \Pr(Y_1 \leq 0) = \frac{1}{2} \). We also have

\[
f(Y_1|Y, \hat{Y}_1 = 1) = \frac{f(Y_1, Y|\hat{Y}_1 = 1)}{f(Y|\hat{Y}_1 = 1)} = \frac{f(Y_1, Y|Y_1 > 0)}{f(Y|Y_1 > 0)} = \frac{\frac{f(Y_1, Y)}{\Pr(Y_1 > 0)}}{\frac{f(Y_1)}{\Pr(Y_1 > 0)}} = \frac{f(Y_1, Y)}{f(Y_1)}\quad Y_1 > 0
\]

\[
f(Y_1|Y, \hat{Y}_1 = 1) = 0, \quad Y_1 \leq 0.
\]

### D.2 Evaluation of the Information Rate with DHD

We evaluate the achievable rate using \( I(X; Y, \hat{Y}_1) = I(X; \hat{Y}_1) + I(X; Y|\hat{Y}_1) \). The distribution of \( \hat{Y}_1 \) is given by:

\[
\Pr(\hat{Y}_1 = 1) = \Pr(Y_1 > T)
\]

\[
= \frac{1}{2} \left( \Pr(Y_1 > T|X = \sqrt{P}) + \Pr(Y_1 > T|X = -\sqrt{P}) \right)
\]

\[
= \frac{1}{2} \left( \int_{y_1 > T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1 + \int_{y_1 > T} G_{y_1}(-g\sqrt{P}, \sigma_1^2)dy_1 \right)
\]

\[
\Pr(\hat{Y}_1 = E) = \Pr(|Y_1| \leq T)
\]

\[
= \frac{1}{2} \left( \Pr(|Y_1| \leq T|X = \sqrt{P}) + \Pr(|Y_1| \leq T|X = -\sqrt{P}) \right)
\]

\[
= \frac{1}{2} \left( \int_{y_1 = -T}^{T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1 + \int_{y_1 = -T}^{T} G_{y_1}(-g\sqrt{P}, \sigma_1^2)dy_1 \right),
\]

and by symmetry, \( \Pr(\hat{Y}_1 = 1) = \Pr(\hat{Y}_1 = -1) \) and \( H(\hat{Y}_1|X = \sqrt{P}) = H(\hat{Y}_1|X = -\sqrt{P}) \). Therefore, we need the conditional distribution \( p(\hat{Y}_1|X = \sqrt{P}) \):

\[
\Pr(\hat{Y}_1 = 1|X = \sqrt{P}) = \Pr(Y_1 > T|X = \sqrt{P}) = \int_{y_1 > T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1
\]

\[
\Pr(\hat{Y}_1 = -1|X = \sqrt{P}) = \Pr(Y_1 < -T|X = \sqrt{P}) = \int_{y_1 < -T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1
\]

\[
\Pr(\hat{Y}_1 = E|X = \sqrt{P}) = 1 - \Pr(\hat{Y}_1 = 1|X = \sqrt{P}) - \Pr(\hat{Y}_1 = -1|X = \sqrt{P}).
\]
This allows us to evaluate $I(X; \hat{Y}_1) = H(\hat{Y}_1) - H(\hat{Y}_1|X)$. For evaluating $I(X; Y|\hat{Y}_1)$ note that

$$h(Y|\hat{Y}_1, X) = h(X + N|\hat{Y}_1, X) = h(N|\hat{Y}_1, X) = h(N) = \frac{1}{2} \log_2(2\pi e\sigma^2),$$

and we need only to evaluate $h(Y|\hat{Y}_1)$: by definition

$$h(Y|\hat{Y}_1) = \Pr(\hat{Y}_1 = 1) h(Y|\hat{Y}_1 = 1) + \Pr(\hat{Y}_1 = E) h(Y|\hat{Y}_1 = E) + \Pr(\hat{Y}_1 = -1) h(Y|\hat{Y}_1 = -1),$$

and note that $h(Y|\hat{Y}_1 = E) = h(Y)$.

Finally,

$$h(Y|\hat{Y}_1 = 1) = - \int_{y=-\infty}^{\infty} f(y|\hat{y}_1 = 1) \log_2(f(y|\hat{y}_1 = 1)) dy$$

$$f_{Y|\hat{Y}_1}(y|\hat{y}_1 = 1) = f(y|y_1 > T) = \frac{f(y, y_1 > T)}{\Pr(Y_1 > T)}$$

$$f_Y(y, y_1 > T) = \frac{1}{2} \left( f(y, y_1 > T|X = \sqrt{P}) + f(y, y_1 > T|X = -\sqrt{P}) \right)$$

$$= \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) \Pr(Y_1 > T|X = \sqrt{P}) + G_y(-\sqrt{P}, \sigma^2) \Pr(Y_1 > T|X = -\sqrt{P}) \right).$$

Evaluating $I(\hat{Y}_1; Y_1|Y)$ we have:

$$I(\hat{Y}_1; Y_1|Y) = H(\hat{Y}_1|Y) - H(\hat{Y}_1|Y_1, Y)$$

$$(a) \ H(\hat{Y}_1|Y)$$

$$= H(\hat{Y}_1) + h(Y|\hat{Y}_1) - h(Y),$$

where (a) is due to the deterministic mapping from $Y_1$ to $\hat{Y}_1$, and $h(Y)$ can be evaluated using (D.3).
D.2.1 DHD when $T \to 0$

As $T \to 0$ we have that $\Pr(\hat{Y}_1 = E) \to 0$ and $\hat{Y}_1$ converges in distribution to a Bernoulli RV with probability $\frac{1}{2}$. Therefore

$$f(Y, \hat{Y}_1 = 1) = \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) \Pr(Y_1 > T|X = \sqrt{P}) \right.$$  

$$\quad + G_y(-\sqrt{P}, \sigma^2) \Pr(Y_1 > T|X = -\sqrt{P}) \right)$$

$$\approx \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) \Pr(Y_1 > 0|X = \sqrt{P}) \right.$$  

$$\quad + G_y(-\sqrt{P}, \sigma^2) \Pr(Y_1 > 0|X = -\sqrt{P}) \right)$$

$$= \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2)P_+ + G_y(-\sqrt{P}, \sigma^2)(1 - P_+) \right),$$

where $P_+ = \Pr(Y_1 > 0|X = \sqrt{P})$. Now, letting $g \to 0$ we have that $P_+ \to \frac{1}{2}$ and therefore

$$f(Y|\hat{Y}_1 = 1) \xrightarrow{g \to 0, T \to 0} f(Y)$$

$$\Rightarrow h(Y|\hat{Y}_1 = 1) \xrightarrow{g \to 0, T \to 0} h(Y).$$

We conclude that as $g \to 0, T \to 0$, then $h(Y|\hat{Y}_1) \to h(Y)$ and therefore the $I(Y_1; \hat{Y}_1|Y)$ becomes

$$I(Y_1; \hat{Y}_1|Y) = H(\hat{Y}_1) + h(Y|\hat{Y}_1) - h(Y) \xrightarrow{g \to 0, T \to 0} 1$$

Using the continuity of $I(Y_1; \hat{Y}_1|Y)$ we conclude that for small values of $g$, as $T$ decreases then $I(Y_1; \hat{Y}_1|Y)$ is bounded from below. This implies that for small $g$ and small $C$ the feasibility is obtained only for a large value $T$, which in turn implies low rate.
D.3 Evaluating the Information Rate with TS-DHD

D.3.1 Evaluating $I(X; Y, \hat{Y}_1)$

We first write

$$I(X; Y, \hat{Y}_1) = I(X; \hat{Y}_1) + I(X; Y|\hat{Y}_1).$$

Evaluating $I(X; \hat{Y}_1) = H(\hat{Y}_1) - H(\hat{Y}_1|X)$ requires the marginal of $\hat{Y}_1$. Using the mapping defined in (4.17) we find the marginal distribution of $\hat{Y}_1$:

$$\Pr(\hat{Y}_1) = \begin{cases} 
1, & (1 - P_{\text{erase}}) \Pr(Y_1 > T) \\
E, & \Pr(|Y_1| \leq T) + P_{\text{erase}} \Pr(|Y_1| > T), \\
-1, & (1 - P_{\text{erase}}) \Pr(Y_1 < -T)
\end{cases},$$

where $P_{\text{erase}} = 1 - P_{\text{no erase}}$ and

$$\Pr(Y_1 > T) = \Pr(Y_1 < -T) = \int_{y_1 = T}^{\infty} \frac{1}{2} \left[ G_{y_1}(\sqrt{P}, \sigma_1^2) + G_{y_1}(-\sqrt{P}, \sigma_1^2) \right] dy_1$$

and

$$\Pr(|Y_1| < T) = \int_{y_1 = -T}^{T} \frac{1}{2} \left[ G_{y_1}(\sqrt{P}, \sigma_1^2) + G_{y_1}(-\sqrt{P}, \sigma_1^2) \right] dy_1.$$

Also, due to symmetry we have that $H(\hat{Y}_1|X = \sqrt{P}) = H(\hat{Y}_1|X = -\sqrt{P})$, and therefore we need only to find the conditional $\Pr(\hat{Y}_1|X = \sqrt{P})$:

$$\Pr(\hat{Y}_1|X = \sqrt{P}) = \begin{cases} 
1, & (1 - P_{\text{erase}}) \Pr(Y_1 > T|X = \sqrt{P}) \\
E, & \Pr(|Y_1| \leq T|X = \sqrt{P}) + P_{\text{erase}} \Pr(|Y_1| > T|X = \sqrt{P}), \\
-1, & (1 - P_{\text{erase}}) \Pr(Y_1 < -T|X = \sqrt{P})
\end{cases},$$

and we note that $f_{Y_1|X}(y_1|x = \sqrt{P}) = G_{y_1}(\sqrt{P}, \sigma_1^2)$.

Next, we need to evaluate $I(X; Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(Y|\hat{Y}_1, X)$. We first note that

$$h(Y|\hat{Y}_1, X) = h(X + N|X, \hat{Y}_1) = h(N|X, \hat{Y}_1) = h(N) = \frac{1}{2} \log_2(2\pi e\sigma_1^2).$$
Lastly, we have

\[ h(Y|\hat{Y}_1) = \Pr(\hat{Y}_1 = 1)h(Y|\hat{Y}_1 = 1) + \Pr(\hat{Y}_1 = E)h(Y|\hat{Y}_1 = E) \]

\[ + \Pr(\hat{Y}_1 = -1)h(Y|\hat{Y}_1 = -1). \]

We note that \( h(Y|\hat{Y}_1 = E) = h(Y) \) and that \( h(Y|\hat{Y}_1 = 1) \) and \( h(Y|\hat{Y}_1 = -1) \) are calculated exactly as in appendix D.2 for the DHD case.

### D.3.2 Evaluating \( I(\hat{Y}_1; Y_1|Y) \)

Begin by writing

\[ I(\hat{Y}_1; Y_1|Y) = h(\hat{Y}_1|Y_1) - h(\hat{Y}_1|Y_1, Y) \]

\[ = h(Y|\hat{Y}_1) + H(\hat{Y}_1) - h(Y) - h(\hat{Y}_1|Y_1) \]

where we used the fact that given \( Y_1 \), \( \hat{Y}_1 \) is independent of \( Y \). All the terms in the above expressions have been calculated in the previous subsection, except \( h(\hat{Y}_1|Y_1) \):

\[ h(\hat{Y}_1|Y_1) = \Pr(\hat{Y}_1 > T)h(\hat{Y}_1|Y_1 > T) + \Pr(|Y_1| \leq T)h(\hat{Y}_1||Y_1| \leq T) \]

\[ + \Pr(Y_1 < -T)h(\hat{Y}_1|Y_1 < -T) \]

\[ = \Pr(\hat{Y}_1 > T)H(P_{\text{erase}}, 1 - P_{\text{erase}}) + \Pr(\hat{Y}_1 < -T)H(P_{\text{erase}}, 1 - P_{\text{erase}}) \]

\[ = (1 - P(|Y_1| \leq T))H(P_{\text{erase}}, 1 - P_{\text{erase}}). \]

### D.4 Gaussian-Quantization Estimate-and-Forward

Here the relay uses the assignment of equation (4.11):

\[ \hat{Y}_1 = Y_1 + N_Q, \quad N_Q \sim \mathcal{N}(0, \sigma_Q^2). \]

We first evaluate

\[ I(X; Y, \hat{Y}_1) = h(Y, \hat{Y}_1) - h(Y, \hat{Y}_1|X) : \]
1. 
\[
h(Y, \hat{Y}_1) = -\int_{y=-\infty}^{\infty} \int_{\hat{y}_1=-\infty}^{\infty} f_{Y,\hat{Y}_1}(y, \hat{y}_1) \log_2 (f_{Y,\hat{Y}_1}(y, \hat{y}_1)) dy \, d\hat{y}_1
\]
\[
f_{Y,\hat{Y}_1}(y, \hat{y}_1) = \frac{1}{2} \left( G_y(\sqrt{P}, \sigma^2) G_{\hat{y}_1}(g\sqrt{P}, \sigma_1^2 + \sigma_Q^2) \\
+ G_y(-\sqrt{P}, \sigma^2) G_{\hat{y}_1}(-g\sqrt{P}, \sigma_1^2 + \sigma_Q^2) \right). \quad \text{(D.8)}
\]

2. We also have
\[
h(Y, \hat{Y}_1|X) = h(X + N, gX + N_1 + N_Q|X)
\]
\[
= h(N, N_1 + N_Q|X)
\]
\[
= h(N) + h(N_1 + N_Q)
\]
\[
= \frac{1}{2} \log_2 \left( (2\pi e \sigma_Q^2) \right).
\]

Lastly we need to evaluate
\[
I(\hat{Y}_1; Y_1|Y) = h(\hat{Y}_1|Y) - h(\hat{Y}_1|Y_1, Y) = h(\hat{Y}_1, Y) - h(Y) - h(\hat{Y}_1|Y_1, Y),
\]
where
\[
h(\hat{Y}_1|Y_1, Y) = h(Y_1 + N_Q|Y_1, Y) = h(N_Q|Y_1, Y) = h(N_Q) = \frac{1}{2} \log_2(2\pi e \sigma_Q^2).
\]
D.5 Approximation of HD-EAF for $\sigma^2 \to \infty$

Using (D.1) and (D.2) we can write

$$R \leq I(X; \hat{Y}_1) = H(\hat{Y}_1) - H(\hat{Y}_1 | X)$$

$$= H\left(\frac{1}{2}P_{\text{no erase}}, 1 - P_{\text{no erase}}, \frac{1}{2}P_{\text{no erase}}\right) - H(P_{\text{no erase}}, 1 - P_{\text{no erase}}, (1 - P_1)P_{\text{no erase}})$$

$$= -P_{\text{no erase}} \log_2 \left(\frac{1}{2}P_{\text{no erase}}\right) - (1 - P_{\text{no erase}}) \log_2 (1 - P_{\text{no erase}})$$

$$+ P_1 P_{\text{no erase}} \log_2 (P_{\text{no erase}}) + (1 - P_{\text{no erase}}) \log_2 (1 - P_{\text{no erase}})$$

$$+ (1 - P_1) P_{\text{no erase}} \log_2 ((1 - P_1)P_{\text{no erase}})$$

$$= -P_{\text{no erase}} \log_2 (P_{\text{no erase}}) + P_{\text{no erase}} + P_1 \log_2 (P_{\text{no erase}}) + (1 - P_1) \log_2 (1 - P_{\text{no erase}})$$

$$= P_{\text{no erase}} (1 + P_1 \log_2 (P_1) + (1 - P_1) \log_2 (1 - P_1))$$

$$= P_{\text{no erase}} (1 - H(P_1, 1 - P_1)).$$

$$I(Y_1; \hat{Y}_1 | Y) = h(\hat{Y}_1 | Y) - h(\hat{Y}_1 | Y_1, Y)$$

$$\approx^\text{(a)} H(\hat{Y}_1) - H(\hat{Y}_1 | Y_1)$$

$$= H\left(\frac{1}{2}P_{\text{no erase}}, 1 - P_{\text{no erase}}, \frac{1}{2}P_{\text{no erase}}\right) - H(P_{\text{no erase}}, 1 - P_{\text{no erase}})$$

$$= -2 P_{\text{no erase}} \log_2 \left(\frac{1}{2}P_{\text{no erase}}\right) - (1 - P_{\text{no erase}}) \log_2 (1 - P_{\text{no erase}})$$

$$+ P_{\text{no erase}} \log_2 (P_{\text{no erase}}) + (1 - P_{\text{no erase}}) \log_2 (1 - P_{\text{no erase}})$$

$$= P_{\text{no erase}},$$

where in (a) we used the fact that $\hat{Y}_1$ and $Y$ become independent as $\sigma^2 \to \infty$, and that given $Y_1$, $\hat{Y}_1$ is independent of $Y$. 

156
REFERENCES


with Side Information at the Decoder”. IEEE Trans. Inform. Theory, IT- 


Telecommunications Conference (NTC), New Orleans, LA, November 1981, 


Networks: An Achievable Rate Region”. IEEE Trans. Inform. Theory, IT- 


[26] F. M. J. Willems. Informationtheoretical Results for the Discrete Memoryless 


