Achievable Rates and Coding Strategies for the Two-Way Relay Channel

von Diplom-Ingenieur
Clemens Schnurr
aus Düsseldorf

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Georg Böck
Gutachter: Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
Gutachter: Prof. Dr.-Ing. Karl-Dirk Kammeyer
Gutachter: Priv.-Doz. Dr.-Ing-hab. Sławomir Stańczak

Tag der wissenschaftlichen Aussprache: 10. September 2008

Berlin 2008
D83

Zusammenfassung

Die vorliegende Arbeit untersucht Übertragungsstrategien für den Zwei-Wege Relais-Kanal. Für diesen Kanal werden neue erreichbare Ratenregionen gezeigt, und es werden Kodierungsstrategien erörtert, mit denen diese Raten erreichbar sind. In der Arbeit werden mehrere Protokolle für die Datenübertragung über einen Zwei-Wege Relais-Kanal vorgestellt und analysiert. Diese basieren auf den bekannten Strategien: Decode-and-Forward und Compress-and-Forward. Im Decode-and-Forward-Protokoll nimmt man an, dass das Relais in der Lage ist, die gesendeten Daten zu dekodieren. Die Arbeit untersucht verschiedene Szenarien für dieses Protokoll, die sich darin unterscheiden, wie die Daten an das Relais übertragen werden. Zu diesem Zweck wird die gesamte Kommunikation in zwei, drei oder vier Phasen unterteilt, in denen jeweils unterschiedliche Knoten senden und empfangen. Für jedes dieser Szenarien kann eine erreichbare Ratenregion angegeben und bewiesen werden. Die Beweise hierfür verwenden zufällige Codes. Dennoch ist es möglich, ein Kodierschema abzuleiten, das für reale Systeme relevant sein könnte. Das Schema ist optimal für Kanäle, die gewisse Symmetrie-Eigenschaften aufweisen. Für allgemeine Kanäle ist die vorgeschlagene Kodierung suboptimal; die Arbeit gewährt jedoch Einsichten, worauf beim Design eines Kodierschemas zu achten ist und durch welche Mechanismen Gewinne bei der Übertragung erzielt werden können. Wenn das Relais nicht dekodieren kann, so besteht die Möglichkeit, den Empfängern eine hinreichend gute Repräsentation des Kanalausgangs am Relais zu übermitteln. Dieser Ansatz, genannt Compress-and-Forward, ermöglicht neue Ratenregionen. Für einige Kanäle ist die so erreichbare Ratenregion größer als die, welche durch Decode-and-Forward erzielt werden kann. In der Arbeit wird eine einfache Compress-and-Forward Strategie in mehreren Schritten erweitert. Im ersten dieser Schritt wird eine komplexere Dekodierstrategie entworfen. Diese Strategie berücksichtigt alle im System auftretenden Abhängigkeiten. Dadurch kann eine größere erreichbare Ratenregion bewiesen werden. Der nächste Schritt zur Erweiterung der Ratenregion berücksichtigt die verschiedenen Informationsflüsse, die im System auftreten können. Dadurch kann die Strategie des Relais flexibel an eine unterschiedliche Qualität der Übertragungskanäle angepasst werden. Dank dieser Anpassungsfähigkeit erreicht dieses Übertragungsprotokoll für allgemeine Kanäle größere Raten im Vergleich zu den bis dahin untersuchten Protokollen. In der Arbeit wird zudem eine Überlagerung von Decode-and-Forward- und Compress-and-Forward-Techniken diskutiert. Durch die Überlagerung kann das Relais einen Teil der gesendeten Nachrichten dekodieren. Die nicht dekodierbare Information wird den Empfängern ähnlich wie im Compress-and-Forward-

Fall übermittelt. Durch die Überlagerung von Decode-and-Forward und Compress-and-Forward wird ein Abtausch der Vor- und Nachteile beider Verfahren ermöglicht. Die erreichbaren Ratenregionen für alle vorgestellten Übertragungsstrategien werden in der Arbeit durch ausführliche Beweise hergeleitet. Die Beweise lassen die Mechanismen erkennen, die zu möglichen Gewinnen bei der Kodierung im Zwei-Wege Relais-Kanal führen. In der Arbeit werden diese Mechanismen ausführlich diskutiert; in der Diskussion werden Design-Kriterien abgeleitet, die bei der Entwicklung von Codes zu berücksichtigen sein werden, um hohe Raten über einen Zwei-Wege Relais-Kanal zu erreichen.

Abstract

This thesis analyzes transmission strategies for the two-way relay channel. New achievable rate regions for this channel are proved and coding strategies are proposed to achieve these new rate regions. We consider several protocols for the two-way relay channel with half duplex nodes which are based on the well known strategies: namely decode-and-forward and compress-and-forward. The decode-and-forward protocol assumes that the relay is able to decode the messages from both terminal nodes. For this protocol several scenarios are considered, which differ in the way the messages are transmitted to the relay. Therefore the overall communication is split into two, three or four phases respectively. For each of these scenarios an achievable rate region is given. The information theoretical proof uses random coding. Nevertheless, a practical coding scheme for the two-phase setup can be derived, which is optimal for certain channels. For general channels, the proposed scheme is suboptimal. Nevertheless, the results provide insight into the question how to design codes and what mechanisms facilitate the gains achievable by using a two-way relay channel. Dropping the assumption that the relay is able to decode the data leads to a new achievable rate region. For the compress-and-forward protocol the output at the relay is compressed and transmitted to the receivers. It turns out that for some channels higher rates can be achieved with a compress-and-forward strategy at the relay as compared to the decode-and-forward approach. A simple compress-and-forward protocol is improved in several steps. First, the achievable rate region is enlarged by a more elaborated decoding procedure. This decoding procedure uses all the known statistical dependencies in the system. Finally the different flows of information occurring in the system are used to propose a protocol which allows to adapt the relaying function more flexible to the channel conditions. Hence for general channels, this strategy can achieve a higher rate compared to the protocols proposed before. Throughout the thesis, all the compress-and-forward protocols are superimposed on a decode-and-forward protocol. As a result, the relay can decode one part of the message. The complementary part of the message is transmitted using the compress-and-forward mechanism. The superposition of decode-and-forward and compress-and-forward allows to balance the advantages and disadvantages of both these protocols. For all the stated achievable rate regions detailed proofs are provided. These proofs give insight into mechanisms that allow for higher rates in the two-way relay channel. The thesis discusses these mechanisms for the proposed protocols and gives insight how a code needs to be designed to achieve the gains.

Contents

Chapter 1

Introduction: The Two-Way Relay Channel

Wireless communication is present everywhere in everyday life today. Cellular voice networks as information medium are commonplace in most people's live; and the number of subscribers is still rapidly increasing. New services such as mobile Internet, office software on handhelds, video transmission to mobile phones or mobile computers via a wireless interface or mobile Internet are emerging into mass markets. The keyword "ubiquitous computing" states the desire to have information accessible everywhere — wherever it is needed and on all devices. All these evolutions create the need for communication techniques which are capable to satisfy the associated demands for connectivity everywhere and high data rates to facilitate the services in a convenient quality.

For communication engineers these trends in todays communication pose a challenge as current technology is not capable to satisfy the demands for connectivity and high data rates as the number of subscribers increases. The network bandwidth in favorable frequency bands is finite since the low frequencies which are technically usable at present are limited. Future systems will operate at higher frequencies. But at these higher frequencies wireless transmission is more sensitive to radio propagation issues; the radio wave propagation becomes akin to the propagation of light. As a consequence classical cellular systems will have coverage problems. To overcome the coverage problem one could try to increase the transmission power. Besides the problems that arise from acceptance by the residents living near the base-stations as well as constraints given by laws and regulations, the increase of transmission power comes along with an increase of interference to other wireless connections. The number of users that can be operated can decrease due to the increased interference; furthermore, interference will diminish the data rate of the system. Therefore other solutions are favorable. Last but not least a larger transmission power leads to a higher energy consumption and therefore diminishes the battery lifetime of the mobile devices. One solution for the sketched problem might be to dramatically increase the number of base-stations. But this leads to a notable increase in infrastructure costs.

Beside cellular networks, wireless networks without infrastructure attract interest in the recent past and are already in use for smaller wireless networks. One example is the well known bluetooth technology, that allows devices to form ad-hoc networks without fixed infrastructure. In the research community wireless ad-hoc networks are discussed also for large networks, which consist of several hundred nodes. Possible applications for these networks are car-to-car-communication where nodes form a dynamic network to share data e.g. data for vehicle telematics. Another example is a sensor network, which consists of several hundred autonomous sensor nodes. These sensor nodes form a network with the task to monitor or measure certain data within an area of interest. Most of these ad-hoc networks use multi-hop protocols to communicate, i.e. the data is not transmitted directly to the receiver, but relayed via one or several intermediate nodes. The nodes in the network cooperate to achieve the best possible performance in the network. The advantage of ad-hoc multi-hop networks is that they have the ability to form a network dynamically, i.e. these networks can react if some nodes enter or leave the system. Furthermore, these networks are capable to compensate the failure of some entities in the networks, e.g. due to bad channel conditions or low battery capacity. Some other node will be used to establish the needed connection between source and sink. This redundancy of nodes makes ad-hoc networks highly robust. Also ad-hoc networks do not need any infrastructure, the nodes are usually configured and maintained by the individual users; from a provider point of view they are cheap.

These advantages of ad-hoc multihop networks could be used to overcome the coverage problem of cellular networks with only limited further infrastructure costs. Instead of a direct transmission from the base-station to the mobile device, other devices could be used to relay the transmission. The relays might either be other mobile terminals that cooperate to increase the connectivity as well as the data rate. Alternatively fixed relays can be alloted over the cells; compared to base-stations these relays can be technically less complex and need no connection to the backbone network of base-stations and access points. The advantages of ad-hoc networks, namely self configuration and robustness against link failure, could be incorporated in the protocol and would be available for the cellular network. Furthermore, a relay that is used to increase the coverage of a cellular network splits up the distance between base-station and receiver. Since the path-loss is super-linear over the distance this might lead to a decrease in transmission power and hence to a decrease of interference and less energy consumption for the mobile terminals.

The above discussion shows that we might gain from using multi-hop and relay transmission techniques for the communication in wireless networks. By using relays the need rises for new transmission techniques for this kind of channel. Of course, single user technology might be used as a first shot, but to achieve all the gain offered by relays we need to understand the channel and the mechanisms which can increase the end-to-end throughput. Even though relay channels are used in practice in many wireless communication links, the understanding of the channel is far from complete. Furthermore, if a relay channel is used in a network as in the discussion above, usually the communication will be bi-directional, i.e. we have a two-way communication, as two network entities exchange messages via a relay. Therefore the channel of interest for the above communication task is a two-way relay channel. Recently, the two-way relay channel became a hot topic in the research of communication systems. Until today, it is unknown what rates can be achieved.

This thesis analyzes this two-way relay channel. New achievable rate regions are given and coding strategies are proposed. In particular the work considers several protocols for the two-way relay channel with half duplex nodes. Details on the protocols as well as the system setup will be given in the subsequent sections. The decode-and-forward protocol assumes that the relay is able to decode the messages from both terminal nodes. For this protocol achievable rate regions are given for the case that the communication is split in two, three or four phases respectively. The information theoretical proof uses random coding. Nevertheless, a practical coding scheme for the two-phase setup can be derived which is optimal for certain channels. For general channels, the work provides insight in how to design codes and what mechanisms facilitate the gains achievable by using a two-way relay channel. Dropping the assumption that the relay is able to decode the data leads to a new achievable rate region. It turns out that for some channels higher rates can be achieved without decoding at the relay as compared to the decode-and-forward approach. The protocol, which compresses the relay's channel output and forwards it to the terminals is improved in several steps: First the achievable rate region is enlarged by a more involved decoding procedure. Finally, the different flows of information occurring in the system are used to propose a protocol, which allows a more flexible adaption of the relaying function to the channel conditions, and therefore for general channels, this strategy can achieve a higher rate compared to the protocols proposed before. For all the stated achievable rate regions detailed proofs are provided. These proofs give insight into mechanisms that offer the gain in the two-way relay channel. The work discusses these mechanisms for all proposed protocols and analyses how a code needs to be designed to facilitate the gains.

Before beginning the analysis, we introduce the two-way relay channel in the subsequent section. This channel is analyzed in detail in the Chapters 2 to 5. First, we will give a general overview of the system considered and relate the work of this thesis to other results in the literature. We give a definition of the system model and introduce concepts as well as terms and definitions that are used throughout the analysis in the next chapters. Thereafter, we define the modes of operation for the channel under consideration that this thesis focuses on. Finally, a summary of the results that are achieved in the thesis related to the two-way relay channel is given together with a short discussion.

1.1 Notation

We use capital letters (X) to indicate random variables. Realizations of random variables are denoted by lower case letters (x). An index is used to differentiate the variables occurring at different terminals, e.g. X_R is the random variable of the relay's channel input while y_1

is a realization of the channel output at terminal 1. To mark a sequence or vector of such variables we use a notation as X^n or y^n for a vector of n random variables X and realizations y respectively, $y^n_{(i)}$ is used to address the i^{th} element of such a vector. We use $p_X(x)$ for the probability distribution function (pdf) of the random variable X. The index X is skipped if the variable is clear from the context or by the argument of the function. $p_1(\cdot)$ and $p_2(\cdot)$ are used for the pdfs induced by the channels, which are fixed. $p^{(n)}(\cdot)$ indicates a pdf of a vector of random variables of length n. $Pr[\cdot]$ is the probability of an event according to the underlying pdf of the random variables in the system. $\mathbb{E}_x\{\cdot\}$ is the expectation operator taken over the statistics of x; we skip the index if it is clear from the context. For alphabets we use a calligraphic font as $\mathcal{X}$, and for other sets such as codebooks we use the same notation. In the discussion on coding schemes we use bold lower case letters as $\boldsymbol{c}$ to address codewords which are vectors in a certain alphabet. For linear codes that are use in the discussion on coding schemes we use bold capital letters as $\boldsymbol{A}$ to address matrices which are used as generator of parity check matrices.

1.2 Two-Way Communication with the Help of a Relay

Two-way communication is one of the fundamental communication scenarios in information theory. In 1961, Shannon introduced the two-way communication channel and stated the problem of communicating as effectively as possible in both directions simultaneously [14]. While single user communication considers the situation that one entity transmits a message to another entity, in two-way communication the receiver has a message for the transmitter as well. Looking at todays communication systems it is evident, that almost all communication links are two-way communication links as even a simple acknowledgment establishes two-way communication. Unfortunately most of todays communication techniques allocate separated resources — such as time or frequencies — for the two directions of communication. This is due to the fact that the two-way communication problem is not easy to tackle. Until today, the question of how much information can be transmitted via a general two-way communication channel remains open.

In [14] the system is also analyzed in a simplified version where a strong restriction was added: the encoders and the decoders at the nodes are separated. From this restriction it follows that the nodes cannot cooperate explicitely but can only exploit the statistical opportunities offered by the channel. Furthermore the use of feedback is prevented by this assumption. This constraint setup is known as the "restricted two-way channel". Shannon was able to state the capacity of this restricted two-way channel. His work on two-way channels is regarded as the first work on multi-user information theory. Compared to the knowledge of the single-user case, we have just begun to understand multi-user information theory.

In this thesis, two-way communication is considered as we think that communication links should make use of the gains offered by the opportunity to transmit information in two ways at the same time. Furthermore, the single-user scenario is nothing but a special case of a two-way communication system, where the rate of one user is set to 0 and maybe restricted by further means, e.g. a channel input alphabet restricted to only one letter. As already mentioned, two-way communication is more the norm than the exception: Acknowledgments, feedback for higher layers of the communication stack, establish two-way communication as well as a simple telephone call, the popular peer-to-peer networks, some data exchange of two users and many more examples.

Beside the two-way channel, another fundamental communication channel is used as a base for the setup considered in this thesis: A relay channel is a communication link where a relay supports the transmission of a message from one node to a third node. The relay channel was introduced in [15] and was discussed in [16], or more recently in [17] and references therein. [16] states upper bounds for the general relay channel and gives achievable rate regions. For certain channels as the degraded relay channel, the reversely degraded relay channel or the relay channel with feedback the capacity regions were proven in [16]. Furthermore this article pointed out two basic relaying strategies, namely decode-and-forward and compress-and-forward. Since then, several special cases of relay channels where considered and some results could be obtained. In the general case the capacity region of the relay channel remains unknown.

The setup we consider in this thesis is motivated by wireless networks. In wireless networks the entities of the network may use relays to communicate with each other. This can be for example due to the setup of the network if relaying via a certain node is required by the communication protocol. Another reason for relaying in wireless networks are channel conditions. First, the channel without a relay between sender and receiver may be bad conditioned due to shadowing, and second, relaying may decrease the power needed to transmit. This originates from the fact that the path loss in wireless channel increases super-linear with the distance. As a relay splits up the distance it follows that the cumulative path loss might be smaller than the path loss of a direct transmission. A third reason to use relays in wireless networks is fading. If the network allows more than one path from the transmitter to the receiver, then the information can be routed via the relays such that the current channel conditions are best. This phenomenon is known as multiuser diversity.

An example for the case of relaying due to the setup of the network is an infrastructure based communication, where two wireless devices exchange information via some router or base station. This is the case in a wireless local area network or some cellular networks. While the classical cellular network is of course also a relay network if the terminal nodes of the communication are in different cells, this setup is not of primary interest for this thesis. The gains we obtain are caused by the non-orthogonality of the wireless channels between relay and terminal nodes. Still the results apply also for this communication setup. An example for relaying to increase throughput is the case where the relay is some normal terminal node, that forwards data in a multi-hop fashion from one terminal to the other or to some base station. Such scenarios may occur in ad-hoc networks and will occur in future cellular networks where terminals are used to forward the data in a multi-hop fashion to increase coverage without further infrastructure costs. Another example is a sensor network. In sensor networks nodes may exchange their measurements in some vicinity. Thereby they could establish some cluster that can be used either for cooperate transmission to some far away sink [18, 19] [1, 2], or to do some in-network calculation [20, 21]. Typically the data exchange within the cluster will proceed via one or several relays. Furthermore the transmission to the sink is usually performed in a multi-hop fashion.

If we consider wireless networks, we need to cope with some technical constraints. Most wireless transceivers cannot transmit and receive at the same time and frequency. Because of technical reasons it is difficult and often impossible to isolate the transmitted signal from the received signal. Therefore the wireless nodes can be modeled as half-duplex nodes.

Following the above observations we combine two-way communication and relaying under the constraint of half-duplex nodes and study the resulting system. We consider a three-node network where one half-duplex node acts as a relay to enable two-way communication between two other nodes. The half-duplex constraint seems to disable the gains obtainable by classical two-way communication, as there is no simultaneous transmission between the terminal nodes possible anymore. Anyhow, by using a relay it turns out that again some gains can be achieved. Since most todays relaying protocols allocate exclusive resources for each link, they suffer from an inherent loss in spectral efficiency. Instead of treating each link as a single-user channel, one can make use of the properties of the wireless medium. Thereby the loss can be reduced significantly.

Most proposals for two-way relaying separate the communication into multiple phases. First, the information is transmitted to the relay node. Then the relay node forwards the information to its destinations. In [22], [23] Gaussian channels are considered, and the relay performs superposition encoding in the second phase. The knowledge of the first phase allows the receiving nodes to perform interference cancellation before decoding so that effectively we achieve interference-free transmission in the second phase. Another interesting approach [24], [25], [26] is based on the network coding principle [27], where the relay node performs an XOR operation on the decoded bit streams. Since a network coding approach operates on the decoded data, it does not deal with channel coding. [28] analyses the two-way relay channel with full-duplex nodes and derives upper and lower bounds on the capacity region. The reference focuses on Gaussian channels and gives achievable rates for decode-and-forward and compress-and-forward system as well as a partial decode-and-forward result.

In this thesis we apply time-division to separate the communication. A division in two phases is assumed for most of the analysis. Details on the system model will be given in the subsequent section. We consider channel coding aspects of the system at hand. In particular, we do an information theoretical analysis for the two fundamental relaying strategies decode-and-forward and compress-and-forward. It turns out that the network coding approach is only a special case of a more general decode-and-forward strategy. Furthermore we give interpretations and shed light on the coding mechanism that facilitate the gains in the system. In particular, for the decode-and-forward strategy we propose a coding framework for certain channels applicable to real systems.

The key focus in this thesis is on the broadcast phase, as this broadcast phase and especially the decoding of the received signal is the origin for the cooperative gains. It turns out that the mechanisms active in the two-way relay channel are closely related to distributed source coding. In particular we see an interesting connection to a joint source and channel coding approach for the broadcast channel based on Slepian-Wolf coding [29].

1.2.1 The Two-Way Relay Channel in this Thesis — an Outline

The treatment of the two-way relay channel in the following chapters can be outlined as follows: In Chapter 2 we consider a decode-and-forward strategy. This is the first intuitive approach to the system at hand one would investigate. It is assumed that the relay receives both messages and is able to decode all the information that is relayed. We are able to state an achievable rate region and give some advises on how one can design codes for this kind of communication system. Furthermore we analyze how the system will change if more than two phases are allowed for the communication.

In Chapter 3 we discuss a new transmission protocol which facilitates a compress-and-forward scheme [16] where the broadcast transmission is designed to make a good enough copy of the channel output at the relay node available to both receivers. In effect, the side information at the receiver can also be used to decode the multiple access channel (MAC). Again, we can state an achievable rate region for the restricted half-duplex two-way relay channel with two transmission phases.

This region is extended in several steps. First the new strategy is superimposed with the decode-and-forward approach. The resulting coding scheme partially decodes the messages of both users at the relay and forwards it to the receivers. The complement of the messages is forwarded using the compressed MAC output as data which is superimposed upon the decode-and-forward data in the broadcast channel (BC) phase. The coding scheme can be interpreted as superposition coding in both phases. The resulting rate region contains the regions obtained by compress-and-forward and decode-and-forward as special cases.

The second extension is a joint decoding mechanism. This is considered in Chapter 4. It turns out, that decoding the relay's transmission without considering the other users MAC transmission might be suboptimal. In fact, a simple example shows that by focusing on the signal transmitted to the receiver the decoding fails while a decoding that focuses on the intended signal using the relay's transmission as just another side information for the decoding succeeds. In this approach the relay's transmission and thereby the compressed MAC output is decoded correctly only as a by-product; the correct decoding is neither required nor forced by the coding or the proof. As a result one could say, that neither the relay nor the receiver care about the data transmitted to them in a direct transmission. This interpretation confirms once again, that coding in networks is not primarily about getting encoded messages through a network but about getting information to the receivers. More concrete, the relay's job changes from making a compressed MAC output available at the receivers to enabling the receiver to decode a message intended for it.

This approach of a joint decoding compress-and-forward coding scheme is superimposed with a decode-and-forward scheme. The superposition yields a region which is a superset of all other regions presented up to that point in this thesis.

Finally in Chapter 5 we consider in detail the different flows of information in the system. As a result we propose a protocol, where the relay generates three different descriptions of its channel output. These are forwarded to the receivers and allow them to decode the message intended for them.

During the analysis we will point out how the coding mechanisms work that facilitate the gains obtained by the protocols. These comments lead the way of a code design for real systems. In Chapter 2 we propose in detail a simple though optimal coding framework for a certain class of channels and the decode-and-forward protocol. For the other protocols we restrict the discussion to mechanisms and point out where difficulties arise and what opportunities are there for a practical code design. A detailed coding framework is beyond the scope of this thesis.

We cannot give a converse for any of the regions. Quite the contrary, during the analysis we will give some remarks on how one could further enlarge the achievable rate region. This leads to a more elaborated coding in the BC phase but its analysis is beyond the scope of this thesis.

1.3 System Model

A two-way relay channel consists of one relay node (labeled by R) and two terminal nodes (referred to as node 1 and node 2). The terminal nodes want to exchange messages with the help of the relay R. We assume a restricted two-way communication [14] so that the transmissions of the terminal nodes in different phases do not depend on any received signal. This constraint simplifies the analysis as its rules out the effect of a feedback and the ability of cooperation by exchange of information. Furthermore it is assumed that all nodes are constrained to operate in half-duplex, meaning that they cannot receive and transmit at the same time.

The goal of communication is to transmit a message w_1 from node 1 to node 2 and w_2 from node 2 to node 1 using the channel between the two nodes and the relay in total $n \in \mathbb{N}$ times. The focus of the analysis is on a two-phase protocol consisting of a MAC phase and a BC phase. The system setup for this two-phase protocol is given here. In Chapter 2 we take a look at protocols with more phases. The adapted system setups needed for these protocols are given in that chapter.

For the bulk of the analysis we assume two phases where $\alpha > 0$ and $\beta > 0$ with $\alpha + \beta = 1$ indicate the timesharing variables between the phases: In the first phase, node 1 transmits the codeword $X_1^{n_1}$ and node 2 transmits the codeword $X_2^{n_1}$ each of length n_1 to the relay using a channel $p_1(y_R|x_1,x_2)$ $n_1 := n_1(n) \in \mathbb{N}$ times with $\frac{n_1}{n} \to \alpha$ as $n \to \infty$. The relay node will receive the signal $Y_R^{n_1}$. In the second phase the relay node transmits $X_R^{n_2}$ to node 1 and node 2 using a channel $p_2(y_1,y_2|x_R)$ $n_2 := n_2(n) \in \mathbb{N}$ times with $\frac{n_2}{n} \to \beta$ as $n \to \infty$. These nodes will receive the signals $Y_1^{n_2}$ and $Y_2^{n_2}$, respectively. Furthermore we have $n_1 + n_2 = n$. All alphabets are discrete and of finite cardinality. All channels are assumed to be memoryless and the channels in the two different phases are assumed to be statistically independent. Therefore, we have a joint probability distribution $p(y_R,y_1,y_2|x_1,x_2,x_R) = p_1(y_R|x_1,x_2)p_2(y_1,y_2|x_R)$ which defines the relay channel considered as follows:

Definition 1.1. A discrete memoryless two-phase two-way relay channel is defined by a family $\left\{p^{(n)} : \mathcal{X}_1^{n_1} \times \mathcal{X}_2^{n_1} \times \mathcal{X}_R^{n_2} \to \mathcal{Y}_R^{n_1} \times \mathcal{Y}_2^{n_2} \times \mathcal{Y}_1^{n_2}\right\}_{n_1\in\mathbb{N},n_2\in\mathbb{N}}$, $n_1 + n_2 = n$ of probability transition functions given by $p^{(n)}\left(y_R^{n_1}, y_1^{n_2}, y_2^{n_2}|x_1^{n_1}, x_2^{n_1}, x_R^{n_2}\right) := \prod_{i=1}^{n_1} p_1\left(y_{R,(i)}^{n_1}|x_{1,(i)}^{n_1}, x_{2,(i)}^{n_1}\right) \prod_{i=1}^{n_2} p_2\left(y_{1,(i)}^{n_2}, y_{2,(i)}^{n_2}|x_{R,(i)}^{n_2}\right)$ for probability functions $p_1 : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_R$ and $p_2 : \mathcal{X}_R \to \mathcal{Y}_1 \times \mathcal{Y}_2$.

1.3.1 An Outer Bound on the Capacity Region

An outer bound on the capacity region of the restricted two-phase two-way relay channel can be obtained by applying a cut set bound [30]. For some simple channels (e.g. bit pipes) this bound is tight and can be achieved with some of the coding techniques proposed in this thesis. We state the bound here without proof. It can be used to see where improvements to the achievable rate regions obtained in this thesis might be possible.

Lemma 1.1. All pairs of achievable rates $[R_1, R_2]$ for the restricted two-phase two-way relay channel satisfy

$$\begin{aligned} R_1 \leq &\quad \min\{\alpha I(X_1;Y_R|X_2,Q), \beta I(X_R;Y_2)\} \\ R_2 \leq &\quad \min\{\alpha I(X_2;Y_R|X_1,Q), \beta I(X_R;Y_1)\} \end{aligned}$$

for some joint probability distribution $p(q)p(x_1|q)p(x_2|q)p_1(y_R|x_1,x_2)p(x_R)p_2(y_1,y_2|x_R)$ and some $\alpha,\beta \geq 0$ with $\alpha + \beta = 1$.

1.4 General Concepts

In this section we specify terms and definitions reused throughout the thesis. These are given here to prevent redundancy and to provide an overview and a clear distinction what will be covered in which of the following chapters, and how the different strategies relate to each other.

1.4.1 Modes of Operation

The channel defined in the previous section can be operated in different modes. In what follows we introduce the modes of operation considered in this thesis. In general there are two different concepts, namely decode-and-forward and compress-and-forward. The compress-and-forward strategy can be applied in different ways involving a different level of complexity. Furthermore the two concepts decode-and-forward and compress-and-forward can be superimposed. The superposition is termed partial decode-and-forward; it is possible to build partial decode-and-forward protocols with all the different compress-and-forward strategies.

1.4.1.1 Decode-and-Forward

For the decode-and-forward protocol we impose the constraint that the relay is able to decode the messages of both receivers. We will see later that this is in fact a restriction, as for the general case the relay could forward a part of the messages without decoding. It follows that in general the relay need not be able to decode the messages. If the relay can decode the messages, then the system can be seen as a sequence of two individual systems: First a MAC transmission is used to transmit the messages to the relay, thereafter a BC transmission enables the receivers to decode the message intended for them. Due to the restriction, that the nodes cannot use the already received signals, the first phase is a classical MAC; the interesting part of the transmission is BC, as here we have side information available. This is due to the fact that both receivers know the message intended for the other receiver. The analysis for this system is given in Section 2.1.

Decode-and-Forward with More than Two Phases In Section 2.3 we allow more than two phases. First we consider a system, which consists of a sequence of three BC transmissions. The first two BC transmissions initiated by the two terminal nodes are used to allow the relay to decode the messages. Furthermore these transmissions allow the other respective terminal node to gather some information, that can be used in the decoding of the message in the concluding BC transmission from the relay. As a consequence, the system now has a direct link between the terminal nodes; this may increase the system performance. In addition to this three-phase protocol we consider a four-phase setup, where a MAC phase is added to transmit the messages to the relay.

1.4.1.2 Compress-and-Forward

The general compress-and-forward protocol assumes that the relay does not try to decode the messages. Its task is to forward a compressed representation of the MAC output to both receivers. In the first simple compress-and-forward protocol which is along the lines of [16], the receivers use their own message as side information to decode the relay's transmission. Thereby the receivers can decode the compressed MAC output. Subsequently, the compressed MAC output is used to decode the other node's transmitted codeword. The compress-and-forward protocol is analyzed in Section 3.1.

1.4.1.3 Partial Decode-and-Forward

The partial decode-and-forward protocol is considered in Section 3.3. In this protocol the relay decodes a part of the messages of the users. Furthermore the MAC output is compressed. The information of both these steps is transmitted to the receives by a superposition code. The receiver can first decode the part of the message which was previously decoded by the relay. Thereafter using the own message as well as the decoded part, the compressed MAC output is recovered. Finally this compressed representative is used to decide, which message was transmitted by the other respective node.

1.4.1.4 Compress-and-Forward with Joint Decoding

In Section 4.1 the decoding at the terminal node is changed in comparison to the first considered compress-and-forward strategy. In the new protocol the receiver does not focus on decoding the relay's transmission anymore. The receiver decodes the message transmitted by the other terminal node directly. The compressed MAC output calculated at the relay can be decoded as a by-product. Compared to the compress-and-forward approach, the receiver uses more dependencies available in the system.

1.4.1.5 Partial Decode-and-Forward with Joint Decoding

Section 4.2 extends the protocol for compress-and-forward with joint decoding in the same way as Section 3.3 extends Section 3.1. In this approach the relay decodes a part of the messages, the complement information is transmitted via compress-and-forward. To decode the compress-and-forward part of the message, the receiver uses a joint decoding mechanism.

1.4.1.6 Compress-and-Forward with Three Information Flows

The compression at the relay is performed jointly for both receivers in the protocols introduced above. In Chapter 5 we extend this approach by using up to three compressed representatives, that are forwarded to one or both of the receivers. Thereby the protocol gives better opportunities for asymmetric systems, where the channels for the different receivers are of different quality. This approach is the most general compress-and-forward approach considered in this thesis. The region of rate pairs achievable by this approach contains all the other rate regions achievable by compress-and-forward protocols. By allowing up to three information flows we gain the ability to balance interference like effects at the relay with interference in the BC.

1.4.2 Some Definitions

Some definitions are used throughout the theses. These definitions apply for the two-phase protocol in all the different modes of operation. In Chapter 2 we take a look at protocols with more phases and we consider the BC of the two-way relay channel separated from the MAC. The adapted definitions needed for these analyses are given in Chapter 2.

Definition 1.2. A $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-code, $n_1 \in \mathbb{N}$, $n_2 \in \mathbb{N}$, $n_1 + n_2 = n$ for the two-phase two-way relay channel consists of an encoder at node $k \in \{1, 2\}$,

$$x_k^{n_1} : \mathcal{W}_k \to \mathcal{X}_k^{n_1}$$

with $\mathcal{W}_k = [1, 2, \ldots, M_k^{(n)}]$, an encoder at the relay node

$$x_R^{n_2} : \mathcal{Y}_R^{n_1} \to \mathcal{X}_R^{n_2}$$

and decoders at node 1 and node 2

$$\begin{aligned} g_1 &: \mathcal{Y}_1^{n_2} \times \mathcal{W}_1 \to \mathcal{W}_2 \\ g_2 &: \mathcal{Y}_2^{n_2} \times \mathcal{W}_2 \to \mathcal{W}_1. \end{aligned}$$

We use $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$ for the messages transmitted by node 1 and node 2 respectively. Furthermore $w := w(w_1, w_2) = [w_1, w_2] \in \mathcal{W} := \mathcal{W}_1 \times \mathcal{W}_2$ is used to indicate the message pair. Note that we assume independent sources in all the theorems and proofs. Furthermore, to make the definition of the average probability of error meaningful, we assume the transmitted messages are drawn independent and identically distributed (i.i.d.) from a uniform distribution over the sets of messages $\mathcal{W}_1$ and $\mathcal{W}_2$.

Definition 1.3. When $w := w(w_1, w_2) = [w_1, w_2] \in \mathcal{W} := \mathcal{W}_1 \times \mathcal{W}_2$ is the message pair transmitted by the two terminal nodes, the receiver 1 is in error if $g_1(y_1^{n_2}, w_1) \neq w_2$. The probability of this error event is denoted by

$$\lambda_1(w) := Pr[g_1(Y_1^{n_2}, w_1) \neq w_2 | w(w_1, w_2) \text{ has been sent}].$$

Accordingly the corresponding error event for the receiver 2 is denoted by

$$\lambda_2(w) := Pr[g_2(Y_2^{n_2}, w_2) \neq w_1 | w(w_1, w_2) \text{ has been sent}].$$

Definition 1.4. The average probability of decoding error at the receivers is given by

$$\mu_1^{(n)} := \frac{\sum_{j \in \mathcal{W}_2} \sum_{k \in \mathcal{W}_1} Pr[g_1(Y_1^{n_2}, k) \neq j | x_1^{n_1}(k), x_2^{n_1}(j)]}{|\mathcal{W}_1||\mathcal{W}_2|} = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \lambda_1(w)$$

for node 1 and

$$\mu_2^{(n)} := \frac{\sum_{j \in \mathcal{W}_2} \sum_{k \in \mathcal{W}_1} Pr[g_2(Y_2^{n_2}, j) \neq k | x_1^{n_1}(k), x_2^{n_1}(j)]}{|\mathcal{W}_1||\mathcal{W}_2|} = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \lambda_2(w)$$

for node 2.

Definition 1.5. Let $\mu_1^{(n)}$ and $\mu_2^{(n)}$ be the average probabilities of decoding errors at node 1 and node 2, respectively. The rate pair $[R_1, R_2]$ is said to be achievable for the two-phase two-way relay channel if there exists a sequence of $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-codes with $\frac{\log M_1^{(n)}}{n} \to R_1$ and $\frac{\log M_2^{(n)}}{n} \to R_2$ such that $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$.

1.4.3 Concept of Typical Sequences and Related Definitions

In the proofs of this thesis we will make extensive use of typical sequences and their properties. Therefore we will now give the definition and state some definitions and related notation, which are used throughout the thesis.

Definition 1.6. Given two random variables $Z_1 \in \mathcal{Z}_1$, $Z_2 \in \mathcal{Z}_2$, let $p_{Z_1}(i) = Pr[Z_1 = i]$ where $i \in \mathcal{Z}_1$ and $p_{Z_1,Z_2}(\hat{i}) = Pr\left[(Z_1, Z_2) = \hat{i}\right]$ where $\hat{i} \in \mathcal{Z}_1 \times \mathcal{Z}_2$. For any $\epsilon > 0$, we define the set of ϵ-typical sequences [31] of length n as[1]

$$\mathcal{T}_\epsilon^{(n)}(Z_1) := \left\{ z_1^n : \forall i \in \mathcal{Z}_1, |N(i|z_1^n) - np_{Z_1}(i)| \leq \frac{\epsilon n p_{Z_1}(i)}{\log(|\mathcal{Z}_1|)} \right\}$$

where $N(i|z_1^n)$ is the number of indices k such that the k^{th} element of the vector $z_1^n = (z_{1,(1)}^n, z_{1,(2)}^n, \ldots, z_{1,(n)}^n)$ equals i, i.e. $z_{1,(k)}^n = i$. Furthermore we define for a given $z_2^n \in \mathcal{Z}_2^n$ the set

$$\mathcal{T}_\epsilon^{(n)}(Z_1|z_2^n) := \left\{ z_1^n : \forall \hat{i} \in \mathcal{Z}_1 \times \mathcal{Z}_2, |N(\hat{i}|z_1^n, z_2^n) - np_{Z_1,Z_2}(\hat{i})| \leq \frac{\epsilon n p_{Z_1,Z_2}(\hat{i})}{\log(|\mathcal{Z}_1||\mathcal{Z}_2|)} \right\}.$$

We will use the properties of sets of typical sequences in the proof of the coding theorems. Some known properties [31] of ϵ-typical sequences are collected in the following lemma without proof:

Lemma 1.2. Let $Z_1 \in \mathcal{Z}_1$, $Z_2 \in \mathcal{Z}_2$ be random variables. For $\delta > 0$, $\epsilon > 0$ and for sufficiently large n it holds that:

- $Pr[Z_1^n \in \mathcal{T}_\epsilon^{(n)}(Z_1)] \geq 1 - \delta$
- For any $z_1^n \in \mathcal{T}_\epsilon^{(n)}(Z_1)$
$$\left| \frac{1}{n} \log p(z_1^n) + H(Z_1) \right| \leq \epsilon$$
- $|\mathcal{T}_\epsilon^{(n)}(Z_1)| \leq 2^{n(H(Z_1)+\epsilon)}$
- $|\mathcal{T}_\epsilon^{(n)}(Z_1)| \geq (1 - \delta) 2^{n(H(Z_1)-\epsilon)}$
- For any $z_2^n \in \mathcal{T}_\epsilon^{(n)}(Z_2)$
$$Pr[Z_1^n \in \mathcal{T}_\epsilon^{(n)}(Z_1|z_2^n) | Z_2^n = z_2^n] \geq 1 - \delta$$

[1] All logarithms in this paper are to the base 2 and we consider entropies and mutual information in bits.

- $(z_1^n, z_2^n) \in \mathcal{T}_\epsilon^{(n)}(Z_1, Z_2)$ implies $z_1^n \in \mathcal{T}_\epsilon^{(n)}(Z_1)$ and $z_2^n \in \mathcal{T}_\epsilon^{(n)}(Z_2)$

Throughout this work we will use indicator functions related to a typical set to simplify the notation. We define the indicator function with general sets; in the proofs within this work the usual application will be to typical sets and product spaces on the alphabets related to these typical sets.

Definition 1.7. Given a set $\mathcal{A}$ and a set $\mathcal{B} \subseteq \mathcal{A}$, define the indicator function $\chi : \mathcal{A} \to \{0, 1\}$ as

$$\chi_{\mathcal{B}}(a) := \begin{cases} 1 & \text{if } a \in \mathcal{B} \\ 0 & \text{otherwise} \end{cases} \tag{1.1}$$

where $a \in \mathcal{A}$.

As a small liberty of notation we use $\chi_{\mathcal{B}}^C(a)$ as a shortcut for $\chi_{\mathcal{B}^C}(a) = 1 - \chi_{\mathcal{B}}(a)$.

1.5 Summary of the Results

Now we will summerize the results of this thesis. First we will give the achievable rate regions obtained in the analysis for the different modes of operation. Thereafter we give a short discussion on further insights obtained by the analysis.

1.5.1 Achievable Rate Regions

The achievable rate regions are stated here without further comments. The reader is referred to the respective chapter for the proof, the ideas of the coding and interpretation of the results.

1.5.1.1 Decode-and-Forward

Theorem (Theorem 2.4)**.** An achievable rate region $\mathcal{R}_{\mathrm{DF}} \subset \mathbb{R}_+^2$ of the two-phase two-way relay channel is given by all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 \leq &\quad \min\{\alpha I(X_1; Y_R|X_2, Q), \beta I(X_R; Y_2)\} \\ R_2 \leq &\quad \min\{\alpha I(X_2; Y_R|X_1, Q), \beta I(X_R; Y_1)\} \\ R_1 + R_2 \leq &\quad \alpha I(X_1, X_2; Y_R|Q) \end{aligned}$$

for some joint probability distribution $p(q)p(x_1|q)p(x_2|q)p_1(y_R|x_1, x_2)p(x_R)p_2(y_1, y_2|x_R)$ and some $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$.

Decode-and-Forward with Three Phases

Theorem (Theorem 2.7)**.** An achievable rate region for the three-phase two-way relay channel using a decode-and-forward protocol is the set of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &< \min\{\alpha I(X_1; Y_{R,1}); \alpha I(X_1; Y_{2,1}) + \gamma I(X_R; Y_{2,3})\} \\ R_2 &< \min\{\beta I(X_2; Y_{R,2}); \beta I(X_2; Y_{1,2}) + \gamma I(X_R; Y_{1,3})\} \end{aligned}$$

for some joint probability distribution $p_1(y_{R,1}, y_{2,1}|x_1)p_2(y_{R,2}, y_{1,2}|x_2)p_R(y_{1,3}, y_{2,3}|x_R)p(x_1, x_2, x_R)$ and some $\alpha, \beta, \gamma \geq 0$ with $\alpha + \beta + \gamma = 1$.

Decode-and-Forward with Four Phases

Theorem (Theorem 2.8)**.** An achievable rate region for the four-phase two-way relay channel using a decode-and-forward protocol is the set of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &< \min\{\alpha I(X_{1,1}; Y_{R,1}) + \gamma I(X_{1,3}; Y_{R,3}|X_{2,3}, Q); \alpha I(X_{1,1}; Y_{2,1}) + \delta I(X_R; Y_{2,4})\} \\ R_2 &< \min\{\beta I(X_{2,2}; Y_{R,2}) + \gamma I(X_{2,3}; Y_{R,3}|X_{1,3}, Q); \beta I(X_{2,2}; Y_{1,2}) + \delta I(X_R; Y_{1,4})\} \\ R_1 + R_2 &< \alpha I(X_{1,1}; Y_{R,1}) + \beta I(X_{2,2}; Y_{R,2}) + \gamma I(X_{1,3}, X_{2,3}; Y_{R,3}|Q) \end{aligned}$$

for some joint probability distribution $p(x_{1,1})p(x_{2,2})p(x_R)p(q)p(x_{1,3}|q)p(x_{2,3}|q)p_1(y_{R,1}, y_{2,1}|x_{1,1})$ $p_2(y_{R,2}, y_{1,2}|x_{2,2})p_R(y_{1,4}, y_{2,4}|x_R)p_M(y_{R,3}|x_{1,3}, x_{2,3})$ and some $\alpha, \beta, \gamma, \delta \geq 0$ with $\alpha + \beta + \gamma + \delta = 1$.

1.5.1.2 Compress-and-Forward

Theorem (Theorem 3.1)**.** An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_1 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &\leq \alpha I(X_1; \hat{Y}_R|X_2, Q) \\ R_2 &\leq \alpha I(X_2; \hat{Y}_R|X_1, Q) \end{aligned}$$

under the constraints

$$\begin{aligned} \alpha\left(H(\hat{Y}_R|X_1, Q) - H(\hat{Y}_R|Y_R)\right) &< \beta I(Y_1; X_R) \\ \alpha\left(H(\hat{Y}_R|X_2, Q) - H(\hat{Y}_R|Y_R)\right) &< \beta I(Y_2; X_R) \end{aligned}$$

for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and for joint probability distributions $p(q)p(x_1|q)p(x_2|q)$ $p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(x_R)p_2(y_1, y_2|x_R)$.

Corollary (Corollary 3.2)**.** An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_2 \subset \mathbb{R}_+^2$ of all rate pairs $[0, R_2]$ satisfying

$$R_2 \leq \alpha I(X_2; \hat{Y}_R|X_1, Q)$$

under the constraint

$$\alpha\left(H(\hat{Y}_R|X_1,Q)-H(\hat{Y}_R|Y_R)\right)<\beta I(Y_1;X_R)$$

and similarly the set $\mathcal{R}_3 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, 0]$ which satisfy

$$R_1 \le \alpha I(X_1;\hat{Y}_R|X_2,Q)$$

under the constraint

$$\alpha\left(H(\hat{Y}_R|X_2,Q)-H(\hat{Y}_R|Y_R)\right)<\beta I(Y_2;X_R)$$

for some $\alpha,\beta>0$ with $\alpha+\beta=1$ and for some joint probability distributions $p(q)p(x_1|q)p(x_2|q)$ $p_1(y_R|x_1,x_2)p(\hat{y}_R|y_R)$ and $p(x_R)p_2(y_1,y_2|x_R)$.

Corollary (Corollary 3.3)**.** An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_{\mathrm{CF}} \subset \mathbb{R}_+^2$ given by the convex hull of $\mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$.

1.5.1.3 Partial Decode-and-Forward

Theorem (Theorem 3.5)**.** An achievable rate region for the two-phase two-way relay channel using a partial decode-and-forward protocol is the set $\mathcal{R}_4 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exists $R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)} \ge 0$ with $R_1^{(1)}+R_1^{(2)}=R_1$, $R_2^{(1)}+R_2^{(2)}=R_2$ satisfying

$$\begin{aligned}
R_1^{(1)} &\le \min\{\alpha I(U_1;Y_R|U_2,Q),\beta I(V;Y_2)\}\\
R_2^{(1)} &\le \min\{\alpha I(U_2;Y_R|U_1,Q),\beta I(V;Y_1)\}\\
R_1^{(1)}+R_2^{(1)} &\le \alpha I(U_1U_2;Y_R|Q)\\
R_1^{(2)} &\le \alpha I(X_1;\hat{Y}_R|X_2,U_1)\\
R_2^{(2)} &\le \alpha I(X_2;\hat{Y}_R|X_1,U_2)
\end{aligned}$$

under the constraints

$$\begin{aligned}
\alpha\left(H(\hat{Y}_R|X_1,U_2)-H(\hat{Y}_R|Y_R)\right)&<\beta I(Y_1;X_R|V)\\
\alpha\left(H(\hat{Y}_R|X_2,U_1)-H(\hat{Y}_R|Y_R)\right)&<\beta I(Y_2;X_R|V)
\end{aligned}$$

for some joint probability distributions $p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1,x_2)p(\hat{y}_R|y_R)$ and $p(v)p(x_R|v)p_2(y_1,y_2|x_R)$ and some $\alpha,\beta>0$ with $\alpha+\beta=1$.

Corollary (Corollary 3.6)**.** An achievable rate region for the two-phase two-way relay channel using a partial decode-and-forward protocol is the set $\mathcal{R}_5 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exists $R_1^{(1)}, R_1^{(2)}, R_2^{(2)} \geq 0$ with $R_1^{(1)} = R_1$, $R_2^{(1)} + R_2^{(2)} = R_2$ satisfying

$$\begin{aligned} R_1^{(1)} &\leq \min\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\} \\ R_2^{(1)} &\leq \min\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\} \\ R_1^{(1)} + R_2^{(1)} &\leq \alpha I(U_1 U_2; Y_R|Q) \\ R_2^{(2)} &\leq \alpha I(X_2; \hat{Y}_R|X_1, U_2) \end{aligned}$$

under the constraint

$$\alpha \left(H(\hat{Y}_R|X_1, U_2) - H(\hat{Y}_R|Y_R) \right) < \beta I(Y_1; X_R|V)$$

and similarly the set $\mathcal{R}_6 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exists $R_1^{(1)}, R_2^{(1)}, R_1^{(2)} \geq 0$ with $R_1^{(1)} + R_1^{(2)} = R_1$, $R_2^{(1)} = R_2$ satisfying

$$\begin{aligned} R_1^{(1)} &\leq \min\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\} \\ R_2^{(1)} &\leq \min\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\} \\ R_1^{(1)} + R_2^{(1)} &\leq \alpha I(U_1 U_2; Y_R|Q) \\ R_1^{(2)} &\leq \alpha I(X_1; \hat{Y}_R|X_2, U_1) \end{aligned}$$

under the constraint

$$\alpha \left(H(\hat{Y}_R|X_2, U_1) - H(\hat{Y}_R|Y_R) \right) < \beta I(Y_2; X_R|V)$$

for some joint probability distributions $p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(v)p(x_R|v)p_2(y_1, y_2|x_R)$ and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.

Corollary (Corollary 3.7)**.** An achievable rate region for the two-phase two-way relay channel using a partial decode-and-forward protocol is the set $\mathcal{R}_{\text{PDF}} \subset \mathbb{R}_+^2$ given by the convex hull of $\mathcal{R}_4 \cup \mathcal{R}_5 \cup \mathcal{R}_6$.

1.5.1.4 Compress-and-Forward with Joint Decoding

Theorem (Theorem 4.1)**.** An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_7 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ satisfying

$$R_1 \leq \max\left\{0, \min\left\{\alpha I(X_1; \hat{Y}_R|X_2, Q), \alpha\left(I(X_1 X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(X_R; Y_2)\right\}\right\}$$

$$R_2 \leq \max\left\{0, \min\left\{\alpha I(X_2; \hat{Y}_R|X_1, Q), \alpha\left(I(X_1 X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(X_R; Y_1)\right\}\right\}$$

for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and for some joint probability distributions $p(q)p(x_1|q)p(x_2|q)$ $p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(x_R)p_2(y_1, y_2|x_R)$.

Corollary (Corollary 4.2)**.** An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_{\text{CF-JD}} \subset \mathbb{R}_+^2$ given by the convex hull of $\mathcal{R}_7$.

1.5.1.5 Partial Decode-and-Forward with Joint Decoding

Theorem (Theorem 4.3)**.** Let $\mathcal{R}_8 \subset \mathbb{R}_+^4$ be the set of all $\left[R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)}\right]$ satisfying

$$
\begin{aligned}
R_1^{(1)} &\leq \min\Big\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\Big\} \\
R_2^{(1)} &\leq \min\Big\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\Big\} \\
R_1^{(1)} + R_2^{(1)} &\leq \alpha I(U_1 U_2; Y_R|Q) \\
R_1^{(2)} &\leq \max\Bigg\{\min\Big\{ \quad \alpha I(X_1; \hat{Y}_R|X_2, U_1), \\
&\qquad \alpha\left(I(X_1X_2; \hat{Y}_R|U_1, U_2) - I(Y_R; \hat{Y}_R|U_1, U_2)\right) + \beta I(X_R; Y_2|V)\Big\}, 0\Bigg\} \\
R_2^{(2)} &\leq \max\Bigg\{\min\Big\{ \quad \alpha I(X_2; \hat{Y}_R|X_1, U_2), \\
&\qquad \alpha\left(I(X_1X_2; \hat{Y}_R|U_1, U_2) - I(Y_R; \hat{Y}_R|U_1, U_2)\right) + \beta I(X_R; Y_1|V)\Big\}, 0\Bigg\}
\end{aligned}
$$

for some joint probability distributions $p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(v)p(x_R|v)p_2(y_1, y_2|x_R)$ and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.
An achievable rate region for the two-phase two-way relay channel using a partial decode-and-forward protocol is the set $\mathcal{R}_{\text{PCF-JD}} \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exists $\left[R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)}\right] \in \text{ConvexHull}(\mathcal{R}_8)$ with $R_1^{(1)} + R_1^{(2)} = R_1$, $R_2^{(1)} + R_2^{(2)} = R_2$.

1.5.1.6 Compress-and-Forward with three Information Flows

Theorem (Theorem 5.1)**.** An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_{\text{CF-JD-3S}} \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ satisfying

$$
\begin{aligned}
R_1 \leq\ & \min\Big\{ \quad \alpha I(X_1; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|X_2, Q); \\
&\qquad \alpha\Big(I(X_1X_2; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q) - I(Y_R; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q)\Big) + \beta I(UV_2; Y_2)\Big\} \\
R_2 \leq\ & \min\Big\{ \quad \alpha I(X_2; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|X_1, Q); \\
&\qquad \alpha\Big(I(X_1X_2; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q) - I(Y_R; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q)\Big) + \beta I(UV_1; Y_1)\Big\} \\
R_1 + R_2 \leq\ & \alpha\Big(I(X_1X_2; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q) + I(X_1X_2; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q) \\
&\quad - I(Y_R; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q) - I(Y_R; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q)\Big) \\
&+ \beta\Big(I(UV_2; Y_2) + I(UV_1; Y_1) - I(V_1; V_2|U)\Big)
\end{aligned}
$$

for some joint probability distributions $p(q)p(x_1|q)p(x_2|q)p(y_R|x_1, x_2)p(\hat{y}_{R,1+2}|y_R)p(\hat{y}_{R,1}|y_R, \hat{y}_{R,1+2})$ $p(\hat{y}_{R,2}|y_R, \hat{y}_{R,1+2})$ and $p(u, v_1, v_2)p(x_R|u, v_1, v_2)p(y_1, y_2|x_R)$ and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.

1.5.2 Discussion

The analysis in the following chapters will show, what gains can be achieved by considering on the one hand the different parts of the network and on the other hand the different abstraction layers of the communication stack jointly and not separated. The results emphasizes that a network has to be considered as a network and in general cannot be split without loss in atoms like BCs and MACs. This insight — which was already used in the network coding approaches — is not restricted to the coding of network information flows. In fact, by considering the atoms jointly the channel coding in these atoms can be adapted and gains become achievable in the overall system. A striking example for this is the joint decoding scheme: Here the relay is not able to decode the transmitted data. Furthermore, the receiver is not able to decode the relay's transmission independently. Only when the receiver considers the MAC transmission and the BC transmission jointly the decoding succeeds.

The need for a joint treatment of channel and network coding is also pointed out in this thesis. Network coding approaches acting on noiseless links between nodes do only half the job while preventing further gains: The mechanisms also active in network coding approaches induce dependencies in the network and can be used in the decoding to increase the link throughput. This is pointed out in the following chapters by simple examples. They are all based upon the XOR encoding being frequently used in network coding. Furthermore, the applied coding mechanisms are well known source coding techniques as e.g. the famous Slepian-Wolf-Coding [32] or the Wyner-Ziv-Coding [33]. Due to the dependencies induced by the network structure, these tools can now be used for channel coding. Thereby the separation between channel coding and network coding is overcome.

The analysis in the following chapters gives some insight on how one should design codes for two-way relay channels. In particular for the decode-and-forward mechanism a simple though very efficient coding framework is proposed, which performs optimal for certain channels. For the other protocols we shed light on the mechanisms at work. It turns out, that for the two-way relay channel those codes are important, which have the property that certain sub-codes of the code have a good performance. For the two-phase case the sub-codes are fixed and can be determined solely by the statistics of the channel. In a protocol with a direct link the effective sub-codes change depending on the signal received in the direct link. For the two-phase protocol the effective codes used in the decoding can therefore be determined offline, while for a protocol with a direct link the effective codes need to be created while decoding. We conclude, that for the two-phase protocol the complexity of the decoding is the same as in a single-user system, if the effective codes together with the mappings used at the relay are calculated offline. The challenge is the design of the code, which is a super-code of several good codes, that are interwoven in a particular way — for both receivers and for the different messages as side information.

For the analysis of the two-way relay channel, the results obtained in this thesis are far from complete. In particular, the approach with three information flows is only a first step. This approach can be enhanced in several directions: For example, there are dependencies in the system that are not used so far for the decoding at the receivers. Furthermore, the auxiliary random variables are constrained to allow a straight forward analysis. Solving these restrictions might lead to further gains. Last but not least, the system is closely related to the general BC. Therefore new insights for this channel might lead to new insights for the considered system. Another possible focus for future research may be the transfer of the results to channels with cost constraints, in particular to Gaussian channels with power constraint inputs. Furthermore, an interesting question is, whether or not one can characterize channels, for which one of the proposed protocols is optimal.

1.5.3 Further Results which are not Part of the Thesis

During the work on this thesis we obtained further interesting results that are not part of this thesis.

- In [3, 4] we study the MAC with correlated binary data. Suppose a set of binary sources produces binary data that is transmitted to a receiver via a Gaussian MAC. The capacity region of such a setup is unknown; an achievable rate region for the problem can be found in [34, 35]. A suboptimal strategy is to encode the data using a distributed source coding scheme (as e.g. proposed in [36]) according to the coding theorem by Slepian and Wolf [37] and then transmit the encoded data to the sink. But results for this strategy in a MAC with correlated sources are quite negative: It turns out that a cooperation of the nodes is inevitable to ensure a sufficient throughput and a good scaling for larger networks. Without any cooperation, the transport capacity of the so called many-to-one or reach-back channel scales too slow leading to the problem of vanishing throughput per node. Any compression scheme is then insufficient to transport the increasing amount of data produced by the nodes (see for instance [38] and references therein). This is true for distributed source coding schemes that exploit the correlation [37, 36], as well as for joint source channel codes that usually assume orthogonal channels [39]. Our work in [3, 4] does not focus on information theoretical results concerning the communication task. Instead we analyze, how one could use the source correlation in a Gaussian MAC with uncoded transmission. We ask how one should place the transmitted symbols in the signal space in a distributed way, such that the receiver will detect the symbols of all the receivers without error with high probability. We consider a transmission scheme based on code-division-multiple-access (CDMA) that exploits the correlation structure of sources. The motivation for this work is to facilitate statistical cooperation in a sensor network scenario as proposed in [40]. We assume the jointly optimum detector and focus on binary sources with arbitrary statistical dependencies. The objective is to characterize signature sequences that minimize the bit error probability for each source. Based on the results of [41], we derive an upper bound on the bit error probability and show that under certain correlation structures, there exist sequences for which the upper bound equals a simple genie-aided lower bound. These sequences are optimal in the sense of minimizing the bit error probability of each source since the lower bound is independent of the choice of sequences. We prove a necessary and sufficient condition for attaining the lower bound, and hence provide insight into the design of sequences for CDMA systems with correlated sources. Finally we give some comments on how to choose sequences if the conditions on the correlation structure are not satisfied and therefore upper and lower bound differ. In relation to the topic of this thesis the results in [3, 4] might be useful for the design of structured codes to improve the performance in the MAC phase of a two-phase two-way relay channel if the source data of the terminals is correlated. Some comments on the use of structured codes for the MAC phase in the two-phase two-way relay channel are given in Section 4.1.4.

- In [5] we analyze the performance of different multiple antenna transmission techniques in wireless networks with interference treated as noise. Our focus is on the impact of simple orthogonal space time codes (STCs) on the so-called network-outage probability. To ensure some quality-of-service, we assume that each connection in a network has to achieve a certain signal-to-interference-and-noise-ratio. Due to channel variations and interference, it might be impossible to maintain the desired SIR on each link permanently. Given some established network topology and channel statistics, one of the most important objectives is then to guarantee a certain outage probability performance of the network. The network is said to be in outage if there exists at least one link, for which the SIR target cannot be satisfied. This event is called network-outage. The network-outage probability is the probability for this event. There is little literature on the performance analysis of multiple antenna systems that are exposed to (unknown) interference from other connections. The work of Blum et al. [42, 43] shows that in scenarios with large interference, standard multiple antenna techniques could fail to achieve the desired performance objectives. In addition, for some systems, it was shown that transmitting with only one antenna is optimal. In [5] analytical results on the network-outage probability are given for some simple networks and different multiple antenna transmission techniques. These results show insufficiency of many traditional space-time coding designs under interference conditions. Simulations suggest that this main result of the analysis in [5] may also hold for general wireless networks, provided that the interference is sufficiently strong. In particular the Alamouti STC is inappropriate for many symbol synchronous networks in which interference is treated as noise since the scheme induces a diversity gain to the interference. Similar results hold for other orthogonal STCs. In many scenarios, transmitting with only one antenna is superior if one considers the increase of complexity due to the STC. In general, receive diversity proves to be give more benefit than transmit diversity by using STCs. Furthermore general orthogonal STCs lead to unequal SIR performance for the different symbols transmitted in one STC symbol. As a consequence, channel knowledge may increase the performance of the code. The results indicate that traditional point-to-point designs might be not suitable in distributed networks with strong interference. Most of the current work does not consider nearby nodes performing similar operations and inducing interference which — although Gaussian distributed for every time instance — may have significantly different impact on the performance than Gaussian noise, if channel statistics are taken into account.

- In co-authored work with Ruben Heras-Evangelio [1, 2] we studied a two-stage relaying scheme in the context of wireless sensor networks. The task is to transmit data from a source to a destination with the help of an array of relays which re-encode the received signal using a distributed space time code. The relays are assumed not to decode the transmitted data. In [2] the pairwise error probability is analyzed and upper bounds are derived for general space time codes. These bounds are used to derive a power allocation to minimize the pairwise error probability. The results show that significant diversity gains can be obtained by the cooperative relaying scheme. In [1] we extend the model by allowing a more flexible placement of the relay nodes. Furthermore a simple hardware model was used to analyze the impact of the power consumption due to the hardware. It turns out that hardware energy consumption favors single user transmission over cooperative relaying if the distance between source and destination is not to large. The reason is, that in this case the fixed energy costs needed to operate the additional nodes are not compensated by the gain offered by the diversity.

Chapter 2

The Two-Way Relay channel with Decode-and-Forward

This chapter considers the two-way relay channel, where we impose a decode-and-forward restriction. This implies that we assume the relay is able to decode the messages of both the receivers. We begin the analysis by assuming a two-phase communication protocol. In the first phase the two nodes transmit their messages to the relay node, which decodes both the messages. This phase is the classical MAC, and the capacity of this channel is known [44] [45]. The second phase is a broadcast from the relay to both receivers. This BC is different from the usual broadcast channel [46, 47, 48], as the receiving nodes know the message intended for the other user. The capacity region of the general BC is unknown, the best achievable rate region is given in [47]. We shall see that the side information available at the receivers simplifies the problem; therefore we are able to state the capacity region of the BC channel, where both receivers know the message intended for the other respective node. The region is given and proven in Section 2.1.2. By using the capacity of the MAC and the the result of Section 2.1.2 we are able to give an achievable rate region for the two-way relay channel obtained with a decode-and-forward protocol.

The theoretical derivation in Section 2.1 uses random coding arguments. Although the random coding arguments do not give an explicit code construction, which yields efficient codes for all channels, it gives rise to a coset structure of single user codes, that can be used for practical code design. In Section 2.2 this idea is pursued and a practical coding scheme based on single user codes is proposed, that is optimal for channels which fulfill a certain symmetry condition that is explained in detail in that section. For general channels this coding scheme is suboptimal; a short discussion shows how one could still design codes from single user codes and which problems arise in more general channels, where the channel disturbance is independent of the channel input. For these channels codes based on lattices can be used. For general channels it seems to be inevitable to use several codes that are interwoven in a special way, i.e. a joint code design for both receivers is necessary.

Section 2.3 extends the previous result by allowing more than two phases. Achievable rate regions are given for three and four phases. The additional phases offer some additional degrees of freedom as follows: The three-phase protocol establishes a direct link between the nodes without violating the half-duplex constraint, while the advantage of the non-orthogonal MAC transmission is sacrificed; in the four phase protocol a trade-off between the benefits of the direct link and the non-orthogonal MAC transmission is permitted. The key feature of all these protocols is the concluding BC phase, where the signal already received in previous phases as well as the message intended for the other user may be used as side information for the decoding at the terminal node. There are further possible extensions, e.g. additional MAC phases, where one of the nodes and the relay transmit to the other node; these extensions are beyond the scope of this thesis.

Some of the results presented in this Chapter and related results have been published in [6, 7, 8, 9, 10]. A similar result as stated in this chapter was recently developed independently and is published in [49].

2.1 A Coding Theorem for the Two-Phase Two-Way Relay channel

In this section we consider the achievable rates for a two-phase protocol if we impose a decode-and-forward constraint at the relay. The setup used in this section is the generic two-way relay setup introduced in Chapter 1. Recall, that we assume half-duplex nodes that cannot transmit and receive at the same time. Furthermore we do not allow the transmitted symbols of any of the terminal nodes to depend on any received signal. For a classical two-way channel this restriction is known as a restricted two-way channel.

In this section we focus on a protocol that consists of only two phases. Furthermore we require that the relay node can successfully decode both messages. This assumption seems reasonable as a first approach, as it simplifies the analysis: By this assumption we can split the system in two sub-systems, namely a MAC transmission and a BC transmission. For the analysis of an achievable rate region we may use timesharing between achievable rate pairs for the two phases. The timesharing result is again achievable by a union bound argument. We can use the sum of the error probabilities for the two sub-sytems as an upper bound on the probability of error for the overall system. The approach to enforce decoding at the relay is motivated by the following observation: In the system all information passes the relay. We will see later in Chapters 3 to 5 that even though all information passes the relay, it may be suboptimal to enforce that the relay is able to decode the data.

With the above restrictions we end up with a multiple access phase where node 1 and node 2 transmit messages w_1 and w_2 to the relay node, and a broadcast phase where the relay forwards the messages to node 2 and 1 respectively. Before considering the complete system with a time-division between these two phases in Section 2.1.5 we will look at the two phases separately.

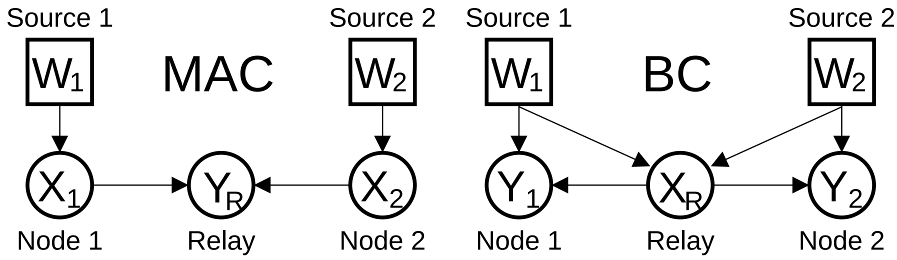

Figure 2.1: Multiple access (MAC) and broadcast (BC) phase of the time division two-way relay channel.

In the MAC phase we have the classical MAC, where the optimal coding strategy and the capacity region C_{MAC} are known [44], [45]. We will restate the capacity region in the next subsection. For the BC phase, we assume that the relay node has successfully decoded the messages w_1 and w_2 in the MAC phase. This is reasonable if we assume rates within the MAC capacity region and a sufficient coding length. Therefore we have a BC where the message w_1 is known at node 1 and the relay node and the message w_2 is known at node 2 and the relay node, as depicted in Figure 2.1. The task of the relay node is to broadcast a signal to node 1 and node 2 which allows both to recover the respective unknown source. This means that node 1 wants to recover source W_2 and node 2 wants to recover source W_1. We will present the information theoretical optimal coding strategy and the capacity region of the bidirectional broadcast channel in Section 2.1.2.

Clearly in the overall system consisting of the MAC and the BC phase a rate pair is achievable if that rate pair is within the achievable rate regions for both the subsystems. Furthermore due to the restriction, which prohibits any kind of feedback and cooperation of the encoders, using the standard cutset bound argument [30] we could conclude that the capacity of the restricted two-phase two-way relay channel with decode-and-forward constraint is indeed given by timesharing between rate pairs in these two regions. The use of the term "capacity" in this context is at the least questionable, since the upper bound on the achievable rate region is more a consequence of the constraints than a consequence or property of the channel. Therefore, we prefer to call the region simply an achievable rate region for the two-phase two-way relay channel.

Recall that X_1, X_2, and X_R denote the input and Y_1, Y_2, and Y_R the output random variables of node 1, node 2, and the relay node respectively. Furthermore, R_1 and R_2 are the rates from node 1 and 2 to the relay node in the MAC phase while they denote the achievable rates between the relay node and node 2 and 1 in the BC phase. All alphabets are assumed to be discrete and of finite cardinality.

2.1.1 Capacity Region of Multiple Access Phase

In this subsection, we restate the well known capacity region of the MAC [44, 45] without proof; the proof can be found in any textbook on multiuser information theory, e.g. [50].

Definition 2.1. A *discrete MAC* consists of two input random variables X_1 and X_2, an output random variable Y_R, and a probability transition function $p(y_R|x_1, x_2)$. X_1 and X_2 take values in the input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively, while Y_R does so in the output alphabet $\mathcal{Y}_R$.

Let $X_k^n = (X_{k,(1)}, X_{k,(2)}, \ldots, X_{k,(n)})$, $k \in \{1, 2\}$, denote the random input sequence of length n at the k^{th} transmitter which takes realizations $x_k^n = (x_{k,(1)}, x_{k,(2)}, \ldots, x_{k,(n)}) \in \mathcal{X}_k^n$ with probability $p(x_k^n)$. Accordingly, let $Y^n = (Y_{R,(1)}, Y_{R,(2)}, \ldots, Y_{R,(n)})$ denote the output sequence of length n which takes realizations $y_R^n = (y_{R,(1)}, y_{R,(2)}, \ldots, y_{R,(n)}) \in \mathcal{Y}_R^n$ with probability $p(y_R^n)$. The MAC is said to be *memoryless* if $p(y_R^n|x_1^n, x_2^n) = \prod_{i=1}^n p(y_{R,(i)}|x_{1,(i)}, x_{2,(i)})$.

Theorem 2.1. The capacity region C_{MAC} of the memoryless MAC is the set of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &\leq I(X_1; Y_R|X_2, U) \\ R_2 &\leq I(X_2; Y_R|X_1, U) \\ R_1 + R_2 &\leq I(X_1, X_2; Y_R|U) \end{aligned}$$

for some probability function $p(u)p(x_1|u)p(x_2|u)p(y_R|x_1, x_2)$, where the set of the auxiliary random variable U has cardinality bounded by $|\mathcal{U}| \leq 3$.

2.1.2 Capacity Region of Broadcast Phase

In this section we consider the BC-phase only. First we need to introduce some notation and present certain simplified preliminaries. The simplification is possible, as we do not consider the MAC phase in this section but assume that the relay has decoded the messages without error.

2.1.2.1 Definitions and Preliminaries

Definition 2.2. A *discrete BC* consists of a random variable X_R taking values in the input alphabet $\mathcal{X}_R$, two random variables Y_1 and Y_2 taking values in the output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively, and a probability transition function $p(y_1, y_2|x)$.

Let $X_R^n = (X_{R,(1)}, X_{R,(2)}, \ldots, X_{R,(n)})$ denote the random input sequence of length n which takes realizations $x_R^n = (x_{R,(1)}, x_{R,(2)}, \ldots, x_{R,(n)}) \in \mathcal{X}_R^n$ with probability $p(x_R^n)$. Accordingly, let $Y_k^n = (Y_{k,(1)}, Y_{k,(2)}, \ldots, Y_{k,(n)})$, $k \in \{1, 2\}$, denote the random output sequence of length n at the k^{th} receiver, which takes realizations $y_k^n = (y_{k,(1)}, y_{k,(2)}, \ldots, y_{k,(n)}) \in \mathcal{Y}_k^n$ with probability $p(y_k^n)$.

The BC is said to be *memoryless* if $p(y_1^n, y_2^n | x_R^n) = \prod_{s=1}^n p(y_{1,(s)}, y_{2,(s)} | x_{R,(s)})$. For a real broadcast situation we impose a "no-collaboration" restriction between the receivers so that we need to consider only the marginal transition probabilities $p(y_1|x_R)$ and $p(y_2|x_R)$.

We now consider a block code of length n. The two independent information sources W_1 and W_2 are random variables, which take values in the message sets

$$\mathcal{W}_1 = \{1, 2, \ldots, M_1^{(n)}\}, \qquad \mathcal{W}_2 = \{1, 2, \ldots, M_2^{(n)}\},$$

according to two separate uniform distributions. Therefore, we have sources with entropy $H(W_1) = \log M_1^{(n)}$ and $H(W_2) = \log M_2^{(n)}$. We collect both sources in the random variable $W = [W_1, W_2]$, which takes realizations $w = [w_1, w_2] \in \mathcal{W}_1 \times \mathcal{W}_2 = \mathcal{W}$.

Definition 2.3. A $(M_1^{(n)}, M_2^{(n)}, n)$-code for the *bidirectional BC* consists of one encoder at the relay node

$$x_R^n : \mathcal{W} \to \mathcal{X}_R^n,$$

and decoders at node 1 and node 2

$$g_1 : \mathcal{Y}_1^n \times \mathcal{W}_1 \to \mathcal{W}_2,$$
$$g_2 : \mathcal{Y}_2^n \times \mathcal{W}_2 \to \mathcal{W}_1.$$

When the sources' output is $w = [w_1, w_2]$, the receiver of node 1 is in error if $g_1(Y_1^n, w_1) \neq w_2$. We denote the probability of this event by

$$\lambda_1(w) = Pr[g_1(Y_1^n, w_1) \neq w_2 \,|\, x_R^n(w = [w_1, w_2]) \text{ has been sent}].$$

Accordingly, we denote the probability that the receiver of node 2 is in error by

$$\lambda_2(w) = Pr[g_2(Y_2^n, w_2) \neq w_1 \,|\, x_R^n(w = [w_1, w_2]) \text{ has been sent}].$$

This allows us to introduce the notation for the average probability of error for the k^{th} node

$$\mu_k^{(n)} = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \lambda_k(w).$$

Definition 2.4. A rate pair $[R_1, R_2]$ is said to be *achievable* for the bidirectional broadcast channel if there exists a sequence of $(M_1^{(n)}, M_2^{(n)}, n)$-codes with $\frac{\log M_1^{(n)}}{n} \to R_1$, $\frac{\log M_2^{(n)}}{n} \to R_2$ while $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. The set of all achievable rate pairs is the *capacity region* of the bidirectional broadcast channel.

2.1.2.2 Coding Theorem

Theorem 2.2. The capacity region $\mathcal{C}_{\text{DF-BC}}$ for sending two sources W_1 and W_2 over the memoryless broadcast channel where the receiving node 1 knows W_1 and node 2 knows W_2 is the set of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &\leq I(X_R; Y_2), \\ R_2 &\leq I(X_R; Y_1), \end{aligned} \tag{2.1}$$

for some probability function $p(x_R)p(y_1, y_2|x_R)$.

Remark 2.1. The region $\mathcal{C}_{\text{DF-BC}}$ is convex due to the concavity of both the mutual information expression, $I(X_R; Y_1)$ and $I(X_R; Y_2)$, as functions of $p(x_R)$ for fixed $p(y_1|x_R)$ and $p(y_2|x_R)$.

Remark 2.2. The result for the capacity given in Theorem 2.2 can be generalized to a capacity region under a power constraint for a Gaussian MIMO broadcast channel. This was done in [10]. A detailed proof will not be given here. The interested reader is refered to [10].

2.1.3 Proof of the Capacity Region for the Broadcast Phase

Now we will proof Theorem 2.2 in two steps. First we will show, how to construct a sequence of codes such that for this sequence, the rate pair corresponding to the codes approaches any given rate pair in the region given by (2.1), while the average error probability of the codes goes to zero and the block length goes to infinity. Thereafter, we proof the converse, i.e. we show that whenever the average error probability of a given sequence of codes goes to zero for a block length $n \to \infty$, then the rate pair corresponding to the codes approaches a rate pair in the region given by (2.1).

2.1.3.1 Proof of Achievability

Proof. We adapt the random coding proof for the degraded broadcast channel of [51] to our context. First, we prove the achievability of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &< I(X_R; Y_2), \\ R_2 &< I(X_R; Y_1), \end{aligned} \tag{2.2}$$

for some probability function $p(x_R)p_2(y_1, y_2|x_R)$. Then we extend this to prove that all points in the closure of the convex hull of (2.2) are achievable, which is exactly the region as stated in Theorem 2.2.

Random Codebook Generation We generate $M_1^{(n)} M_2^{(n)}$ independent codewords $X_R^n(w)$, $w = [w_1, w_2]$ of length n with $M_1^{(n)} = 2^{\lfloor nR_1 \rfloor}$ and $M_2^{(n)} = 2^{\lfloor nR_2 \rfloor}$ according to $\prod_{s=1}^{n} p(x_{R,(s)})$. The random code is revealed to both receivers and the relay.

Encoding To send the pair $w = [w_1, w_2]$ with $w_k \in \mathcal{W}_k$, $k \in \{1, 2\}$, the relay sends the corresponding codeword $x_R^n(w)$.

Decoding The receiving nodes will use typical set decoding. For a strict definition of the decoding sets we choose parameters ϵ_1, ϵ_2 for the typical sets as $\epsilon_1 < \frac{I(X_R;Y_1)-R_2}{3}$ and $\epsilon_2 < \frac{I(X_R;Y_2)-R_1}{3}$ respectively. Knowing w_1 the decoder at node 1 decides that w_2 was transmitted if $x_R^n(w_1, w_2)$ is the only codeword such that $\left(x_R^n(w_1, w_2), y_1^n\right) \in \mathcal{T}_{\epsilon_1}^{(n)}(X_R, Y_1)$. Accordingly, the decoder at receiver 2 chooses w_1 if $x_R^n(w_1, w_2)$ is the only codeword such that $\left(x_R^n(w_1, w_2), y_2^n\right) \in \mathcal{T}_{\epsilon_2}^{(n)}(X_R, Y_2)$. If there is no or no unique codeword $x_R^n(w_1, \cdot)$ for receiver 1 or $x_R^n(\cdot, w_2)$ for receiver 2, the decoder maps on the index 1 (to keep the definition of the decoder consistent).

When $x_R^n(w)$ with $w = [w_1, w_2]$ has been sent, and y_1^n and y_2^n have been received we say that the decoder at node 1 is in error if either $x^n(w)$ is not in $\mathcal{T}_{\epsilon_1}^{(n)}(X_R, Y_1)$ for the received signal y_1^n (occurring with probability $P_{e,1}^{(1)}(w)$) or if $x_R^n(w_1, \hat{w}_2)$ with $\hat{w}_2 \neq w_2$ is in $\mathcal{T}_{\epsilon_1}^{(n)}(X_R, Y_1)$ (occurring with $P_{e,1}^{(2)}(w)$). We define the error events at node 2 in an analogous way; these events for receiver 2 occur with probability $P_{e,2}^{(1)}(w)$ and $P_{e,2}^{(2)}(w)$ respectively.

Analysis of the Probability of Error From the union bound we have

$$\lambda_k(w) \leq P_{e,k}^{(1)}(w) + P_{e,k}^{(2)}(w)$$

with

$$P_{e,k}^{(1)}(w) = \sum_{y_k^n \in \mathcal{Y}_k^n} p(y_k^n | x_R^n(w)) \chi^C_{\mathcal{T}_{\epsilon_k}^{(n)}(X_R, Y_k)}(x_R^n(w), y_k^n)$$

for $k \in \{1, 2\}$ and

$$P_{e,1}^{(2)}(w) = \sum_{y_1^n \in \mathcal{Y}_1^n} p(y_1^n | x_R^n(w)) \sum_{\hat{w}_2 \neq w_2} \chi_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R, Y_1)}(x_R^n(w_1, \hat{w}_2), y_1^n),$$

and

$$P_{e,2}^{(2)}(w) = \sum_{y_2^n \in \mathcal{Y}_2^n} p(y_2^n | x_R^n(w)) \sum_{\hat{w}_1 \neq w_1} \chi_{\mathcal{T}_{\epsilon_2}^{(n)}(X_R, Y_2)}(x_R^n(\hat{w}_1, w_2), y_2^n),$$

For uniformly distributed messages W_1 and W_2 we define

$$P_{e,k}^{(m)} = \frac{1}{|\mathcal{W}_1| \, |\mathcal{W}_2|} \sum_{w \in \mathcal{W}_1 \times \mathcal{W}_2} P_{e,k}^{(m)}(w)$$

for $m \in \{1, 2\}$ so that $\mu_k^{(n)} \leq P_{e,k}^{(1)} + P_{e,k}^{(2)}$. Next, we average over all codebooks, i.e. $\mathbb{E}_{x_R^n}\{\mu_k^{(n)}\} \leq \mathbb{E}_{x_R^n}\{P_{e,k}^{(1)} + P_{e,k}^{(2)}\}$.

In the following, we show that if $R_2 < I(X, Y_1)$, we have $\mathbb{E}_{x_R^n}\{\mu_1\} \to 0$ as $n \to \infty$. The analogous result that if $R_1 < I(X, Y_2)$, we have $\mathbb{E}_{x_R^n}\{\mu_2\} \to 0$ as $n \to \infty$ follows immediately.

We have

$$
\begin{aligned}
\mathbb{E}_{x_R^n}\left\{P_{e,1}^{(1)}\right\} =& \frac{1}{|\mathcal{W}_1|\,|\mathcal{W}_2|}\sum_{w\in\mathcal{W}_1\times\mathcal{W}_2}\mathbb{E}_{x_R^n}\left\{P_{e,1}^{(1)}(w)\right\} \\
\overset{\text{for any}}{\underset{\text{fixed } w}{=}}& \sum_{y_1^n\in\mathcal{Y}_1^n}\mathbb{E}_{x_R^n}\left\{p\left(y_1^n|x_R^n(w)\right)\chi^C_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}\left(x_R^n(w),y_1^n\right)\right\} \\
=& \sum_{y_k^n\in\mathcal{Y}_1^n}\sum_{x_R^n\in\mathcal{X}_R^n}p(x_R^n)p(y_1^n|x_R^n)\chi^C_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}(x_R^n,y_1^n) \\
=& \mathbb{E}_{x_R^n,y_1^n}\left\{\chi^C_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}(x_R^n,y_1^n)\right\}\underset{n\to\infty}{\longrightarrow}0.
\end{aligned}
$$

The last term goes to 0 exponentially fast by the law of large numbers and the definition of the typical set. This can be seen by noting that it is the probability of the event that two sequences drawn according to a joint probability distribution are not jointly typical.

For the calculation of $\mathbb{E}_{x_R^n}\left\{P_{e,1}^{(2)}\right\}$ we use the fact that for $w=[w_1,w_2]\neq[w_1,\hat{w}_2]$ the random variable $p\left(y_1^n|X_R^n(w)\right)$ is independent of the random variable $\chi_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}(x_R^n(w_1,\hat{w}_2),y_1^n)$.

$$
\begin{aligned}
\mathbb{E}_{x_R^n}\left\{P_{e,1}^{(2)}\right\} =& \frac{1}{|\mathcal{W}_1||\mathcal{W}_2|}\sum_{w\in\mathcal{W}_1\times\mathcal{W}_2}\mathbb{E}_{x_R^n}\left\{P_{e,1}^{(2)}(w)\right\} \\
\overset{\text{for any}}{\underset{\text{fixed } w}{=}}& \sum_{y_1^n\in\mathcal{Y}_1^n}\mathbb{E}_{x_R^n}\left\{p\left(y_1^n|x_R^n(w)\right)\sum_{\substack{\hat{w}_2=1\\ \hat{w}_2\neq w_2}}^{|\mathcal{W}_2|}\chi_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}\left(x_R^n(w_1,\hat{w}_2),y_1^n\right)\right\} \\
=& \sum_{y_1^n\in\mathcal{Y}_1^n}\sum_{\substack{\hat{w}_2=1\\ \hat{w}_2\neq w_2}}^{|\mathcal{W}_2|}\mathbb{E}_{x_R^n}\left\{p\left(y_1^n|x_R^n(w)\right)\right\}\mathbb{E}_{x_R^n}\left\{\chi_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}\left(x_R^n(w_1,\hat{w}_2),y_1^n\right)\right\} \\
=& \sum_{y_1^n\in\mathcal{Y}_1^n}\sum_{\substack{\hat{w}_2=1\\ \hat{w}_2\neq w_2}}^{|\mathcal{W}_2|}p(y_1^n)\mathbb{E}_{x_R^n}\left\{\chi_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}\left(x_R^n(w_1,\hat{w}_2),y_1^n\right)\right\} \\
=& (|\mathcal{W}_2|-1)\sum_{y_1^n\in\mathcal{Y}_1^n}\sum_{x_R^n\in\mathcal{X}_R^n}p(x_R^n)p(y_1^n)\chi_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}(x_R^n,y_1^n)
\end{aligned}
$$

For $(x_R^n,y_1^n)\in\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)$ and sufficiently large n we have by the properties of the typical set

$$
\begin{aligned}
\mathbb{E}_{x_R^n}\left\{P_{e,1}^{(2)}\right\} &= (|\mathcal{W}_2|-1)\sum_{y_1^n\in\mathcal{Y}_1^n}\sum_{x_R^n\in\mathcal{X}_R^n}p(x_R^n)p(y_1^n)\chi_{\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)}(x_R^n,y_1^n) \\
&\leq (|\mathcal{W}_2|-1)|\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)|2^{-n(H(Y_1)-\epsilon_1)}2^{-n(H(X_R)-\epsilon_1)}.
\end{aligned}
$$

Furthermore

$$|\mathcal{T}_{\epsilon_1}^{(n)}(X_R,Y_1)|\leq 2^{n(H(X_R,Y_1)+\epsilon_1)}$$

and

$$(|\mathcal{W}_2|-1)\leq 2^{nR_2}.$$

Therefore

$$\mathbb{E}_{x_R^n}\left\{P_{e,1}^{(2)}\right\} \le 2^{n(R_2+3\epsilon_1-I(X_R;Y_1))}$$

which goes to 0 for $n \to \infty$ as we choose $\epsilon_1 < \frac{I(X_R;Y_1)-R_2}{3}$.

Hence, whenever $R_1 < I(X, Y_2)$ and $R_2 < I(X, Y_1)$, the average probability of error for both receivers, averaged over codebooks and codewords, gets arbitrarily small for sufficiently large block length n. Moreover, if $R_1 < I(X; Y_2)$ and $R_2 < I(X; Y_1)$ we can choose ϵ and n such that we have $\mathbb{E}_{x_R^n}\left\{\mu_1^{(n)} + \mu_2^{(n)}\right\} < \epsilon$. Since the average probabilities of error over the codebooks is small, there exists at least one codebook $C^\star$ with small average probabilities of error $\mu_1^{(n)} + \mu_2^{(n)} < \epsilon$. This proves the achievability of any rate pair satisfying the equations (2.2).

A Note on Average vs. Maximum Error Probability In the definition of achievable rates and in the definition of capacity we used the average probability of error and not the maximum probability of error. In single user systems this difference is not significant. Simply speaking, one can always take the good performing codewords of the random average probability of error code while dropping the bad ones without sacrificing too much. One can prove that this is always possible. Moreover, in single user systems the reduction of codewords is sub-exponential; hence the rate reduction is not evident for large block length.

This argument does not always hold in multiuser systems. In fact, in general the derivation of a maximum error code from a average error code is not possible without a loss in rate [52]. The problem lies in the fact that the reduced code needs to have a special structure. In the current setup it is still possible to find a large enough subset of the code, which codewords perform as needed. But this code cannot be decoded at the receiver. To show where the problem occurs we give the usual proving technique here and point out where it fails.

The idea of a code construction for performance under a maximum error criterion starting from a code for average probability of error is that the encoder uses only codewords $x_R^n(w)$ of the code $C^\star$ with an index in the set of codewords $w \in \mathcal{Q}^\star$, which have a maximum error $\lambda_k(w) < 8\epsilon$, $k \in \{1, 2\}$ for both receivers. Suppose we have a codebook $C^\star$ with small average probabilities of error $\mu_1^{(n)} + \mu_2^{(n)} < \epsilon$. This implies that we have $\mu_1^{(n)} < \epsilon$ and $\mu_2^{(n)} < \epsilon$. Next, we define sets

$$\mathcal{Q} = \{w \in \mathcal{W} : \lambda_1(w) < 8\epsilon \text{ and } \lambda_2(w) < 8\epsilon\},$$

and

$$\mathcal{R}_k = \{w \in \mathcal{W} : \lambda_k(w) \ge 8\epsilon\},$$

$k \in \{1, 2\}$. Therefore, $\mathcal{Q}$ contains messages with a small probability of error for the code $C^\star$ and for both receivers, while $\mathcal{R}_k$ contains messages with a large probability of error for receiver k. Since

$$\epsilon > \frac{1}{|\mathcal{W}|}\sum_{w\in\mathcal{W}} \lambda_k(w) \ge \frac{|\mathcal{R}_k|}{|\mathcal{W}|}8\epsilon,$$

we can bound the cardinality $|\mathcal{R}_k| < \frac{|\mathcal{W}|}{8}$ for $k \in \{1, 2\}$. It follows from $\mathcal{W} = \mathcal{Q} \cup \mathcal{R}_1 \cup \mathcal{R}_2$ that

$$|\mathcal{Q}| \geq |\mathcal{W}| - |\mathcal{R}_1| - |\mathcal{R}_2| > \frac{3}{4}|\mathcal{W}|.$$

Now, let $\mathcal{T}$ be the set of w_1 having the property that for each w_1 there are at least $\frac{1}{2}M_2^{(n)}$ choices of w_2 so that $[w_1, w_2] \in \mathcal{Q}$. Therefore, for $w_1 \in \mathcal{T}$ there are at most $M_2^{(n)}$ choices $w_2 \in \mathcal{W}_2$ and for $w_1 \notin \mathcal{T}$ there are less than $\frac{1}{2}M_2^{(n)}$ choices $w_2 \in \mathcal{W}_2$ such that $[w_1, w_2] \in \mathcal{Q}$. Accordingly, we have

$$|\mathcal{T}|M_2^{(n)} + |\mathcal{W}_1 \setminus \mathcal{T}|\frac{1}{2}M_2^{(n)} > |\mathcal{Q}| > \tfrac{3}{4}M_1^{(n)}M_2^{(n)}.$$

It follows that $|\mathcal{T}| > \frac{1}{2}M_1^{(n)}$ where we used $|\mathcal{W}_1 \setminus \mathcal{T}| = M_1^{(n)} - |\mathcal{T}|$. As a consequence there exists an index set $\mathcal{Q}_1^\star \subset \mathcal{W}_1$ with $\frac{1}{2}M_1^{(n)}$ indices w_1, to each of which we can find an index set $\mathcal{Q}_2^\star(w_1) \subset \mathcal{W}_2$ with $\frac{1}{2}M_2^{(n)}$ indices w_2 so that we have for each $w_1 \in \mathcal{Q}_1^\star$ and $w_2 \in \mathcal{Q}_2^\star(w_1)$ a maximum error $\lambda_k(w_1, w_2) < 8\epsilon$ for $k \in \{1, 2\}$.

Note that the index set has no Cartesian structure. This is where the problem will occur. From the above arguments it follows that there exist bijective mappings

$$\Phi : \mathcal{W}^\star \to \mathcal{Q}^\star,$$

$$\Phi_1 : \mathcal{W}_1^\star \to \mathcal{Q}_1^\star,$$
$$\Phi_2^{w_1} : \mathcal{W}_2^\star \to \mathcal{Q}_2^\star(w_1)$$

for each $w_1 \in \mathcal{Q}_1^\star$ where

$$\Phi(w_1, w_2) = \left[\Phi_1(w_1), \Phi_2^{w_1}(w_2)\right]$$

with sets $\mathcal{W}^\star = \mathcal{W}_1^\star \times \mathcal{W}_2^\star$, $\mathcal{W}_k^\star = \left\{1, 2, \dots, \frac{1}{2}M_k^{(n)}\right\}$ for $k \in \{1, 2\}$, and

$$\mathcal{Q}^\star = \left\{[w_1, w_2] \in \mathcal{W} : w_1 \in \mathcal{Q}_1^\star, w_2 \in \mathcal{Q}_2^\star(w_1)\right\} \subset \mathcal{Q}.$$

Furthermore, there exist inverse mappings

$$\Psi_k : \mathcal{Q}^\star \to \mathcal{W}_k^\star, \quad k \in \{1, 2\},$$

with

$$w = \left[\Psi_1(\Phi(w)), \Psi_2(\Phi(w))\right].$$

Using codewords $x_R^n(w)$ of the codebook $\mathcal{C}^\star$ and corresponding decoders $g_1(y_1^n, w_1)$ and $g_2(y_2^n, w_2)$ the above arguments allow us to define a $(\frac{1}{2}M_1^{(n)}, \frac{1}{2}M_2^{(n)}, n)$-code as follows: The encoder $\tilde{x}_R^n : \mathcal{W}^\star \to \mathcal{X}_R^n$ is given by

$$\tilde{x}_R^n(w) = x_R^n(\Phi(w)).$$

The decoders

$$\tilde{g}_1 : \mathcal{Y}_1^n \times \mathcal{W}_1^\star \to \mathcal{W}_2^\star$$

and

$$\tilde{g}_2 : \mathcal{Y}_2^n \times \mathcal{W}_2^\star \to \mathcal{W}_1^\star$$

are defined as

$$\tilde{g}_1(y_1^n, w_1) = \tilde{\Psi}_2\left(\Phi_1(w_1), g_1\left(y_1^n, \Phi_1(w_1)\right)\right)$$

and

$$\tilde{g}_2(y_2^n, w_2) = \tilde{\Psi}_1\left(g_2\left(y_2^n, \Phi_2^{w_1}(w_2)\right), \Phi_2^{w_1}(w_2)\right)$$

with the mappings $\tilde{\Psi}_k : \mathcal{W} \to \mathcal{W}_k^\star$ given by

$$\tilde{\Psi}_k = \begin{cases} \Psi_k(w), & \text{if } w \in \mathcal{Q}^\star \\ 1, & \text{if } w \notin \mathcal{Q}^\star \end{cases}$$

for $k \in \{1,2\}$.

The code has a maximum error performance as needed for receiver 1. The problem occurs at receiver 2. To decode the codeword this receiver uses the side information. But the side information was re-indexed depending on the message for receiver 2, i.e. the receiver is not able to calculate $\Phi_2^{w_1}(w_2)$ which is needed in the decoding process. To facilitate the decoding at receiver 2 the mapping $\Phi_2^{w_1}(w_2)$ needs to be independent of w_1. This induces the need of a Cartesian structure of the good codewords. In general this requirement can only be satisfied with a rate loss compared to the average probability of error code [52].

The above only shows, that this way does not lead to a code with arbitrarily small maximum probability of error. It might still be possible to construct such a code. The used random coding proof seems inadequate for this task.

Achivability of the Closure of the Rate Region Let $\mathcal{R}(p(x_R))$ denote the rate region which we achieve with the input distribution $p(x_R)$. Since the cardinality of the input set $\mathcal{X}_R$ is finite, the rate region $\bigcup_{p(x_R)} \mathcal{R}(p(x_R))$ is bounded.

The achievability of the closure of the rate region is a consequence of the definition of achievability: What is needed in this step of the proof is the construction of a sequence of codes such that their rate pair converges to a point on the boundary. We know that we have such sequences for any rate point in the interior of the region. In particular there exist such sequences for rate points arbitrarily close to the boundary. The idea now is to choose a sequence of rate pairs that converges to the boundary of the rate region and choose some codes from the sequences of codes corresponding to these rate pairs. As a consequence the rate pair of the resulting new sequence of codes converges to the boundary.

For any rate pair $[I(X_R, Y_2) - \frac{\epsilon}{m}, I(X_R, Y_1) - \frac{\epsilon}{m}]$, $\epsilon > 0$, $m \in \mathbb{N}$, there exists a sequence of

$(2^{\lfloor n(I(X_R,Y_2)-\frac{\epsilon}{m})\rfloor}, 2^{\lfloor n(I(X_R,Y_1)-\frac{\epsilon}{m})\rfloor}, n)$-codes such that $\mu_{k,m}^{(n)} \to 0$, $k \in \{1,2\}$, when $n \to \infty$. Therefore, for any m there exists $n_{0,m}$ such that we have $\mu_{k,m}^{(n)} < \frac{1}{m}$ for $n > n_{0,m}$. Now, let $m^{(n)} = \max\{m : n > n_{0,m}\}$, which denotes the largest m such that $\mu_{k,m}^{(n)} < \frac{1}{m}$ holds. Since $\mu_{k,m}^{(n)} \to 0$, it follows that $m^{(n)} \to \infty$ when $n \to \infty$ so that for the sequence of $(2^{\lfloor n(I(X_R,Y_2)-\frac{\epsilon}{m^{(n)}})\rfloor}, 2^{\lfloor n(I(X_R,Y_1)-\frac{\epsilon}{m^{(n)}})\rfloor}, n)$-codes we have $\frac{1}{n}\lfloor n(I(X_R, Y_k) - \frac{\epsilon}{m^{(n)}})\rfloor \to I(X_R, Y_k)$ with $\mu_k^{(n)} < \frac{1}{m^{(n)}} \to 0$, $k \in \{1,2\}$, when $n \to \infty$. Therefore, the rate pair $[I(X_R, Y_2), I(X_R, Y_1)]$ is achievable and $\mathcal{R}(p(x_R))$ is closed. □

2.1.3.2 Proof of the Converse

Proof. We have to show that any given sequence of $(M_1^{(n)}, M_2^{(n)}, n)$-codes with $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ satisfies $\frac{1}{n}H(W_1) \le I(X;Y_2)$ and $\frac{1}{n}H(W_2) \le I(X;Y_1)$ for a joint distribution $p(x_R)p(y_1,y_2|x_R)$. For a fixed block length n the joint distribution

$$p(w_1, w_2, x_R^n, y_1^n, y_2^n) = \frac{1}{|\mathcal{W}_1|}\frac{1}{|\mathcal{W}_2|}p(x_R^n|w_1, w_2)\prod_{i=1}^{n} p(y_{1,(i)}|x_{(i)})p(y_{2,(i)}|x_{R,(i)})$$

on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{X}_R^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ is well-defined. In what follows the mutual information and entropy expressions are calculated with respect to this distribution.

Lemma 2.3. We can adapt Fano's inequality for our context as

$$H(W_2|Y_1^n, W_1) \le \mu_1^{(n)} \log|\mathcal{W}_2| + 1 = n\epsilon_1^{(n)}, \tag{2.3}$$

with $\epsilon_1^{(n)} = \frac{\log|\mathcal{W}_2|}{n}\mu_1^{(n)} + \frac{1}{n} \to 0$ for $n \to \infty$ as $\mu_1^{(n)} \to 0$.

Proof. From Y_1^n and W_1 node 1 decodes the index W_2 of the transmitted codeword $X_R^n(W_1, W_2)$. We define the event of an error at node 1 as

$$E_1 = \begin{cases} 1, & \text{if } g_1(Y_1^n, W_1) \neq W_2, \\ 0, & \text{if } g_1(Y_1^n, W_1) = W_2. \end{cases}$$

Therefore we have for the mean probability of error $\mu_1^{(n)} = Pr[E_1 = 1]$. We can extend $H(E_1, W_2|Y_1^n, W_1)$ in two different ways using the chain rule for entropies:

$$\begin{aligned} H(E_1, W_2|Y_1^n, W_1) &= H(W_2|Y_1^n, W_1) + H(E_1|Y_1^n, W_1, W_2) \\ &= H(E_1|Y_1^n, W_1) + H(W_2|E, Y_1^n, W_1). \end{aligned} \tag{2.4}$$

Since E_1 is a function of W_1, W_2, and Y_1^n, we have $H(E_1|Y_1^n, W_1, W_2) = 0$. Furthermore, since

E_1 is a binary valued random variable, $H(E_1|Y_1^n, W_1) \leq H(E_1) \leq 1$ and we have

$$\begin{aligned} H(W_2|Y_1^n, W_1, E_1) &= Pr[E_1 = 0]H(W_2|Y_1^n, W_1, E_1 = 0) + Pr[E_1 = 1]H(W_2|Y_1^n, W_1, E_1 = 1) \\ &\leq (1 - \mu_1^{(n)})0 + \mu_1^{(n)} \log(|\mathcal{W}_2| - 1) \\ &\leq \mu_1^{(n)} \log |\mathcal{W}_2|. \end{aligned} \tag{2.5}$$

It follows that

$$\begin{aligned} H(W_2|Y_1^n, W_1) &= H(W_2|E, Y_1^n, W_1) + H(E_1|Y_1^n, W_1) \\ &\leq 1 + \mu_1^{(n)} \log |\mathcal{W}_2|. \end{aligned} \tag{2.6}$$

This concludes the proof of the lemma. □

With the above lemma, we can bound the entropy $H(W_2)$ as follows

$$\begin{aligned} H(W_2) &= H(W_2|W_1) \\ &= I(W_2; Y_1^n|W_1) + H(W_2|Y_1^n, W_1) \\ &\leq I(W_2; Y_1^n|W_1) + n\epsilon_1^{(n)} \\ &\leq I(W_1, W_2; Y_1^n) + n\epsilon_1^{(n)} \\ &\leq I(X^n; Y_1^n) + n\epsilon_1^{(n)} \\ &\leq H(Y_1^n) - H(Y_1^n|X^n) + n\epsilon_1^{(n)} \end{aligned} \tag{2.7}$$

where the equations and inequalities follow from the independence of the messages, the definition of mutual information, Lemma 2.3, the chain rule for mutual information, the positivity of mutual information and the data processing inequality.

If we divide the inequality by n we get the rate

$$\begin{aligned} \frac{1}{n}H(W_2) &\leq \frac{1}{n}\sum_{i=1}^{n}\Big(H(Y_{1,(i)}|Y_1^{i-1}) - H(Y_{1,(i)}|Y_1^{i-1}, X_R^n)\Big) + \epsilon_1^{(n)} \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\left(H(Y_{1,(i)}) - H(Y_{1,(i)}|X_{R,(i)})\right) + \epsilon_1^{(n)} \\ &= \frac{1}{n}\sum_{i=1}^{n} I(Y_{1,(i)}; X_{R,(i)}) + \epsilon_1^{(n)} \end{aligned} \tag{2.8}$$

using the memoryless property and again standard arguments. A similar derivation for the source rate $\frac{1}{n}H(W_1)$ gives the bound

$$\frac{1}{n}H(W_1) \leq \frac{1}{n}\sum_{i=1}^{n} I(Y_{2,(i)}; X_{R,(i)}) + \epsilon_2^{(n)}$$

with

$$\epsilon_2^{(n)} = \frac{\log|\mathcal{W}_1|}{n}\mu_2^{(n)} + \frac{1}{n} \to 0$$

for $n \to 0$ as $\mu_2^{(n)} \to 0$. In words: The rates of the sources are bounded by averages of the mutual informations calculated at the empirical distribution in column i of the codebook. Therefore, we can rewrite these inequalities with an auxiliary random variable U, where $U = i \in \mathcal{U} = \{1, 2, \ldots, n\}$ with probability $\frac{1}{n}$.

$$\begin{aligned}
\frac{1}{n}H(W_2) &\leq \tfrac{1}{n}\sum_{i=1}^{n} I(Y_{1,(i)}; X_{R,(i)}) + \epsilon_1^{(n)} \\
&= \sum_{i=1}^{n} Pr(U = i)I(Y_{1,(i)}; X_{R,(i)}|U = i) + \epsilon_1^{(n)} \\
&= I(Y_{1,U}; X_{R,U}|U) + \epsilon_1^{(n)} \\
&= I(Y_1; X_R|U) + \epsilon_1^{(n)}
\end{aligned} \tag{2.9}$$

and accordingly $\frac{1}{n}H(W_1) \leq I(Y_2; X_R|U) + \epsilon_2^{(n)}$ with $\epsilon_k^{(n)} \to 0$, $k \in \{1, 2\}$, when $n \to \infty$, where $Y_k = Y_{k,U}$ and $X_R = X_{R,U}$ are new random variables whose distribution depend on U in the same way as the distributions of $Y_{k,(i)}$ and $X_{R,(i)}$ depend on i. Now in the current coding scenario $U \to X_R \to Y_k$, $k \in \{1, 2\}$ forms a Markov chain and therefore $I(Y_k; X_R|U) \leq I(Y_k; X_R)$. This completes the proof of the converse and the proof of the capacity region of the bidirectional broadcast channel. □

2.1.4 Discussion and Example

The coding principles are similar to the network coding approach where we would have implemented a bitwise XOR operation on the decoded messages at the relay node [24], [25], [26]. In fact a slight change in the above proof reveals that — without any rate loss — the encoder could also operate on the modulo sum of the two messages represented in an appropriate field. The difference to the usual network coding approach lies in the fact that we use the side information in the channel decoding, while the standard XOR approach inverts the modulo operation after the decoding. Therefore for the network coding approach the achievable rates in the BC phase are limited by the worst receiver, i.e $R_1, R_2 \leq \min\{I(X; Y_1), I(X; Y_2)\}$ for some input distribution $p(x)$. For our coding scheme each achievable rate depends only on the common input distribution and its own channel distribution. This means that we can find the optimal input distribution for each channel separately, which achieves the maximal achievable rate for that link (equal to the single link capacity). Though this input distribution need not be optimal for the other channel.[1] Therefore, we see that the network coding approach using XOR on the decoded messages at the relay and after the channel decoding at the terminal nodes

[1]It is curious that if we transfer the result to scalar Gaussian channels we will see that one input distribution will maximize both links simultaneously.

achieves the capacity of the broadcast phase if and only if for the maximizing input distribution $p^\star(x) = \arg\max_{p(x)} \max\{I(X;Y_1), I(X;Y_2)\}$ we have $I(X;Y_1) = I(X;Y_2)$.

Example 2.1 (The Binary Symmetric Broadcast Channel)**.** *For the binary symmetric broadcast channel, let p_1 and p_2 denote the probability that the relay input $X \in \{0,1\}$ is complemented at the output $Y_1 \in \{0,1\}$ and $Y_2 \in \{0,1\}$ of node 1 and 2 respectively. From [30, Chapter 8.1.4] we know that an uniform input distribution maximizes the binary symmetric channel. Therefore, the broadcast capacity region for the binary symmetric channel is given by*

$$C_{\mathrm{BC}} = [0, 1 - H(p_2)] \times [0, 1 - H(p_1)], \tag{2.10}$$

which includes the region $[0, 1 - \max\{H(p_1), H(p_2)\}]^2 \subset \mathbb{R}_+^2$, which is achievable using XOR at the relay node according to [24].

2.1.4.1 A Note on Coding Mechanisms in the BC phase

Before putting together the pieces to get an achievable rate region for the two-phase two-way relay channel, we take a deeper look at the coding mechanism, which permits the seemingly interference free transmission in the BC phase. In the following sections and chapters we will give a similar discussion to point out where the proposed schemes differ, and what the key features in the different approaches are and how they could be used for practical coding schemes. A practical coding scheme that follows directly from the mechanism of the coding is proposed in Section 2.2. It facilitates the use of single user codes. The following discussion will give the motivation for this scheme.

In what follows suppose a code for the BC phase is given which facilitates the required performance, i.e. the code has a sufficiently low error probability for both the receivers. The first thing to note is that the BC phase uses only one code for both users together. This code has a Cartesian structure because of the constraint that the relay is able to decode both messages. This is depicted in Figure 2.2 on the left. Each square represents a codeword for one message pair. Therefore, in the general case there is one codeword for each pair of messages (w_1, w_2). If a message pair is transmitted, the encoder chooses the codeword corresponding to that message pair and transmits it to both receivers. In the figure such a codeword is indicated by $\bullet$.

The right side of Figure 2.2 shows what happens at the receiver. The receiver knows its own message, e.g. node 1 knows the message w_1. Now, for this decoder the possible choices of codewords are reduced by the side information, i.e. the decoder decodes the message using only a sub-code of the original code determined by the side information. The codewords of this sub-code $\mathcal{C}(w_1)$ are marked by squares filled with vertical lines in the figure. Similarly decoder 2 uses a sub-code $\mathcal{C}(w_2)$ marked by squares filled with horizontal lines in the figure. The transmitted codeword belongs to both these sub-codes. If the overall code for the BC phase has a good performance than — in average — all the sub-codes $\mathcal{C}(w_1)$ are good codes for the channel to receiver 1, while all the sub-codes $\mathcal{C}(w_2)$ are good codes for the channel to receiver

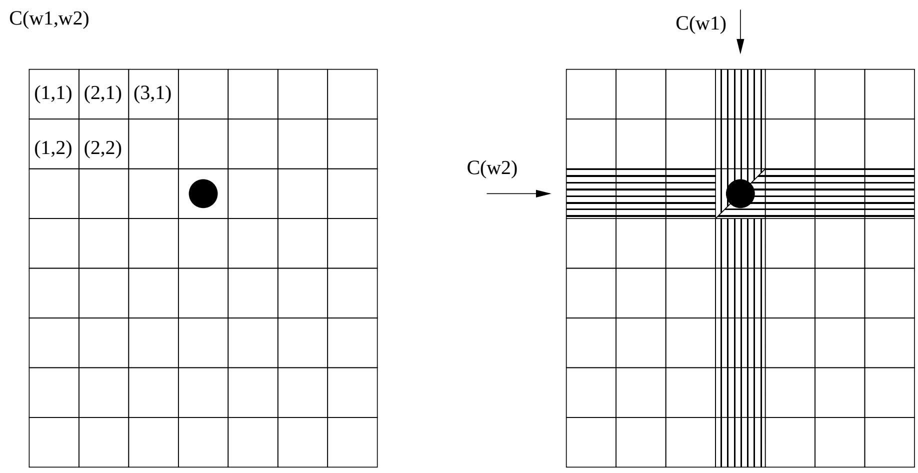

Figure 2.2: Coding mechanisms in the BC phase: The left hand side of this figure shows the Cartesian structure of the code used by the relay to encode the two messages. On the right hand side the decoding at the receivers is shown. Each square represents a codeword for one message pair. The transmitted codeword for the message pair (w_1, w_2) is indicated by $\bullet$. Both receivers use a sub-code of the relay's code for the decoding. These sub-codes depend on the side information; the sub-codes for the actual side information are marked with vertical and horizontal lines. The transmitted codeword belongs to the sub-codes of both the messages.

2. Note that some of the codes may be bad codes, as the proof considers only the average probability of error.

To construct a code for the BC phase one could use this interpretation and start with a set of codes for both users. These sets of codes need to be interwoven as it is depicted in Figure 2.2: The codeword $C(w_1, w_2)$ needs to be a codeword of both codes, $C(w_1)$ and $C(w_2)$. An important thing to note is, that as the decoder does not care about codewords which are not contained in the code $C(w_1)$ for the given side information w_1, the same codeword may be used for different pairs of messages, i.e. we can have $c(w_1, w_2) = c(\hat{w}_1, \hat{w}_2)$ for $w_1 \neq \hat{w}_1$, $w_2 \neq \hat{w}_2$. As the channel is independent of the message, this indicates that one could use the same set of codewords for all side informations — say we use the set of codewords from the code $C(w_1)$ for all possible side information if the number of messages M_1 for receiver 2 is not greater than the number of messages M_2 for receiver 1. Only the encoder and decoder mapping need to be different for every side information. This induces codes $C(w_2)$ for the second user. If the number of codewords $M_1 < M_2$, this code construction leads to sub-codes containing different codewords for receiver 2. Therefore using this code construction one has to ensure that in average the resulting sub-codes for receiver 2 satisfy the needed performance requirement.

Now, compare the coding mechanism with that of the XOR coding scheme [24]. In the XOR coding scheme the decoded messages are combined with an XOR operation. Therefore the shorter message is padded with some predefined symbols, e.g. with zeros. Note that for the XOR scheme it turns out that we have $M_1 = M_2$ if we operate at maximum sum rate. Therefore in this case applying the idea of the coding scheme one may use the same codewords in every sub-code $C(w_1)$, e.g. with a mapping cyclically shifted for all the different side information. This leads to sub-codes $C(w_2)$ that consist all of the same codewords, again with a mapping which is cyclically shifted. The cyclical shift in turn can be expressed by an XOR operation of the messages in an adequate representation, e.g. binary.

Another way of looking at the problem is to assume that we already calculated the XOR operation on the messages $w_1 \oplus w_2$. Lets assume that the number of messages $M_1 < M_2$ and therefore lets say the binary representation of w_1 was padded with zeros to allow the XOR operation. Instead of decoding the complete code used by the relay to encode the resulting message $w_1 \oplus w_2$ consisting of M_2 different codewords and invert the XOR operation after decoding, one can now as well decode in a sub-code of this code. Indeed for a given side information w_2 there are only M_1 possible codewords to choose from. Using the random coding argument, one can show that there exist codes such that in average over the resulting sub-codes for receiver 2 the error probability goes to zero as the block length goes to infinity. Only a small change in the proof of achievability is required to show that the relay may also operate on the XOR sum of both messages, and still the same rate pairs are achievable. The important difference to the standard XOR approach is, that now the decoder uses the side information to restrict the number of possible codewords before decoding. This allows the rate of both the nodes to be chosen according only to its own respective channel, i.e. the rate for the node with the better channel is not restricted by the weaker channel.

2.1.5 Time Division between MAC and BC

Using the above capacity results for the MAC and BC phase we can now state an achievable rate region by concatenating both phases. The achievable rate region follows by timesharing between both phases. The rate pair needs to be achievable for both the sub-systems.

Recall the setup given in Chapter 1: Node 1 wants to transmit message w_1 with rate nR_1 in n channel uses of the two-way relay channel to node 2. Simultaneously, node 2 wants to transmit message w_2 with rate nR_2 in n channel uses to node 1. Then let n_1 and $n_2 = n - n_1$ denote the number of channel uses in the MAC phase and BC phase with the property $\frac{n_1}{n} \to \alpha \in [0,1]$ and $\frac{n_2}{n} \to \beta = 1 - \alpha$ as $n \to \infty$ respectively. With a sufficient block length n (respectively n_1 and n_2) we can achieve a two-way transmission of messages w_1 and w_2 with arbitrary small decoding error if rate pairs $[R_1^{\text{MAC}}, R_2^{\text{MAC}}] \in C_{\text{MAC}}$ and $[R_1^{\text{DF-BC}}, R_2^{\text{DF-BC}}] \in C_{\text{DF-BC}}$ exist so that we have

$$\begin{aligned} nR_1 &\leq \min\{n_1 R_1^{\text{MAC}}; n_2 R_1^{\text{DF-BC}}\}, \\ nR_2 &\leq \min\{n_1 R_2^{\text{MAC}}; n_2 R_2^{\text{DF-BC}}\}. \end{aligned}$$

This argumentation holds as we can apply a union bound to bound the probability of error from above by the sum of the error probabilities of the two phases. As the error probability of the two phases goes to zero as $n_1, n_2 \to \infty$, the sum goes to zero as $n \to \infty$. For the code for the two-

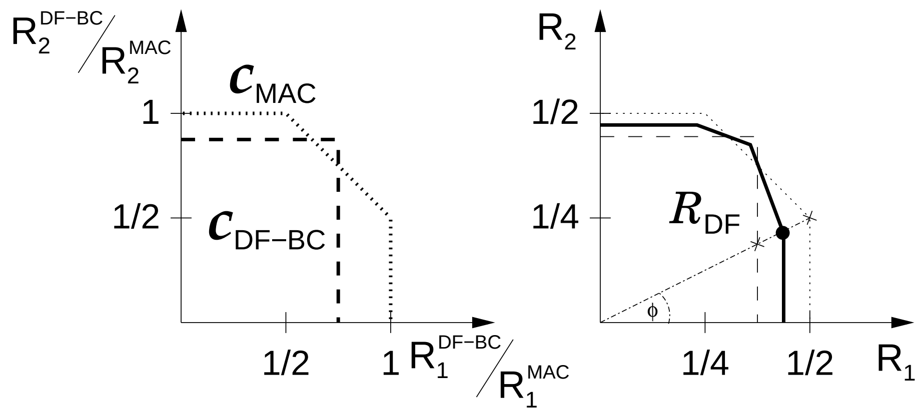

Figure 2.3: The left figure shows the capacity regions $\mathcal{C}_{\text{MAC}}$ (dotted line) and $\mathcal{C}_{\text{DF-BC}}$ (dashed line), the right figure shows the corresponding achievable rate region $\mathcal{R}_{\text{DF}}$ (solid line). The dashed-dotted line shows exemplary for one angle ϕ the achievable rate pair ($\bullet$) on the boundary of $\mathcal{R}_{\text{DF}}$ with the optimal time-division between the two rate tuples ($\times$) on the boundary of $\mathcal{C}_{\text{MAC}}$ and $\mathcal{C}_{\text{DF-BC}}$.

phase two-way relay channel we use the encoder of the code used in the proof of the achievable rate for the MAC. The decoder is that of the bidirectional BC. The encoder at the relay is a concatenation of the decoder for the MAC and the encoder of the BC. Using appropriate codes from the sequences of codes used in the proofs for MAC an bidirectional BC we can get for all n a code for the two-phase two-way relay channel such that for the resulting sequence of codes the probability of error goes to 0 as $n \to \infty$. Therefore the rate pair $[R_1, R_2]$ is achievable.

As a consequence, an achievable rate region of the two-way relay channel is given by the set of all rate pairs $[R_1, R_2]$ which are achievable with some time-devision parameter $\alpha, \beta \in [0, 1]$ as $n \to \infty$. We collect the previous consideration in the following theorem.

Theorem 2.4. The achievable rate region $\mathcal{R}_{\text{DF}} \subset \mathbb{R}_+^2$ of the two-phase two-way relay channel is given by all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 \leq &\ \min\{\alpha I(X_1; Y_R|X_2, Q), \beta I(X_R; Y_2)\} \\ R_2 \leq &\ \min\{\alpha I(X_2; Y_R|X_1, Q), \beta I(X_R; Y_1)\} \\ R_1 + R_2 \leq &\ \alpha I(X_1, X_2; Y_R|Q) \end{aligned} \tag{2.11}$$

for some joint probability distribution $p(q)p(x_1|q)p(x_2|q)p_1(y_R|x_1, x_2)p(x_R)p_2(y_1, y_2|x_R)$ and some $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$.

Since the capacity region of the BC phase $\mathcal{C}_{\text{DF-BC}}$ is larger than the region achieved by applying interference cancellation [22, 23] or by performing an XOR operation on the decoded messages at the relay node [24, 25, 26], the achievable rate region $\mathcal{R}_{\text{DF}}$ includes the region which can be achieved by interference cancellation and network coding.

Remark 2.3. The region $\mathcal{R}_{\text{DF}}$ is convex. This can be seen by noting that $[R_1, R_2] \in \mathcal{R}_{\text{DF}}$ if and only if for some fixed α, $\beta = 1 - \alpha$ we have $[R_1, R_2] = \alpha[R_1^{\text{MAC}}, R_2^{\text{MAC}}]$ with $[R_1^{\text{MAC}}, R_2^{\text{MAC}}] \in$

$\mathcal{C}_{\mathrm{MAC}}$ and $[R_1, R_2] = (1-\alpha)[R_1^{\text{DF-BC}}, R_2^{\text{DF-BC}}]$ with $[R_1^{\text{DF-BC}}, R_2^{\text{DF-BC}}] \in \mathcal{C}_{\text{DF-BC}}$. This is a direct consequence of the definition of $\mathcal{R}_{\mathrm{DF}}$, $\mathcal{C}_{\mathrm{MAC}}$, and $\mathcal{C}_{\text{DF-BC}}$. Now define two sets

$$\mathcal{S}_1 = \{[[R_1^{\mathrm{MAC}}, R_2^{\mathrm{MAC}}], 0, 0] | [R_1^{\mathrm{MAC}}, R_2^{\mathrm{MAC}}] \in \mathcal{C}_{\mathrm{MAC}}\}$$

and

$$\mathcal{S}_2 = \{[0, 0, R_1^{\text{DF-BC}}, R_2^{\text{DF-BC}}] | [R_1^{\text{DF-BC}}, R_2^{\text{DF-BC}}] \in \mathcal{C}_{\text{DF-BC}}\}.$$

Furthermore define $\mathcal{S}_3 = \mathrm{convexHull}\{\mathcal{S}_1 \cup \mathcal{S}_2\}$. The set $\mathcal{S}_3$ is convex. Now define $\mathcal{S}_5$ as the intersection of $\mathcal{S}_3$ with a plane given by $\mathcal{S}_4 = \{[a, b, c, d] | a = c, b = d\}$, i.e. $\mathcal{S}_5 = \mathcal{S}_3 \cap \mathcal{S}_4$. By construction $\mathcal{S}_5$ is convex. The set contains all points $[a, b, c, d]$ such that $[a, b] = \alpha[R_1^{\mathrm{MAC}}, R_2^{\mathrm{MAC}}]$ with $[R_1^{\mathrm{MAC}}, R_2^{\mathrm{MAC}}] \in \mathcal{C}_{\mathrm{MAC}}$ and $[a, b] = (1-\alpha)[R_1^{\text{DF-BC}}, R_2^{\text{DF-BC}}]$ with $[R_1^{\text{DF-BC}}, R_2^{\text{DF-BC}}] \in \mathcal{C}_{\text{DF-BC}}$. Therefore we have $\mathcal{R}_{\mathrm{DF}} = \{[R_1, R_2] | \exists c, d : [R_1, R_2, c, d] \in \mathcal{S}_5\}$. Therefore we conclude that $\mathcal{R}_{\mathrm{DF}}$ is convex. An alternative proof can be given using arguments analogous to that used in Remark 2.11 below.

We close this section with looking briefly at an example with binary channels. In Figure 2.3 the capacity regions $\mathcal{C}_{\mathrm{MAC}}$ and $\mathcal{C}_{\text{DF-BC}}$ as well as the achievable rate region $\mathcal{R}_{\mathrm{DF}}$ are depicted for a symmetric binary erasure multiple access channel [30, Example 14.3.3] and a binary symmetric broadcast channel, cf. equations (2.10). The boundary of the achievable rate region can be obtained geometrically if one takes for any angle $\phi \in [0, \pi/2]$ half of the harmonic mean between the boundary rate tuples of the capacity regions where we have $\tan\phi = \frac{R_2^{\mathrm{MAC}}}{R_1^{\mathrm{MAC}}} = \frac{R_2^{\text{DF-BC}}}{R_1^{\text{DF-BC}}}$.

2.2 A Practical Coding Scheme for the Broadcast Phase

In the last section we derived the capacity region for the broadcast phase of a two-way relay channel with decode-and-forward constraint. The proofs of Theorem 2.2 rely on random coding arguments, hence provide only limited insight into the problem of designing practical codes for the bidirectional relay channel. So, this section complements the previous work by presenting a simple coding scheme that achieves or closely approaches the asymptotic capacity bounds of Theorem 2.2. Again, we focus on the second (broadcast) phase of the overall communication protocol of the two-way relay channel with decode-and-forward at the relay. In doing so, we neglect potential decoding errors in the first phase, which is equivalent to the assumption that the relay node has perfect knowledge about the messages of both terminals. This knowledge can be used to enhance the performance of the broadcast channel [6][24, 22, 26, 53].

To start the discussion on the design of practical codes for the broadcast phase of the two-way relay channel with decode-and-forward at the relay, we first review Theorem 2.2:

Theorem (Theorem 2.2)**.** The capacity region $\mathcal{C}_{\text{DF-BC}}$ for sending two sources W_1 and W_2 over the memoryless broadcast channel where the receiving node 1 knows W_1 and node 2 knows W_2

is the set of all rate pairs $[R_1, R_2]$ satisfying

$$R_1 \leq I(X_R; Y_2),$$
$$R_2 \leq I(X_R; Y_1),$$

for some probability function $p(x_R)p(y_1, y_2|x_R)$.

This result can be interpreted as follows: The performance for each terminal is in many cases close to the performance of a single-user channel. The possible performance loss compared to the single-user channel is a potentially suboptimal channel input distribution, which may be needed to achieve some rate tuples. It is interesting to point out that in case of scalar Gaussian channels, there is no loss compared to the marginal single-user channels, as the Gaussian input distribution maximizes jointly both the mutual information expressions.

In this section, we address the problem of designing practical codes for this communication channel, which is referred to as the *bidirectional broadcast channel*[2]. More precisely, we provide guidelines for the use of well-developed single-user codes in the bidirectional broadcast channel, with the goal of achieving or closely approaching the performance bounds of Theorem 2.2. In particular, in the case of finite alphabet channels with the channel distortion being independent of the channel input, the performance of our coding scheme only depends on the performance of the base codes on the corresponding single-user channels. In other words, if the involved single-user codes provide the best possible performance, so does the proposed coding scheme for the bidirectional broadcast channel. In case of arbitrary channels, however, further performance gains may be achieved by better exploiting the distortion characteristics of the channel.

The main advantage of the proposed coding scheme is that it only involves simple operations on single-user codes designed for some types of single-user channels. Consequently, from the practical point of view, the proposed scheme is very attractive as it merely requires minor modifications of the traditional single-user coding schemes. In particular, it can be easily generalized to the case of additional soft-information. Finally, note that the assumption of channels with finite input/output alphabets is not restrictive in view of practical system implementation, since in practical systems usually finite input alphabets are used; the channel output is quantized to a finite alphabet, possibly with additional soft-information attached to every symbol, which can be used in the proposed scheme as well.

2.2.1 A Coding Scheme for Symmetric Marginal Channels

In this section, we propose a coding scheme for the bidirectional broadcast channel. This scheme is quite general and works for several coding techniques. After introducing some notation and basic assumptions, we present an example of a concrete realization of the scheme

[2]We use the term "bidirectional" because each receiver knows the messages intended for the other receiver. From this point on, "terminal" and "user" are used interchangeably.

that is based on linear block codes. For brevity, we focus on finite code lengths. An extension to infinite code length that is relevant in view of convolutional coding is straightforward, but requires more cumbersome notation.

2.2.1.1 Symmetric Marginal Channels

We consider a broadcast channel with finite input and output alphabets. If the probability transition function of a broadcast channel is $p(y_1, y_2|x_R)$, then the channels from the relay node to both the terminals with the conditional marginal distributions $p(y_1|x_R)$ and $p(y_2|x_R)$ are referred to as *marginal channels*. A key assumption for our analysis is the symmetry property defined below.

Definition 2.5 (Symmetric marginal channel)**.** A marginal channel is called symmetric if the following holds.

(i) For both users $k \in \{1, 2\}$ the channel has an output alphabet $\mathcal{Y}_k = \{0, 1, \ldots, M-1, e\}$ of cardinality $M + 1$ and a common input alphabet $\mathcal{X}_R = \mathcal{Y}_k \backslash \{e\}$ of cardinality M, where e is the erasure symbol. We assume that an addition $+$ is defined such that $(\mathcal{X}_R, +)$ forms an Abelian group with neutral element 0. The addition is extended to $\mathcal{Y}_k, k \in \{1, 2\}$, by defining $a + e = e + a = e$ for any $a \in \mathcal{X}_R$. Furthermore there is no inverse element of the erasure symbol e.

(ii) For some integer L, the channel output vectors of length L are given by $y_k^L = x_R^L + z_k^L \in \mathcal{Y}_k^L$, where $k \in \{1, 2\}$ and $z_k^L \in \mathcal{Y}_k^L$ has some distribution on $\mathcal{Y}_k^L$ independent of $x_R^L \in \mathcal{X}_R^L$. The addition is according to the definition above.

Remark 2.4. The channel parameter L is introduced to make the result more general. By allowing for $L > 1$ the noise may be structured. It is only required that for sequences of length $L < \infty$ the noise vectors are i.i.d. and independent of the channel input. A simple example for $L = 2$ is a binary channel, where every second bit is inverted with some probability while the other bits are transmitted without distortion.

Remark 2.5. Note that both conditions are satisfied by many important channels, of which the most prominent one is the binary symmetric channel with erasure. For usual error correction code design it is often assumed that the channel satisfies the above conditions. This is motivated by the assumption that modulation and demodulation are used often in combination with some scrambling; therefore one can abstract from the real physical channel that may be highly non-symmetric. This assumption is often suboptimal, especially if it is the structure of the channel coding in combination with the channel characteristics which are used for performance gains, as in this scheme.

In this section, the marginal channels are assumed to be symmetric in the sense of the above definition. Section 2.2.4.1 illustrates potential consequences of dropping this assumption. For simplicity, in what follows we slightly abuse the notation by using the symbols in $\boldsymbol{y}_k = \boldsymbol{x}_R + \boldsymbol{z}_k$ for the vectors of channel output, channel input and channel distortion, even if we consider inputs of length $N = aL$ for some $a \in \mathbb{N}$.

2.2.1.2 Encoding and Decoding

Our coding scheme is based on two given base codes[3], say codes C_1 and C_2 that are defined over the channel input alphabet $\mathcal{X}_R$ with encoders E_1 and E_2 as well as decoders D_1 and D_2, respectively. It is assumed $C_k \subseteq \mathcal{X}_R^N, k \in \{1, 2\}$, where N is a multiple of the channel parameter L. Adequate codes and encoders can be found by simply concatenating several codewords of some given codes to generate new codes with a code length which is for instance the least common multiple of the length of both the codes and the channel parameter L. Both base codes may have different coding rates R_1 and R_2 (in bits per code symbol) so that the encoders

$$E_k : \{0, 1, \dots, 2^{R_k N} - 1\} \to C_k \subseteq \mathcal{X}_R^N$$

generate codewords $\boldsymbol{c}_k \in \mathcal{X}_R^N, k \in \{1, 2\}$. For each code, say code $k \in \{1, 2\}$, the decoder

$$D_k : \mathcal{Y}_k^N \to \{0, 1, \dots, 2^{R_k N} - 1\}$$

is assumed to decode a received word $\boldsymbol{y}_k = \boldsymbol{c}_k + \boldsymbol{z}_k$ correctly iff the distortion word $\boldsymbol{z}_k$ is in the set of correctable errors $\mathcal{E}_k^{cor}$, i.e. iff $\boldsymbol{z}_k \in \mathcal{E}_k^{cor}$.

Now, the coding and thereby the code C for the bidirectional broadcast channel is defined as follows:

Encoding: Suppose that $w_k \in \{0, 1, \dots, 2^{R_i N} - 1\}$ is a given message of user $k \in \{1, 2\}$. Then, the encoder at the relay node is a mapping

$$\psi : \{0, 1, \dots, 2^{R_1 N} - 1\} \times \{0, 1, \dots, 2^{R_2 N} - 1\} \to C \subseteq \mathcal{X}_R^N$$

with the encoding rule given by

$$\psi(w_1, w_2) = E_1(w_2) + E_2(w_1)\,.$$

As before, the addition is symbol-wise and defined over the Abelian group $\mathcal{X}_R$. In words, the two encoders generate $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ using the encoders on the information generated for the other user respectively, i.e. E_1 on the information w_2 and E_2 on the information w_1. The resulting codewords are added and $\boldsymbol{x}_R = \psi(w_1, w_2) = \boldsymbol{c}_1 + \boldsymbol{c}_2$ is transmitted via the broadcast channel to users 1 and 2, which observe $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, respectively.

[3] We use the term "code" as a set of codewords. Encoder and decoder are not part of the code, but the encoder may define the code. Encoder, decoder and code together form a coding scheme.

Decoding: The decoders at the terminals are mappings

$$\phi_1 : \{0, 1, \ldots, 2^{R_1 N} - 1\} \times \mathcal{Y}^N \to \{0, 1, \ldots, 2^{R_2 N} - 1\}$$

$$\phi_2 : \{0, 1, \ldots, 2^{R_2 N} - 1\} \times \mathcal{Y}^N \to \{0, 1, \ldots, 2^{R_1 N} - 1\}$$

The decoding rules are as follows: Upon receiving $\boldsymbol{y}_1$, the first user uses its perfectly known side information w_1 to generate $\boldsymbol{c}_2 = E_2(w_1)$. Then, it calculates $\hat{\boldsymbol{c}}_1 = \boldsymbol{y}_1 - \boldsymbol{c}_2$, where "−" denotes the addition of the inverse element, and declares the estimate of w_2 to be $\hat{w}_2 = D_1(\hat{\boldsymbol{c}}_1)$. The decoder of the second user is defined accordingly. In our setup, w_k can be interpreted as side information perfectly known to user $k \in \{1, 2\}$.

Remark 2.6 (Codes are constrained to use the same input alphabet)**.** Note that both codes use the same alphabet $\mathcal{X}_R$. Although this assumption may appear as a significant restriction, it does not impact the generality of the analysis, since we do not require the use of all alphabet symbols. Therefore, starting at codes defined on different alphabets with a possibly lower cardinality, the codes can be transformed to a common alphabet. The restriction is necessary and reasonable since the marginal channels use the same input.

In fact, from a practical point of view, the receiver side, i.e. the channel output alphabet, poses more problems. For instance, it is not clear what should be done if different modulations are used on a marginal channel in a real system. In this case, it may be required to restrict the alphabet at the receiver from some super-alphabet to the needed modulation alphabet before detection. A crucial point here is that the subtraction now needs to be done in signal space without knowing the received constellation point. However, using a scheme similar to the proposed one, where addition is defined in a modulo like manner based on some lattice on the signal space as it is done for lattice codes [54, 55, 56], one arrives at a practical scheme for the broadcasting in the two-way relay channel. This is nothing but a straightforward generalization of the proposed scheme to coding in signal space, where the modulo addition is used to fulfill a possibly given power restriction. However, the theorems of this section concerning the performance of the codes do not generalize to such channels. The resulting coding is in general suboptimal. A more elaborated discussion on this issue is given in Section 2.2.4.2.

Remark 2.7 (Joint network-channel coding)**.** The coding at the relay node can be seen as a joint network-channel coding. Instead of using the traditional network code with a modulo addition performed on data symbols as proposed e.g. by [24, 26, 53], our scheme follows the lines of the information theoretical analysis above. As a result, the modulo addition is performed on the codeword symbols. An important difference to the traditional approach is that we can easily handle the case of marginal channels of different quality, as the base codes may have different coding rates. For the reasons mentioned in what follows, the proposed scheme is referred to as *joint network-channel coding* and the code C as *joint network-channel code*. The code we use is in essence a nested code [54, 57] which allows different interpretations of the transmitted information for both receivers.

Note that except for the code length, the two base codes can be chosen independently. Indeed, as far as the coding performance is concerned, it does not even matter whether they consist of the same or have completely different codewords. Also note that in general, the resulting encoding mapping ψ is not necessarily bijective so that there may be no way to decode messages without side information.

2.2.1.3 Examples

Example 2.2 (Identical Linear Codes)**.** *Consider two identical linear codes $C_1 = C_2 = \hat{C}$. In this case, the resulting joint network-channel code is the same as the base code so that $C = \hat{C}$. Only the encoder ψ and the decoders ϕ_k, $k \in \{1, 2\}$, are different as they depend on the side information. Note that without side information, no information symbol can be recovered. Due to the linearity of all the operations, the addition and subtraction can be performed on the information symbols without any loss of performance. Therefore, the modulo addition on the data symbols as e.g. proposed by [24, 26, 53] can be seen as a special case of the proposed coding scheme. Another interpretation is that the side information is used for data compression. In this case, the encoding is in fact a classical Slepian-Wolf encoding [37], where the transmitter needs to compress (w_1, w_2) for two users that have side information w_1 and w_2, respectively [58]. The compressed data is broadcasted using an error correction code that fits for both the marginal channels. Following this interpretation there is no side information used for channel coding.*

Example 2.3 (Codes with different Codewords)**.** *Now consider two base codes having different codewords. In this case, the resulting joint network-channel code may have more codewords than either of the two base codes. This may be true for instance if the two channels are of different quality. To illustrate this, consider the following binary codes $C_1 = \{000, 111\}$ and $C_2 = \{000, 110, 101, 011\}$, in which case the encoding mapping ψ is bijective, meaning that there is no compression using side information. Because of the increased number of codewords, both users may not be able to correct the transmission error without side-information; for the example at hand, all tuples in $\{0, 1\}^3$ are codewords, and therefore the error correction or detection is impossible without side information. Error correction decoding (or error detection for the second user) becomes possible by restricting the decoding to a subset of codewords with the help of the side information. In this example, there is no compression at all since $|C| = |C_1||C_2|$, but the side information is used in a similar way as the already decoded information is used in the decoding of the weaker signals in interference cancellation schemes (see for instance [59]).*

2.2.2 Analysis of the Coding Scheme

2.2.2.1 A Conservation Law for the Coding Scheme

The following theorem shows that the performance characteristics of the base codes in the corresponding marginal single-user channels are preserved under the proposed coding scheme.

Theorem 2.5. The proposed coding scheme with encoder-decoder pair $(\psi, \phi_k), k \in \{1, 2\}$, and code C has exactly the same performance as the base coding scheme consisting of the encoder-decoder pair (E_k, D_k) and base code C_k in the corresponding marginal (single-user) channel.

Proof. User 1 observes $\boldsymbol{y}_1 = \boldsymbol{x}_R + \boldsymbol{z}_1$ and computes

$$\boldsymbol{y}_1 - \boldsymbol{c}_2 = \boldsymbol{x}_R + \boldsymbol{z}_1 - \boldsymbol{c}_2 = \boldsymbol{c}_1 + \boldsymbol{c}_2 + \boldsymbol{z}_1 - \boldsymbol{c}_2 = \boldsymbol{c}_1 + \boldsymbol{z}_1 \,.$$

By assumption, this can be decoded correctly iff $\boldsymbol{z}_1 \in \mathcal{E}_1^{cor}$. Now, since the distribution of $\boldsymbol{z}_1$ is the same as in the marginal channel, the performance of the code is the same as in the marginal channel. The same reasoning holds for the second user, completing the proof. □

2.2.2.2 An Optimality Property of the Scheme

An immediate consequence of Theorem 2.5 is that in our setting the performance of the proposed coding scheme is entirely determined by the performance of the base codes in the corresponding marginal channels. The following theorem proves an optimality property of the proposed procedure in the sense that it enables us to construct a coding scheme whose performance for both users is at least as good as any other coding scheme for the bidirectional broadcast channel.

Theorem 2.6. For the considered symmetric channel, we can always find two base codes of length N with encoders $E_1(w_2)$, $E_2(w_1)$ and decoders $D_1(\boldsymbol{y}_1), D_2(\boldsymbol{y}_2)$ such that the resulting code with encoder $\psi(w_1, w_2)$ and decoders $\phi_1(w_1), \phi_2(w_2)$ has a probability of error for both the users at least as low as any given code of length N with encoder $\theta(w_1, w_2)$ as well as decoders $\rho_1(w_1, \boldsymbol{y}_1)$ and $\rho_2(w_2, \boldsymbol{y}_2)$. This is true even if a given code with encoder $\theta(w_1, w_2, q)$ permits a certain randomness depending on some randomization parameter $q \in Q$ as long as q is independent of w_k.

Remark 2.8. The parameter $q \in Q$ is introduced to make the result more general. Due to the parameter the result holds for coding schemes which use some kind of common randomness, e.g. some dither, scrambling, or random interleaving.

Proof. Suppose that $q \in Q$ is a randomization parameter independent of w_1 and w_2 and that some code for a symmetric channel with encoder $\theta(w_1, w_2, q)$ is given. For deterministic encoding, let the cardinality of Q be $|Q| = 1$. Based on this code and encoder, we can define codes for the marginal channels of both the users by fixing the side information and the randomization parameter, until we find good marginal codes. More precisely: For an arbitrary pair (w_1, q) of side information w_1 and the randomization parameter q, define a code for user 1 with encoder $\hat{E}_{1,w_1,q}(w_2) = \theta(w_1, w_2, q)$ and decoder $\hat{D}_{1,w_1,q}(\boldsymbol{y}_1) = \rho_1(w_1, \boldsymbol{y}_1, q)$. This encoder defines a code of length N. Check the probability of error for this coding scheme on the marginal channel. If the probability of error of this code and this user is at least as low as that of the given code, fix encoder $E_1(w_2) = \hat{E}_{1,w_1,q}(w_2)$ and decoder $D_1(\boldsymbol{y}_1) = \hat{D}_{1,w_1,q}(\boldsymbol{y}_1)$, and then use this code as base code for user 1. Otherwise, check another pair (w_1, q) of side information w_1 and randomization parameter q. As q is independent of w_1 and w_2, and the decoder is the same as in the original code, there exists at least one such code. The same is repeated for user 2. By Theorem 2.5, the code with encoder $\psi(w_1, w_2)$ constructed based on the single user encoders $E_1(w_2)$ and $E_1(w_2)$ has an error probability at least as low as the given code for both the users, if the decoding $\phi_k(w_k)$ use $D_k(\boldsymbol{y}_k)$ after subtracting the encoded side information. Furthermore, the code is of length N as required by the theorem. □

This proves the optimality of the proposed scheme. By using the coding not on the data but, as proposed by the information theoretical result, on the channel input, we can use different coding rates for both the users and may — provided that we find good single user codes — achieve any point in the capacity region given by Theorem 2.2.

2.2.3 Interpretation and Example

In general, the joint network-channel code utilizes more codewords than the users are able to differentiate. Furthermore, since $\max\{|C_1|, |C_2|\} \leq |C| \leq |C_1||C_2|$ holds with strict inequalities for many base codes, the proposed scheme is in general neither full compression nor pure channel coding. In fact, depending on the choice of the base codes, one achieves a certain point on a tradeoff curve between these two extreme points. By eliminating all the effects caused by the simultaneous transmission, the proposed joint network-channel coding restricts the error correction decoding at the receiver to a sub-code, which is the corresponding base code. By the symmetry of the channel, the decoder of this error correction code "sees" only the distortion caused by the marginal channel. As the code is designed for this channel, the performance will be as desired.

Another interpretation of the coding/decoding scheme, which is more along the lines of the information theoretical result, is that of a set of codes[4]. The direct part of the proof of Theorem 2.2 indicates that one should design several codes with encoder and decoder for each user, all corresponding to the marginal channel; one triplet of encoder, code and decoder for each potential side information.

Following these requirements, we need $2^{R_1 N}$ encoders E_{1,w_1} to encode w_2 and $2^{R_2 N}$ encoders

[4]Note that the codes in the set need not to be disjoint. It may even happen that some of these codes are identical and only the decoder and encoder differ. In fact, the design is a classical coset code design as proposed by [36], except that our scheme may use the same coset with different mappings several times.

E_{2,w_2} to encode w_1. Furthermore, only one codeword should be transmitted. Therefore the encoders and codes need to be interwoven with each other such that $E_{1,w_1}(w_2) = E_{2,w_2}(w_1)$. If this equation holds then given (w_1, w_2) the output of both encoders E_{1,w_1} and E_{2,w_2} is the same, in spite of the codeword carrying different information for different receivers.

In general, the requirements on coding that follow from Theorem 2.2 seem to be hard to satisfy. Nevertheless, the proposed coding scheme is a simple recipe to realize codes with the required properties: The codes and corresponding encoder mappings for different side information are created by shifting one base code and the encoder mapping in $\mathcal{X}_R^N$. The shift is realized by adding the encoded side information. The resulting codes are coset-codes of the original code. By the commutativity of the addition this design guarantees the desired interconnection $E_{1,w_1}(w_2) = E_{2,w_2}(w_1)$ of the codes. At the decoder, instead of decoding the shifted code, we can as well invert the shift and decode the base codes. If needed, this also indicates how soft-information should be handled: Shift the soft-information in the same way as the code.

2.2.3.1 Example

As an example for the proposed coding scheme, we present a code construction starting from two linear block-codes of length N. In this case, the encoders E_1 and E_2 of the codes are given as two generator matrices $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ and the codewords are calculated e.g. for the first user by $\boldsymbol{c}_1 = \boldsymbol{w}_2\boldsymbol{G}_1$ from data words $\boldsymbol{w}_k \in \mathcal{X}^{R_k N}$, where the coding rates R_k are given in information symbols per code symbol.

Assume a binary Hamming code for the first user with the generator matrix given by

$$\boldsymbol{G}_1 = \begin{bmatrix} 1001110 \\ 0101011 \\ 0011101 \end{bmatrix}$$

which has a minimum code distance of 3 and a code for user 2 with the generator matrix given by

$$\boldsymbol{G}_2 = \begin{bmatrix} 1010101 \\ 0101011 \end{bmatrix}$$

which has a minimum distance of 4. Furthermore, note that the corresponding check matrices $\boldsymbol{H}_k$ satisfy $\boldsymbol{G}_k\boldsymbol{H}_k^T = 0$ so that they are the generator matrices of the dual codes.

In the case of linear block codes, the encoding at the relay is simply

$$\boldsymbol{x} = [\boldsymbol{w}_2\boldsymbol{w}_1]\begin{bmatrix} \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \end{bmatrix}$$

We do not require the rows of $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ to be linearly independent. In the example, the overall code has a minimum distance of 2. For instance, the codeword 1010101 corresponds to $w_2 = 000$ and $w_1 = 10$, while 1010011 is the codeword for $w_2 = 101$ and $w_1 = 00$. However, due to subtracting the codeword generated from the side information, the codes used for decoding at the terminal have a code distance of 3 and 4, respectively. This is because all the codeword pairs with distance 2, as those in the example above, belong to different cosets for both users. This in turn implies that they will only appear with different side information.

Error detection can be performed via syndrome calculation. If this is done before subtracting the encoded side information, this reveals the coset structure on which the scheme is based: The first terminal uses the side information to calculate the new syndrome for error-free transmission $\boldsymbol{s}_1 = \boldsymbol{w}_1\boldsymbol{G}_2\boldsymbol{H}_1^T$. Now use the matrix $\boldsymbol{H}_1$ to check the received codeword $\boldsymbol{y}_1$ with this syndrome $\boldsymbol{y}_1\boldsymbol{H}_1^T \overset{?}{=} \boldsymbol{s}_1$ before subtracting the encoded side information. If the equation is fulfilled, there is no error. If not, the difference between the left and right hand side may be used to locate the error. Error detection at terminal 2 works accordingly. As an alternative to subtracting the encoded side information, the decoding can shift the mapping defined by the generator matrix $\boldsymbol{G}_k$ according to the coset indicated by the side information. Note that now, not only the syndrome but all the information needs to be taken into account: The coset-code may have an all-zero-syndrome, as for example the coset of code $C_{1,\boldsymbol{w}_1}$ for user 1 with side information $\boldsymbol{w}_1 = 01$. Still, the mapping that is used in the encoding is different from the original mapping given by $\boldsymbol{G}_1$.

2.2.3.2 Concluding Remark

The proposed framework for designing coding schemes for the symmetric bidirectional broadcast channel is a strong and easy-to-apply tool for symmetric marginal channels. Our framework is optimal in the sense that for every given coding scheme, we can construct a new coding scheme which performs at least as good as the given one. Furthermore, the scheme has the advantage that we can use well-developed single user codes. As a consequence, we can conclude that there is no need to develop special codes for the symmetric bidirectional broadcast channel.

2.2.4 Discussion of Effects in General Channels

Now we leave the assumption of symmetric marginal channels behind and look at more general channels. It turns out that for these channels the performance of the coding scheme degrades. The next two subsections will point out why this happens and what needs to be done to circumvent this problem. Unfortunately, we are not able to give a coding scheme that can be build up from single user codes in a simple way for general channels without performance loss.

2.2.4.1 Non-symmetric Channels

In the previous section, the case of symmetric channels was considered. This does not include a channel in which the channel distortion word $\boldsymbol{z}_k$ depends on the channel input. A simple counter example presented below shows that in this case, the performance of the base codes may degrade if we use the proposed scheme.

Counter-example 2.4. *Assume a binary symmetric channel to user* 2 *with crossover probability such that a three bit repetition code achieves the desired probability of error, i.e. we have* $C_2 = \{000, 111\}$. *The non-symmetric channel to user* 1 *outputs* 3 *random bits, whenever the binary sum of the input of three bits equals* 1. *If the sum is* 0*, the channel is error-free. A* 3*-bit code for this channel may simply use all the four even weight sequences of length* 3 *and will achieve zero error probability. Therefore, a possible code is* $C_2 = \{000, 011, 110, 101\}$. *The joint network-channel code resulting from the proposed scheme is now* $C = \{0, 1\}^3$. *If we use this code to broadcast the information to both users, user* 1 *cannot achieve the desired performance. Half of the codewords of this code are inappropriate for the channel to user* 1. *Whenever* $\boldsymbol{c}_2 = 111$*, the output of the channel to user* 1 *is random. Even with side information, user* 1 *cannot decide which data word is correct. This is because the coset code* $C_{1,w_1} = \{111, 100, 001, 010\}$ *for this side information does not perform as desired for the given channel. The second user will still achieve the desired performance as its channel is symmetric.*

In the counter example above, the coding may be changed such that the performance of user 1 is increased at the cost of user 2. As the base code C_1 of user 1 is the only set of codewords of cardinality 4 that has the desired performance, in order to protect the performance of user 1, the new base code $\tilde{C}_2$ for user 2 needs to consist only of codewords of the base-code C_1. This is sufficient to obtain coset codes such that $C_{1,w_1} = C_1$ for any w_1 as the code C_1 is a linear code. Therefore, $\tilde{C}_2$ may consist of any two codewords of even weight and three bit length, as e.g. $\tilde{C}_2 = \{000, 011\}$. This degrades the performance of user 2 as the code distance now is at most 2. User 2 may as well use up to 4 codewords without further degradation of its performance.

A similar tradeoff effect can be noticed in the information theoretical result of Theorem 2.2: Whenever the marginal channels do not match concerning their optimizing input distribution, there is a tradeoff between the performance of the two users. In the proposed coding scheme the needed match is that of cosets of codes: The coset codes, which are created by the side information, need to match the channel such that the desired performance is achieved on average. By the symmetry assumption, we assure that all cosets have the same performance. In case of non-symmetric channels, an optimal code design might become far more difficult. A simple but potentially suboptimal solution is to combine the coding with a pseudo noise sequence known to the transmitter and receiver. By randomly scrambling the channel input, the channel seen by the code becomes symmetric. In doing so, one looses the possibility to use certain structure of the channel noise to enhance the decoding performance.

2.2.4.2 Additive Noise Channels with Non-Discrete Alphabets

In practice, wireless channels are not constrained to have finite discrete input and output alphabets. In fact, their inputs and outputs can have arbitrary values and channel distortion often can be modeled in terms of additive noise. In this section, we briefly consider such channels. Note that the discussion in this section should only point out where problems may arise if we use the above scheme on e.g. Gaussian channels. A detailed treatment or a general solution is out of the scope of this work.

First of all note that as far as the results of this section are concerned, we do not need the restriction to discrete finite alphabets. In fact, the following two requirements must hold:

1. There must be an Abelian group $(\mathcal{X}_R, +)$ defined on the input and the output alphabet. This is often fulfilled e.g. by channels that have input and output in $\mathbb{C}$ or $\mathbb{R}$.

2. The channel distortion is additive noise only, which is independent of the channel input and is added according to the definition of $(\mathcal{X}_R, +)$.

These requirements hold for AWGN channels as well as for many more general additive noise channels, where the noise is not necessarily Gaussian.

For a channel that fulfills these two conditions, all the results of this section hold, provided that the described coding is performed in signal-space, that is on the channel input/output alphabet according to $(\mathcal{X}_R, +)$. The problem is that in practice, we have some additional constraints such as power constraints on the channel input. The use of the coding scheme is in general suboptimal if e.g. power constraints are given. In fact, compared to the point-to-point transmission on the marginal channels for which the base codes are designed and that serve as performance-reference for the results in this section, the required transmission power is increased by the use of the coding scheme. Therefore the scheme is not optimal in view of the information theoretical result of Theorem 2.2 adapted to this case by restricting the probability distribution $p(x_R)$ to fulfill the power constraint. The information theoretical result [10] states that we should be able to achieve for both users the performance of some[5] transmission on the marginal channels using the same power as in the broadcast channel.

Now the question is whether there is a simple solution to the coding problem in additive noise channels? From a practical point of view, a good but in general suboptimal solution could be as follows: Define a new Abelian group $(\hat{\mathcal{X}}_R^T, \oplus)$, where $\hat{\mathcal{X}}_R^T$, $T \in \mathbb{N}$ is such that every $\hat{\boldsymbol{x}}_R{}^T \in \hat{\mathcal{X}}_R^T$ and every sequence of $\hat{\boldsymbol{x}}_R^T$ fulfills the given constraint. For example given a power constraint $E\{|x_R|^2\} \leq P$ and $\mathcal{X}_R \subset \mathbb{C}$, one may use a modulo like operation on real and imaginary part such that $\Re(x_R) \in [-\sqrt{\frac{P}{2}}, +\sqrt{\frac{P}{2}}]$ and $\Im(x_R) \in [-\sqrt{\frac{P}{2}}, +\sqrt{\frac{P}{2}}]$. It follows that every symbol $x_R \in \hat{\mathcal{X}}_R$ fulfills the constraint[6]. More advanced schemes may consider more than one symbol $T > 1$ to fulfill the constraints or use a more complex group, e.g. based on lattices. In addition, one needs two base codes with $\mathcal{C}_k \in \hat{\mathcal{X}}_R^{nT}$, $k \in \{1, 2\}$, for some $n \in \mathbb{N}$. Now add the two codewords $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ according to the Abelian group $(\hat{\mathcal{X}}_R^T, \oplus)$. The resulting symbol sequence fulfills the given constraints.

The next problem to cope with is that the noise is not added in accordance with $(\hat{\mathcal{X}}_R^T, \oplus)$. Furthermore, noise (e.g. Gaussian noise) is not restricted to $\hat{\mathcal{X}}_R$, and therefore decoding must be

[5] Note that in the general case, the performance of the set of codes used on the marginal channel may be suboptimal for that single user channel.

[6] There may be symbols $x_R \notin \hat{\mathcal{X}}_R$ that fulfill the constraint as well.

performed in the unconstrained receive alphabet $\mathcal{Y}_k$, $k \in \{1, 2\}$. As in the discussion of the last sections, the decoder now may decode the coset code induced by the side information. Nonetheless, inverting the shift from the base code to the coset code is in general not a simple subtraction of the encoded data, as the noise is not added according to $(\hat{\mathcal{X}}_R^T, \oplus)$. Therefore decoding may be more complex than before. In particular, the base decoder may be not applicable. Furthermore, not all coset codes may achieve the same performance as the base code. Therefore the results of this section do not apply for this class of channels, if a constraint to the input of the channel is given.

As a toy example, consider an AWGN channel with real input and real output and a constraint on the amplitude per symbol $x_R \in [-A, A)$. Following the above proposal, we use $C_1 = \{-A, +A - \epsilon\}, \epsilon > 0$ and $C_2 = \{-A, 0\}$ and add the encoded symbols according to $c = ((c_1 + c_2 + A) \mod 2A) - A$. The code of user 1 has a good performance for $c_2 = 0$, but for $c_2 = -A$ we get the coset code $C_{1,c_2=-A} = \{0, -\epsilon\}$, which has a bad performance if ϵ is small.

Obviously, this example is constructed to fail; for application in real systems however, the proposed coding in signal space may often be a good option, especially if one uses good lattices for the coding that are able to control or prevent the problems pointed out in this subsection.

Nested lattice codes are the counterpart of linear codes in signal space [54, 55, 56] and can be used for the coding in the BC phase of the two way relay channel. These codes can prevent the problems pointed out by the above discussion. It is known that nested lattice codes together with a lattice decoder can achieve capacity in the AWGN channel. For a more elaborate discussion on these codes the reader is referred to the literature, e.g.[55] and references therein. In fact we can use two such lattice codes for AWGN channels for the two Gaussian channels to both the receivers. The need of a common input alphabet transfers to the constraint, that now both codes use the same fundamental region Ω of the corse lattice Λ, that is used for shaping.

For a single user nested lattice code the codewords are formed by a fine lattice Λ_1. For the transmission the codeword $\boldsymbol{c}_1 \in \Lambda_1$ and a dither $\boldsymbol{d}$ are added modulo the fundamental region Ω the corse lattice Λ. The dither is drawn at random according to a uniform distribution over a fundamental region Ω of Λ. The transmitted signal is therefore $\boldsymbol{x}_1 = \boldsymbol{c}_1 + \boldsymbol{d} \mod {}_\Omega\Lambda$. At the receiver, the received signal is scaled and the known dither is subtracted. All operations are performed modulo this fundamental region Ω. To perform a lattice decoding the received signal is mapped to the fundamental region Ω of the corse lattice Λ. By the scaling the probability of error is minimized. It can be shown that there exists codes and lattices to achieve the capacity in the AWGN channel [54, 55, 56].

We can use two such codes as base codes in the BC phase of a two-way relay channel with decode-and-forward. For the encoding, the relay transmits $\boldsymbol{x}_r = \boldsymbol{c}_1 + \boldsymbol{c}_2 + \boldsymbol{d} \mod {}_\Omega\Lambda$, where $\boldsymbol{c}_k$ is a codeword for the code of receiver k. The dither d is again drawn at random according to a uniform distribution over a fundamental region Ω of Λ. The receiver k can now decode by treating the other codeword as part of an effective dither $\tilde{\boldsymbol{d}}_k$, i.e. receiver 2 uses an effective dither $\tilde{\boldsymbol{d}}_2 = \boldsymbol{d} + \boldsymbol{c}_1 \mod {}_\Omega\Lambda$ while receiver 1 uses $\tilde{\boldsymbol{d}}_1 = \boldsymbol{d} + \boldsymbol{c}_2 \mod {}_\Omega\Lambda$. The performance for both receivers will not degrade compared to the single user channel, as $\tilde{\boldsymbol{d}}_k$ has a uniform distribution over the fundamental region Ω of Λ and is independent of $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ [55, Lemma 1]. Therefore it follows, that as in the case of symmetric marginal channels the performance of both receivers is the same as the performance of the single user codes upon which the new code is based. Therefore for these codes the result of Theorem 2.5 applies.

In the single user code with a lattice decoder the dither is used to have a noise term in the transformed modulo lattice additive noise channel which is independent of the channel input. In the BC code for the two-way relay channel the dither ensures that the performance of the different cosets of the lattice code have the same performance. While in the single user AWGN channel with ML decoding the dither may be neglected, it is mandatory in our setup. The reason is, that we perform a modulo addition at the relay to merge the information for both receivers into the transmitted signal. By using the dither in effect we make the performance of both receivers independent of the message transmitted to the other respective receiver.

It is unlikely that the other results of this section hold for lattice codes and general additive noise channels. One reason is, that the proposed lattice codes do not perform optimal for the single user channel. Even though the codes can achieve capacity if the coding length goes to infinity, there are codes for the AWGN channel, that have a lower probability of error. Furthermore a code for the single user channel has no restriction to use a certain fundamental region or even a lattice. The structural properties make the nested lattice codes a nice choice for a practical coding scheme for the BC phase of a two-way channel. Simultaneously these structural properties impose some restrictions on the choice of the codes we can base upon. Therefore the resulting coding scheme might be suboptimal. For the optimal performance, joint tranmission schemes need to be developed.

2.3 Achievable Rates for a Decode-and-Forward System with More Than Two Phases

In the above sections we discussed the two-way relay channel with a decode-and-forward constraint and the assumption that we have two phases to transmit the message to the receiver. In this section we extend the model by allowing more than two phases. We focus on systems where we can use side information in a BC phase from the relay to the terminal nodes. Therefore the additional freedom is used only to transmit the messages to the relay and to use the direct path between the terminals.

The treatment in this section serves the purpose to show how such additional degrees of freedom can be used and how the coding changes. We provide only a sketch of the proofs, serving the purpose to see what needs to be changed and where gains can be achieved. The treatment is incomplete as we restrict ourselves to setups where the messages are transmitted to the terminals via a BC. Protocols where the relay transmits a signal via a MAC to one of the terminals are not considered. Furthermore, restrictions similar to the ones used in the above sections are used. Namely we assume the following:

- We assume that the transmissions of the terminal nodes in different phases do not depend on any received signal, i.e. we do not allow the nodes to use any feedback mechanisms in any of the phase. If any such information is gathered e.g. during the broadcast of the relay or the transmission of the other node, it is not used in any of the following phases. The relay node and the terminal nodes are restricted to half-duplex, i.e. the nodes may use only information received in those phases in which they are not transmitting.
- The relay node is assumed to be able to decode the messages of both the nodes. This restriction is in general suboptimal. The restriction does not mean, that the information flow is necessarily via the relay as it is in the two-phase case. Even the case that there is no channel from the relay to the nodes is handled by the theorems of this section.

We consider two scenarios: A three-phase setup and a four-phase setup. In the three-phase setup we make use of the direct link between the terminal nodes. This link can be used either to transmit the message directly to the other terminal or to provide additional side information for the final decoding at the receiver. The three phases are used as three BC transmissions. First, node 1 transmits to the relay and node 2, than node 2 transmits to the relay and node 1, and finally the relay broadcasts to both terminal nodes.

The four-phase setup consists of two broadcasts and a MAC phase to enable the decoding at the relay: First node 1 transmits to the relay and node 2, then node 2 transmits to the relay and node 1, thereafter both nodes transmit to the relay using a MAC. The fourth phase is again the broadcast from the relay to both the terminal nodes. Therefore, the four-phase protocol can be seen as a combination of the three-phase and the two-phase protocols.

2.3.1 An Achievable Rate Region for a Three-Phase Relay Channel

The goal is to transmit a message w_1 from node 1 to node 2 and w_2 from node 2 to node 1 using the medium between the two nodes and the relay a total of $n \in \mathbb{N}$ times. We start the discussion by adapting the definitions to this setup.

We assume three phases where $1 \geq \alpha, \beta, \gamma \geq 0$, $\alpha + \beta + \gamma = 1$ indicate timesharing between the phases: In the first phase, node 1 transmits the codeword $X_1^{n_1}$ of length n_1 to the relay and node 2 using a channel $p_1(y_{R,1}, y_{2,1}|x_1)$ $n_1 := n_1(n) \in \mathbb{N}$ times with $\frac{n_1}{n} \to \alpha$ as $n \to \infty$. These nodes will receive the signals $Y_{R,1}^{n_1}$ and $Y_{2,1}^{n_1}$ respectively. In the second phase, node 2 transmits the codeword $X_2^{n_2}$ of length n_2 to the relay and node 1 using a channel $p_1(y_{R,2}, y_{1,2}|x_2)$ $n_2 := n_2(n) \in \mathbb{N}$ times with $\frac{n_2}{n} \to \beta$ as $n \to \infty$. The received signals are $Y_{R,2}^{n_2}$ and $Y_{1,2}^{n_2}$ respectively. In the third phase the relay node transmits $X_R^{n_3}$ to node 1 and node 2 using a channel $p_2(y_{1,3}, y_{2,3}|x_R)$ $n_3 := n_3(n) \in \mathbb{N}$ times with $\frac{n_3}{n} \to \gamma$ as $n \to \infty$. The terminal nodes receive the signals $Y_{1,3}^{n_3}$ and $Y_{2,3}^{n_3}$ respectively. All channels are assumed to be memoryless and the channels in the three phases are assumed to be independent. Therefore we have a joint probability distribution $p(y_{R,1}, y_{2,1}, y_{R,2}, y_{1,2}, y_{1,3}, y_{2,3}|x_1, x_2, x_R) = p_1(y_{R,1}, y_{2,1}|x_1)p_2(y_{R,2}, y_{1,2}|x_2)p_R(y_{1,3}, y_{1,3}|x_R)$ which defines the considered relay channel as follows:

Definition 2.6. A discrete memoryless three-phase two-way relay channel is defined by a family

$$\left\{p^{(n)} : \mathcal{X}_1^{n_1} \times \mathcal{X}_2^{n_2} \times \mathcal{X}_R^{n_3} \to \mathcal{Y}_{R,1}^{n_1} \times \mathcal{Y}_{2,1}^{n_1} \times \mathcal{Y}_{R,2}^{n_2} \times \mathcal{Y}_{1,2}^{n_2} \times \mathcal{Y}_{1,3}^{n_3} \times \mathcal{Y}_{2,3}^{n_3}\right\}_{n_1\in\mathbb{N},n_2\in\mathbb{N},n_3\in\mathbb{N}}$$

with $n_1 + n_2 + n_3 = n$. The family consists of probability transition functions given by

$$p^{(n)}(y_{R,1}^{n_1}, y_{2,1}^{n_1}, y_{R,2}^{n_2}, y_{1,2}^{n_2}, y_{1,3}^{n_3}, y_{2,3}^{n_3}|x_1^{n_1}, x_2^{n_2}, x_R^{n_3}) := \prod_{i=1}^{n_1} p_1(y_{R,1,(i)}, y_{2,1,(i)}|x_{1,(i)}) \prod_{i=1}^{n_2} p_2(y_{R,2,(i)}, y_{1,2,(i)}|x_{2,(i)}) \prod_{i=1}^{n_3} p_R(y_{1,3,(i)}, y_{2,3,(i)}|x_{R,(i)})$$

for probability functions $p_1 : \mathcal{X}_1 \to \mathcal{Y}_{R,1}\times\mathcal{Y}_{2,1}$, $p_2 : \mathcal{X}_2 \to \mathcal{Y}_{R,2}\times\mathcal{Y}_{1,2}$ and $p_R : \mathcal{X}_R \to \mathcal{Y}_{1,3}\times\mathcal{Y}_{2,3}$.

Definition 2.7. A $(M_1^{(n)}, M_2^{(n)}, n_1, n_2, n_3)$-code for the three-phase two-way relay channel under a decode-and-forward protocol consists of an encoder at node one

$$x_1^{n_1} : \mathcal{W}_1 \to \mathcal{X}_1^{n_1}$$

with $\mathcal{W}_1 = [1, 2, \dots, M_1^{(n)}]$, an encoder at node two

$$x_2^{n_2} : \mathcal{W}_2 \to \mathcal{X}_2^{n_2}$$

with $\mathcal{W}_2 = [1, 2, \dots, M_2^{(n)}]$, an encoder at the relay node

$$x_R^{n_3} : \mathcal{W}_1 \times \mathcal{W}_2 \to \mathcal{X}_R^{n_3},$$

a decoder at node one and node two

$$g_1 : \mathcal{Y}_{1,2}^{n_2} \times \mathcal{Y}_{1,3}^{n_3} \times \mathcal{W}_1 \to \mathcal{W}_2$$

$$g_2 : \mathcal{Y}_{2,1}^{n_1} \times \mathcal{Y}_{2,3}^{n_3} \times \mathcal{W}_2 \to \mathcal{W}_1$$

and a decoder at the relay node

$$g_R : \mathcal{Y}_{R,1}^{n_1} \times \mathcal{Y}_{R,2}^{n_2} \to \mathcal{W}_1 \times \mathcal{W}_2.$$

Definition 2.8. When $w := w(w_1, w_2) = [w_1, w_2] \in \mathcal{W} := \mathcal{W}_1 \times \mathcal{W}_2$ is the message pair transmitted by the two terminal nodes, the message w_2 is decoded in error if $g_1(y_{1,2}^{n_2}, y_{1,3}^{n_3}, w_1) \neq w_2$ or if $g_R(y_{R,1}^{n_1}, y_{R,2}^{n_2}) \neq (\tilde{w}_1, w_2)$ for some $\tilde{w}_1 \in \mathcal{W}_1$. The probability of this error event is denoted by

$$\lambda_1(w) := Pr\left[g_1(Y_{1,2}^{n_2}, Y_{1,3}^{n_3}, w_1) \neq w_2 \vee g_R(y_{R,1}^{n_1}, y_{R,2}^{n_2}) \neq (\tilde{w}_1, w_2)|w(w_1, w_2) \text{ has been sent}\right].$$

Accordingly the corresponding error event for the message w_1 is denoted by

$$\lambda_2(w) := Pr\left[g_2(Y_{2,1}^{n_1}, Y_{2,3}^{n_3}, w_2) \neq w_1 \vee g_R(y_{R,1}^{n_1}, y_{R,2}^{n_2}) \neq (w_1, \tilde{w}_2)|w(w_1, w_2) \text{ has been sent}\right].$$

Note that the definition for the error is with respect to the messages rather than with respect to the decoder. The reason for this is that we need to capture the constraint of decoding at the relay in the definition of achievable rates.

Definition 2.9. The average probability of decoding error is given by

$$\mu_1^{(n)} := \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \lambda_1(w)$$

for message w_2 and

$$\mu_2^{(n)} := \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \lambda_2(w)$$

for message w_1.

Definition 2.10. Let $\mu_1^{(n)}$ and $\mu_2^{(n)}$ be the average probabilities of decoding error for message w_2 and w_1, respectively. The rate pair $[R_1, R_2]$ is said to be achievable for the three-phase two-way relay channel under a decode-and-forward protocol if there exists a sequence of $(M_1^{(n)}, M_2^{(n)}, n_1, n_2, n_3)$-codes with $\frac{\log M_1^{(n)}}{n} \to R_1$ and $\frac{\log M_2^{(n)}}{n} \to R_2$ such that $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$.

Theorem 2.7. An achievable rate region for the three-phase two-way relay channel using a decode-and-forward protocol is the set of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &< \min\{\alpha I(X_1; Y_{R,1}); \alpha I(X_1; Y_{2,1}) + \gamma I(X_R; Y_{2,3})\} \\ R_2 &< \min\{\beta I(X_2; Y_{R,2}); \beta I(X_2; Y_{1,2}) + \gamma I(X_R; Y_{1,3})\} \end{aligned} \tag{2.12}$$

for some joint probability distribution $p_1(y_{R,1}, y_{2,1}|x_1)p_2(y_{R,2}, y_{1,2}|x_2)p_R(y_{1,3}, y_{2,3}|x_R)p(x_1, x_2, x_R)$ and some $\alpha, \beta, \gamma \geq 0$ with $\alpha + \beta + \gamma = 1$.

Remark 2.9. Due to the factorization of the channel, there is no rate loss if we restrict the probability distribution of the input $p(x_1, x_2, x_R)$ such that X_1, X_2 and X_R are independent.

Remark 2.10 (Necessity of the feedback constraint)**.** The restriction of the nodes not to use any feedback mechanism may not seem necessary for the considered setup. To see that it is indeed necessary, consider a toy setup: Suppose that there is no channel from node 1 to the relay, while all other channels are error and interference free and offer some rate to transmit. For this setup $I(X_1; Y_{R,1}) = 0$ and therefore R_1 is zero as well. Without the restriction it would of course be possible to transmit some data from node 1 to node 2 in the first phase. In the second phase node 2 could forward the data received from node 1 to the relay. Therefore the decode-and-forward requirement would be satisfied. If we do not allow such cooperation between the nodes, than using a cutset bound [30] slightly adapted to the considered setup the given region can be shown to be optimal. The adaption is needed, as the proof in [30] assumes, that the messages in the network are independent, while here we transmit the same message to the relay and to the terminal node. Furthermore the encoders in [30] may use all the received signals for the encoding.

Remark 2.11 (Convexity of the rate region)**.** Looking at the rate region it is not immediately clear whether the region is convex or not. For fixed timesharing parameters the region is obviously convex. It remains to show that there exist parameters and probability distributions such that if $[R_1^{(1)}; R_2^{(1)}]$ and $[R_1^{(2)}; R_2^{(2)}]$ are achievable with possibly different timesharing parameters, then also a convex combination of both is within the achievable rate region. Indeed the proof is not that difficult, so we will only sketch it for the given region once. It turns out, that the the weighted addition can be encapsulated in some auxiliary variable Q. Note that all terms in the above rate region can be conditioned on Q, if we change $p(x_1, x_2, x_R)$ to $p(x_1, x_2, x_R|q)p(q)$. This will not change the region. Furthermore using the observation in Remark 2.9, we can add three variables Q_1,Q_2, Q_3 to the expressions and change $p(x_1, x_2, x_R)$ to $p(x_1|q_1)p(x_2|q_2)p(x_3|q_3)p(q_1)p(q_2)p(q_3)$. The minimum operation can now be split up into two inequalities. We receive for R_1^* being a convex combination of $R_1^{(1)}$ and $R_1^{(1)}$

$$R_1^* = aR_1^{(1)} + (1 - a)R_1^{(1)}$$

the resulting inequalities

$$R_1^* \leq a\alpha^{(1)} I(X_1^{(1)}; Y_{R,1}^{(1)}|Q_1^{(1)}) + (1 - a)\alpha^{(2)} I(X_1^{(2)}; Y_{R,1}^{(2)}|Q_1^{(2)})$$

and

$$\begin{aligned} R_1^* \leq a\alpha^{(1)} I(X_1^{(1)}; Y_{2,1}^{(1)}|Q_1^{(1)}) + a\gamma^{(1)} I(X_R^{(1)}; Y_{2,3}^{(1)}|Q_3^{(1)}) + \\ (1 - a)\alpha^{(2)} I(X_1^{(2)}; Y_{2,1}^{(2)}|Q_1^{(2)}) + (1 - a)\gamma^{(2)} I(X_R^{(2)}; Y_{2,3}^{(2)}|Q_3^{(2)}). \end{aligned}$$

Now

$$a\alpha^{(1)} I(\cdot|Q_1^{(1)}) + (1 - a)\alpha^{(2)} I(\cdot|Q_1^{(2)}) = (a\alpha^{(1)} + (1 - a)\alpha^{(2)}) I(\cdot|\tilde{Q}_1),$$

where $\tilde{Q}_1$ with $|\tilde{Q}_1| = |Q_1^{(1)}| + |Q_1^{(1)}|$ is used to include the weighted sum in the expectation over a appropriately constructed probability distribution. Note that we do not change the channel if we switch to $\tilde{Q}_1$, but only the input distribution to the channel. The above steps can be performed for the other inequalities and for R_2 where the same variables $\tilde{Q}_1$, $\tilde{Q}_2$, $\tilde{Q}_3$ can be used. We used three auxiliary variables, as the needed random variable might be different for the different phases. We see that the resulting vector $[R_1, R_2]$ is in the region for the timesharing parameters $\alpha^* = (a\alpha^{(1)} + (1-a)\alpha^{(2)})$, $\beta^* = (a\beta^{(1)} + (1-a)\beta^{(2)})$, and $\gamma^* = (a\gamma^{(1)} + (1-a)\gamma^{(2)})$. Similar arguments can be used for the proof of the convexity of other regions in this thesis.

Proof. For $\gamma = 0$ the result follows immediately by interpreting marginal channels of the two broadcast channels as a compound channel [60, 29] used n_1 and n_2 times respectively. Therefore in what follows we assume $\gamma > 0$. Let $R_1, R_2, p(x_1, x_2, x_R), \alpha, \beta, \gamma \geq 0$ with $\alpha+\beta+\gamma = 1$ be given such that the inequalities in (2.12) are strict. The achievability of the closure is a consequence of the definition of achievability and will not be repeated here. It can be proved analogous to the arguments in the proof of Theorem 2.2 in Section 2.1.3.

Random Codebook Generation Let $n_1 = \lfloor \alpha n \rfloor$, $n_2 = \lfloor \beta n \rfloor$ and $n_3 = n - n_1 - n_2 \leq \lceil \gamma n \rceil + 1$. We generate $M_1^{(n)} = 2^{\lfloor nR_1 \rfloor}$ independent codewords $X_1^{n_1}(w_1)$ of length n_1 drawn according to $\prod_{i=1}^{n_1} p(x_{1,(i)})$ where $p(x_1) = \sum_{x_2 \in \mathcal{X}_2} \sum_{x_R \in \mathcal{X}_R} p(x_1, x_2, x_R)$ is the marginal probability distribution. Similarly, we generate $M_2^{(n)} = 2^{\lfloor nR_2 \rfloor}$ independent codewords $X_2^{n_2}(w_2)$ of length n_2 drawn according to $\prod_{i=1}^{n_2} p(x_{2,(i)})$ and $M_1^{(n)} M_2^{(n)}$ independent codewords $X_R^{n_3}(w)$, $w = [w_1, w_2]$ of length n_3 drawn according to $\prod_{i=1}^{n_3} p(x_{R,(i)})$. The random code is revealed to both terminal nodes and the relay.

Encoding Depending on the message to transmit w_1 node 1 sends the corresponding codeword $x^{n_1}(w_1)$ using the channel n_1 times. Node 2 sends $x^{n_2}(w_2)$ to transmit the message w_2. To send the decoded pair $w = [w_1, w_2]$ with $w_k \in \mathcal{W}_k$, $k \in \{1, 2\}$, the relay sends the corresponding codeword $x_R^{n_3}(w)$.

Decoding The receiving nodes will use typical set decoding. For a strict definition of the decoding sets we choose parameter for the typical sets as $\epsilon_1 < \frac{\alpha I(X_1, Y_{R,1}) - R_1}{3\alpha}$, $\epsilon_2 < \frac{\beta I(X_2, Y_{R,2}) - R_2}{3\beta}$, $\epsilon_3 < \frac{\gamma I(X_R, Y_{2,3}) + \alpha I(X_1, Y_{2,1}) - R_1}{6\alpha}$, and $\epsilon_4 < \frac{\gamma I(X_R, Y_{2,3}) + \alpha I(X_1, Y_{2,1}) - R_1}{6\gamma}$. For the first two phases the relay decides that w_1 and w_2 are transmitted, if $x_1^{n_1}(w_1)$ and $x_2^{n_2}(w_2)$ are the only codewords jointly typical with the received signals $y_{R,1}^{n_1}$ and $y_{R,2}^{n_2}$ respectively, i.e. $\left(x_1^{n_1}(w_1), y_{R,1}^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, Y_{R,1})$ and $\left(x_2^{n_2}(w_2), y_{R,2}^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(X_2, Y_{R,2})$. Knowing w_2, the decoder at node 2 decides that w_1 was transmitted, if this is the unique w_1 such that $\left(x_1^{n_1}(w_1), y_{2,1}^{n_1}\right) \in \mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, Y_{2,1})$ and simultaneously $\left(x_R^{n_3}(w_1, w_2), y_{2,3}^{n_3}\right) \in \mathcal{T}_{\epsilon_4}^{(n_3)}(X_R, Y_{2,3})$. Decoding at receiver 1 works in an analogous way. To keep the definition of the decoder consistent the decoders map to the default message $w_1 = 1$ and $w_2 = 1$ if no or more than one codewords is found in their respective decoding sets.

Analysis of the Probability of Error Now we bound the probability of error for the decoding. We give the proof for the transmission of message w_1 to receiver 2. The proof for the message w_2 follows from analogous arguments. We have

$$\mu_2^{(n)} = P_{e,2}^{(1)} \tilde{P}_{e,2}^{(2)} + P_{e,2}^{(1)}(1 - \tilde{P}_{e,2}^{(2)}) + (1 - P_{e,2}^{(1)}) \tilde{P}_{e,2}^{(2)}$$

where $\tilde{P}_{e,2}^{(2)}$ is the average probability of the event, that the decoding at the terminal node fails and $P_{e,2}^{(1)}$ is the average probability for a decoding error in the decoding of w_1 at the relay. Therefore we can bound $\mu_2^{(n)}$ form above as

$$\mu_2^{(n)} \leq P_{e,2}^{(1)} + P_{e,2}^{(2)}$$

where $P_{e,2}^{(2)}$ is the average probability for a decoding error at node 2 given that the relay decoded without error. We can split the analysis of the overall error probability into the analysis of the error in decoding at the relay and of the error in the decoding at the terminal node. As we use the random coding argument, in the analysis, we will average over the random codebook and consider $\mathbb{E}_{x_1^{n_1},x_2^{n_2},x_R^{n_3}}\left\{\mu_k^{(n)}\right\} \leq \mathbb{E}_{x_1^{n_1},x_2^{n_2},x_R^{n_3}}\left\{P_{e,k}^{(1)} + P_{e,k}^{(2)}\right\}$. We show that we have $\mathbb{E}_{x_1^{n_1},x_2^{n_2},x_R^{n_3}}\left\{\mu_1^{(n)}\right\} \to 0$. We will then conclude, that there exists at least one codebook with small average probability of error for the decoding of w_1 and w_2.

Decoding at the relay First consider the decoding at the relay. The relay is in error if either $\left(x_1^{n_1}(w_1), y_{R,1}^{n_1}\right) \notin \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, Y_{R,1})$ or if there exists a $\hat{w}_1 \neq w_1$ with $\left(x_1^{n_1}(\hat{w}_1), y_{R,1}^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, Y_{R,1})$. Using the union bound it is sufficient to show that both these error events occur with arbitrarily small probability as $n \to \infty$.

By the law of large numbers the probability that $\left(x_1^{n_1}(w_1), y_{R,1}^{n_1}\right) \notin \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, Y_{R,1})$ for sequences $\left(x_1^{n_1}(w_1), y_{R,1}^{n_1}\right)$ drawn according to a joint probability distribution can be made arbitrarily small by choosing n (and as a consequence n_1) big.

The probability of $\left(x_1^{n_1}(\hat{w}_1), y_{R,1}^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, Y_{R,1})$ for $\hat{w}_1 \neq w_1$ averaged over all codewords and the random codebook can be bounded as follows:

$$\begin{aligned}
&\sum_{y_{R,1}^{n_1}\in\mathcal{Y}_{R,1}^{n_1}} \mathbb{E}_{x_1^{n_1}}\left\{p\left(y_{R,1}^{n_1}|x_1^{n_1}(w_1)\right)\sum_{\substack{\hat{w}_1=1\\ \hat{w}_1\neq w_1}}^{|\mathcal{W}_1|}\chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(X_1,Y_{R,1})}(x_1^{n_1}(\hat{w}_1), y_{R,1}^{n_1})\right\}\\
&\qquad = (|\mathcal{W}_1|-1)\sum_{y_{R,1}^{n_1}\in\mathcal{Y}_{R,1}^{n_1}}\sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} p(x_1^{n_1})p(y_{R,1}^{n_1})\chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(X_1,Y_{R,1})}(x_1^{n_1}, y_{R,1}^{n_1})\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \leq 2^{n(R_1+3\alpha\epsilon_1-\alpha I(X_1,Y_{R,1}))+I(X_1,Y_{R,1})}.
\end{aligned}$$

The last step follows from the properties of the typical set analogous to the procedure in the proof of Theorem 2.2 and the choice of $\alpha n - 1 \leq n_1 \leq \alpha n$. Now for $n \to \infty$

$$2^{n(R_1+3\alpha\epsilon_1-\alpha I(X_1,Y_{R,1}))+I(X_1,Y_{R,1})} \to 0$$

as we choose $\epsilon_1 < \frac{\alpha I(X_1,Y_{R,1})-R_1}{3\alpha}$. We conclude that $\mathbb{E}_{x_R^n}\left\{P_{e,k}^{(1)}\right\}$ can be made arbitrarily small for n large.

Decoding at the terminal node For the calculation of the probability $\mathbb{E}\{P_{e,2}^{(2)}\}$ we assume that the relay has decoded w_1 and w_2 without error. Furthermore node 2 received some $y_{2,1}^{n_1}$ drawn according to $p\left(y_{2,1}^{n_1}|x_1^{n_1}(w_1)\right)$.

The error that may occur at node 2 is characterized by several possible events:

- E_1: $x_1^{n_1}(w_1)$ is not jointly typical with $y_{2,1}^{n_1}$,
- E_2: $x_R^{n_3}(w_1, w_2)$ is not jointly typical with $y_{2,3}^{n_3}$,
- E_3: $x_2^{n_1}(\tilde{w}_1)$ is jointly typical with $y_{2,1}^{n_1}$ for some $\tilde{w}_1 \neq w_1$, or
- E_4: $x_R^{n_3}(\tilde{w}_1, w_2)$ is jointly typical with $y_{2,3}^{n_3}$ for some $\tilde{w}_1 \neq w_1$.

For these events we can calculate the probability P_{E_i} where we take into account the random codebook generation as well as the joint randomness in the system. We can bound the average probability of error for the decoding from above by

$$\mathbb{E}\{P_{e,2}^{(2)}\} \leq P_{E_1} + P_{E_2} + M_1 P_{E3} P_{E4},$$

where the average is over the random codebook as well as over the transmitted symbols. The last term follows form the observation that an error occurs if E_3 and E_4 happen simultaneously. The factor M_1 attributes the fact, that this may happen for each wrong message $\tilde{w}_1 \neq w_1$.

Clearly $P_{E_1} \to 0$ if $n_1 \to \infty$ and $P_{E_2} \to 0$ if $n_3 \to \infty$ which follows from the definition of strong typicality and the law of large numbers. Analogous to the proceeding above we have

$$P_{E_3} \leq 2^{n(3\alpha\epsilon_3 - \alpha I(X_1,Y_{2,1})) + I(X_1,Y_{2,1})}$$

and

$$P_{E_4} \leq 2^{n(3\gamma\epsilon_4 - \gamma I(X_R,Y_{2,3})) + 2\epsilon_4}$$

and therefore

$$M_2 P_{E3} P_{E4} \leq 2^{n(R_1 - \alpha I(X_1,Y_{2,1}) + 3\alpha\epsilon_3 - \gamma I(X_R,Y_{2,3}) - 3\gamma\epsilon_4) + I(X_1,Y_{2,1}) + 2\epsilon_4}.$$

We can now conclude that $\mathbb{E}\left\{P_{e,2}^{(2)}\right\} \to 0$ as $n \to \infty$ by the choice of $\epsilon_3 < \frac{\gamma I(X_R,Y_{2,3}) + \alpha I(X_1,Y_{2,1}) - R_1}{6\alpha}$ and $\epsilon_4 < \frac{\gamma I(X_R,Y_{2,3}) + \alpha I(X_1,Y_{2,1}) - R_1}{6\gamma}$.

This proves that $\mathbb{E}_{x_1^{n_1}, x_2^{n_2}, x_R^{n_3}}\left\{\mu_1^{(n)}\right\} \to 0$ for $n \to \infty$. Similarly $\mathbb{E}_{x_1^{n_1}, x_2^{n_2}, x_R^{n_3}}\left\{\mu_2^{(n)}\right\} \to 0$. Therefore $\mathbb{E}_{x_1^{n_1}, x_2^{n_2}, x_R^{n_3}}\left\{\mu_1^{(n)} + \mu_2^{(n)}\right\} \to 0$ and we can conclude that there is at least one codebook such that $\mu_1^{(n)}$ and $\mu_2^{(n)}$ can be made arbitrarily small by choosing n large enough. The closure of the region is achievable using similar arguments as in the proof of Theorem 2.2. Therefore the claim is proved. □

2.3.2 An Achievable Rate Region for a Four-Phase Relay Channel

To maintain brevity without any loss of information we assume that all the definitions given in Section 2.3.1 are extended in the obvious way to the four phase protocol. In this setup we assume four phases where $1 \geq \alpha, \beta, \gamma, \delta \geq 0$, $\alpha + \beta + \gamma + \delta = 1$ indicate timesharing between the phases. The first two phases are broadcasts, each from one terminal node to the relay and the other respective terminal node. The third phase is a MAC phase, and the forth phase is a BC phase.

Theorem 2.8. An achievable rate region for the four-phase two-way relay channel using a decode-and-forward protocol is the set of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &< \min\{\alpha I(X_{1,1}; Y_{R,1}) + \gamma I(X_{1,3}; Y_{R,3}|X_{2,3}, Q); \alpha I(X_{1,1}; Y_{2,1}) + \delta I(X_R; Y_{2,4})\} \\ R_2 &< \min\{\beta I(X_{2,2}; Y_{R,2}) + \gamma I(X_{2,3}; Y_{R,3}|X_{1,3}, Q); \beta I(X_{2,2}; Y_{1,2}) + \delta I(X_R; Y_{1,4})\} \\ R_1 + R_2 &< \alpha I(X_{1,1}; Y_{R,1}) + \beta I(X_{2,2}; Y_{R,2}) + \gamma I(X_{1,3}, X_{2,3}; Y_{R,3}|Q) \end{aligned} \tag{2.13}$$

for some joint probability distribution $p(x_{1,1})p(x_{2,2})p(x_R)p(q)p(x_{1,3}|q)p(x_{2,3}|q)p_1(y_{R,1}, y_{2,1}|x_{1,1})$ $p_2(y_{R,2}, y_{1,2}|x_{2,2})p_R(y_{1,4}, y_{2,4}|x_R)p_M(y_{R,3}|x_{1,3}, x_{2,3})$ and some $\alpha, \beta, \gamma, \delta \geq 0$ with $\alpha + \beta + \gamma + \delta = 1$.

Remark 2.12 (Two-phase and three-phase protocols are special cases)**.** The region includes the region of Theorem 2.4 and the region of Theorem 2.7 as special cases. In fact, ignoring the minimum operation and stating the region in a five dimensional space, the region can be seen to be the convex hull of the regions of Theorem 2.4 and Theorem 2.7 written in this way. Furthermore, due to the minimum operation we conclude that this region is a super set to the regions in Theorem 2.4, Theorem 2.7, and the convex combination of both.

Sketch of proof. As the proof does not use any new arguments and does not give new insight to the problem solution, we only provide a sketch of the proof. We use a random coding argument as in the proofs above. We start with strict inequalities. In each of the first three phases both nodes transmit the whole message by transmitting some $x_{1,1}^{n_1}(w_1)$, $x_{2,2}^{n_2}(w_2)$, $x_{1,3}^{n_3}(w_1)$ and $x_{2,3}^{n_3}(w_2)$. The codewords are drawn according to $p(x_{1,3}^{n_3}|q^{n_3})$ and $p(x_{2,3}^{n_3}|q^{n_3})$ for a fixed q^{n_3}, which is part of the codebook and which is drawn according to $p(q^{n_3})$.

As in the proof of Theorem 2.7 we can split the analysis of the probability of error in two parts:

- $P_k^{(1)}$ is the probability of error for the decoding at the relay.
- $P_k^{(2)}$ is the probability of error for decoding at the receivers given that the relay decoded correctly.

Note, that the analysis of $P_k^{(2)}$ is essentially the same as in the proof of Theorem 2.7. It is therefore left to show that the conditions given in the theorem are sufficient to decode both messages at the relay.

For the bounding of $P_k^{(2)}$ we have to specify the decoding at the relay. The relay node uses typical set decoding, i.e. the relay decides that $w = [w_1, w_2]$ was transmitted if this is the unique w such that $\left(x_{1,1}^{n_1}(w_1), y_{R,1}^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X_{1,1}, Y_{R,1})$, $\left(x_{2,2}^{n_2}(w_2), y_{R,2}^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(X_2, Y_{R,2})$, and

$\left(x_{1,3}^{n_3}(w_1), x_{2,3}^{n_3}(w_2), y_{R,3}^{n_3}\right) \in \mathcal{T}_{\epsilon_3}^{(n_3)}(X_{1,3}, X_{2,3}, Y_{R,2})$. The parameters for the decoding sets are chosen as

$$\begin{aligned}
\epsilon_1 &< \min\left\{ \frac{\alpha I(X_{1,1};Y_{R,1})+\gamma I(X_{1,3};Y_{R,3}|X_{2,3},Q)-R_1}{6\alpha}; \frac{\alpha I(X_{1,1};Y_{R,1})+\beta I(X_{2,2};Y_{R,2})+\gamma I(X_{1,3},X_{2,3};Y_{R,3}|Q)-R_1-R_2}{9\alpha} \right\}, \\
\epsilon_2 &< \min\left\{ \frac{\beta I(X_{2,2};Y_{R,2})+\gamma I(X_{2,3};Y_{R,3}|X_{1,3},Q)-R_2}{6\beta}; \frac{\alpha I(X_{1,1};Y_{R,1})+\beta I(X_{2,2};Y_{R,2})+\gamma I(X_{1,3},X_{2,3};Y_{R,3}|Q)-R_1-R_2}{9\beta} \right\}, \\
\epsilon_3 &< \min\left\{ \frac{\alpha I(X_{1,1};Y_{R,1})+\gamma I(X_{1,3};Y_{R,3}|X_{2,3},Q)-R_1}{12\gamma}; \frac{\beta I(X_{2,2};Y_{R,2})+\gamma I(X_{2,3};Y_{R,3}|X_{1,3},Q)-R_2}{12\gamma}; \right. \\
&\qquad \left. \frac{\alpha I(X_{1,1};Y_{R,1})+\beta I(X_{2,2};Y_{R,2})+\gamma I(X_{1,3},X_{2,3};Y_{R,3}|Q)-R_1-R_2}{24\gamma} \right\}.
\end{aligned} \tag{2.14}$$

An error occurs, if the sequences are not jointly typical for the correct w. The probability of this event can be made arbitrary small by choosing n large. The second event that leads to an error is that there exists some $\tilde{w} \neq w$ such that the codewords are jointly typical with the received signal.

We split this event in three sub-events:

- E_1: $\tilde{w}_1 = w_1, \tilde{w}_2 \neq w_2$,
- E_2: $\tilde{w}_1 \neq, w_1\ \tilde{w}_2 = w_2$,
- E_3: $\tilde{w}_1 \neq, w_1\ \tilde{w}_2 \neq w_2$.

For n sufficiently large, the probability of E_1 averaged over all codewords and over the random codebook can be bounded from above by

$$\begin{aligned}
\mathbb{E}\{Pr[E_1]\} &\leq 2^{nR_2+n_2(3\epsilon_2-I(Y_{R,2};X_{2,2}))+n_3(6\epsilon_3-I(Y_{R,3};X_{2,3}|X_{1,3},Q))} \\
&\leq 2^{n(R_2+3\beta\epsilon_2-\beta I(Y_{R,2};X_{2,2})+6\gamma\epsilon_3-\gamma I(Y_{R,3};X_{2,3}|X_{1,3},Q))+I(Y_{R,2};X_{2,2})+I(Y_{R,3};X_{2,3}|X_{1,3},Q)}
\end{aligned}$$

using the properties of the typical set. Therefore $\mathbb{E}\{Pr[E_1]\}$ goes to zero for $n \to \infty$ by the choice of ϵ_2 and ϵ_3. $\mathbb{E}\{Pr[E_2]\}$ can be bounded in a similar way. Furthermore we have

$$\begin{aligned}
&\mathbb{E}\{Pr[E_3]\} \leq 2^{n(R_1+R_2)+n_1(3\epsilon_1-I(Y_{R,1};X_{1,1}))+n_2(3\epsilon_2-I(Y_{R,2};X_{2,2}))+n_3(8\epsilon_3-I(Y_{R,3};X_{2,3}X_{1,3}|Q))} \\
&\leq 2^{n(R_1+R_2+3\alpha\epsilon_1-\alpha I(Y_{R,1};X_{1,1})+3\beta\epsilon_2-\beta I(Y_{R,2};X_{2,2})+8\gamma\epsilon_3-\gamma I(Y_{R,3};X_{2,3}|X_{1,3},Q))+I(Y_{R,2};X_{2,2})+I(Y_{R,1};X_{1,1})+I(Y_{R,3};X_{2,3}|X_{1,3},Q)}.
\end{aligned}$$

Using a union bound argument we conclude that there exists a codebook such that $P_k^{(1)} \to 0$. Therefore the decoding at the relay will succeed with high probability, whenever the conditions are fulfilled with strict inequality. The achievability of the closure of the set follows from the definition of achievability. □

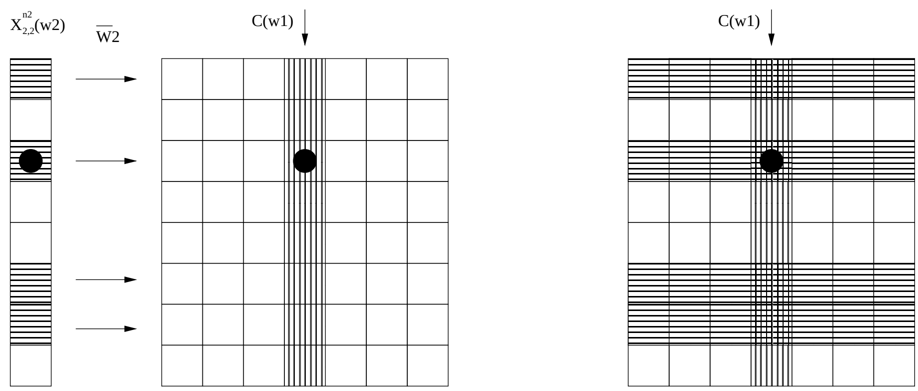

Figure 2.4: Coding mechanisms in the BC phase for more than two phases: The figure shows the decoding mechanism at receiver 1. On the left, the code $X_{2,2}^{n_2}(w_2)$ used by node 2 in its broadcast transmission is shown. The transmitted codeword is indicated by •. The receiver can not decode the message w_2 from the signal received in the broadcast transmission of phase 2. Nonetheless it can determine a subset $\bar{\mathcal{W}}_2$ of messages that could have been sent. The figure on the right of this code $X_{2,2}^{n_2}(w_2)$ displays the Cartesian structured code used by the relay's encoder. The codeword transmitted by the relay is again indicated by •. The known message w_1 determines a sub-code $C(w_1)$ of this code. Note that each codeword in the set $\bar{\mathcal{W}}_2$ corresponds to one row in the relay's code. In the figure on the right hand side, the codewords of this relay's code, which match these restrictions, are marked with vertical and horizontal lines respectively. The effective code used in the decoding for receiver 1 consists of the codewords marked with crossing lines.

2.3.3 A Note on Coding Mechanisms in the BC Phase for More Than Two Phases

The coding mechanisms in the BC phase for the three phase and for the four phase protocol are very similar. In fact, the same code can be used for both protocols under the condition that the following three parameters of the system stay the same:

1. the code in the first two phases,
2. the channel in the direct link,
3. the channel in the BC phase.

The only purpose of the MAC phase is to allow the relay to decode the data. As for the two-phase protocol, the code used by the relay's encoder possesses a Cartesian structure. The message known at the receiver restricts the possible transmitted codewords in the decoding process for the receiver to a subset of the codewords of the relay's code in the same way as it was discussed in Section 2.1.4.1.

Compared to the two-phase protocol the receiver now has additional side information due to the direct link between the terminal nodes. This additional side information imposes another restriction on the possible transmitted messages: The receiver cannot determine which message was sent solely by looking at the signal received via the direct link. Nevertheless the receiver can exclude some messages if the corresponding codewords are not jointly typical with the received signal. This is shown in Figure 2.4. The decoding at the receiver now combines both these restrictions leading to a sub-code of the code determined by the known message. As a consequence, the sub-code used in the decoding contains less codewords compared to the code for the two-phase protocol.

For a practical coding scheme one has to design interwoven single user codes as it was done for the two-phase protocol. But these codes need not be good codes for the marginal channel in the BC phase, as they need not be decodable without the additional side information. Now, the single user codes need to fulfill another constraint. These single user codes are super-codes of a set of single user sub-codes. These subcodes consist of less than $2^{n_{BC}I(X_R;Y_{k,BC})}$ codewords each, where n_{BC} is the block length of the BC code and $Y_{k,BC}$ is the random variable induced at the receiver k by the BC transmission form the relay. Note that for a transmitted message, say w_2, there may occur more than one such sub-code, as the direct link need not restrict the options for the receiver concerning w_2 to the same subset of $\bar{\mathcal{W}}_2 \subset \mathcal{W}_2$ in every transmission. It depends on the statistics of the direct link as well as on the code used for this transmission which and how many sub-codes may occur. In comparison to the coding for the two-phase protocol this means, that the codes used in the decoding cannot be determined offline anymore. The effective code used in the decoding depends on the signal received via the direct line and may change in each transmission, even if the known message is the same. On average, each of the sub-codes used in the decoding at the terminal node needs to be a good code for the marginal channel in the BC phase.

A simple though possibly infeasible or suboptimal code design for this setup may use single user codes interwoven as in the two-phase protocol. These single user codes have the property that — in average with respect to the statistics of the direct link, as we only consider the average probability of error — all subsets of size $2^{n_{BC}I(X_R;Y_{k,BC})}$ of the single user code are good codes for the channel from the relay to the receiver. As in the two-phase protocol, it is possible to use the XOR operated messages as an input for the encoder at the relay. In this case we have only one set of codewords, which needs to have the property, that in average the subsets of size $2^{n_{BC}I(X_R;Y_{k,BC})}$ are good codes for the channel to receiver k. Note, that if $R_1 \neq R_2$ and the XOR operation is used, not all possible subsets of the original set of codewords may occur for both receivers. The receiver with the lower rate already uses subsets of the complete set of codewords, even if the additional side information by the direct link is not present. By the additional side information the effective codes can only be subsets of these subsets and therefore not all subsets of the original set of codewords will occur.

Alternatively one can consider both transmissions as two independent transmissions of the same message that are both not decodable by themselves. This interpretation suggests the use of a turbo-like [61] mechanism with iterative decoding: Alternate between the decoding of the two codes and use the result of the last decoding step as soft information for the next decoding step. This is possible, as in both transmissions the same message is transmitted. In this case the known message restricts the relay's code in the same way as in the two-phase protocol and a similar code design can be applied.

2.4 Concluding Remarks

This chapter addresses the two-way relay channel with half-duplex nodes and a decode-and-forward constraint. We give an achievable rate region for a two-phase protocol for this kind of channel. The key ingredient of the scheme is the coding for the BC phase. In this phase the message available at the receiver can be used as side information. This enables a transmission that is de-facto interference free, implying that both receivers can achieve a rate as if the other receiver were not present. The only drawback stems from the common input distribution to the channel. This distribution may be suboptimal for one or both the single user channels. Starting at the achievability proof we derived a coding scheme that facilitates this special broadcast scenario where the receivers know the message for the other respective node. Finally, we extended the approach to use the direct link between both terminals and gave achievable rate regions for a protocol with three and four phases. In these protocols additional side information gathered by the direct link transmission further improves the decoding in the BC phase.

In all the coding schemes discussed in this section the messages known at the relay are the key to a quasi-interference free broadcast transmission: these messages can be used to restrict the relay's code to a sub-code, that is decodable for the receiver. The signal received via the direct link further restricts the code. This restriction is again determined by means of the decoded messages that index the codewords.

In the next section we will see that the relay does not need to decode the messages. We are still able to use the side information. While in the two phase protocol the restriction was a deterministic restriction, as the decoder knows one of the two transmitted messages, in the three phase protocol the restriction becomes stochastic. Though the restriction is still via the messages, the restricted set is partially determined by the decoder via typicality, by using the signal received in the direct link. In the next section we go one step further in this direction: We will focus on a two-phase channel without decoding at the relay. Still the relay's code used in the BC phase is restricted by the known message, although only in a stochastic sense. The relay's codeword depends on the signal received by the relay. This in turn depends on the codewords transmitted in the MAC phase and thereby on the known message at the receiver.

Chapter 3

The Two-Way Relay Channel with Compress-and-Forward

In the last chapter we considered the two-way relay channel with the constraint that the relay is able to decode the messages sent by both sources. In this chapter we drop this assumption and ask, whether it might be advantageous not to decode the messages. As a result, in this chapter we propose another coding strategy and a corresponding rate region for the two-way relay channel. Note that we maintain the constraint that all the nodes operate in a half-duplex mode. Furthermore we assume a two-phase protocol, i.e. we have a system without a direct link between the nodes.

The approach not to decode the messages at the relay seems counter-intuitive at first glance, as all information passes the relay. Still, it turns out that there are channels where this approach can enlarge the achievable rate region. A simple example shows that indeed performance gains can be achieved:

Example 3.1. *Consider a channel where the channel output in the first phase is the XOR sum of two binary inputs. The sum rate of this channel using it as a classical MAC is restricted to one bit per channel use. Now, assume that we have a channel in the second phase, which can error free transmit one bit of information per channel use to both receivers. Then, the protocol proposed in Chapter 2 cannot achieve the rate pair* $[0.5bit; 0.5bit]$ *since this rate pair would require the use of the BC for* 50% *of the time, which in turn would lead to a maximum sum rate of* 0.5 *bits in the MAC. It is easy to see that we can do better than that by simply transmitting the channel output to both receivers. Using this protocol, the overall channel can be interpreted as a special case of a restricted two-way channel [14]. The capacity region of this channel is known [14] and in the special case of the example it is given by*

$$\begin{aligned} R_1 &\leq 0.5bit, \\ R_2 &\leq 0.5bit. \end{aligned}$$

Therefore, the rate pair $[0.5bit; 0.5bit]$ *is achievable. This example shows that, at least when* *the sum-rate constraint of the MAC-phase limits the overall transmission rate, there might be a better way than decoding at the relay.*

The strategy proposed in this chapter follows the line of the classical compress-and-forward strategy in the paper by Cover et al. [16]. It includes the strategy used in the example above as a special case. The channel output at the relay is compressed and forwarded to the receivers. The receivers decode the message transmitted by the relay and therefore have an estimate of the MAC output. This estimate is used to decode the message intended for the receiver.

In Section 3.1 an achievable rate region is given using the compress-and-forward approach sketched in the example above. We will prove that gains can be achieved using the known message for both: decoding of the BC transmission and decoding of the MAC phase. The result has some interesting properties, namely the rate region has some strange non-continuity if one of the users is idle, i.e. has no data to transmit. This is discussed in detail in Section 3.1.3, where this is analyzed and thereby the base for further extensions is laid, which will be given later in Chapter 5.

In contrast to the achievable rate region $\mathcal{R}_{\mathrm{DF}}$ stated in Theorem 2.4, the new achievable rate region requires some auxiliary random variables. These variables are not directly determined by the system itself. To have a complete characterization of the achievable rate region, the cardinality of these auxiliary random variables needs to be bounded from above. We do this bounding in Section 3.2 for the variables that occur in the main theorem of this chapter (Theorem 3.1). For the bounding of the cardinality of further auxiliary random variables, which occur in other theorems in this thesis, the interested reader is referred to the appendix.

It will turn out that the new coding scheme may degrade the performance of the BC phase while it enhances the performance of the MAC phase. Therefore, in Section 3.3 we superimpose the new scheme and the decode-and-forward coding proposed in Chapter 2 to facilitate a tradeoff between both these approaches. The resulting superposition coding yields an achievable rate region which contains the region $\mathcal{R}_{\mathrm{DF}}$ stated in Theorem 2.4 as well as the region of Theorem 3.1 as special cases. Furthermore, the new region also contains the rate pairs that can be achieved by timesharing between the decode-and-forward and the compress-and-forward strategies.

An extension to more than two phases as it was done in Section 2.3 is possible and might further extend the achievable rate region. The usage of the direct link makes additional side information available at the receiver. This side information, i.e. the received signal, can also be used in the decoding of both phases in essentially the same way the codeword transmitted in the MAC phase of the two-phase protocol is used. The analysis of this extension does not yield any new insight to the solution of the problem and is not considered in more detail in this thesis.

3.1 A Compress-and-Forward Coding Theorem

In this section we propose an achievable rate region that is attained by a compress-and-forward strategy. Instead of decoding the MAC output the relay forwards a quantized representation of this output to both receivers. As in the proof of Theorem 2.2 the message known by the receiver is used to enhance the decoding performance in the BC phase. This is possible as the MAC output depends on the codeword transmitted by the receiver. Having decoded the compressed MAC output, the receiver uses the known message a second time. By interpreting the compressed MAC output as the output of a two-way channel between both the terminal nodes, the side information can be used in the decoding of the message intended for the receiver.

The results of this section where published in [11] and [12]. For the following discussion, recall the system model and the definitions given in Chapter 1.

3.1.1 Coding Theorem

Theorem 3.1. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_1 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &\leq \alpha I(X_1; \hat{Y}_R | X_2, Q) \\ R_2 &\leq \alpha I(X_2; \hat{Y}_R | X_1, Q) \end{aligned} \tag{3.1}$$

under the constraints

$$\begin{aligned} \alpha\left(H(\hat{Y}_R|X_1, Q) - H(\hat{Y}_R|Y_R)\right) &< \beta I(Y_1; X_R) \\ \alpha\left(H(\hat{Y}_R|X_2, Q) - H(\hat{Y}_R|Y_R)\right) &< \beta I(Y_2; X_R) \end{aligned} \tag{3.2}$$

for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and some joint probability distributions $p(q)p(x_1|q)p(x_2|q)$ $p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(x_R)p_2(y_1, y_2|x_R)$.

Remark 3.1. Because of the Markov chain $\hat{Y}_R - Y_R - X_k Q$ we have $H(\hat{Y}_R|X_k, Q) - H(\hat{Y}_R|Y_R) = I(\hat{Y}_R; Y_R|X_k, Q)$. Written in this way, the result is more similar to the compress-and-forward result of Cover et al. [16].

Remark 3.2 (Cardinalities of random variables)**.** To achieve any point in the stated region it is sufficient to consider only random variables Q and $\hat{Y}_R$ with cardinalities restricted to $|\mathcal{Q}| \leq 4$ and $|\hat{\mathcal{Y}}_R| \leq |\mathcal{Y}_R| + 3$. This can be shown using the Fenchel-Bunt extension of Caratheodory's theorem [62]. A proof of this claim is given in the Section 3.2.

Remark 3.3 (Connections to Wyner-Ziv coding and the restricted two-way channel)**.** The idea of the proof is to convey a quantized version of the output of the MAC at the relay to both receivers using the Wyner-Ziv coding mechanisms [33] in connection with the tools used in Tuncel's proof to transmit data to receivers that have some correlated side information [29].

By forwarding a compressed version of the relay's channel output to both receivers, a virtual restricted two-way channel with output $\hat{Y}_R$ at both receivers is established. This channel determines the rates achievable in the overall communication for probability distributions constrained by (3.2). This can be seen by comparing (3.1) with the result in [14]. The targeted rate in turn is a kind of distortion requirement for the quality of the the compressed channel output, i.e. $\alpha I(X_1; \hat{Y}_R|X_2, Q)$, $\alpha I(X_2; \hat{Y}_R|X_1, Q)$ can be read as two quality (or inverse distortion) measures while R_1 and R_2 is the respective reconstruction quality needed. With this interpretation the inequalities (3.2) define an achievable source region for transmitting a compressed variable Y_R under some distortion constraint over a BC to receivers that have some side information available. In the classical Wyner-Ziv setup [33], the reconstruction of Y_R is a function of the side information and the decoded variable $\hat{Y}_R$. In our setup, there is no gain in making such a step because of the nature of the "distortion constraint": The reconstruction cannot be improved with respect to the mutual information expression in (3.1) even if we replace $\hat{Y}_R$ by some $Z_k = g(\hat{Y}_R, X_k)$.

Remark 3.4 (More than one quantized variables)**.** A more general approach and a seemingly natural extension to this result could utilize two quantized variables $\hat{Y}_{R,1}$ and $\hat{Y}_{R,2}$, i.e. a virtual two-way channel with different outputs. In the general case of such a setup, the BC phase becomes more difficult. To facilitate the side-information in the BC, the two variables should be correlated, which leads to the problem of transmitting correlated data over a BC where the receivers have some correlated side-information. Without side-information the problem is treated by Han et al. [63], giving an achievable rate region. To the best of our knowledge, the problem was not yet considered with additional correlated side information at the receiver. At the end of this section we give some arguments that there might be some gains possible with this more general approach. These arguments will lead to an extension of the scheme along the lines sketched above. The resulting achievable rate region is stated and discussed in Chapter 5.

Remark 3.5 (Convexity of $\mathcal{R}_1$)**.** The region $\mathcal{R}_1$ is convex. To see that it is convex for fixed α and β note, that one can add Q as a condition to all entropy and mutual information terms without changing the region. If we allow for different timesharing parameters α and β, then we can use arguments analogous to that in Remark 2.11 to prove, that the region is convex.

3.1.2 Proof of the Coding Theorem

Proof. The proof of the theorem will be as follows: First we assume that the inequalities in the theorem are strict. For these cases we give a construction of a random codebook, define encoders and decoders. Thereafter we define the events that lead to an error while decoding. We use the union bound to bound the probability of a decoding error from above by using these events. We will show, that every event that leads to an error appears with a vanishing probability if the length of the code goes to infinity. Therefore the probability of error goes to zero given that the inequalities in the theorem are strict. Finally we show, that the closure of the rate region is achievable. This will conclude the proof.

Suppose the inequalities in (3.1) are strict for the probability distributions[1] $p(q)p(x_1|q)$ $p(x_2|q)p_1(y_R|x_1,x_2)p(\hat{y}_R|y_R)$, $p(x_R|q)p_2(y_1,y_2|x_R)$, some $\alpha,\beta > 0$ with $\alpha+\beta = 1$, and a rate pair $[R_1,R_2]$. We will first show how to construct a $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-code for a fixed n such that for the sequence of this $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-codes the probability of error goes to zero and the rate pair of the codes goes to $[R_1,R_2]$ as $n \to \infty$.

Some remarks on the variable Q: We do not use Q to switch between different codebooks as in traditional timesharing, but we use this parameter to create the now seemingly dependent variables X_1 and X_2. The traditional way of proving the achievability for a fixed q and then arguing with timesharing will not give the region in the theorem, as this region only forces the "average" codebook to fulfill the constraints, while the traditional timesharing argument, i.e. the convex combination of achievable rate points requires every codebook used in the timesharing to fulfill the constraints.

3.1.2.1 Random Codebook Generation

For a given n set $n_1 = \lfloor \alpha n \rfloor$, $n_2 = \lceil \beta n \rceil$.

- Choose one q^{n_1} drawn according to the probability $\prod_{s=1}^{n_1} p(q_{(s)}^{n_1})$.

- Choose $2^{\lfloor nR_1 \rfloor}$ i.i.d. codewords $x_1^{n_1}$ each according to the probability $\prod_{s=1}^{n_1} p(x_{1,(s)}^{n_1}|q_{(s)}^{n_1})$. Label these $x_1^{n_1}(w_1)$, $w_1 \in \{1,2,\ldots,2^{\lfloor nR_1 \rfloor}\}$.

- Choose $2^{\lfloor nR_2 \rfloor}$ i.i.d. codewords $x_2^{n_1}$ each according to the probability $\prod_{s=1}^{n_1} p(x_{2,(s)}^{n_1}|q_{(s)}^{n_1})$. Label these $x_2^{n_1}(w_2)$, $w_2 \in \{1,2,\ldots,2^{\lfloor nR_2 \rfloor}\}$.

- Choose $\epsilon_q \in (0, \min\{\epsilon_q^{(1)}, \epsilon_q^{(2)}, \epsilon_q^{(3)}, \epsilon_q^{(4)}\})$ where $\epsilon_q^{(1)} := \frac{1}{2\alpha+\beta}\Big(\beta I(X_R;Y_1) - \alpha\Big(H(\hat{Y}_R|X_1,Q)$ $-H(\hat{Y}_R|Y_R)\Big)\Big)$, $\epsilon_q^{(2)} := \frac{1}{2\alpha+\beta}\left(\beta I(X_R;Y_2) - \alpha\left(H(\hat{Y}_R|X_2,Q) - H(\hat{Y}_R|Y_R)\right)\right)$, $\epsilon_q^{(3)} := \frac{I(X_2;\hat{Y}_R|X_1,Q) - \frac{R_2}{\alpha}}{3}$ and $\epsilon_q^{(4)} := \frac{I(X_1;\hat{Y}_R|X_2,Q) - \frac{R_1}{\alpha}}{3}$.

- For each $i \in \{1,2,\ldots,2^{\lceil \alpha n R_Q \rceil}\}$, $R_Q = I(Y_R;\hat{Y}_R|Q) + \epsilon_q$, draw one codeword $\hat{y}_R^{n_1}(i)$ according to $\prod_{s=1}^{n_1} p(\hat{y}_{R,(s)}^{n_1}|q_{(s)}^{n_1})$ and one codeword $x_R^{n_2}(i)$ according to $\prod_{s=1}^{n_2} p(x_{R,(s)}^{n_2})$. The $2^{\lceil \alpha n R_Q \rceil}$ codeword pairs are drawn i.i.d..

This constitutes a random codebook $C^{(n)} = \{q^{n_1}\} \cup C_{x_1}^{(n)}(q^{n_1}) \cup C_{x_2}^{(n)}(q^{n_1}) \cup C_{\hat{y}_R}^{(n)}(q^{n_1}) \cup C_{x_R}^{(n)}$ where $C_{x_1}^{(n)}(q^{n_1})$ is the ordered set of codewords $x_1^{n_1}(1), \ldots x_1^{n_1}(2^{\lfloor nR_1 \rfloor})$ drawn conditioned on a given q^{n_1}, and the ordered sets $C_{x_2}^{(n)}(q^{n_1})$, $C_{\hat{y}_R}^{(n)}(q^{n_1})$, $C_{x_R}^{(n)}$ are defined accordingly for the remaining codewords.

[1] All probabilities in the proof will be calculated for these given distributions.

3.1.2.2 Decoding Sets

For the decoding we use typical set decoding. The coding and decoding steps are specified below. For a strict definition of the decoding sets we choose parameter for the typical sets as $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon_4 \in (0, \frac{\epsilon_q}{6})$. The missing parameters for the receiver 2 are chosen in an analogous way.

3.1.2.3 Coding

i To transmit message w_1 node 1 sends $x_1^{n_1}(w_1)$.

ii To transmit message w_2 node 2 sends $x_2^{n_1}(w_2)$.

iii Upon receiving $y_R^{n_1}$ the relay looks for the first i such that $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})$. If such an i is found the relay transmits $x_R^{n_2}(i)$. If no such i is found the relay chooses[2] $i = 1$ and transmits $x_R^{n_2}(1)$. This induces a mapping $f : \mathcal{Y}_R^{n_1} \to C_{\hat{y}_R}^{(n)}(q^{n_1})$ as $\hat{y}_R^{n_1}(i) = f(y_R^{n_1})$.

iv Upon receiving $y_1^{n_2}$ node 1 looks for the unique i such that $x_R^{n_2}(i)$ and the received signal $y_1^{n_2}$ are jointly typical, and simultaneously the side-information $x_1^{n_1}(w_1)$ and $\hat{y}_R^{n_1}(i)$ are jointly typical given q^{n_1}, i.e. $\left(x_R^{n_2}(i), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, Y_1)$ and $\left(x_1^{n_1}(w_1), \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, \hat{Y}_R|q^{n_1})$. This enables node 1 to recover $\hat{y}_R^{n_1}(i)$. If no or more than one such i is found, the decoding is aborted and $w_2 = 1$ is chosen.

v Knowing $\hat{y}_R^{n_1}(i)$ and $x_1^{n_1}(w_1)$ node 1 decides that w_2 was transmitted if $x_2^{n_1}(w_2)$ is the only codeword such that $x_2^{n_1}(w_2)$, $\hat{y}_R^{n_1}(i)$ and $x_1^{n_1}(w_1)$ are jointly typical given q^{n_1}, i.e. we have $\left(x_1^{n_1}(w_1), x_2^{n_1}(w_2), \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R|q^{n_1})$. If no or more than one such codeword is found, the decoding is aborted and $w_2 = 1$ is chosen.

vi The decoding at node 2 is performed in an analogous way.

3.1.2.4 Error Events

Now we show that the average probability of error goes to zero in the average of all random codebooks, more precisely we show that for any given ϵ there exists an $n^{(0)}$ such that $\mathbb{E}\{\mu_k^{(n)}\} < \epsilon$, $k \in \{1, 2\}, n > n^{(0)}$, where the expectation is taken over the random codebook. This in turn implies that for any ϵ we can find $n^{(0)}$ such that $\mathbb{E}\{\mu_1^{(n)} + \mu_2^{(n)}\} < 2\epsilon$, $n > n^{(0)}$, and therefore there is at least one codebook with $\mu_1^{(n)} + \mu_2^{(n)} < 2\epsilon$, $n > n^{(0)}$, and therefore $\mu_1^{(n)} < 2\epsilon$ and $\mu_2^{(n)} < 2\epsilon$ for $n > n^{(0)}$.

We bound the average probability of error $\mathbb{E}\{\mu_1^{(n)}\}$ from above by the union bound using six events E_j, $j \in \{1, 2, \ldots, 6\}$, whose union is a superset of the error event. Therefore we have

[2] This is done to have a well defined error probability. Equivalently one could declare an error at the relay, but this induces a much more cumbersome notation in the definition of the error probability. Similar arguments apply when choosing $w_2 = 1$ in coding step iv and iv.

$\mathbb{E}\{\mu_1^{(n)}\} \leq \sum_{j=1}^{6} \mathbb{E}\{Pr[E_j]\}$. The average error probability $\mathbb{E}\{\mu_2^{(n)}\}$ can be bounded in an analogous way.

In what follows we summarize the definition of the error events for receiver 1:

- E_1: Suppose a codebook is given and $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted. E_1 is the event that there does not exist an $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_Q \rceil}\}$ such that $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R | q^{n_1})$.

- E_2: Suppose a codebook is given and transmits $x_R^{n_2}(i)$. E_2 is the event that $\left(x_R^{n_2}(i), y_1^{n_2}\right) \notin \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R^{n_2}, Y_1^{n_2})$.

- E_3: Suppose a codebook is given, $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted and the relay chooses some i such that $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i) = f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R | q^{n_1})$. E_3 is the event that $\left(x_1^{n_1}(w_1), \hat{y}_R^{n_1}(i)\right) \notin \mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, \hat{Y}_R | q^{n_1})$.

- E_4: Suppose a codebook is given, $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted, the relay chooses some i and $x_R^{n_2}(i)$ is transmitted. E_4 is the event that $\exists j \neq i : \left(x_1^{n_1}(w_1), \hat{y}_R^{n_1}(j)\right) \in \mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, \hat{Y}_R | q^{n_1})$ and $\left(x_R^{n_2}(j), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n)}(X_R^{n_2}, Y_1^{n_2})$.

- E_5: Suppose a codebook is given, $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted and the relay chooses some i such that $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i) = f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R | q^{n_1})$. E_5 is the event that $\left(x_1^{n_1}(w_1), x_2^{n_1}(w_2), \hat{y}_R^{n_1}(i)\right) \notin \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R | q^{n_1})$.

- E_6: Suppose a codebook is given, $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted and the relay chooses some i such that $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i) = f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R | q^{n_1})$. E_6 is the event that $\left(x_1^{n_1}(w_1), x_2^{n_1}(\hat{w}_2), \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_4}^{(\alpha n)}(X_1, X_2, \hat{Y}_R | q^{n_1})$ with $\hat{w}_2 \neq w_2$.

To see that these error events capture all events that may lead to an error we step through the coding procedure and verify, that all possible causes of an error where captured: First note that the probability of error can be bounded from above by

$$Pr[E] \leq Pr[E^{\mathrm{v}} \cup \bar{E}^{\mathrm{iv}} \cup \bar{E}^{\mathrm{iii}}] + Pr[E^{\mathrm{iv}}] + Pr[E^{\mathrm{iii}}].$$

Here E^{v} is the event that coding step v fails, i.e. that no or more than one codeword $x_2^{n_1}(w_2)$ is found. Accordantly E^{iv} and E^{iii} are the event that the coding step iv and iii failed respectively. A bar indicates the complementary event.

We break down the events even further: In coding step iii it may turn out that there is no typical $\hat{y}_R^{n_1}(i)$ for the received $y_R^{n_1}$. This does not yield an error immediately, but it may lead to an error in later decoding. To simplify the error calculation we treat this as an error captured by event E_1 and for the following considerations about error events we can assume that we have $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R | q^{n_1})$. Therefore E_1 is not an intrinsic error event, but it is used to simplify the definitions and the calculation of the errors that may happen in the coding steps iv and v.

Coding step iv fails, if either i is not found, or if $j \neq i$ is found. The correct i is not found if either $\left(x_R^{n_2}(i), y_1^{n_2}\right) \notin \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, Y_1)$, captured by E_2, or if $\left(x_1^{n_1}(w_1), \hat{y}_R^{n_1}(i)\right) \notin \mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, \hat{Y}_R|q^{n_1})$, captured by E_3. The event that $j \neq i$ is found in step iv is captured by E_4.

Coding step v fails, if either w_2 is not found, or if $\hat{w}_2 \neq w_2$ is found. These events are captured by E_5 and E_6 respectively. Clearly no other events lead to an error for the decoding process at receiver 1.

We will now prove for each event E_j, $j \in \{1, 2, \ldots, 6\}$, that there exists an $n^{(j)}$ such that $\mathbb{E}\{Pr[E_j]\} < \frac{\epsilon}{6}$ for $n \geq n^{(j)}$. This in turn implies that for $n \geq \max_{j\in\{1,2,\ldots,6\}} n^{(j)} =: n^{(0)}$ we have $\mathbb{E}\{\mu_k^{(n)}\} < \epsilon$, $k \in \{1, 2\}$.

3.1.2.5 Bounding the Probability of the Error Events

We now bound the probability of the error events averaged over the codebooks $C^{(n)}$ for a code of length n and the transmitted messages $[w_1, w_2] \in \mathcal{W}$.

Error event E_1 The averaged probability for the error event E_1 is

$$\mathbb{E}\{Pr[E_1]\} = \frac{1}{|\mathcal{W}|} \sum_{(w_1,w_2)\in\mathcal{W}} \sum_{C^{(n)}} p(C^{(n)}) \sum_{y_R^{n_1}: y_R^{n_1} \in \mathcal{J}(C^{(n)})} p\left(y_R^{n_1}|x_1^{n_1}(w_1), x_2^{n_1}(w_2)\right)$$

with

$$\mathcal{J}(C^{(n)}) = \Big\{ y_R^{n_1} \in \mathcal{Y}_R^{n_1} : \nexists i \in \left\{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\right\}, \hat{y}_R^{n_1}(i) \in C_{\hat{y}_R}^{(n)}(q^{n_1}) \text{ such that } \left(y_R^{n_1}, \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1}) \Big\}.$$

This can be rewritten as

$$\mathbb{E}\{Pr[E_1]\} = \sum_{C^{(n)}} p(C^{(n)}) \sum_{y_R^{n_1}: y_R^{n_1} \in \mathcal{J}(C^{(n)})} p\left(y_R^{n_1}|x_1^{n_1}(1), x_2^{n_1}(1)\right)$$

as the codewords are drawn i.i.d.. Furthermore, we can simplify the expression by summing over $x_1^{n_1}(w_1)$, $w_1 \neq 1$, $x_2^{n_1}(w_2)$, $w_2 \neq 1$ and $x_R^{n_2}(i)$, and dropping the index of the remaining codewords $x_1^{n_1}(1)$ and $x_2^{n_1}(1)$:

$$\mathbb{E}\{Pr[E_1]\} = \sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} \sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(x_1^{n_1}, x_2^{n_1}, q^{n_1}, y_R^{n_1}) \sum_{C_{\hat{y}_R}^{(n)}(q^{n_1}): y_R^{n_1} \in \mathcal{J}\left(C_{\hat{y}_R}^{(n)}(q^{n_1})\right)} p\left(C_{\hat{y}_R}^{(n)}(q^{n_1})\right).$$

Here we used

$$\mathcal{J}\left(C_{\hat{y}_R}^{(n)}(q^{n_1})\right) = \left\{y_R^{n_1} \in \mathcal{Y}_R^{n_1} : \nexists i \in \left\{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\right\}, \hat{y}_R^{n_1}(i) \in C_{\hat{y}_R}^{(n)}(q^{n_1}) \text{ such that} \right.$$
$$\left. \left(y_R^{n_1}, \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R | q^{n_1})\right\}.$$

With the indicator function χ on the typical set $\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})$ it follows that

$$\mathbb{E}\{Pr[E_1]\} = \sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{q^{n_1} \in Q^{n_1}} p(q^{n_1}, y_R^{n_1}) \left(1 - \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|q^{n_1}) \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1})\right)^{2^{\lceil \alpha n R_Q \rceil}}.$$

The term in the bracket is the probability that for a given q^{n_1} and a given $y_R^{n_1}$, a random $\hat{y}_R^{n_1}$ drawn according to $p(\hat{y}_R^{n_1}|q^{n_1})$ is not jointly typical with $y_R^{n_1}$. The exponent $2^{\lceil \alpha n R_Q \rceil}$ accounts for the fact, that there are $2^{\lceil \alpha n R_Q \rceil}$ possible $\hat{y}_R^{n_1}$ in the code, and an error occurs if none of them is jointly typical with $y_R^{n_1}$.

Now we can bound

$$\sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|q^{n_1}) \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1})$$

from below by using properties of the typical set: For $(y_R^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})$ and sufficiently large n, i.e. for some $n > n^{(1,1)}$ we have

$$\begin{aligned} p(\hat{y}_R^{n_1}|y_R^{n_1}, q^{n_1}) &= \frac{p(\hat{y}_R^{n_1}, y_R^{n_1}|q^{n_1})}{p(y_R^{n_1}|q^{n_1})} \\ &\leq p(\hat{y}_R^{n_1}|q^{n_1}) \frac{2^{-n_1\left(H(Y_R, \hat{Y}_R|Q) - 2\epsilon_1\right)}}{2^{-n_1\left(H(Y_R|Q) + 2\epsilon_1\right)} 2^{-n_1\left(H(\hat{Y}_R|Q) + 2\epsilon_1\right)}} \\ &= p(\hat{y}_R^{n_1}|q^{n_1}) 2^{\lfloor \alpha n \rfloor \left(I(Y_R; \hat{Y}_R|Q) + 6\epsilon_1\right)} \\ &\leq p(\hat{y}_R^{n_1}|q^{n_1}) 2^{\alpha n \left(I(Y_R; \hat{Y}_R|Q) + 6\epsilon_1\right)}. \end{aligned}$$

Therefore

$$\begin{aligned} &\sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|q^{n_1}) \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}) \\ &\qquad \geq \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|y_R^{n_1}, q^{n_1}) 2^{-\alpha n \left(I(Y_R; \hat{Y}_R|Q) + 6\epsilon_1\right)} \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}) \end{aligned}$$

and

$$\mathbb{E}\{Pr[E_1]\} \leq \sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}, y_R^{n_1}) \Big(1 - \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1} | y_R^{n_1}, q^{n_1}) 2^{-\alpha n \left(I(Y_R;\hat{Y}_R|Q)+6\epsilon_1\right)} \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1})\Big)^{2^{\alpha n R_Q}}.$$

This can be bounded form above [30, Lemma 13.5.3] by

$$\mathbb{E}\{Pr[E_1]\} \leq 1 - \sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(q^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1}) \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}) + \exp\left(-2^{\alpha n \left(R_Q - I(Y_R;\hat{Y}_R|Q)-6\epsilon_1\right)}\right).$$

Now, $R_Q = I(Y_R; \hat{Y}_R|Q) + \epsilon_q$ and $\epsilon_q > 6\epsilon_1$. Therefore the term $\exp\left(-2^{\alpha n \left(R_Q - I(Y_R;\hat{Y}_R|Q)-6\epsilon_1\right)}\right)$ can be made arbitrarily small for n large. In particular for a given $\epsilon > 0$ we can find $n^{(1,2)}$ such that $\exp\left(-2^{\alpha n \left(R_Q - I(Y_R;\hat{Y}_R|Q)-6\epsilon_1\right)}\right) < \frac{\epsilon}{12}$ for all $n > n^{(1,2)}$.

The remaining term

$$1 - \sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(q^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1}) \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}) \tag{3.3}$$

is the probability that $(y_R^{n_1}, \hat{y}_R^{n_1}) \notin \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})$ for sequences $q^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1}$ drawn according to the joint probability distribution $p(q^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1})$. By the law of large numbers this probability goes to zero. In particular by Lemma 1.2 for a given $\epsilon > 0$ we can find $n^{(1,3)}$ such that (3.3) is smaller than $\frac{\epsilon}{12}$ for all $n > n^{(1,3)}$. We can now choose $n^{(1)} \geq \max\left\{n^{(1,1)}, n^{(1,2)}, n^{(1,3)}\right\}$ and the probability of error for the first error event can be bounded by $\mathbb{E}\{Pr[E_1]\} < \frac{\epsilon}{6}$ for $n \geq n^{(1)}$.

Error event E_2 For the fixed i chosen by the relay the averaged probability for the error event E_2 is

$$\begin{aligned} \mathbb{E}\{Pr[E_2]\} &= \sum_{C^{(n)}} p(C^{(n)}) \sum_{y_1^{n_2} \in \mathcal{Y}_1^{n_2}} p(y_1^{n_2} | x_R^{n_2}(i)) \chi^C_{\mathcal{T}_{\epsilon_2}^{(n_2)}(X_R,Y_1)}(x_R^{n_2}(i), y_1^{n_2}) \\ &= \sum_{x_R^{n_2} \in \mathcal{X}_R^{n_2}} \sum_{y_1^{n_2} \in \mathcal{Y}_1^{n_2}} p(x_R^{n_2}, y_1^{n_2}) \chi^C_{\mathcal{T}_{\epsilon_2}^{(n_2)}(X_R,Y_1)}(x_R^{n_2}, y_1^{n_2}). \end{aligned}$$

This is the probability that a pair of sequences $(x_R^{n_2}, y_1^{n_2})$ drawn according to $p(x_R^{n_2}, y_1^{n_2})$ is not in $\mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, Y_1)$. The probability for this goes to zero as $n \to \infty$ by the law of large numbers and the definition of the typical set. Therefore for any given ϵ we can find $n^{(2)}$ such that for $\mathbb{E}\{P(E_2)\} < \frac{\epsilon}{6}$ for $n \geq n^{(2)}$.

Error event E_3 This part of the proof could be done along the lines to the proof for error event E_5 below. But in fact as for the parameters associated with the typical sets in these coding steps we have $\epsilon_3 = \epsilon_4$, it follows that whenever E_5 does not appear, E_3 will not appear as well, because of the definition of the typical sets. Therefore we have $\mathbb{E}\{Pr[E_3]\} \leq \mathbb{E}\{Pr[E_5]\}$ for any n and we can simply choose some $n^{(3)} \geq n^{(5)}$ and have $\mathbb{E}\{Pr[E_3]\} < \frac{\epsilon}{6}$ for $n \geq n^{(3)}$ given that $\mathbb{E}\{Pr[E_5]\} < \frac{\epsilon}{6}$ for $n \geq n^{(5)}$.

Error event E_4 The averaged probability for the error event E_4 is bounded from above by

$$\mathbb{E}\{Pr[E_4]\} \leq \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}) Pr[E_{4,1}] Pr[E_{4,2}] 2^{\lceil \alpha n R_Q \rceil}$$

Here $E_{4,1}$ is the event that for two sequences $x_1^{n_1}, \hat{y}_R^{n_1}$ drawn independently of each other given q^{n_1} we have $(x_1^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_3}^{(\alpha n)}(X_1, \hat{Y}_R | q^{n_1})$. $x_1^{n_1}$ and $\hat{y}_R^{n_1}$ are drawn at random according to $p(x_1^{n_1}|q^{n_1})$ and $p(\hat{y}_R^{n_1}|q^{\alpha n})$ respectively to capture the averaging over the random codebooks. $E_{4,2}$ is the event that for two sequences $x_R^{n_2}, y_1^{n_2}$ drawn independently of each other we have $(x_R^{n_2}, y_1^{n_2}) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, Y_1)$. The factor $2^{\lceil \alpha n R_Q \rceil}$ accounts for the fact that we can use a union bound and the error occurs if at least one $j \neq i$ is found fulfilling the requirements.

For sufficiently large n, we have

$$\begin{aligned} Pr[E_{4,1}] = & \sum_{(x_1^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{X}_1^{n_1} \times \mathcal{Y}_R^{n_1}} p(x_1^{n_1}|q^{n_1}) p(\hat{y}_R^{n_1}|q^{n_1}) \chi_{\mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, \hat{Y}_R|q^{n_1})}(x_1^{n_1}, \hat{y}_R^{n_1}) \\ & \leq |\mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, \hat{Y}_R|q^{n_1})| 2^{-n_1(H(X_1|Q)-2\epsilon_3)} 2^{-n_1\left(H(\hat{Y}_R|Q)-2\epsilon_3\right)} \end{aligned}$$

due to the properties of the typical set. Furthermore, it follows from these properties that for sufficiently large n

$$|\mathcal{T}_{\epsilon_3}^{(n_1)}(X_1, \hat{Y}_R|q^{n_1})| \leq 2^{n_1\left(H(X_1, \hat{Y}_R|Q)+2\epsilon_3\right)}.$$

$Pr[E_{4,2}]$ can be bounded in a similar way. As a consequence there exists $n^{(4,1)}$ such that the above bounds for $Pr[E_{4,1}]$, $Pr[E_{4,2}]$ hold for all $n > n^{(4,1)}$. Therefore we have for $n > n_{4,1}$

$$\begin{aligned} \mathbb{E}\{Pr[E_4]\} \leq & \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}) 2^{-n_1\left(I(X_1;\hat{Y}_R|Q)-6\epsilon_3\right)} 2^{-n_2(I(X_R;Y_1)-6\epsilon_2)} 2^{\alpha n R_Q+1} \\ & \leq 2^{-n\left(\alpha\left(I(X_1;\hat{Y}_R|Q)-R_Q-6\epsilon_3\right)+\beta(I(X_R;Y_1)-6\epsilon_2)\right)+1+I(X_1;\hat{Y}_R|Q)+6\epsilon_2} \\ & = 2^{-n\left(\alpha\left(I(X_1;\hat{Y}_R|Q)-I(Y_R;\hat{Y}_R|Q)\right)+\beta I(X_R;Y_1)-\tilde{\epsilon}\right)+1+I(X_1;\hat{Y}_R|Q)+6\epsilon_2} \end{aligned}$$

with

$$\tilde{\epsilon} = \alpha\epsilon_q + \beta 6\epsilon_2 + \alpha 6\epsilon_3.$$

This term goes to zero if

$$\alpha\left(I(X_1;\hat{Y}_R|Q)-I(Y_R;\hat{Y}_R|Q)\right)+\beta I(X_R;Y_1)-\tilde{\epsilon}=\\ \beta I(X_R;Y_1)-\alpha\left(H(\hat{Y}_R|X_1,Q)-H(\hat{Y}_R|Y_R)\right)-\tilde{\epsilon}>0.$$

Now

$$\beta I(X_R;Y_1)-\alpha\left(H(\hat{Y}_R|X_1,Q)-H(\hat{Y}_R|Y_R)\right)>0$$

because of the constraints fulfilled by assumption, and

$$\tilde{\epsilon}<\beta I(X_R;Y_1)-\alpha\left(H(\hat{Y}_R|X_1,Q)-H(\hat{Y}_R|Y_R)\right)$$

due to the choice of the parameters ϵ_q, ϵ_2, and ϵ_3. Therefore for any given ϵ we can find $n^{(4)}>n^{(4,1)}$ such that for $\mathbb{E}\{Pr[E_4]\}<\frac{\epsilon}{6}$ for $n\geq n^{(4)}$.

Error event E_5 The averaged probability for the error event E_5 is bounded from above by

$$\mathbb{E}\{Pr[E_5]\}\leq\frac{1}{|\mathcal{W}|}\sum_{(w_1,w_2)\in\mathcal{W}}\sum_{C^{(n)}}p\left(C^{(n)}\right)\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}p\left(y_R^{n_1}|x_1^{n_1}(w_1),x_2^{n_1}(w_2)\right)\\ \times\chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}\left(y_R^{n_1},f(y_R^{n_1})\right)\chi^C_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}\left(x_1^{n_1}(w_1),x_2^{n_1}(w_2),f(y_R^{n_1})\right).$$

In this formula $f:\mathcal{Y}_R^{n_1}\to C_{\hat{y}_R}^{(n)}(q^{n_1})$ is the mapping induced by the relay when choosing i upon receiving $y_R^{n_1}$. We use $f(y_R^{n_1})$ here instead of $\hat{y}_R^{n_1}(i)$ as all randomness needed for this calculation of the probability of the event E_5 is induced by the channel via $y_R^{n_1}$ and the random coding, which determines the mapping $f(\cdot)$. Furthermore, by restricting the sum via the indicator function on the set $\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})$ we assume that the coding mechanism at the relay found a typical sequence $\hat{y}_R^{n_1}(i)$.

We can rewrite the upper bound as

$$\mathbb{E}\{Pr[E_5]\}\leq\sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}}\sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}}\sum_{q^{n_1}\in\mathcal{Q}^{n_1}}\sum_{C_{\hat{y}_R}^{(n)}(q^{n_1})}\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}p\left(C_{\hat{y}_R}^{(n)}(q^{n_1})\right)p\left(q^{n_1},x_1^{n_1},x_2^{n_1},y_R^{n_1}\right)\\ \times\chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}\left(y_R^{n_1},f(y_R^{n_1})\right)\chi^C_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}\left(x_1^{n_1}(w_1),x_2^{n_1}(w_2),f(y_R^{n_1})\right)\\ =\sum_{q^{n_1}\in\mathcal{Q}^{n_1}}\sum_{C_{\hat{y}_R}^{(n)}(q^{n_1})}\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}p\left(y_R^{n_1},q^{n_1}\right)p\left(C_{\hat{y}_R}^{(n)}(q^{n_1})\right)\chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}\left(y_R^{n_1},f(y_R^{n_1})\right)\\ \sum_{(x_1^{n_1},x_2^{n_1})\in\mathcal{X}_1^{n_1}\times\mathcal{X}_2^{n_1}}p(x_1^{n_1},x_2^{n_1}|y_R^{n_1},q^{n_1})\chi^C_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}\left(x_1^{n_1}(w_1),x_2^{n_1}(w_2),f(y_R^{n_1})\right),$$

where $C_{\hat{y}_R}^{(n)}(q^{n_1})$ is the part of the codebook containing $\hat{y}_R^{n_1}(i), i\in\left\{1,2,\ldots,2^{\lceil\alpha nR_Q\rceil}\right\}$ and therefore defines the mapping $f(y_R^{n_1})$. The last sum is the probability that for a given $y_R^{n_1}$, q^{n_1} and for $(x_1^{n_1}, x_2^{n_1})$ drawn according to $p(x_1^{n_1}, x_2^{n_1}|y_R^{n_1}, q^{n_1})$ the triple $x_1^{n_1}, x_2^{n_1}, f(y_R^{n_1})$ is not jointly typical, i.e. $\left(x_1^{n_1}, x_2^{n_1}, f(y_R^{n_1})\right) \notin \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R|q^{n_1})$. Now, $\left(y_R^{n_1}, f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})$ implies $\left(y_R^{n_1}, f(y_R^{n_1}), q^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R, Q)$. Furthermore for any $\left(y_R^{n_1}, f(y_R^{n_1}), q^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R, Q)$ and for $(x_1^{n_1}, x_2^{n_1})$ drawn according to $p(x_1^{n_1}, x_2^{n_1}|y_R^{n_1}, q^{n_1})$ we have by the properties of the typical set that

$$Pr\left[(x_1^{n_1}, x_2^{n_1}) \in \mathcal{T}_{\epsilon_1}^{(n_1)}\left(X_1, X_2|y_R^{n_1}, f(y_R^{n_1}), q^{n_1}\right)|y_R^{n_1}, f(y_R^{n_1}), q^{n_1}\right]$$

can be made arbitrarily close to 1 by choosing n large. Here we used the fact that $(x_1^{n_1}, x_2^{n_1})$ are independent of $\hat{y}_R^{n_1}$ given $y_R^{n_1}$ and q^{n_1}.

Furthermore, $\left(x_1^{n_1}, x_2^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, X_2|y_R^{n_1}, f(y_R^{n_1}), q^{n_1})$ implies that we have $\left(x_1^{n_1}, x_2^{n_1}, y_R^{n_1}, f(y_R^{n_1}), q^{n_1}\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, X_2, Y_R, \hat{Y}_R, Q)$ and therefore $\left(x_1^{n_1}, x_2^{n_1}, f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X_1, X_2, \hat{Y}_R|q^{n_1})$. Now, as we choose $\epsilon_1 = \epsilon_4$ we can conclude that $\mathbb{E}\{Pr[E_{5,1}]\}$ can be made arbitrarily small by choosing n large. In particular for a given $\epsilon > 0$ we can find $n^{(5)}$ such that $\mathbb{E}\{Pr[E_5]\} < \frac{\epsilon}{6}$ for all $n > n^{(5)}$.

Error event E_6 The probability of this event can be bounded from above by

$$\mathbb{E}\{Pr[E_6]\} \leq \frac{1}{|\mathcal{W}|} \sum_{(w_1,w_2)\in\mathcal{W}} \sum_{\substack{\hat{w}_2\in\mathcal{W}_2\\ \hat{w}_2\neq w_2}} \sum_{C^{(n)}} p\left(C^{(n)}\right) \sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} p\left(y_R^{n_1}|x_1^{n_1}(w_1), x_2^{n_1}(w_2)\right)$$
$$\times \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}\left(y_R^{n_1}, f(y_R^{n_1})\right)\chi_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}\left(x_1^{n_1}(w_1), x_2^{n_1}(\hat{w}_2), f(y_R^{n_1})\right)$$

where $f : \mathcal{Y}_R^{n_1} \to C_{\hat{y}_R}^{(n)}(q^{n_1})$ is the mapping induced by the relay when choosing i upon receiving $y_R^{n_1}$. This can be bounded from above as

$$\mathbb{E}\{Pr[E_6]\} \leq \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} \sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} \sum_{\hat{y}_R^{n_1}\in\hat{\mathcal{Y}}_R^{n_1}} \sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} 2^{nR_2}2^{\lceil\alpha nR_Q\rceil}p(q^{n_1}, x_1^{n_1}, x_2^{n_1})$$
$$\times p(y_R^{n_1}|x_1^{n_1})p(\hat{y}_R^{n_1}|q^{n_1})\chi_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1})\chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}).$$

For $(y_R^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})$ and for n sufficiently large we have

$$\begin{aligned} p(\hat{y}_R^{n_1}|y_R^{n_1}, q^{n_1}) &= \frac{p(\hat{y}_R^{n_1}, y_R^{n_1}|q^{n_1})}{p(y_R^{n_1}|q^{n_1})} \\ &\geq p(\hat{y}_R^{n_1}|q^{n_1})\frac{2^{-n_1\left(H(Y_R,\hat{Y}_R|Q)+2\epsilon_1\right)}}{2^{-n_1\left(H(Y_R|Q)-2\epsilon_1\right)}2^{-n_1\left(H(\hat{Y}_R|Q)-2\epsilon_1\right)}} \\ &= p(\hat{y}_R^{n_1}|q^{n_1})2^{n_1\left(I(Y_R;\hat{Y}_R|Q)-6\epsilon_1\right)}. \end{aligned}$$

Therefore we have

$$\mathbb{E}\{Pr[E_6]\} \leq 2^{\alpha n(\epsilon_q+6\epsilon_1)+1+I(Y_R;\hat{Y}_R|Q)} 2^{nR_2} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} \sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} \sum_{\hat{y}_R^{n_1}\in\hat{\mathcal{Y}}_R^{n_1}} \sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} p(q^{n_1}, x_1^{n_1}, x_2^{n_1})$$
$$\times p(y_R^{n_1}|x_1^{n_1}) p(\hat{y}_R^{n_1}|y_R^{n_1}) \chi_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) \chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,\hat{Y}_R|q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}).$$

The sum in this upper bound can be seen to be the probability that for sequences $x_1^{n_1}$, $x_2^{n_1}$, $y_R^{n_1}$, $\hat{y}_R^{n_1}$, q^{n_1} drawn according to the distributions $p(q^{n_1})$, $p(y_R^{n_1}, \hat{y}_R^{n_1}, x_1^{n_1}|q^{n_1})$ and $p(x_2^{n_1}|q^{n_1})$ we have that $(y_R^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, \hat{Y}_R|q^{n_1})$ while at the same time $(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R|q^{n_1})$. It follows that

$$\mathbb{E}\{Pr[E_6]\} \leq 2^{\alpha n(\epsilon_q+6\epsilon_1)+1+I(Y_R;\hat{Y}_R|Q)} 2^{nR_2} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} \sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} \sum_{\hat{y}_R^{n_1}\in\hat{\mathcal{Y}}_R^{n_1}} p(q^{n_1}) p(\hat{y}_R^{n_1}, x_1^{n_1}|q^{n_1}) p(x_2^{n_1}|q^{n_1})$$
$$\times \chi_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}).$$

For sufficiently large n this can be bounded from above by

$$\begin{aligned}\mathbb{E}\{Pr[E_6]\} &\leq 2^{\alpha n(\epsilon_q+6\epsilon_1)+1+I(Y_R;\hat{Y}_R|Q)} 2^{nR_2} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1}) 2^{-n_1\left(I(X_2;\hat{Y}_R|X_1,Q)-6\epsilon_4\right)} \\ &\leq 2^{n\left(R_2-\alpha I(X_2;\hat{Y}_R|X_1,Q)\right)} 2^{\alpha n(\epsilon_q+6\epsilon_1+6\epsilon_4)+1+I(Y_R;\hat{Y}_R|Q)+I(X_2;\hat{Y}_R|X_1,Q)} \\ &\leq 2^{n\left(R_2-\alpha\left(I(X_2;\hat{Y}_R|X_1,Q)+3\epsilon_q\right)\right)+1+I(Y_R;\hat{Y}_R|Q)+I(X_2;\hat{Y}_R|X_1,Q)}\end{aligned}$$

using the properties of the typical set.

By assumption $R_2 < \alpha I(X_2; \hat{Y}_R|X_1, Q)$ and we choose $\epsilon_q < \frac{I(X_2;\hat{Y}_R|X_1,Q)-\frac{R_2}{\alpha}}{3}$. Therefore the probability of this event can be made arbitrarily small for n large. In particular for a given $\epsilon > 0$ we can find $n^{(6)}$ such that for all $n > n^{(6)}$ we have $\mathbb{E}\{Pr[E_6]\} < \frac{\epsilon}{6}$ and such that $n > n^{(6)}$ is sufficiently large to ensure the inequalities used in this part of the proof. Therefore the probability of error for the sixth error event can be bounded by $\mathbb{E}\{Pr[E_6]\} < \frac{\epsilon}{6}$ for $n \geq n^{(6)}$.

3.1.2.6 The Case $R_2 = I(X_2; \hat{Y}_R|X_1, Q) = 0$

The case $I(X_2; \hat{Y}_R|X_1, Q) = 0$ needs a special treatment as it is not captured by the above arguments which assume strict inequality in (3.1). In this case we can simply set the number of messages $M_2 = 1$ and the error probability of the receiver 1 is 0 by definition. In the calculation of the error probability of receiver 2 neither R_2 nor $I(X_2; \hat{Y}_R|X_1, Q)$ appear but in the definition of ϵ_q. In this case the definition can be changed by removing the requirement $\epsilon_q < \frac{I(X_2;\hat{Y}_R|X_1,Q)-\frac{R_2}{\alpha}}{3}$ as this requirement is only needed to ensure the low error probability of receiver 1 and is therefore not necessary for this case. The changed code and the above steps of the proof for receiver 2 yield a sequence of $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-codes. This sequence of codes has the property, that its probability of error goes to zero and the rate of the codes goes to $[R_1, R_2]$ as $n \to \infty$ given

$$\begin{aligned} R_1 &< \alpha I(X_1; \hat{Y}_R | X_2, Q) \\ R_2 &= \alpha I(X_2; \hat{Y}_R | X_1, Q) = 0 \end{aligned}$$

and

$$\alpha\left(H(\hat{Y}_R|X_1, Q) - H(\hat{Y}_R|Y_R)\right) < \beta I(Y_1; X_R|Q)$$
$$\alpha\left(H(\hat{Y}_R|X_2, Q) - H(\hat{Y}_R|Y_R)\right) < \beta I(Y_2; X_R|Q)$$

for the probability distributions $p(q)p(x_1|q)p(x_2|q)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$, $p(x_R|q)p_2(y_1, y_2|x_R)$ and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.

Analogous arguments apply for $R_1 = I(X_1; \hat{Y}_R|X_2, Q) = 0$. The achievability of $[R_1, R_2] = [0, 0]$ is obvious from the definition.

3.1.2.7 The Achievable Set is Closed

The above proves that any $[R_1, R_2]$ with

$$R_1 < \alpha I(X_1; \hat{Y}_R|X_2, Q)$$
$$R_2 < \alpha I(X_2; \hat{Y}_R|X_1, Q)$$

is achievable as long as the constraints (3.2) are fulfilled. Left is the proof that the achievable rate region is closed. This follows from the definition of achievability: Given a rate pair $[R_1, R_2]$ with $R_1 > 0$, $R_2 > 0$ on the boundary of $\mathcal{R}_1$, then for any rate pair $[R_1 - \frac{\epsilon}{m}, R_2 - \frac{\epsilon}{m}]$, $\epsilon > 0$, $m \in \mathbb{N}$ there exists a sequence of $\left(2^{\lfloor n(R_1 - \frac{\epsilon}{m})\rfloor}, 2^{\lfloor n(R_2 - \frac{\epsilon}{m})\rfloor}, n_1, n_2\right)$-codes such that $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. Therefore for any m there exists $n_0(m)$ such that $\mu_k^{(n)} < \frac{1}{m}$, $k \in \{1, 2\}$ for $n > n_0(m)$. Let $m^{(n)} = \max\{m : n > n_0(m)\}$. Because $\mu_k^{(n)} \to 0$ as $n \to \infty$ we have $m^{(n)} \to \infty$. So we can construct a sequence of $\left(2^{\lfloor n(R_1 - \frac{\epsilon}{m^{(n)}})\rfloor}, 2^{\lfloor n(R_2 - \frac{\epsilon}{m^{(n)}})\rfloor}, n_1, n_2\right)$-codes with $\frac{1}{n}\lfloor n(R_1 - \frac{\epsilon}{m^{(n)}})\rfloor \to R_1$, $\frac{1}{n}\lfloor n(R_2 - \frac{\epsilon}{m^{(n)}})\rfloor \to R_2$, $\mu_k^{(n)} < \frac{1}{m} \to 0$, $i \in \{1, 2\}$ as $n \to \infty$. Therefore by the definition of achievability the rate pair $[R_1, R_2]$ is achievable. An analogous argument applies for rate pairs on the boundary where one of the rates is 0. It follows that the set of achievable rates is closed. This completes the proof. □

3.1.3 Boundary effects: Rate Region for the One-Way Case

The treatment of the case $R_2 = I(X_2; \hat{Y}_R|X_1, Q) = 0$ at the end of the proof shows that the calculation of the error probability for the two receivers separates, even though the coding does not. In the proof above we changed only the allowed range of ϵ_q in the code such that the claimed region is achievable. But in fact we can choose ϵ_q even more freely in the case that one of the rates is 0. Furthermore the constraint in (3.2) corresponding to the user which has no message to transmit is not needed anymore. Doing so yields a region which is similar to the region of the one-way compress-and-forward relay channel, but with half-duplex nodes and without a direct channel. Furthermore the node which does not need to convey any information now helps the transmission of the other node.

Corollary 3.2. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_2 \subset \mathbb{R}_+^2$ of all rate pairs $[0, R_2]$ satisfying

$$R_2 \leq \alpha I(X_2; \hat{Y}_R | X_1, Q)$$

under the constraint

$$\alpha \left(H(\hat{Y}_R | X_1, Q) - H(\hat{Y}_R | Y_R) \right) < \beta I(Y_1; X_R)$$

and similarly the set $\mathcal{R}_3 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, 0]$ which satisfy

$$R_1 \leq \alpha I(X_1; \hat{Y}_R | X_2, Q)$$

under the constraint

$$\alpha \left(H(\hat{Y}_R | X_2, Q) - H(\hat{Y}_R | Y_R) \right) < \beta I(Y_2; X_R)$$

for some joint probability distributions $p(q)$ $p(x_1|q)$ $p(x_2|q)$ $p_1(y_R|x_1, x_2)$ $p(\hat{y}_R|y_R)$ and $p(x_R)$ $p_2(y_1, y_2|x_R)$ and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.

Proof. The rate pairs claimed in the corollary where one of the rates is set to 0 follow directly from the proof of Theorem 3.1 by observing, that the terminal with rate 0 does not need to transmit any information. Therefore, its only purpose is to help the transmission of the other terminal. For the following treatment we assume $R_1 = 0$. The proof for $R_2 = 0$ is analogous. From the proof of Theorem 3.1 it follows immediately, that — as by the simple decoder $\forall k \in \mathcal{W}_2$, $y_1^{n_2} \in \mathcal{Y}_1^{n_2}$ $g_1(y_1^{n_2}, k) = 1$, we have $\mu_1^{(n)} = 0$ — we do only need to restrict ϵ_q in the code by $\epsilon_q \in (0, \min\{\epsilon_q^{(1)}, \epsilon_q^{(3)}\})$. Furthermore, for the same reason there is no need for the constraint $\alpha(H(\hat{Y}_R|X_2, Q) - H(\hat{Y}_R|Y_R)) < \beta I(Y_2; X_R)$. □

Now we can join the three regions to a new achievable rate region by convex combination of rate pairs from the three regions.

Corollary 3.3. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_{\mathrm{CF}} \subset \mathbb{R}_+^2$ given by the convex hull of $\mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$.

Remark 3.6 (Timesharing of codes)**.** The code construction in Theorem 3.1 and Corollary 3.2 is very similar. It seems, that one could include the convex hull operation in some time sharing variable. But in fact we already use a timesharing variable in Theorem 3.1, which is explicitly not used to timeshare between codebooks, as this would lead to more constraints compared to the stated region. In contrast to these arguments, additional constraints are now accepted, with the benefit that for the corresponding sub-codebooks yielding the new constraints a less restricted set of probability distributions is allowed. The decoding of the resulting code can be seen as decoding up to three codes interleaved with each other separately, each having its own constraints. The additional timesharing over different codes could be captured by an additional timesharing variable, but this would lead to much more complicated expressions to cope with the constraints active in the different timesharing phases: The code resulting from the timesharing operation can be interpreted as using up to three heavily constraint dependent random variables $\hat{Y}_R^{(1)}$, $\hat{Y}_R^{(2)}$, and $\hat{Y}_R^{(1,2)}$ for the quantized representation of the MAC output. $\hat{Y}_R^{(1)}$ is transmitted only to receiver 1, $\hat{Y}_R^{(2)}$ is transmitted only to receiver 2, and $\hat{Y}_R^{(1,2)}$ is transmitted to both the receivers; the BC transmission is performed separately for these three variables.

As these assumptions are rather restrictive, this leads to the conjecture that a more general approach with more than one variable $\hat{Y}_R$ could lead to a larger region for some channels. For the general case of such a setup, the BC phase becomes much more difficult. As the three variables all depend on the MAC output and to facilitate the side-information in the BC, the variables will be correlated. The resulting coding problem is similar to the problem of transmitting correlated data over a BC where the receivers have some correlated side-information. Without side-information the problem of transmitting correlated data over a BC is treated by Han et al. [63], giving an achievable rate region. To the best of our knowledge, the problem was not yet considered with additional correlated side information at the receiver. Chapter 5 focuses on extending the rate region by using approaches as the one discussed in this remark.

3.1.4 A Note on Coding Mechanisms

In contrast to the encoding for the decode-and-forward system, the relay now has no knowledge about any of the two messages. Therefore the code used at the relay to encode the data has no Cartesian structure. This is displayed in Figure 3.1

Still, the known messages can be used to restrict the number of codewords for the receiver. As the MAC output depends on the transmitted message, so does the quantized representative of the MAC output. Therefore the message rules out some of the indices i that will not occur if this message was transmitted. As a consequence, each message — say w_1 — determines a sub-code $C(w_1)$ of the relay's code, which is the effective code used in the decoding for receiver 1. This sub-code consists of less then $2^{\beta n I(X_R;Y_1)}$ codewords and therefore can be decoded at the terminal node. Note, that different sub-codes $C(w_1)$ and $C(\hat{w}_1)$ need not be disjoint. Furthermore, the two codes $C(w_1)$ and $C(w_2)$ for a message pair (w_1, w_2) may have more then one codeword in common. This is a consequence of the missing Cartesian structure. As a consequence of the fact that the relay cannot decode it follows that the same codeword could be sent by the relay for the messages pairs (w_1, w_2) and $(\hat{w}_1, \hat{w}_2)$ as long as $\hat{w}_1 \neq w_1$ and $\hat{w}_2 \neq w_2$. If this were not true for all message pairs, the relay would be able to decode. At the receivers, the different known messages will change the effective code used in the decoding. Therefore the result of

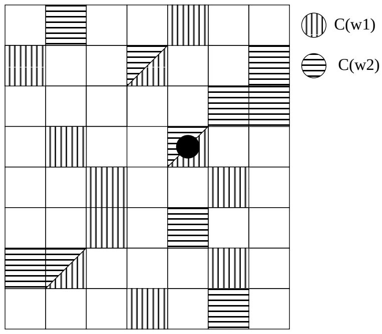

Figure 3.1: Coding mechanisms in the BC phase: The relay's code has no Cartesian structure in the compress-and-forward system. The transmitted codeword for the quantization index i is indicated by $\bullet$. Both receivers use a sub-code of the relay's code for the decoding. These sub-codes depend on the messages w_1 and w_2 known at the receivers. In the figure the sub-codes for the some possible pair of side information are marked with vertical and horizontal lines. The transmitted codeword belongs to the sub-codes of both the messages. Note, that the same codeword could be send by the relay for messages $\hat{w}_1 \neq w_1$ and $\hat{w}_2 \neq w_2$. In this case both the effective codes used by the receivers will be different and therefore the decoding will result in a different decoded message even though the sent codeword is the same.

the decoding will be different, even though the transmitted codeword is the same in both cases.

The restriction of the relay's code to the sub-code $C(w_1)$ stems from the stochastic dependency of the MAC output from the transmitted codeword $x_1^{n_1}(w_1)$: Only a subset of the possible MAC outputs in $\mathcal{Y}_R^{n_1}$ is jointly typical with the transmitted codeword $x_1^{n_1}(w_1)$. As the mapping $f(\cdot)$ used at the relay for the compression is fixed, only a subset of the $2^{\lceil \alpha n R_Q \rceil}$ quantized MAC outputs in $C_{\hat{y}_R}^{(n)}(q^{n_1})$ will occur at the relay. This subset in turn corresponds to a subset of codewords $x_R^{n_2}$ which form the effective code $C(w_1)$ for the receiver 1 in the BC phase. As the subsets are determined solely by the known message and the statistical properties of the channels, the codes corresponding to the different side information can be calculated offline. This reduces the complexity for the decoding at the terminal nodes.

For a practical code design one needs to control the statistical dependency of the MAC and the quantization at the relay, i.e. one has to control the subset of indices i that can occur in the quantization if w_1 is transmitted. The corresponding codewords $x_R^{n_2}(i)$ should form a good code for the channel $p(y_1|x_R)$. This needs to be true in average for all messages w_1. Furthermore, in average an analogous requirement needs to be fulfilled for all messages w_2 and the channel $p(y_2|x_R)$. A simple but potentially suboptimal code design could consider the BC code independent of the quantization. In this case the BC code should have the property that every subset of size $2^{\beta n I(X_R;Y_1)}$ needs to be a good code for the channel to receiver 1, while every subset of the relay's code of size $2^{\beta n I(X_R;Y_2)}$ needs to be a good code for the channel to receiver 2. Furthermore, the quantization has to ensure that for each w_1 (and w_2 respectively) one of at most $2^{\beta n I(X_R;Y_1)}$ (and $2^{\beta n I(X_R;Y_2)}$ respectively) quantization indices will occur with high probability.

3.2 Bounding Auxiliary Variables

We can bound the required cardinality of $\mathcal{Q}$ and $\hat{\mathcal{Y}}$ in Theorem 3.1 by using the Fenchel-Bunts extension of Caratheodorys theorem [62], which says:

Theorem 3.4 (Fenchel-Bunts extension of Caratheodorys theorem [62])**.** If $\mathcal{S} \subset \mathbb{R}^n$ has no more than n connected components (in particular, if $\mathcal{S}$ is connected), then any $x \in \text{ConvexHull}(\mathcal{S})$ can be expressed as a convex combination of n elements of $\mathcal{S}$.

We can make use of this theorem by interpreting the auxiliary variables as factors in a convex combination in an appropriate defined space. For the variable Q this can be done in a simple way by interpreting the constraints (3.2) as two coordinates in a four dimensional space — in addition to the two rates (3.1). For the variable $\hat{Y}$ we need a slightly less intuitive space, such that implicit constraints due to the mixed appearance of conditioned and non-conditioned probability distributions in the formulas are fulfilled by the result of the convex combination. This is done along the lines of the bounding of the auxiliary variables in many information theoretical results. An example can be found in the appendix of [33].

3.2.1 The Cardinality of Q in Theorem 3.1

We start by bounding the required cardinality of $\mathcal{Q}$. The achievable rate region $\mathcal{R}_1$ can be rewritten as the set of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &\leq \alpha \sum_{q\in\mathcal{Q}} p(q) I(X_1; \hat{Y}_R | X_2, Q = q) \\ R_2 &\leq \alpha \sum_{q\in\mathcal{Q}} p(q) I(X_2; \hat{Y}_R | X_1, Q = q) \end{aligned} \tag{3.4}$$

with the constraint

$$\begin{aligned} 0 &< \sum_{q\in\mathcal{Q}} p(q) \Big(\beta I(Y_1; X_R | Q = q) - \alpha \Big(H(\hat{Y}_R | X_1, Q = q) - H(\hat{Y}_R | Y_R, Q = q)\Big)\Big) \\ 0 &< \sum_{q\in\mathcal{Q}} p(q) \Big(\beta I(Y_2; X_R | Q = q) - \alpha \Big(H(\hat{Y}_R | X_2, Q = q) - H(\hat{Y}_R | Y_R, Q = q)\Big)\Big) \end{aligned} \tag{3.5}$$

for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and some joint probability distributions $p(q)p(x_1|q)p(x_2|q)$ $p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(x_R)p_2(y_1, y_2|x_R)$. It is save to add Q as a condition to all terms in (3.5) due to the constraints on the probability distribution; this additional conditioning will not change the value of the entropy and mutual information expressions. The above formulation gives rise to an interpretation as a convex combination in a four dimensional space.

Define for given channels $p_1(y_R|x_1,x_2)$, $p_2(y_1,y_2|x_R)$ and variables α and β

$$\mathcal{S} = \bigcup_{p(x_1)p(x_2)p(\hat{y}_R|y_R)p(x_R)} \Big\{[\delta_1(p), \delta_2(p), \delta_3(p), \delta_4(p)] \\ |p = p(x_1)p(x_2)p_1(y_R|x_1,x_2)p(\hat{y}_R|y_R)p(x_R)p_2(y_1,y_2|x_R)\Big\},$$

where the union is over the compact set of all possible $p(x_1)p(x_2)p(\hat{y}_R|y_R)p(x_R)$ and where

$$\begin{aligned}
\delta_1(p) &= I(X_1;\hat{Y}_R|X_2)\\
\delta_2(p) &= I(X_2;\hat{Y}_R|X_1)\\
\delta_3(p) &= \beta I(Y_1;X_R) - \alpha\left(H(\hat{Y}_R|X_1) - H(\hat{Y}_R|Y_R)\right)\\
\delta_4(p) &= \beta I(Y_2;X_R) - \alpha\left(H(\hat{Y}_R|X_2) - H(\hat{Y}_R|Y_R)\right).
\end{aligned}$$

Furthermore let $C = \text{ConvexHull}(\mathcal{S})$ and let $\bar{C} = \{[\delta_1,\delta_2,\delta_3,\delta_4] \in C : \delta_3 > 0, \delta_4 > 0\}$.

Now, the achievable rate region can be stated as

$$\mathcal{R}_1 = \left\{[R_1,R_2] \in \mathbb{R}_+^2 : \exists[\delta_1,\delta_2,\delta_3,\delta_4] \in \bar{C} \text{ with } \delta_1 \geq R_1, \delta_2 \geq R_2\right\}.$$

The set $\mathcal{S}$ is connected, as it is the continuous image of a continuous compact set. Therefore all points in C can be expressed as a convex combination of at most $\dim\{\mathcal{S}\} = 4$ elements of $\mathcal{S}$. Furthermore, as $\bar{C} \subset C$ all points in $\bar{C}$ can be expressed as a convex combination of at most $\dim\{\mathcal{S}\} = 4$ elements of $\mathcal{S}$. Therefore we can bound the required cardinality of $\mathcal{Q}$ from above by 4 by noting that the weighted sums in (3.4) and (3.5) can be interpreted as a convex combination of $|\mathcal{Q}|$ elements of $\mathcal{S}$.

Applying the above arguments to Corollary 3.2 it follows immediately, that in this case a cardinality $|\mathcal{Q}| \leq 2$ is sufficient to achieve all points in any of the two regions $\mathcal{R}_2$ and $\mathcal{R}_3$.

3.2.2 The Cardinality of $\hat{\mathcal{Y}}$ in Theorem 3.1

The achievable rate region $\mathcal{R}_1$ can be rewritten as the set of all rate pairs $[R_1,R_2]$ satisfying

$$\begin{aligned}
R_1 &\leq \alpha \sum_{\hat{y}_R \in \hat{\mathcal{Y}}_R} p(\hat{y}_R)\left(H(X_1|X_2,Q) - H(X_1|X_2,Q,\hat{Y}_R = \hat{y}_R)\right)\\
R_2 &\leq \alpha \sum_{\hat{y}_R \in \hat{\mathcal{Y}}_R} p(\hat{y}_R)\left(H(X_2|X_1,Q) - H(X_2|X_1,Q,\hat{Y}_R = \hat{y}_R)\right)
\end{aligned} \tag{3.6}$$

with

$$\begin{aligned}
0 &< \sum_{\hat{y}_R \in \hat{\mathcal{Y}}_R} p(\hat{y}_R)\left(\beta I(Y_1;X_R|Q) - \alpha\left(H(Y_R|X_1,Q) - H(Y_R|X_1,Q,\hat{Y}_R = \hat{y}_R)\right)\right)\\
0 &< \sum_{\hat{y}_R \in \hat{\mathcal{Y}}_R} p(\hat{y}_R)\left(\beta I(Y_2;X_R|Q) - \alpha\left(H(Y_R|X_2,Q) - H(Y_R|X_2,Q,\hat{Y}_R = \hat{y}_R)\right)\right)
\end{aligned} \tag{3.7}$$

for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and for some joint probability distributions $p(q)p(x_1|q)p(x_2|q)$ $p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(x_R)p_2(y_1, y_2|x_R)$.

Note that this is not a simple convex combination in a four dimensional space. In contrast to the bounding of the cardinality of Q we cannot add $\hat{Y}_R$ as a further condition to the other entropy expressions without changing their values. Therefore in this case, we have to cope with the problem of marginal versus conditioned probability distributions, i.e. we need to ensure that

$$\sum_{\hat{y}_R \in \hat{\mathcal{Y}}_R} p(\hat{y}_R)p(y_R|\hat{y}_R) = p(y_R). \tag{3.8}$$

This condition is sufficient, because all other random variables are independent of $\hat{Y}_R$ given Y.

For the following derivation we need an auxiliary probability vector, to build up a space, such that the convex combination works as needed. This auxiliary probability vector plays the role of a wild card for the probability distribution conditioned on $\hat{y}_R$, i.e. a wild card for $p(y_R|\hat{y}_R)$. Furthermore the auxiliary probability vector adds some dimensions to the space, in which the convex combination takes place, as by doing so we can ensure that the marginal distribution fits to the conditioned distribution, i.e. (3.8) is fulfilled.

We start with the needed definition and specify the notation:

Definition 3.1. For $n \in \mathbb{N}$, let Δ_n be the simplex of probability n-vectors.

Now, let $s_1 \in \Delta_{|\mathcal{Y}_R|}$ be a wild card distribution for $p(Y_R|\hat{Y}_R = \hat{y}_R)$. We address the elements of the probability vector by a $(\cdot)$-notation, i.e. $s_1(i)$ is the i^{th} element in s_1.

For given channels $p_1(y_R|x_1, x_2)$, $p_2(y_1, y_2|x_R)$ and fixed $p = p(q)p(x_1|q)p(x_2|q)p(x_R)$ let the set $\mathcal{S}(p)$ be given by

$$\mathcal{S}(p) = \bigcup_{s_1 \in \Delta_{|\mathcal{Y}_R|}} \{[\delta_1, \delta_2, \delta_3, \delta_4, s_1]\}$$

where the union is over all $s_1 \in \Delta_{|\mathcal{Y}_R|}$ and we have

$$\delta_1 = \alpha\Big(H(X_1|X_2, Q) + \sum_{i,x_1,x_2,q} s_1(i)p(x_1, x_2, q|Y_R = i)$$
$$\times \Big(\log\Big(\sum_j s_1(j)p(x_2, q|Y_R = j)\Big) - \log\Big(\sum_j s_1(j)p(x_1, x_2, q|Y_R = j)\Big)\Big)\Big)$$

$$\delta_2 = \alpha\Big(H(X_2|X_1, Q) + \sum_{i,x_1,x_2,q} s_1(i)p(x_1, x_2, q|Y_R = i)$$
$$\times \Big(\log\Big(\sum_j s_1(j)p(x_1, q|Y_R = j)\Big) - \log\Big(\sum_j s_1(j)p(x_1, x_2, q|Y_R = j)\Big)\Big)\Big)$$

$$\delta_3 = \beta I(Y_1; X_R) - \alpha\Big(H(Y_R|X_1, Q) - \sum_{i,x_1,q} s_1(i)p(x_1, q|Y_R = i) \\ \times \Big(\log\Big(\sum_j s_1(j)p(x_1, q|Y_R = j)\Big) - \log\Big(s_1(i)p(x_1, q|Y_R = i)\Big)\Big)\Big)$$

$$\delta_4 = \beta I(Y_2; X_R) - \alpha\Big(H(Y_R|X_1, Q) - \sum_{i,x_1,q} s_1(i)p(x_2, q|Y_R = i) \\ \times \Big(\log\Big(\sum_j s_1(j)p(x_2, q|Y_R = j)\Big) - \log\Big(s_1(i)p(x_2, q|Y_R = i)\Big)\Big)\Big).$$

Here we use the common convention $0 \log 0 = 0$ justified by continuity since $x \log x \to 0$ as $x \to 0$.

Now, let $C(p) = \text{ConvexHull}(\mathcal{S}(p))$. Furthermore let

$$\bar{C}(p) = \{[\delta_1, \delta_2, \delta_3, \delta_4, s_1] \in C(p) : \delta_3 > 0, \delta_4 > 0, \forall i \ s_1(i) = p(Y_R = i)\}.$$

Now, the achievable rate region can be stated as

$$\mathcal{R}_1 = \bigcup_p \{[R_1, R_2] : \exists[\delta_1, \delta_2, \delta_3, \delta_4, s_1] \in \bar{C}(p) \text{ with } \delta_1 \geq R_1, \delta_2 \geq R_2\}.$$

The set $\mathcal{S}(p)$ is connected, as it is the continuous image of the continuous compact set $\Delta_{|\mathcal{Y}_R|}$. Therefore all points in $C(p)$ can be expressed as a convex combination of at most $\dim\{\mathcal{S}(p)\} = |\mathcal{Y}_R| + 3$ elements of $\mathcal{S}(p)$. As $\bar{C}(p) \subset C(p)$ all points in $\bar{C}(p)$ can be expressed as a convex combination of at most $\dim\{\mathcal{S}(p)\} = |\mathcal{Y}_R| + 3$ elements of $\mathcal{S}(p)$. Therefore all points in the achievable rate region can be achieved with $|\hat{\mathcal{Y}}_R| \leq |\mathcal{Y}_R| + 3$.

Applying the above arguments to Corollary 3.2 it follows immediately, that in this case a cardinality $|\hat{\mathcal{Y}}_R| \leq |\mathcal{Y}_R| + 1$ is sufficient to achieve all points in any of the two regions $\mathcal{R}_2$ and $\mathcal{R}_3$.

3.3 A Partial-Decode-and-Forward Coding Theorem

3.3.1 Coding Theorem

The region $\mathcal{R}_{\text{CF}}$ of Corollary 3.3 and the region $\mathcal{R}_{\text{DF}}$ stated in Theorem 2.4 might be different as shown in Example 3.1. Furthermore, neither the region $\mathcal{R}_{\text{CF}}$ need to be a superset of $\mathcal{R}_{\text{DF}}$ nor need this be true the other way around. In this section we prove an extended region, that contains both these regions as special cases as well as the convex combination of both of them. In the achievability proof we use a superposition code following the lines of [16]: The relay partially decodes the message and transmits the decoded messages to the receivers using a technique similar to the one used in the proof of Theorem 2.2. The missing information needed to decode the complete message is relayed to the receivers by a compress-and-forward mechanism similar to the one used in the proof of Theorem 3.1.

Theorem 3.5. An achievable rate region for the two-phase two-way relay channel using a partial decode-and-forward protocol is the set $\mathcal{R}_4 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exist $R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)} \geq 0$ with $R_1^{(1)} + R_1^{(2)} = R_1$, $R_2^{(1)} + R_2^{(2)} = R_2$ satisfying

$$\begin{aligned} R_1^{(1)} &\leq \min\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\} \\ R_2^{(1)} &\leq \min\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\} \\ R_1^{(1)} + R_2^{(1)} &\leq \alpha I(U_1 U_2; Y_R|Q) \\ R_1^{(2)} &\leq \alpha I(X_1; \hat{Y}_R|X_2, U_1) \\ R_2^{(2)} &\leq \alpha I(X_2; \hat{Y}_R|X_1, U_2) \end{aligned} \tag{3.9}$$

under the constraints

$$\begin{aligned} \alpha\left(H(\hat{Y}_R|X_1, U_2) - H(\hat{Y}_R|Y_R)\right) &< \beta I(Y_1; X_R|V) \\ \alpha\left(H(\hat{Y}_R|X_2, U_1) - H(\hat{Y}_R|Y_R)\right) &< \beta I(Y_2; X_R|V) \end{aligned} \tag{3.10}$$

for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and for some joint probability distributions $p(q)p(u_1|q)p(u_2|q)$ $p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(v)p(x_R|v)p_2(y_1, y_2|x_R)$.

Remark 3.7 (Cardinalities of random variables)**.** To achieve any point in the stated region it is sufficient to consider only random variables Q, U_1, U_2, V, and $\hat{Y}_R$ with cardinalities restricted to $|Q| \leq 7$, $|\mathcal{U}_1| \leq |\mathcal{X}_1||Q| + 3$, $|\mathcal{U}_2| \leq |\mathcal{X}_2||Q| + 3$, $\mathcal{V} \leq |\mathcal{X}_R| + 1$, and $|\hat{\mathcal{Y}}_R| \leq |\mathcal{Y}_R| + 3$. This can be shown using the Fenchel-Bunt extension of Caratheodory's theorem [62]. A proof of this claim is given in the appendix.

Remark 3.8 (Relaxing the constraints for $I(Y_k; X_R|V) = 0$)**.** We have $H(\hat{Y}_R|X_1, U_2) - H(\hat{Y}_R|Y_R) = I(Y_R, \hat{Y}_R|X_1, U_2)$ due to the Markov chain constraint. Furthermore we have $\beta I(Y_1; X_R|V) > \alpha I(Y_R, \hat{Y}_R|X_1, U_2) \geq \alpha I(X_2; \hat{Y}_R|X_1, U_2) \geq R_2^{(2)}$. As in the proof of the theorem the calculation of the probability of error for the two receiver separates for the compress-and-forward part, this leads to the conclusion, that in case of $I(Y_1; X_R|V) = 0$ the corresponding strict inequality in (3.10) can be relaxed to a non-strict inequality. A similar argument will be used to strengthen the above result in Corollary 3.6 below.

Remark 3.9 (Decode-and-forward and compress-and-forward as special cases)**.** The result contains the results stated in the Theorems 2.4 and 3.1 as special cases. The region $\mathcal{R}_{\mathrm{DF}}$ is obtained by choosing $|\hat{\mathcal{Y}}_R| = 1$ $U_1 = X_1$, $U_2 = X_2$, and $V = X_R$, and noting that in this case the strict inequality constraints (3.10) can be relaxed; we obtain the region $\mathcal{R}_1$ by choosing $U_1 = U_2 = Q$ and $|\mathcal{V}| = 1$.

Remark 3.10 (Interpretation of auxiliary variables)**.** The idea behind the coding theorem is to build a hybrid scheme superimposing both approaches, decode-and-forward and compress-and-forward as it was done in [16] for the classical relay channel. To get some meaning to the variables and the rates in the theorem, one may consider $R_1^{(1)}, R_2^{(1)}$ as rates, which are achieved by decoding at the relay. For this part of the coding scheme U_1 and U_2 play the role of the channel input of a (virtual) MAC with $p(y_R|u_1, u_2)$, while V is the channel input of a (virtual) BC with $p(y_1, y_2|v)$, both induced by the real channels and the random coding. On the other hand $R_1^{(2)}, R_2^{(2)}$ are rates, that are achieved by forwarding a compressed version of the channel output at the relay. Since we may assume that we already decoded the decode-and-forward part of the message, the compress-and-forward operations can use this information as side information at the relay and at the receivers.

Remark 3.11 (Convexity of $\mathcal{R}_4$)**.** The region $\mathcal{R}_4$ is convex. To see that it is convex for fixed α and β note, that one can add Q as a condition to all entropy and mutual information terms without changing the region. If we allow for different timesharing parameters α and β, then we can use arguments analogous to that in Remark 2.11 to prove, that the region is convex.

3.3.2 Proof of the Coding Theorem

Proof. Suppose we have strict inequalities in (3.9) for the probability distributions[3] $p(q)p(u_1|q)$ $p(u_2|q)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$, $p(v)p(x_R|v)p_2(y_1, y_2|x_R)$, some $\alpha, \beta > 0$ with $\alpha + \beta = 1$, and a rate pair $[R_1, R_2]$ with $R_1^{(1)} + R_1^{(2)} = R_1$, $R_2^{(1)} + R_2^{(2)} = R_2$. We will first show how to construct a $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-code for a fixed n such that for the sequence of these $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-codes the probability of error goes to zero and the rate of the codes goes to $[R_1, R_2]$ as $n \to \infty$. Let $M_1^{(n)} = 2^{\lfloor nR_1^{(1)} \rfloor + \lfloor nR_1^{(2)} \rfloor}$ and $M_2^{(n)} = 2^{\lfloor nR_2^{(1)} \rfloor + \lfloor nR_2^{(2)} \rfloor}$ with $R_1 = R_1^{(1)} + R_1^{(2)}$, $R_2 = R_2^{(1)} + R_2^{(2)}$.

3.3.2.1 Random Codebook Generation:

For a given n set $n_1 = \lfloor \alpha n \rfloor$, $n_2 = \lceil \beta n \rceil$.

- Label the messages $w_1 \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(1)} \rfloor + \lfloor nR_1^{(2)} \rfloor}\}$ as $w_1(w_1^{(1)}, w_1^{(2)})$, $w_1^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(1)} \rfloor}\}$, $w_1^{(2)} \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(2)} \rfloor}\}$.
- Label the messages $w_2 \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(1)} \rfloor + \lfloor nR_2^{(2)} \rfloor}\}$ as $w_2(w_2^{(1)}, w_2^{(2)})$, $w_2^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(1)} \rfloor}\}$, $w_2^{(2)} \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(2)} \rfloor}\}$.
- Choose one q^{n_1} drawn according to the probability $\prod_{s=1}^{n_1} p(q_{(s)}^{n_1})$.
- Choose $2^{\lfloor nR_1^{(1)} \rfloor}$ i.i.d. codewords $u_1^{n_1}$ each drawn according to the probability distribution $\prod_{s=1}^{n_1} p(u_{1,(s)}^{n_1}|q_{(s)})$. Label these $u_1^{n_1}(w_1^{(1)})$, $w_1^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(1)} \rfloor}\}$.
- For each $w_1^{(1)}$ choose $2^{\lfloor nR_1^{(2)} \rfloor}$ i.i.d. codewords $x_1^{n_1}$ each drawn according to the probability distribution $\prod_{s=1}^{n_1} p(x_{1,(s)}^{n_1}|u_{1,(s)}^{n_1}(w_1^{(1)}))$. Label these $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $w_1^{(2)} \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(2)} \rfloor}\}$.

[3] All probabilities in the proof will be calculated for these given distributions.

- Choose $2^{\lfloor nR_2^{(1)} \rfloor}$ i.i.d. codewords $u_2^{n_1}$ each drawn according to the probability distribution $\prod_{s=1}^{n_1} p(u_{2,(s)}^{n_1}|q_{(s)}^{n_1})$. Label these $u_2^{n_1}(w_2^{(1)})$, $w_2^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(1)} \rfloor}\}$.

- For each $w_2^{(1)}$ choose $2^{\lfloor nR_2^{(2)} \rfloor}$ i.i.d. codewords $x_2^{n_1}$ each drawn according to the probability distribution $\prod_{s=1}^{n_1} p(x_{2,(s)}^{n_1}|u_{2,(s)}^{n_1}(w_2^{(1)}))$. Label these $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$, $w_2^{(2)} \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(2)} \rfloor}\}$.

- Let $\epsilon_q^{(1)} = \frac{1}{2\alpha+\beta}\Big(\beta I(X_R; Y_1|V) - \alpha\Big(H(\hat{Y}_R|X_1, U_2) - H(\hat{Y}_R|Y_R)\Big)\Big)$, $\epsilon_q^{(2)} = \frac{1}{2\alpha+\beta}\Big(\beta I(X_R; Y_2|V) - \alpha\Big(H(\hat{Y}_R|X_2, U_1) - H(\hat{Y}_R|Y_R)\Big)\Big)$, $\epsilon_q^{(3)} = \frac{I(X_2;\hat{Y}_R|X_1,U_2) - \frac{R_2^{(2)}}{\alpha}}{3}$, and $\epsilon_q^{(4)} = \frac{I(X_1;\hat{Y}_R|X_2,U_1) - \frac{R_1^{(2)}}{\alpha}}{3}$. Choose $\epsilon_q \in (0, \min\{\epsilon_q^{(1)}, \epsilon_q^{(2)}, \epsilon_q^{(3)}, \epsilon_q^{(4)}\})$.

- For each pair $(u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)}))$, $w_1^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(1)} \rfloor}\}$, $w_2^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(1)} \rfloor}\}$, draw $2^{\lceil \alpha n R_Q \rceil}$, $R_Q = I(Y_R; \hat{Y}_R|U_1, U_2, Q) + \epsilon_q$ i.i.d. codewords $\hat{y}_R$ according to the probability distribution $\prod_{s=1}^{n_1} p(\hat{y}_{R,(s)}^{n_1}|u_{1,(s)}^{n_1}(w_1^{(1)}), u_{2,(s)}^{n_1}(w_2^{(1)}))$. Label these $\hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$, $i \in \{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\}$.

- Draw i.i.d. $2^{\lfloor nR_1^{(1)} \rfloor + \lfloor nR_2^{(1)} \rfloor}$ codewords v^{n_2} according to $\prod_{s=1}^{n_2} p(v_{(s)}^{n_2})$. Label these $v^{n_2}(w_1^{(1)}, w_2^{(1)})$, $w_1^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(1)} \rfloor}\}$, $w_2^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(1)} \rfloor}\}$.

- For each $v^{n_2}(w_1^{(1)}, w_2^{(1)})$, $w_1^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_1^{(1)} \rfloor}\}$, $w_2^{(1)} \in \{1, 2, \ldots, 2^{\lfloor nR_2^{(1)} \rfloor}\}$, draw $2^{\lceil \alpha n R_Q \rceil}$ i.i.d. codewords $x_R^{n_2}$ according to $\prod_{s=1}^{n_2} p(x_{R,(s)}^{n_2}|v_{(s)}^{n_2}(w_1^{(1)}, w_2^{(1)}))$. Label these $x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)})$, $i \in \{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\}$

This constitutes a random codebook $C^{(n)}$ given by $C^{(n)} = \{q^{n_1}\} \cup C_{u_1}^{(n)}(q^{n_1}) \cup C_{u_2}^{(n)}(q^{n_1}) \cup \bigcup_{u_1^{n_1} \in C_{u_1}^{(n)}(q^{n_1})} \bigcup_{u_2^{n_1} \in C_{u_2}^{(n)}(q^{n_1})} \Big(C_{x_1}^{(n)}(u_1^{n_1}) \cup C_{x_2}^{(n)}(u_2^{n_1}) \cup C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1})\Big) \cup C_v^{(n)} \cup \bigcup_{v^{n_2} \in C_v^{(n)}} C_{x_R}^{(n)}(v^{n_2})$ where $C_{u_1}^{(n)}(q^{n_1})$ is the ordered set of codewords $u_1^{n_1}(1), \ldots u_1^{n_1}(2^{\lfloor nR_1^{(1)} \rfloor})$ drawn conditioned on a given q^{n_1}, and the ordered sets $C_{u_2}^{(n)}(q^{n_1})$, $C_{x_1}^{(n)}(u_1^{n_1})$, $C_{x_2}^{(n)}(u_2^{n_1})$, $C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1})$, $C_v^{(n)}$,and $C_{x_R}^{(n)}(v^{n_2})$ are defined accordingly for the remaining codewords.

3.3.2.2 Decoding Sets

For the decoding we will use typical set decoding. For a strict definition of the decoding sets we choose parameter for the typical sets as $\epsilon_2 = \epsilon_4 = \epsilon_5 = \epsilon_6 \in (0, \frac{\epsilon_q}{6})$, $\epsilon_1 \in \Big(0, \min\Big\{\frac{\alpha I(U_1;Y_R|U_2,Q) - R_1^{(1)}}{6}, \frac{\alpha I(U_2;Y_R|U_1,Q) - R_2^{(1)}}{6}, \frac{\alpha I(U_1U_2;Y_R|Q) - R_1^{(1)} - R_2^{(1)}}{8}\Big\}\Big)$, and $\epsilon_3 \in \Big(0, \min\Big\{\frac{\beta I(V;Y_1) - R_2^{(1)}}{3}, \frac{\beta I(V;Y_2) - R_1^{(1)}}{3}\Big\}\Big)$. The missing parameters for receiver 2 are chosen in an analogous way.

3.3.2.3 Coding

i To transmit message w_1 node 1 sends $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$.

ii To transmit message w_2 node 2 sends $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$.

iii Upon receiving $y_R^{n_1}$ the relay looks for the unique $w_1^{(1)}, w_2^{(1)}$ such that $\left(y_R^{n_1}, u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, U_1, U_2|q^{n_1})$. If no unique $w_1^{(1)}, w_2^{(1)}$ is found the relay chooses[4] $w_1^{(1)} = w_2^{(1)} = 1$.

iv Knowing $w_1^{(1)}, w_2^{(1)}$ the relay looks for the first i such that the pair $y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$ is jointly typical, i.e. $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$. If such an i is found the relay transmits $x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)})$. If no such i is found the relay chooses $i = 1$ and transmits $x_R^{n_2}(1|w_1^{(1)}, w_2^{(1)})$. This induces a mapping $f : \mathcal{Y}_R^{n_1} \to \hat{\mathcal{Y}}_R^{n_1}$ as $f(y_R^{n_1}) := \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$.

v Upon receiving $y_1^{n_2}$ and knowing its own side information $w_1^{(1)}$, node 1 looks for the unique $w_2^{(1)}$ such that $\left(y_1^{n_2}, v^{n_2}(w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_3}^{(n_2)}(Y_1, V)$. If no unique $w_2^{(1)}$ is found node 1 chooses $w_2^{(1)} = 1$.

vi Knowing the already decoded $w_2^{(1)}$ and its own side information $w_1^{(1)}$ and $w_1^{(2)}$, node 1 looks for the unique i such that $x_R^{n_2}(i|w_2^{(1)}, w_1^{(1)})$ and the received signal $y_1^{n_2}$ are jointly typical, and simultaneously the transmitted codeword $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$ and $\hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$ are jointly typical, i.e we have $\left(x_R^{n_2}(i|w_2^{(1)}, w_1^{(1)}), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_4}^{(n_2)}\left(X_R, Y_1|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)$ and simultaneously for the same i we have $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_5}^{(n_1)}\left(X_1, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)}), q^{n_1}\right)$. This enables node 1 to recover $\hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$. If no or more than one such i is found, node 1 chooses $i = 1$.

vii Knowing $\hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$, $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $u_1^{n_1}(w_1^{(1)})$, and $u_2^{n_1}(w_2^{(1)})$ receiver 1 decides that the message $w_2(w_2^{(1)}, w_2^{(2)})$ was transmitted if $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ is the only codeword jointly typical with $\hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$ and $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, i.e. $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(w_2^{(2)}|w_2^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1, X_2, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)}), q^{n_1}\right)$. If no or more than one such codeword is found node 1 chooses $w_2(w_2^{(1)}, w_2^{(2)}) = 1$.

viii The decoding at node 2 is performed in an analogous way.

3.3.2.4 Error Events

Now, we show that the average probability of error goes to zero in the average of all random codebooks, more precisely we show that for any given ϵ there exists an $n^{(0)}$ such that $\mathbb{E}\{\mu_k^{(n)}\} < \epsilon$, $k \in \{1, 2\}, n > n^{(0)}$, where the expectation is over the random codebook. This in turn implies that for each ϵ we can find $n^{(0)}$ such that $\mathbb{E}\{\mu_1^{(n)} + \mu_2^{(n)}\} < 2\epsilon$, $n > n^{(0)}$, and therefore there is at least one codebook with $\mu_1^{(n)} + \mu_2^{(n)} < 2\epsilon$, $n > n^{(0)}$, and therefore $\mu_1^{(n)} < 2\epsilon$ and $\mu_2^{(n)} < 2\epsilon$ for $n > n^{(0)}$.

We bound the average error probability $\mathbb{E}\{\mu_1^{(n)}\}$ from above by the union bound using ten events E_j, $j \in \{1, 2, \ldots, 10\}$, whose union is a superset of the error event. Therefore we have $\mathbb{E}\{\mu_1^{(n)}\} \leq \sum_{j=1}^{10} \mathbb{E}\{Pr[E_j]\}$. The average error probability $\mathbb{E}\{\mu_2^{(n)}\}$ can be bounded in an analogous way. In what follows we summarize the definition of the error events for receiver 1:

[4]This is done to have a well defined error probability. Equivalently one could declare an error at the relay, but this induces a much more cumbersome notation in the definition of the error probability. Similar arguments apply for a similar default choice in the other coding steps.

- E_1: Suppose a codebook is given and $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ are transmitted. E_1 is the event that $\left(y_R^{n_1}, u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right) \notin \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, U_1, U_2|q^{n_1})$.

- E_2: Suppose a codebook is given and $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ are transmitted. E_2 is the event that there exists a pair $(\hat{w}_1^{(1)}, \hat{w}_2^{(1)}) \neq (w_1^{(1)}, w_2^{(1)})$ such that $\left(y_R^{n_1}, u_1^{n_1}(\hat{w}_1^{(1)}), u_2^{n_1}(\hat{w}_2^{(1)})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, U_1, U_2|q^{n_1})$.

- E_3: Suppose a codebook is given, $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ are transmitted, and the messages $(w_1^{(1)}, w_2^{(1)})$ are known at the relay. E_3 is the event that $\nexists i \in \{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\}$: $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$.

- E_4: Suppose a codebook is given, $x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)})$ is transmitted, and $w_1^{(1)}$ is known at receiver 1. E_4 is the event that we have $\left(y_1^{n_2}, v^{n_2}(w_1^{(1)}, w_2^{(1)})\right) \notin \mathcal{T}_{\epsilon_3}^{(n_2)}(Y_1, V)$.

- E_5: Suppose a codebook is given, $x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)})$ is transmitted, and $w_1^{(1)}$ is known at receiver 1. E_5 is the event that we have $\left(y_1^{n_2}, v^{n_2}(w_1^{(1)}, \hat{w}_2^{(1)})\right) \in \mathcal{T}_{\epsilon_3}^{(n_2)}(Y_1, V)$ for some $\hat{w}_2^{(1)} \neq w_2^{(1)}$.

- E_6: Suppose a codebook is given, $x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)})$ is transmitted. E_6 is the event that we have $\left(x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)}), y_1^{n_2}\right) \notin \mathcal{T}_{\epsilon_4}^{(n_2)}\left(X_R, Y_1|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)$.

- E_7: Suppose a codebook is given, $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ are transmitted, the relay chooses some i with $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)}) = f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$, and the messages $(w_1^{(1)}, w_2^{(1)})$ are known at receiver 1. E_7 is the event that we have $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \notin \mathcal{T}_{\epsilon_5}^{(n_1)}\left(X_1, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)}), q^{n_1}\right)$.

- E_8: Suppose a codebook is given, the codewords $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, and $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ are transmitted, the relay chooses some i, $(w_1^{(1)}, w_2^{(1)})$ is known at receiver 1, and $x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)})$ is transmitted. E_8 is the event that $\exists j \neq i$ such that $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), \hat{y}_R^{n_1}(j|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_5}^{(n_1)}\left(X_1, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)}), q^{n_1}\right)$ and for the same j we have $\left(x_R^{n_2}(j|w_1^{(1)}, w_2^{(1)}), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_4}^{(n)}\left(X_R, Y_1|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)$.

- E_9: Suppose a codebook is given, $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ are transmitted, and the relay chooses some i with $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)}) = f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$. E_9 is the event that we have $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(w_2^{(2)}|w_2^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)}) = f(y_R^{n_1})\right) \notin \mathcal{T}_{\epsilon_6}^{(n_1)}\big(X_1, X_2, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)}), q^{n_1}\big)$.

- E_{10}: Suppose a codebook is given, $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$, $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ are transmitted, the relay chooses some i with $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)}) = f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$, and $(w_1^{(1)}, w_2^{(1)})$ is known at receiver 1. E_{10} is the event that for some $\hat{w}_2^{(2)} \neq w_2^{(2)}$ we have $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(\hat{w}_2^{(2)}|w_2^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1, X_2, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$.

To see that these error events capture all events that may lead to an error we step through the coding procedure and verify, that all possible causes of an error are captured.

The probability of error for node 1 can be bounded from above by

$$\begin{aligned}Pr[E] \leq &+Pr[E^{\mathrm{vii}} \cup \bar{E}^{\mathrm{vi}} \cup \bar{E}^{\mathrm{v}} \cup \bar{E}^{\mathrm{iv}} \cup \bar{E}^{\mathrm{iii}}] + Pr[E^{\mathrm{iv}} \cup \bar{E}^{\mathrm{iii}}] \\ &+Pr[E^{\mathrm{iii}}] + Pr[E^{\mathrm{v}} \cup \bar{E}^{\mathrm{iii}}] + Pr[E^{\mathrm{vi}} \cup \bar{E}^{\mathrm{v}} \cup \bar{E}^{\mathrm{iv}} \cup \bar{E}^{\mathrm{iii}}].\end{aligned}$$

Here E^{n} is the event that coding step n fails. A bar indicates the complementary event.

Coding step iii is a multiple access decoding of two virtually transmitted symbols $u_1^{n_1}$ and $u_2^{n_1}$. In this coding step it may turn out that $w_1^{(1)}$, $w_2^{(1)}$ is not found, either because we have $\left(y_R^{n_1}, u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right) \notin \mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R, U_1, U_2|q^{n_1})$ or because the result of the typical set decoding is not unique. This is captured by E_1 and E_2 respectively.

In coding step iv there might be no typical $\hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$ for the received $y_R^{n_1}$ given $w_1^{(1)}$, $w_2^{(1)}$. This does not yield an error immediately, but it may lead to an error in later decoding. To simplify the error calculation we treat this as an error captured by event E_3 and for the following considerations about error events we can assume that we have $\left(y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)}) = f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}\left(Y_R, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$. Similar arguments apply to E_1 and E_2. E_1, E_2 and E_3 are not intrinsic error events but is used to simplifies the definitions and the calculation of the errors that may happen in the coding steps vi and vii.

Coding step v is a coding for a virtual broadcast channel with input V and outputs Y_1, Y_2 where the receiving nodes have side information w_1 and w_2 respectively. In this coding step an error occurs if either $\left(y_1^{n_2}, v^{n_2}(w_1^{(1)}, w_2^{(1)})\right) \notin \mathcal{T}_{\epsilon_3}^{(n_2)}(Y_1, V)$ or if the decoding is not unique. This is captured by E_4 and E_5 respectively. For the following coding steps we may assume, that $w_1^{(1)}$ and $w_2^{(1)}$ are known at the receiving node.

In coding step vi the receiver cannot find the correct i, if either we have $\left(x_R^{n_2}(i|w_2^{(1)}, w_1^{(1)}), y_1^{n_2}\right) \notin \mathcal{T}_{\epsilon_4}^{(n_2)}\left(X_R, Y_1|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)$, which is captured by event E_6, or if $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \notin \mathcal{T}_{\epsilon_5}^{(n_1)}\left(X_1, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$, captured by E_7. The solution is not unique, i.e. $j \neq i$ is found in step vi, if $\exists j \neq i : \left(x_R^{n_2}(j|w_2^{(1)}, w_1^{(1)}), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_4}^{(n_2)}\left(X_R, Y_1|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)$ and simultaneously $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), \hat{y}_R^{n_1}(j|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_5}^{(n_1)}\left(X_1, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$. This is captured by E_8.

Coding step vii fails, if either the correct $w_2^{(2)}$ is not found by the typical set decoding, i.e. $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(w_2^{(2)}|w_2^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \notin \mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1, X_2, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$ or if for $\hat{w}_2^{(2)} \neq w_2^{(2)}$ we have $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(\hat{w}_2^{(2)}|w_2^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1, X_2, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$. These events are captured by E_9 and E_{10} respectively.

Clearly no other events lead to an error for the decoding process at receiver 1. We will now prove for each event E_j, $j \in \{1, 2, \ldots, 10\}$, that there exists an $n^{(j)}$ that $\mathbb{E}\{Pr[E_j]\} < \frac{\epsilon}{10}$ for $n \geq n^{(j)}$. This in turn implies that for $n \geq \max_j n^{(j)} = n^{(0)}$ we have $\mathbb{E}\{\mu_k^{(n)}\} < \epsilon$, $k \in \{1, 2\}$.

3.3.2.5 Bounding the Probability of the Error Events

We now bound the probability of the error events averaged over the codebooks $C^{(n)}$ of length n and the transmitted messages $[w_1, w_2] \in \mathcal{W}_1 \times \mathcal{W}_2$.

Error event E_1 The averaged probability for the error event E_1 is

$$\begin{aligned}\mathbb{E}\{Pr[E_1]\} &= \frac{1}{|\mathcal{W}|}\sum_{(w_1,w_2)\in\mathcal{W}}\sum_{C^{(n)}}p(C^{(n)})\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} p\left(y_R^{n_1}|x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(w_2^{(2)}|w_2^{(1)})\right)\\ &\quad\times\chi^C_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,U_1,U_2|q^{n_1})}\left(y_R^{n_1}, u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)\\ &= \sum_{q^{n_1}\in\mathcal{Q}^{n_1}}\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}\sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}}\sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} p(q^{n_1}, y_R^{n_1}, u_1^{n_1}, u_2^{n_1})\chi^C_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,U_1,U_2|q^{n_1})}\left(y_R^{n_1}, u_1^{n_1}, u_2^{n_1}\right).\end{aligned}$$

This probability can be made arbitrarily small by choosing n_1 large enough by the properties of the typical set.

Error event E_2 For sufficient large n we have

$$\begin{aligned}\mathbb{E}\{Pr[E_2]\} &\le \sum_{q^{n_1}\in\mathcal{Q}^{n_1}}\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}\sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}}\sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} p(q^{n_1})\chi_{\mathcal{T}_{\epsilon_1}^{(n_1)}(Y_R,U_1,U_2|q^{n_1})}\left(y_R^{n_1}, u_1^{n_1}, u_2^{n_1}\right)\\ &\quad\times\Big(2^{nR_2^{(1)}}p(y_R^{n_1}, u_1^{n_1}|q^{n_1})p(u_2^{n_1}|q^{n_1}) + 2^{nR_1^{(1)}}p(y_R^{n_1}, u_2^{n_1}|q^{n_1})p(u_1^{n_1}|q^{n_1})\\ &\qquad + 2^{nR_1^{(1)}}2^{nR_2^{(1)}}p(y_R^{n_1}|q^{n_1})p(u_1^{n_1}|q^{n_1})p(u_2^{n_1}|q^{n_1})\Big)\\ &\le \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1})2^{n_1(H(Y_R,U_1,U_2|Q)+2\epsilon_1)}\Big(2^{nR_2^{(1)}}2^{-n_1(H(Y_R,U_1|Q)-2\epsilon_1)}2^{-n_1(H(U_2|Q)-2\epsilon_1)}\\ &\qquad + 2^{nR_2^{(1)}}2^{-n_1(H(Y_R,U_2|Q)-2\epsilon_1)}2^{-n_1(H(U_1|Q)-2\epsilon_1)}\\ &\qquad + 2^{n(R_1^{(1)}+R_2^{(1)})}2^{-n_1(H(Y_R|Q)-2\epsilon_1)}2^{-n_1(H(U_1|Q)-2\epsilon_1)}2^{-n_1(H(U_1|Q)-2\epsilon_1)}\Big)\\ &\le 2^{n\left(R_2^{(1)}-\alpha I(U_2;U_1Y_R|Q)+6\epsilon_1\right)} + 2^{n\left(R_1^{(1)}-\alpha I(U_1;U_2Y_R|Q)+6\epsilon_1\right)} + 2^{n\left(R_1^{(1)}+R_2^{(1)}-\alpha I(U_1U_2;Y_R|Q)+8\epsilon_1\right)}.\end{aligned}$$

Now because U_1 and U_2 are independent given Q this goes to zero for sufficient large n if

$$\begin{aligned}R_1^{(1)} + 6\epsilon_1 &< \alpha I(U_1;Y_R|U_2,Q)\\ R_2^{(1)} + 6\epsilon_1 &< \alpha I(U_2;Y_R|U_1,Q)\\ R_1^{(1)} + R_2^{(1)} + 8\epsilon_1 &< \alpha I(U_1U_2;Y_R|Q)\end{aligned}$$

as assumed in the code design and by the choice of the parameter ϵ_1.

Error event E_3 The averaged probability for the error event E_3 is

$$\mathbb{E}\{Pr[E_3]\} = \frac{1}{|\mathcal{W}|}\sum_{(w_1,w_2)\in\mathcal{W}}\sum_{C^{(n)}}p(C^{(n)})\sum_{y_R^{n_1}:y_R^{n_1}\in\mathcal{J}(C^{(n)})} p\left(y_R^{n_1}|x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(w_1^{(2)}|w_1^{(1)})\right)$$

with

$$\mathcal{J}(C^{(n)}) = \Big\{ y_R^{n_1} \in \mathcal{Y}_R^{n_1} : \nexists i \in \left\{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\right\}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)}) \in C^{(n)} \text{ such that} \\ \left(y_R^{n_1}, \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R, \hat{Y}_R | u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)}), q^{n_1}\right)\Big\}.$$

This can be rewritten as

$$\mathbb{E}\{Pr[E_3]\} = \sum_{C^{(n)}} p(C^{(n)}) \sum_{y_R^{n_1} : y_R^{n_1} \in \mathcal{J}(C^{(n)})} p\left(y_R^{n_1} | x_1^{n_1}(1|1), x_1^{n_1}(1|1)\right)$$

as the codewords where drawn i.i.d.. Now, we can simplify the expression by dropping some of the indices as

$$\mathbb{E}\{Pr[E_3]\} = \sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{u_1^{n_1} \in \mathcal{U}_1^{n_1}} \sum_{u_2^{n_1} \in \mathcal{U}_2^{n_1}} \sum_{x_1^{n_1} \in \mathcal{X}_1^{n_1}} \sum_{x_2^{n_1} \in \mathcal{X}_2^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(u_1^{n_1}, u_2^{n_1}, x_1^{n_1}, x_2^{n_1}, q^{n_1}, y_R^{n_1}) \\ \sum_{C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1}) : y_R^{n_1} \in \mathcal{J}\left(C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1})\right)} p\left(C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1})\right)$$

where

$$\mathcal{J}\left(C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1})\right) = \Big\{ y_R^{n_1} \in \mathcal{Y}_R^{n_1} : \nexists i \in \left\{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\right\}, \hat{y}_R^{n_1}(i|1, 1) \in C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1}) \text{ such that} \\ \left(y_R^{n_1}, \hat{y}_R^{n_1}(i|1, 1)\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R | u_1^{n_1}, u_2^{n_1}, q^{n_1})\Big\}.$$

Now, we can eliminate $\mathcal{J}\left(C_{\hat{y}_R}^{(n)}(u_1^{n_1}, u_2^{n_1})\right)$ in the above expression by using the indicator function χ on the typical set $\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R | u_1^{n_1}, u_2^{n_1}, q^{n_1})$:

$$\mathbb{E}\{Pr[E_3]\} = \sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{u_1^{n_1} \in \mathcal{U}_1^{n_1}} \sum_{u_2^{n_1} \in \mathcal{U}_2^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1}, y_R^{n_1}) \\ \left(1 - \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1} | u_1^{n_1}, u_2^{n_1}, q^{n_1}) \chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R | u_1^{n_1}, u_2^{n_1}, q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1})\right)^{2^{\lceil \alpha n R_Q \rceil}}$$

We can bound

$$\sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1} | u_1^{n_1}, u_2^{n_1}, q^{n_1}) \chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R | u_1^{n_1}, u_2^{n_1}, q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1})$$

from below by using properties of the typical set: For $(y_R^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R | u_1^{n_1}, u_2^{n_1}, q^{n_1})$ and sufficiently large n, i.e. for some $n > n^{(3,1)}$ we have

$$\begin{aligned}
p(\hat{y}_R^{n_1}|y_R^{n_1}, u_1^{n_1}, u_2^{n_1}, q^{n_1}) &= \frac{p(\hat{y}_R^{n_1}, y_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1})}{p(y_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1})} \\
&\leq p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1}) \frac{2^{-n_1\left(H(Y_R,\hat{Y}_R|U_1,U_2,Q)-2\epsilon_2\right)}}{2^{-n_1\left(H(Y_R|U_1,U_2,Q)+2\epsilon_2\right)}2^{-n_1\left(H(\hat{Y}_R|U_1,U_2,Q)+2\epsilon_2\right)}} \\
&= p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1})2^{\lfloor \alpha n \rfloor\left(I(Y_R;\hat{Y}_R|U_1,U_2,Q)+6\epsilon_2\right)} \\
&\leq p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1})2^{\alpha n\left(I(Y_R;\hat{Y}_R|U_1,U_2,Q)+6\epsilon_2\right)}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1})\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1},q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}) \\
&\qquad \geq \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|y_R^{n_1}, u_1^{n_1}, u_2^{n_1}, q^{n_1})2^{-\alpha n\left(I(Y_R;\hat{Y}_R|U_1,U_2,Q)+6\epsilon_2\right)}\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1},q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1})
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\{Pr[E_3]\} \leq &\sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{u_1^{n_1} \in \mathcal{U}_1^{n_1}} \sum_{u_2^{n_1} \in \mathcal{U}_2^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1}, y_R^{n_1}) \\
&\left(1 - \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|y_R^{n_1}, u_1^{n_1}, u_2^{n_1}, q^{n_1})2^{-\alpha n\left(I(Y_R;\hat{Y}_R|U_1,U_2,Q)+6\epsilon_2\right)}\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1},q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1})\right)^{2^{\alpha n R_Q}}.
\end{aligned}$$

This can be bounded from above [30, Lemma 13.5.3] by

$$\begin{aligned}
\mathbb{E}\{Pr[E_3]\} \leq 1 - &\sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{u_1^{n_1} \in \mathcal{U}_1^{n_1}} \sum_{u_2^{n_1} \in \mathcal{U}_2^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1}, y_R^{n_1}) \\
&\sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1}|y_R^{n_1}, u_1^{n_1}, u_2^{n_1}, q^{n_1})\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1},q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}) \\
&\qquad + \exp\left(-2^{\alpha n\left(R_Q - I(Y_R;\hat{Y}_R|U_1,U_2,Q)-6\epsilon_2\right)}\right).
\end{aligned}$$

Now, since $R_Q = I(Y_R; \hat{Y}_R|U_1, U_2, Q) + \epsilon_q$ and $\epsilon_q > 6\epsilon_2$, the last term can be made arbitrarily small for n large. In particular for a given $\epsilon > 0$ we can find $n^{(3,2)}$ such that the last term in the sum $\exp\left(-2^{\alpha n\left(R_Q - I(Y_R;\hat{Y}_R|U_1,U_2,Q)-6\epsilon_2\right)}\right) < \frac{\epsilon}{20}$ for all $n > n^{(3,2)}$.

The remaining term

$$1 - \sum_{y_R^{n_1} \in \mathcal{Y}_R^{n_1}} \sum_{u_1^{n_1} \in \mathcal{U}_1^{n_1}} \sum_{u_2^{n_1} \in \mathcal{U}_2^{n_1}} \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1})\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1},q^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}) \quad (3.11)$$

is the probability that $(y_R^{n_1}, \hat{y}_R^{n_1}) \notin \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R|u_1^{n_1}, u_2^{n_1}, q^{n_1})$ for sequences $q^{n_1}, u_1^{n_1}, u_2^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1}$

drawn according to the joint probability distribution $p(q^{n_1}, u_1^{n_1}, u_2^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1})$. By the law of large numbers this probability goes to zero.

In particular by Lemma 1.2 for a given $\epsilon > 0$ we can find $n^{(3,3)}$ such that (3.11) is smaller than $\frac{\epsilon}{20}$ for all $n > n^{(3,3)}$. We can now choose $n^{(3)} \geq \max\{n^{(3,1)}, n^{(3,2)}, n^{(3,3)}\}$ and the probability of error for the error event E_3 can be bounded by $\mathbb{E}\{Pr[E_3]\} < \frac{\epsilon}{10}$ for $n \geq n^{(3)}$.

Error event E_4

$$\begin{aligned}\mathbb{E}\{Pr[E_4]\} &= \frac{1}{|\mathcal{W}|}\sum_{(w_1,w_2)\in\mathcal{W}}\sum_{C^{(n)}} p(C^{(n)}) \sum_{y_1^{n_2}\in\mathcal{Y}_1^{n_2}} p\left(y_1^{n_2}|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)\chi^C_{\mathcal{T}_{\epsilon_3}^{(n_2)}(V,Y_1)}(v^{n_2}(w_1^{(1)}, w_2^{(1)}), y_1^{n_2})\\ &= \sum_{v^{n_2}\in\mathcal{V}^{n_2}}\sum_{y_1^{n_2}\in\mathcal{Y}_1^{n_2}} p(v^{n_2}, y_1^{n_2})\chi^C_{\mathcal{T}_{\epsilon_3}^{(n_2)}(V,Y_1)}(v^{n_2}, y_1^{n_2})\end{aligned}$$

This is the probability that two codewords $(v^{n_2}, y_1^{n_2})$ drawn according to $p(v^{n_2}, y_1^{n_2})$ are not in $\mathcal{T}_{\epsilon_3}^{(n_2)}(V, Y_1)$. The probability for this goes to zero as $n \to \infty$ by the law of large numbers and the definition of the typical set. Therefore for any given ϵ we can find $n^{(4)}$ such that $\mathbb{E}\{Pr[E_4]\} < \frac{\epsilon}{10}$ for $n \geq n^{(4)}$.

Error event E_5 For sufficiently large n we have

$$\begin{aligned}\mathbb{E}\{Pr[E_5]\} &\leq \sum_{(v^{n_2}, y_1^{n_2})\in\mathcal{V}^{n_2}\times\mathcal{Y}_1^{n_2}} p(y_1^{n_2})p(v^{n_2})2^{\lfloor nR_2^{(1)}\rfloor}\chi_{\mathcal{T}_{\epsilon_3}^{(n_2)}(V,Y_1)}(v^{n_2}, y_1^{n_2})\\ &\leq 2^{n(R_2^{(1)}-\beta I(V,Y_1)+3\epsilon_3)+3\epsilon_3}\end{aligned}$$

by the properties of the typical set. This goes to zero as by assumption $R_2^{(1)} + 3\epsilon_3 < \beta I(V, Y_1)$.

Error event E_6

$$\begin{aligned}\mathbb{E}\{Pr[E_6]\} &= \frac{1}{|\mathcal{W}|}\sum_{(w_1,w_2)\in\mathcal{W}}\sum_{C^{(n)}} p(C^{(n)}) \sum_{y_1^{n_2}\in\mathcal{Y}_1^{n_2}} p\left(y_1^{n_2}|x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)})\right)\\ &\quad\times\chi^C_{\mathcal{T}_{\epsilon_2}^{(n_2)}(X_R,Y_1|v^{n_2}(w_1^{(1)},w_2^{(1)}))}\left(x_R^{n_2}(i|w_1^{(1)}, w_2^{(1)}), y_1^{n_2}\right)\\ &= \sum_{v^{n_2}\in\mathcal{V}^{n_2}}\sum_{x_R^{n_2}\in\mathcal{X}_R^{n_2}}\sum_{y_1^{n_2}\in\mathcal{Y}_1^{n_2}} p(v^{n_2}, x_R^{n_2}, y_1^{n_2})\chi^C_{\mathcal{T}_{\epsilon_2}^{(n_2)}(X_R,Y_1|v^{n_2})}\left(x_R^{n_2}, y_1^{n_2}\right)\end{aligned}$$

This is the probability that for codewords $(v^{n_2}, x_R^{n_2}, y_1^{n_2})$ drawn according to $p(v^{n_2}, x_R^{n_2}, y_1^{n_2})$ we have $(x_R^{n_2}, y_1^{n_2}) \notin \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, Y_1|v^{n_2})$. The probability for this goes to zero as $n \to \infty$ by the law of large numbers and the definition of the typical set. Therefore for any given ϵ we can find $n^{(6)}$ such that for $\mathbb{E}\{Pr[E_6]\} < \frac{\epsilon}{10}$ for $n \geq n^{(6)}$.

Error event E_7 This part of the proof could be done analogous to the proof for error event E_9 below. But in fact as $\epsilon_5 = \epsilon_6$ whenever E_9 does not appear, E_7 will not appear as well, because of the definition of the typical sets. Therefore we have $\mathbb{E}\{Pr[E_7]\} \leq \mathbb{E}\{Pr[E_9]\}$ for any n and we can simply set $n^{(7)} \geq n^{(9)}$ and have $\mathbb{E}\{Pr[E_7]\} < \frac{\epsilon}{10}$ for $n \geq n^{(7)}$ given that $\mathbb{E}\{Pr[E_9]\} < \frac{\epsilon}{10}$ for $n \geq n^{(9)}$.

Error event E_8

$$\mathbb{E}\{Pr[E_8]\} \leq \sum_{v^{n_2}\in\mathcal{V}^{n_2}} \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1})p(v^{n_2})Pr[E_{8,1}]Pr[E_{8,2}]2^{\lceil \alpha n R_Q \rceil}$$

Here $E_{8,1}$ is the event that given $u_1^{n_1}, u_2^{n_1}$, and q^{n_1} for two sequences $x_1^{n_1}, \hat{y}_R^{n_1}$ drawn independent of each other we have $(x_1^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_5}^{(n_1)}(X_1, \hat{Y}_R|u_1^{n_1}, u_2^{n_1}, q^{n_1})$. For the calculation of this probability $x_1^{n_1}$ and $\hat{y}_R^{n_1}$ are drawn according to $p(x_1^{n_1}|u_1^{n_1}, q^{n_1}) = p(x_1^{n_1}|u_1^{n_1}, u_2^{n_1} q^{n_1})$ and $p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{\alpha n})$ respectively to capture the averaging over the random codebooks and the known side information. $E_{8,2}$ is the event, that given v^{n_2} for two sequences $x_R^{n_2}, y_1^{n_2}$ drawn independent of each other we have $(x_R^{n_2}, y_1^{n_2}) \in \mathcal{T}_{\epsilon_4}^{(n_2)}(X_R, Y_1|v^{n_2})$. The factor $2^{\lceil \alpha n R_Q \rceil}$ accounts for the fact that we can use a union bound and the error occurs if at least one $j \neq i$ is found fulfilling the requirements.

For sufficiently large n, we have

$$\begin{aligned} Pr[E_{8,1}] = &\sum_{(x_1^{n_1}, \hat{y}_R^{n_1})\in\mathcal{X}_1^{n_1}\times\in\mathcal{Y}_R^{n_1}} p(x_1^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1})p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1}, q^{n_1})\chi_{\mathcal{T}_{\epsilon_5}^{(n_1)}(X_1,\hat{Y}_R|u_1^{n_1},u_2^{n_1},q^{n_1})}(x_1^{n_1}, \hat{y}_R^{n_1}) \\ &\leq |\mathcal{T}_{\epsilon_5}^{(n_1)}(X_1, \hat{Y}_R|u_1^{n_1}, u_2^{n_1}, q^{n_1})| 2^{-n_1\left(H(X_1|U_1,U_2,Q)-2\epsilon_5\right)} 2^{-n_1\left(H(\hat{Y}_R|U_1,U_2,Q)-2\epsilon_5\right)} \end{aligned}$$

due to the properties of the typical set. Furthermore, it follows from these properties that for sufficiently large n

$$|\mathcal{T}_{\epsilon_5}^{(n_1)}(X_1, \hat{Y}_R|u_1^{n_1}, u_2^{n_1}, q^{n_1})| \leq 2^{n_1\left(H(X_1,\hat{Y}_R|U_1,U_2,Q)+2\epsilon_5\right)}.$$

$Pr[E_{8,2}]$ can be bounded in a similar way. As a consequence, there exists $n^{(8,1)}$ such that the above properties hold for all $n > n^{(8,1)}$ and for both, $Pr[E_{8,1}]$ and $Pr[E_{8,2}]$. We have for $n > n^{(8,1)}$

$$\begin{aligned} \mathbb{E}\{Pr[E_8]\} \leq &\sum_{v^{n_2}\in\mathcal{V}^{n_2}} \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1})p(v^{n_2})2^{-n_1\left(I(X_1;\hat{Y}_R|U_1,U_2,Q)-6\epsilon_5\right)} \\ &\qquad 2^{-n_2\left(I(X_R;Y_1|V)-6\epsilon_4\right)} 2^{\alpha n R_Q+1} \\ &\leq 2^{-n\left(\alpha\left(I(X_1;\hat{Y}_R|U_1,U_2,Q)-R_Q-6\epsilon_5\right)+\beta\left(I(X_R;Y_1|V)-6\epsilon_4\right)\right)+1+I(X_1;\hat{Y}_R|U_1,U_2,Q)+6\epsilon_4} \\ &= 2^{-n\left(\alpha\left(I(X_1;\hat{Y}_R|U_1,U_2,Q)-I(Y_R;\hat{Y}_R|U_1,U_2,Q)\right)+\beta I(X_R;Y_1|V)-\tilde{\epsilon}\right)+1+I(X_1;\hat{Y}_R|U_1,U_2,Q)+6\epsilon_4} \end{aligned}$$

with

$$\tilde{\epsilon} = \alpha\epsilon_q + \beta 6\epsilon_4 + \alpha 6\epsilon_5.$$

This upper bound goes to zero if

$$\alpha\left(I(X_1;\hat{Y}_R|U_1,U_2,Q)-I(Y_R;\hat{Y}_R|U_1,U_2,Q)\right)+\beta I(X_R;Y_1|V)-\tilde{\epsilon}=$$
$$\beta I(X_R;Y_1|V)-\alpha\left(H(\hat{Y}_R|X_1,U_2)-H(\hat{Y}_R|Y_R)\right)-\tilde{\epsilon}>0.$$

Now

$$\beta I(X_R;Y_1|V)-\alpha\left(H(\hat{Y}_R|X_1,U_2)-H(\hat{Y}_R|Y_R)\right)>0$$

because of the constraints (3.10) fulfilled by assumption, and

$$\tilde{\epsilon}<\beta I(X_R;Y_1)-\alpha\left(H(\hat{Y}_R|X_1,U_2)-H(\hat{Y}_R|Y_R)\right)$$

due to the choice of the parameters ϵ_q, ϵ_4, and ϵ_5. Therefore for any given ϵ we can find an $n^{(8)}>n^{(8,1)}$ such that for $\mathbb{E}\{Pr[E_8]\}<\frac{\epsilon}{10}$ for $n\geq n^{(8)}$.

Error event E_9

$$\mathbb{E}\{Pr[E_9]\}\leq\frac{1}{|\mathcal{W}|}\sum_{(w_1,w_2)\in\mathcal{W}}\sum_{C^{(n)}}p(C^{(n)})\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}p\left(y_R^{n_1}|x_1^{n_1}(w_1^{(2)}|w_1^{(1)}),x_2^{n_1}(w_2^{(2)}|w_2^{(1)})\right)$$
$$\times\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R,\hat{Y}_R|u_1^{n_1}(w_1^{(1)}),u_2^{n_1}(w_2^{(1)})\right)}\left(y_R^{n_1},f(y_R^{n_1})\right)\chi^C_{\mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1,X_2,\hat{Y}_R|u_1^{n_1}(w_1^{(1)}),u_2^{n_1}(w_2^{(1)})\right)}\left(x_1^{n_1}(w_1),x_2^{n_1}(w_2),f(y_R^{n_1})\right)$$

where $f(y_R^{n_1})=\hat{y}_R^{n_1}(i|w_1^{(1)},w_2^{(1)})$ is the mapping induced by the relay when choosing i upon receiving $y_R^{n_1}$ and knowing $(w_1^{(1)},w_2^{(1)})$. This can be rewritten as

$$\mathbb{E}\{Pr[E_9]\}=\sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}}\sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}}\sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}}\sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}}\sum_{q^{n_1}\in\mathcal{Q}^{n_1}}\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}\sum_{C_{\hat{y}_R}^{(n)}(u_1^{n_1},u_2^{n_1})}p\left(C_{\hat{y}_R}^{(n)}(u_1^{n_1},u_2^{n_1})\right)$$
$$p(q^{n_1},u_1^{n_1},u_2^{n_1},x_1^{n_1},x_2^{n_1},y_R^{n_1})$$
$$\times\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}\left(y_R^{n_1},f(y_R^{n_1})\right)\chi^C_{\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1,X_2,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}\left(x_1^{n_1},x_2^{n_1},f(y_R^{n_1})\right)$$
$$=\sum_{q^{n_1}\in\mathcal{Q}^{n_1}}\sum_{u_1^{n_1},u_2^{n_1}}\sum_{C_{\hat{y}_R}^{(n)}(u_1^{n_1},u_2^{n_1})}p\left(C_{\hat{y}_R}^{(n)}(u_1^{n_1},u_2^{n_1})\right)$$
$$\sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}}p(y_R^{n_1},u_1^{n_1},u_2^{n_1},q^{n_1})\chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}\left(y_R^{n_1},f(y_R^{n_1})\right)$$
$$\sum_{(x_1^{n_1},x_2^{n_1})\in\mathcal{X}_1^{n_1}\times\mathcal{X}_2^{n_1}}p(x_1^{n_1},x_2^{n_1}|y_R^{n_1},q^{n_1},u_1^{n_1},u_2^{n_1})\chi^C_{\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1,X_2,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}\left(x_1^{n_1},x_2^{n_1},f(y_R^{n_1})\right)$$

Here $C_{\hat{y}_R}^{(n)}(u_1^{n_1},u_2^{n_1})$ is the part of the codebook containing $\hat{y}_R^{n_1}(i|w_1^{(1)},w_2^{(1)}), i\in\{1,2,\ldots,2^{\lceil\alpha nR_Q\rceil}\}$ and therefore defines the mapping $f(y_R^{n_1})$.

The last sum is the probability that for given $y_R^{n_1},q^{n_1},u_1^{n_1},u_2^{n_1}$ for $(x_1^{n_1},x_2^{n_1})$ drawn according to $p(x_1^{n_1},x_2^{n_1}|u_1^{n_1},u_2^{n_1},y_R^{n_1},q^{n_1})$ we have $\left(x_1^{n_1},x_2^{n_1},f(y_R^{n_1})\right)\notin\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1,X_2,\hat{Y}_R|u_1^{n_1},u_2^{n_1},q^{n_1})$. Now,

$\left(y_R^{n_1}, f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R | u_1^{n_1}, u_2^{n_1}, q^{n_1})$ implies by the definition of the typical set that we have $\left(y_R^{n_1}, f(y_R^{n_1}), q^{n_1}, u_1^{n_1}, u_2^{n_1}\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R, Q, U_1, U_2)$. Furthermore it follows from the properties of the typical set, that for any $\left(y_R^{n_1}, f(y_R^{n_1}), q^{n_1}, u_1^{n_1}, u_2^{n_1}\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R, Q, U_1, U_2)$ and for $(x_1^{n_1}, x_2^{n_1})$ drawn according to $p(x_1^{n_1}, x_2^{n_1} | y_R^{n_1}, q^{n_1}, u_1^{n_1}, u_2^{n_1})$ we have

$$Pr\left[(x_1^{n_1}, x_2^{n_1}) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(X_1, X_2 | y_R^{n_1}, f(y_R^{n_1}), q^{n_1}, u_1^{n_1}, u_2^{n_1}\right) | y_R^{n_1}, f(y_R^{n_1}), q^{n_1}, u_1^{n_1}, u_2^{n_1}\right]$$

can be made arbitrarily close to 1 by choosing n large. Here we used the fact that $(x_1^{n_1}, x_2^{n_1})$ are independent of $\hat{y}_R^{n_1}$ given $y_R^{n_1}$, $u_1^{n_1}$, $u_2^{n_1}$, and q^{n_1}.

Now $(x_1^{n_1}, x_2^{n_1}) \in \mathcal{T}_{\epsilon_2}^{(n_1)}\left(X_1, X_2 | y_R^{n_1}, f(y_R^{n_1}), q^{n_1}, u_1^{n_1}, u_2^{n_1}\right)$ implies that we have $\left(x_1^{n_1}, x_2^{n_1}, y_R^{n_1}, f(y_R^{n_1}), q^{n_1}, u_1^{n_1}, u_2^{n_1}\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(X_1, X_2, Y_R, \hat{Y}_R, Q, U_1, U_2)$ and therefore it follows that $\left(x_1^{n_1}, x_2^{n_1}, f(y_R^{n_1})\right) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(X_1, X_2, \hat{Y}_R | q^{n_1}, u_1^{n_1}, u_2^{n_1})$. As we choose $\epsilon_2 = \epsilon_6$ we can conclude that $\mathbb{E}\{Pr[E_{9,1}]\}$ can be made arbitrarily small by choosing n large. In particular for a given $\epsilon > 0$ we can find $n^{(9)}$ such that $\mathbb{E}\{Pr[E_9]\} < \frac{\epsilon}{10}$ for all $n > n^{(9)}$.

Error event E_{10} The probability of this event can be bounded from above by

$$\begin{aligned}\mathbb{E}\{Pr[E_{10}]\} \leq \frac{1}{|\mathcal{W}|} \sum_{(w_1,w_2)\in\mathcal{W}} \sum_{\hat{w}_2^{(2)}\neq w_2^{(2)}} \sum_{C^{(n)}} p(C^{(n)}) \sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} p\left(y_R^{n_1} | x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(w_2^{(2)}|w_2^{(1)})\right) \\ \times \chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}\left(Y_R,\hat{Y}_R|u_1^{n_1}(w_1^{(1)}),u_2^{n_1}(w_2^{(1)})\right)}\left(y_R^{n_1}, f(y_R^{n_1})\right)\chi_{\mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1,X_2,\hat{Y}_R|u_1^{n_1}(w_1^{(1)}),u_2^{n_1}(w_2^{(1)})\right)}\left(x_1^{n_1}(w_1), x_2^{n_1}(\hat{w}_2), f(y_R^{n_1})\right),\end{aligned}$$

where $f(y_R^{n_1}) = \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$ is a mapping induced by the relay when choosing i upon receiving $y_R^{n_1}$ and knowing $(w_1^{(1)}, w_2^{(1)})$. We can rewrite the upper bound as

$$\begin{aligned}\mathbb{E}\{Pr[E_{10}]\} \leq \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} \sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} 2^{nR_2^{(2)}} 2^{\lceil\alpha n R_Q\rceil} p(q^{n_1}, u_1^{n_1}, u_2^{n_1}, x_1^{n_1}, x_2^{n_1}) \\ \sum_{\hat{y}_R^{n_1}\in\hat{\mathcal{Y}}_R^{n_1}} p(\hat{y}_R^{n_1} | u_1^{n_1}, u_2^{n_1}) \chi_{\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1,X_2,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) \\ \sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} p(y_R^{n_1}|x_1^{n_1}) \chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}(y_R^{n_1}, \hat{y}_R^{n_1}).\end{aligned}$$

Now for $(y_R^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R | u_1^{n_1}, u_2^{n_1})$ and for n sufficiently large we have

$$\begin{aligned} p(\hat{y}_R^{n_1} | y_R^{n_1}, u_1^{n_1}, u_2^{n_1}) &= \frac{p(\hat{y}_R^{n_1}, y_R^{n_1} | u_1^{n_1}, u_2^{n_1})}{p(y_R^{n_1} | u_1^{n_1}, u_2^{n_1})} \\ &\geq p(\hat{y}_R^{n_1} | u_1^{n_1}, u_2^{n_1}) \frac{2^{-n_1\left(H(Y_R,\hat{Y}_R|U_1,U_2)+2\epsilon_2\right)}}{2^{-n_1\left(H(Y_R|U_1,U_2)-2\epsilon_2\right)} 2^{-n_1\left(H(\hat{Y}_R|U_1,U_2)-2\epsilon_2\right)}} \\ &= p(\hat{y}_R^{n_1} | u_1^{n_1}, u_2^{n_1}) 2^{n_1\left(I(Y_R;\hat{Y}_R|U_1,U_2)-6\epsilon_2\right)}. \end{aligned}$$

It follows that

$$\mathbb{E}\{Pr[E_{10}]\} \leq 2^{\alpha n(\epsilon_q+6\epsilon_2)+1+I(Y_R;\hat{Y}_R|U_1,U_2)} 2^{nR_2^{(2)}} \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1},u_1^{n_1},u_2^{n_1})$$
$$\sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} \sum_{\hat{y}_R^{n_1}\in\hat{\mathcal{Y}}_R^{n_1}} \sum_{y_R^{n_1}\in\mathcal{Y}_R^{n_1}} p(x_2^{n_1}|u_2^{n_1})p(x_1^{n_1}|u_1^{n_1})p(\hat{y}_R^{n_1}|y_R^{n_1},u_1^{n_1},u_2^{n_1})p(y_R^{n_1}|x_1^{n_1})$$
$$\times \chi_{\mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}(y_R^{n_1},\hat{y}_R^{n_1}) \chi_{\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1,X_2,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}(x_1^{n_1},x_2^{n_1},\hat{y}_R^{n_1}).$$

The last sum can be seen to be the probability that for sequences $x_1^{n_1}, x_2^{n_1}, y_R^{n_1}, \hat{y}_R^{n_1}, q^{n_1}, u_1^{n_1}, u_2^{n_1}$ drawn according to $p(q^{n_1}, u_1^{n_1}, u_2^{n_1})$, $p(y_R^{n_1}, \hat{y}_R^{n_1}, x_1^{n_1}|u_1^{n_1}, u_2^{n_1})$, and $p(x_2^{n_1}|u_2^{n_1})$ we have $(y_R^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_2}^{(n_1)}(Y_R, \hat{Y}_R|u_1^{n_1}, u_2^{n_1})$ and $(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_6}^{(n_1)}(X_1, X_2, \hat{Y}_R|u_1^{n_1}, u_2^{n_1})$ simultaneously. Therefore we can write

$$\mathbb{E}\{Pr[E_{10}]\} \leq 2^{\alpha n(\epsilon_q+6\epsilon_2)+1+I(Y_R;\hat{Y}_R|U_1,U_2)} 2^{nR_2^{(2)}}$$
$$\times \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} \sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} \sum_{\hat{y}_R^{n_1}\in\hat{\mathcal{Y}}_R^{n_1}} p(q^{n_1},u_1^{n_1},u_2^{n_1})p(\hat{y}_R^{n_1},x_1^{n_1}|u_1^{n_1},u_2^{n_1})$$
$$\times p(x_2^{n_1}|u_2^{n_1}) \chi_{\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1,X_2,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}(x_1^{n_1},x_2^{n_1},\hat{y}_R^{n_1}).$$

For sufficiently large n this can be bounded from above by

$$\mathbb{E}\{Pr[E_{10}]\} \leq 2^{\alpha n(\epsilon_q+6\epsilon_1)+1+I(Y_R;\hat{Y}_R|U_1,U_2)} 2^{nR_2^{(2)}}$$
$$\times \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1},u_1^{n_1},u_2^{n_1}) 2^{-n_1\left(I(X_2;\hat{Y}_R|X_1,U_2)-6\epsilon_6\right)}$$
$$\leq 2^{n\left(R_2^{(2)}-\alpha I(X_2;\hat{Y}_R|X_1,U_2)\right)} 2^{\alpha n(\epsilon_q+6\epsilon_2+6\epsilon_6)+1+I(Y_R;\hat{Y}_R|U_1,U_2,Q)+I(X_2;\hat{Y}_R|X_1,U_2)}$$
$$\leq 2^{n\left(R_2^{(2)}-\alpha(I(X_2;\hat{Y}_R|X_1,U_2)+3\epsilon_q)\right)+1+I(Y_R;\hat{Y}_R|U_1,U_2,Q)+I(X_2;\hat{Y}_R|X_1,U_2)}.$$

By assumption $R_2^{(2)} < \alpha I(X_2; \hat{Y}_R|X_1, U_2)$ and we choose $\epsilon_q < \frac{I(X_2;\hat{Y}_R|X_1,U_2)-\frac{R_2^{(2)}}{\alpha}}{3}$. Therefore the probability of this event can be made arbitrarily small for n large. In particular for a given $\epsilon > 0$ we can find $n^{(10)}$ such that for all $n > n^{(10)}$ we have $\mathbb{E}\{Pr[E_{10}]\} < \frac{\epsilon}{10}$ and such that the $n > n^{(10)}$ is sufficiently lage to ensure the inequalities used in this part of the proof. Therefore the probability of error for the tenth error event can be bounded by $\mathbb{E}\{Pr[E_{10}]\} < \frac{\epsilon}{10}$ for $n \geq n^{(10)}$.

3.3.2.6 Cases, where the Assumed Strict Inequalities are not Possible

The case $I(X_2; \hat{Y}_R|X_1, U_2) = 0$ needs a special treatment as it is not captured in the above proof: In this case the error probability for the compress-and-forward part of the proof at receiver 1, i.e. the error probability for decoding a wrong $w_2^{(2)}$ is 0 by definition, i.e. we do not need to consider the error events E_8 and E_{10}. In the calculation of the other error events and in the calculation of the error probability of receiver 2 neither $R_2^{(2)}$ nor $I(X_2; \hat{Y}_R|X_1, U_2)$

is restricted but in the definition of ϵ_q. This definition can in this case be changed by removing the requirement $\epsilon_q < \epsilon_q^{(3)} = \frac{I(X_2;\hat{Y}_R|X_1,U_2)-\frac{R_2^{(2)}}{\alpha}}{3}$ as this requirement is only needed to ensure the low probability for error event E_{10} for receiver 1 and is therefore not necessary for this case. The changed code and the above steps of the proof for receiver 2 yield a sequence of $\left(M_1^{(n)}, M_2^{(n)}, n_1, n_2\right)$-codes such that the probability of error goes to zero and the rate of the codes goes to $[R_1, R_2]$ as $n \to \infty$ given the above strict inequalities hold except for the one restricting $R_2^{(2)}$, and we have $R_2^{(2)} = I(X_2; \hat{Y}_R|X_1, U_2) = 0$ for the probability distributions $p(q)p(x_1|q)p(x_2|q)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$, $p(x_R|q)p_2(y_1, y_2|x_R)$, and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$. Analogous arguments apply for $R_1 = I(X_1; \hat{Y}_R|X_2, U_1) = 0$.

Using similar arguments the cases $R_1^{(1)} = \beta I(V; Y_2) = 0$ and $R_2^{(1)} = \beta I(V; Y_1) = 0$ can be handled, as in these cases the decoding of $w_1^{(1)}$ or $w_2^{(1)}$ respectively at the receiver cannot be in error and therefore the event E_3 does not need to be considered. It follows that the choice of ϵ_3 can be relaxed (or is not necessary at all). Similarly $R_1^{(1)} = \alpha I(U_1; Y_R|U_2, Q) = 0$, $R_2^{(1)} = \alpha I(U_2; Y_R|U_1, Q) = 0$, and $R_1^{(1)} + R_2^{(1)} = \alpha I(U_1 U_2; Y_R|Q) = 0$ can be treated: A small change calculation of E_2 needs to be applied to cope with the fact that for one or both of the codewords $w_1^{(1)}$, $w_2^{(1)}$ there is no possible wrong decision. This again leads to the possibility either to relax ϵ_1 such that it can still be chosen as $\epsilon_1 > 0$, or to discard ϵ_1 as the connected decoding set is not necessary anymore. The achievability of $[R_1, R_2] = [0, 0]$ is obvious from the definition.

3.3.2.7 The Achievable Set is Closed

The above proves that any $[R_1, R_2]$ with

$$\begin{aligned} R_1^{(1)} &< \min\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\} && (3.12)\\ R_2^{(1)} &< \min\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\} && (3.13)\\ R_1^{(1)} + R_2^{(1)} &< \alpha I(U_1 U_2; Y_R|Q) && (3.14)\\ R_1^{(2)} &< \alpha I(X_1; \hat{Y}_R|X_2, U_1) && (3.15)\\ R_2^{(2)} &< \alpha I(X_2; \hat{Y}_R|X_1, U_2) && (3.16) \end{aligned}$$

is achievable as long as the constraints (3.10) are fulfilled. We conclude the proof by showing that the achievable rate region is closed. But this follows from the definition of achievability: Let $[R_{1,0}, R_{2,0}]$ be some rate pair on the boundary of the set with $R_{1,0} > 0$, $R_{2,0} > 0$. For any rate pair $[R_{1,0} - \frac{\epsilon}{m}, R_{2,0} - \frac{\epsilon}{m}]$, $\epsilon > 0$, $m \in \mathbb{N}$ there exists a sequence of $\Big(2^{\lfloor n(R_{1,0}^{(1)} - \frac{\epsilon}{2m})\rfloor + \lfloor n(R_{1,0}^{(2)} - \frac{\epsilon}{2m})\rfloor}$, $2^{\lfloor n(R_{2,0}^{(1)} - \frac{\epsilon}{2m})\rfloor + \lfloor n(R_{2,0}^{(2)} - \frac{\epsilon}{2m})\rfloor}, n_1, n_2\Big)$-codes such that $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. Therefore for any m there exists $n_0(m)$ such that $\mu_k^{(n)} < \frac{1}{m}$, $k \in \{1, 2\}$ for $n > n_0(m)$. Let $m^{(n)} = \max\{m : n > n_0(m)\}$. Because $\mu_k^{(n)} \to 0$ as $n \to \infty$ we have $m^{(n)} \to \infty$. So we can construct a sequence of $\Big(2^{\lfloor n(R_{1,0}^{(1)} - \frac{\epsilon}{2m^{(n)}})\rfloor + \lfloor n(R_{1,0}^{(2)} - \frac{\epsilon}{2m^{(n)}})\rfloor}, 2^{\lfloor n(R_{2,0}^{(1)} - \frac{\epsilon}{2m^{(n)}})\rfloor + \lfloor n(R_{2,0}^{(2)} - \frac{\epsilon}{2m^{(n)}})\rfloor}, n_1, n_2\Big)$-codes with $\frac{1}{n}\lfloor n(R_{1,0}^{(1)} - \frac{\epsilon}{2m^{(n)}})\rfloor + \lfloor n(R_{1,0}^{(2)} - \frac{\epsilon}{2m^{(n)}})\rfloor \to R_{1,0}$, $\frac{1}{n}\lfloor n(R_{2,0}^{(1)} - \frac{\epsilon}{2m^{(n)}})\rfloor + \lfloor n(R_{2,0}^{(2)} - \frac{\epsilon}{2m^{(n)}})\rfloor \to R_{2,0}$, $\mu_k^{(n)} < \frac{1}{m} \to 0$, $k \in \{1, 2\}$ as $n \to \infty$. Therefore by the definition of achievability the rate pair $[R_{1,0}, R_{2,0}]$ is achievable. With analogous arguments rate pairs on the boundary of the region where one of the rates is 0 can be achieved. This proves that the set of achievable rates is closed. □

3.3.3 Asymmetric Strategies

We can make use of the boundary effects that needed a special treatment in the above proof. These boundary effects may lead to more relaxed constraints and therefore allow a larger rate for one of the users. Using the boundary effects means: One of the messages is forwarded using both, the compress-and-forward as well as the decode-and-forward strategy, while the other node's message is fully decoded.

Corollary 3.6. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_5 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exists $R_1^{(1)}, R_1^{(2)}, R_2^{(2)} \geq 0$ with $R_1^{(1)} = R_1$, $R_2^{(1)} + R_2^{(2)} = R_1$ satisfying

$$\begin{aligned} R_1^{(1)} &\leq \min\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\} && (3.17)\\ R_2^{(1)} &\leq \min\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\} && (3.18)\\ R_1^{(1)} + R_2^{(1)} &\leq \alpha I(U_1 U_2; Y_R|Q) && (3.19)\\ R_2^{(2)} &\leq \alpha I(X_2; \hat{Y}_R|X_1, U_2) && (3.20) \end{aligned}$$

under the constraint

$$\alpha(H(\hat{Y}_R|X_1, U_2) - H(\hat{Y}_R|Y_R)) < \beta I(Y_1; X_R|V)$$

and similarly the set $\mathcal{R}_6 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exists $R_1^{(1)}, R_2^{(1)}, R_1^{(2)} \geq 0$ with $R_1^{(1)} + R_1^{(2)} = R_1$, $R_2^{(1)} = R_2$ satisfying

$$\begin{aligned} R_1^{(1)} &\leq \min\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\} && (3.21)\\ R_2^{(1)} &\leq \min\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\} && (3.22)\\ R_1^{(1)} + R_2^{(1)} &\leq \alpha I(U_1 U_2; Y_R|Q) && (3.23)\\ R_1^{(2)} &\leq \alpha I(X_1; \hat{Y}_R|X_2, U_1) && (3.24) \end{aligned}$$

under the constraint

$$\alpha(H(\hat{Y}_R|X_2, U_1) - H(\hat{Y}_R|Y_R)) < \beta I(Y_2; X_R|V)$$

for some joint probability distributions $p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(v)p(x_R|v)p_2(y_1, y_2|x_R)$, and for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.

Proof. The rate pairs claimed in the corollary where one of the compress-and-forward sub-rates is implicitly set to 0 follow directly from the proof of Theorem 3.5 by observing, that the terminal with rate 0 does not need to transmit any information via the compress-and-forward mechanism. For the following treatment we assume $R_1^{(2)} = 0$. The treatment of

$R_2^{(2)} = 0$ is analogous. From the proof of Theorem 3.5 it follows immediately, that — as the events E_8 and E_{10} do not lead to an error for receiver 2, and even more the coding steps vi and vii can be skipped for this decoder — we do only need to restrict ϵ_q in the code by $\epsilon_q \in \left(0, \min\{\epsilon_q^{(1)}, \epsilon_q^{(3)}\}\right)$ with $\epsilon_q^{(1)} = \frac{1}{2\alpha+\beta}\left(\beta I(X_R; Y_1|V) - \alpha\left(H(\hat{Y}_R|X_1, U_2) - H(\hat{Y}_R|Y_R)\right)\right)$ and $\epsilon_q^{(3)} = \frac{I(X_2;\hat{Y}_R|X_1,U_2)-\frac{R_2^{(2)}}{\alpha}}{3}$. Furthermore for the same reason there is no requirement for the constraint $\alpha\left(H(\hat{Y}_R|X_2, U_2) - H(\hat{Y}_R|Y_R)\right) < \beta I(X_R; Y_2|V)$. □

Now we can join the three regions to a new achievable rate region by convex combination of rate pairs from the three regions.

Corollary 3.7. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_{\mathrm{PDF}} \subset \mathbb{R}_+^2$ given by the convex hull of $\mathcal{R}_4 \cup \mathcal{R}_5 \cup \mathcal{R}_6$.

3.4 Concluding remarks

In this chapter the two-way relay channel with a compress-and-forward strategy is studied. We proposed an achievable rate region for a two-phase protocol. The base upon which the scheme is build is the observation, that the side information available at the receiver can be used to enhance the performance in both, the MAC phase and the BC phase. This is due to the fact that the output of the MAC is caused by the transmitted signals. Therefore the transmitted signal restricts the possible outcome of the MAC output at the relay, and thereby reduces the number of possible occurring codewords transmitted in the BC phase for both receivers.

The coding mechanism analyzed in Section 3.1.4 gives rise to further improvements: Recall Figure 3.1. Suppose for now, that the receiver 1 would know w_2 in addition to its side information w_1 when decoding the relay's codeword. This restricts the codewords that may occur even further. The codeword $x_R(i)$ transmitted by the relay is in both subsets of the relay's code, in the code $C(w_1)$ and in the code $C(w_2)$. In Figure 3.1 this is true for three codewords. Therefore the receiver knows that $x_R(i) \in C(w_1) \cap C(w_2)$. The receiver would only need to decide which of these three codewords has been transmitted. Now, suppose, that $M_2 > 1$; the receiver does not know w_2 but it knows the codewords $x_2^{n_1}(w_2)$, $w_2 \in \mathcal{W}_2$ used by the other transmitter. Therfore the receiver knows the restrictions $C(w_2)$ for each of the transmitted messages $w_2 \in \mathcal{W}_2$. Therefore one could change the coding paradigm and decode only one of the codewords in $\bigcup_{w_2}(C(w_1) \cap C(w_2))$. Unfortunately this will not reduce the number of codewords in the code: $\bigcup_{w_2} C(w_2)$ includes all the codewords used by the relay. Otherwise the relay could dismiss the codeword not in $\bigcup_{w_2} C(w_2)$ without any loss in performance, as it would never be used.

Now, recall, that the restriction to the sub-codes $C(w_1)$ and $C(w_2)$ is due to statistical dependencies in the MAC. Furthermore, note that some sequence of length n_1 in the MAC output alphabet $\mathcal{Y}_R^{n_1}$ might be jointly typical with $x_1^{n_1}(w_1)$ and it might be jointly typical with $x_2^{n_1}(w_2)$. But this still does not imply, that it is jointly typical with the pair $\left(x_1^{n_1}(w_1), x_2^{n_1}(w_2)\right)$. Therefore we can construct a better restriction say a set $C(w_1, w_2) \subseteq C(w_1) \cap C(w_2)$ of codewords that may occur at the relay if (w_1, w_2) was transmitted. The union of these restricted codes $\bigcup_{w_2} C(w_1, w_2) \subseteq \bigcup_{w_2}(C(w_1) \cap C(w_2)) = C(w_1)$ in turn can be used at the receiver 1 to decode the quantization index i transmitted by the relay and the message transmitted by node 2 jointly. This joint decoding can improve the performance of the two-way relay channel even more. In effect one cannot only disable the sum constraint in the MAC, as it was done in this chapter. Even more, one can use the effect that causes this constraint to improve the performance in the overall system. The next chapter will analyze this joint decoding mechanism in detail. As a result an achievable rate region is stated, which extends the region $\mathcal{R}_1$ of Theorem 3.1.

Chapter 4

The Two-Way Relay Channel with Compress-and-Forward and Joint Decoding

In this chapter we extend the compress-and-forward region given in Chapter 3. In Theorem 3.1 the goal is to transmit a good enough representation of the MAC output to both receivers. The receiver decodes the message transmitted by the relay and therefore gets the quantized MAC output. This is used to decode the MAC transmission of the other user. The receivers use their own transmitted messages as side information in all decoding steps. Indeed, for some channels one can do better than decoding the relay's transmission, i.e. there are gains possible compared to the strategy proposed in Theorem 3.1. The following toy example gives some intuition on how one could improve the coding:

Example 4.1. *The MAC output in this example is the XOR sum of two binary inputs. The channel from the relay to receiver* 1 *is a lossless channel which allows transmission of one bit per channel use. The channel to receiver* 2 *is noisy. Now suppose that node* 2 *transmits uncoded. Node* 1 *uses some code of rate smaller than* 1*, e.g. some repetition code (In this toy example this code is in fact used to fight the noise in the BC phase as we will see below.). Clearly the relay cannot decode the messages. Now consider the following strategy: The relay forwards the MAC output to both receivers. Using the strategy of Theorem 3.1 receiver* 1 *can decode the relay's transmitted codeword without error using its own message as side information, but receiver* 2 *cannot. Note, that the relay's transmission code contains all binary codewords of length* n*. Using only its own transmitted signal as side information for the decoding of the relay's transmission does — by the construction of the example — not restrict the number of possible transmitted codewords. This way of decoding is not very intuitive: Looking at the problem at hand one automatically uses the knowledge of the code of node* 1*. Inverting the XOR operation the restricted transmission code of the relay is clearly the same as the one used by node* 1 *for the MAC transmission.*

This toy example shows: Decoding is possible, if the receiver 2, in addition to its own transmitted codeword, also uses knowledge about the MAC transmission code of the other node. Doing so results in a joint decoding of the MAC codeword of node 1 and the relay's transmission. In fact the decoding changes to not decoding the relay's transmission explicitly. The focus is now on the message intended for the receiver using the received signal as side information and the knowledge, how the MAC code of the other node restricts the relay's transmission code. The correct decoding of the relay's transmission is only a by-product. The goal of the relay also changes: now the goal is not to convey a MAC output to the receivers, but to help the receiver to decode the message intended for it. As in the toy example it might not be possible to decode the relay's codeword without simultaneously decoding the message intended for the receiver.

The new strategy proposed in this section does — unlike the strategy of Theorem 3.1 — not explicitly decode the transmitted signal x_R and therefore the compressed channel output $\hat{y}_R$ at the receivers. In fact it uses the dependency structure between the variables that occur in the system. Decoding of the message is now performed by a joint decoding over all involved random variables. This strategy will lead to a new region which can be shown to be an extension of the region given in Theorem 3.1.

In Section 4.1 we state an achievable rate region for the two-phase two-way relay channel with compress-and-forward and joint decoding at the receiver. Section 4.1.2 gives the proof for this region. We shed light on the coding mechanisms used in the decoding in Section 4.1.3. There, some hints are given for the design of practical codes. In Section 4.2 we extend the achievable rate region of Theorem 4.1 in Section 4.1 by superimposing the scheme on a decode-and-forward scheme. The resulting region $\mathcal{R}_{\text{PCF-JD}}$ stated in Theorem 4.3 includes all regions for the two-phases two-way relay channel given in this thesis up to that point.

Some of the results of this section where published in [13].

4.1 An Achievable Rate Region with Joint Decoding

In this section we propose an achievable rate region that is attained by a compress-and-forward strategy with joint decoding at the receiver. As in Chapter 3, the relay forwards a quantized representation of this output to both receivers. Again the message known by the receiver is used to enhance the decoding performance in the BC phase. This is possible as the MAC output depends on the codeword transmitted by the receiver. In addition to this side information the receiver uses its knowledge about the codewords used by the other node in the MAC transmission and about the statistical dependencies due to the MAC. As a consequence the receiver does not aim to decode the relay's transmission but focuses on the decoding of the message transmitted by the other node directly. The decoding of the relay's transmission is a by-product of the joint decoding. For the following discussion, recall the system model and the definitions given in Chapter 1.

4.1.1 Coding Theorem

Theorem 4.1. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_7 \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned} R_1 &\le \max\Big\{0, \min\Big\{\alpha I(X_1; \hat{Y}_R|X_2, Q), \alpha\Big(I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\Big) + \beta I(X_R; Y_2)\Big\}\Big\} \\ R_2 &\le \max\Big\{0, \min\Big\{\alpha I(X_2; \hat{Y}_R|X_1, Q), \alpha\Big(I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\Big) + \beta I(X_R; Y_1)\Big\}\Big\} \end{aligned} \tag{4.1}$$

for some $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and for some joint probability distributions $p(q)p(x_1|q)p(x_2|q)$ $p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(x_R)p_2(y_1, y_2|x_R)$.

Remark 4.1 (Cardinalities of random variables)**.** To achieve any point in the given region it is sufficient to consider only random variables Q and $\hat{Y}_R$ with cardinalities restricted to $|\mathcal{Q}| \le 3$ and $|\hat{\mathcal{Y}}_R| \le |\mathcal{Y}_R| + 2$. This can be shown using the Fenchel-Bunt extension of Caratheodory's theorem [62]. A proof of this claim is given in the appendix.

Remark 4.2 (Extension of the compress-and-forward region)**.** To see that the region $\mathcal{R}_7$ is an extension of the region $\mathcal{R}_1$ given in Theorem 3.1 we plug in some rate pair and rewrite the new rate constraint the same way as the constraints (3.2) for $\mathcal{R}_1$: For Theorem 4.1 we have

$$\begin{aligned} \alpha\left(I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(X_R; Y_1) - R_2 \ge 0 \\ \alpha\left(I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(X_R; Y_2) - R_1 \ge 0 \end{aligned} \tag{4.2}$$

while the constraints (3.2) can be rewritten as

$$\begin{aligned} \alpha\left(I(X_1; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(Y_1; X_R) > 0 \\ \alpha\left(I(X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(Y_2; X_R) > 0. \end{aligned} \tag{4.3}$$

Now $\alpha\left(I(X_1X_2; \hat{Y}_R|Q) - I(X_1; \hat{Y}_R|Q)\right) - R_2 = \alpha I(X_2; \hat{Y}_R|X_1, Q) - R_2 \ge 0$ for any distribution that fulfills the requirements of any of the rate regions. Therefore the additional rate constraints of Theorem 4.1 are less strict than the constraints (3.2).

Remark 4.3 (Loss due to quantization)**.** The expressions (4.1) split into a part for the MAC and a part for the BC. The rate constraints are similar to those in the decode-and-forward region (2.11); the difference is, that we trade the sum constraint for the MAC in for a penalty term for the BC. Now, an interesting question is, whether we can eliminate the sum constraint without a penalty to the BC rate region.

We have $I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q) \le 0$ due to the data processing inequality. The second expression in (4.1) is only dependent on the broadcast channel, if in addition to the requirement of the theorem we had $p(x_1, x_2, y_R, \hat{y}_R|q) = p(x_1, x_2|q)p(\hat{y}_R|x_1, x_2)p_1(y_R|x_1, x_2)$. We have by the requirements of the theorem $p(y_R, \hat{y}_R|x_1, x_2) = p(\hat{y}_R|y_R)p_1(y_R|x_1, x_2)$. Furthermore we now would need $p(y_R, \hat{y}_R|x_1, x_2) = p_1(y_R|x_1, x_2)p(\hat{y}_R|x_1, x_2)$. So the second expression in the rate requirements is only dependent on the broadcast channel if and only if $\forall x_1, x_2, y_R : p_1(y_R|x_1, x_2) \neq 0$ we have $p(\hat{y}_R|x_1, x_2) = p(\hat{y}_R|y_R)$. This is easy to fulfill if we chose $\hat{Y}_R$ independent of X_1, X_2, and Y_R but in this case the first expression in the rate constraint is 0. The first expression in the rate constraint can be nonzero if the multiple access channel is such, that for $K > 1$ disjoint subsets in the channel input alphabet $\bar{\mathcal{X}}_k \subset \mathcal{X}_1 \times \mathcal{X}_2$, $k \in \{1, 2, \dots, K\}$, $\bigcup_k \bar{\mathcal{X}}_k = \mathcal{X}_1 \times \mathcal{X}_2$ and for $L, 1 < L \leq K$ disjoint subsets in the channel output alphabet $\bar{\mathcal{Y}}_l \subset \mathcal{Y}_R$, $\bigcup_l \bar{\mathcal{Y}}_l = \mathcal{Y}_R$ we have $\forall (x_1, x_2) \in \bar{\mathcal{X}}_k, y_R \in \bar{\mathcal{Y}}_l, l \neq f(k)$ $p_1(y_R|x_1, x_2) = 0$, where $f(\cdot)$ is some fixed mapping. The required compression at the relay is in general very lossy and will degrade the MAC performance. In fact the positive values in $p(y_R|x_1, x_2)$ do not matter for the performance using that kind of compression. The toy example given in the introduction uses this scheme and suffers no performance loss due to the special deterministic structure of the MAC.

Remark 4.4 (Non-convexity of the region)**.** In the general case the region given in Theorem 4.1 in non-convex. In comparison to the compress-and-forward region $\mathcal{R}_1$ it turns out, that the region $\mathcal{R}_7$ contains the special case of idle users which was treated in Corollary 3.2 as new regions $\mathcal{R}_2$ and $\mathcal{R}_3$. In fact we can interpret the region as the union of one two-dimensional region, where we have no idle users, and two one-dimensional regions, where one of the users is idle, i.e. has a rate of 0. Setting one of the user's rate to 0 we gain some more freedom, as in this case the probability distribution is less constraint, i.e. we allow $\alpha \left(I(X_1 X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q) \right) + \beta I(X_R; Y_1) < 0$ if $R_2 = 0$. This increased freedom may be used to achieve a bigger rate R_1. Therefore the region $\mathcal{R}_7$ may be non-convex. Clearly the convex hull of the region $\mathcal{R}_7$ is achievable by timesharing.

Corollary 4.2. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_{\text{CF-JD}} \subset \mathbb{R}_+^2$ given by the convex hull of $\mathcal{R}_7$.

Remark 4.5 (Timesharing of codes)**.** If the convexification and therefore timesharing of codes is needed, the resulting code does not fulfill the requirements of Theorem 4.1. As a consequence the typical set decoding used in the proof need not work on the overall code. The decoding needs to use the special structure of the code, i.e. ignoring some of the symbols in the decoding for one of the users. The coding operations of the resulting code can be seen as decoding up to three codes interleaved with each other separately, each having its own constraints. In fact, as already discussed in Remark 3.6 the timesharing could be included in some time sharing variable with similar difficulties. As pointed out in Remark 3.6 the BC phase gets more complicated if we use more then one random variable for the compression at the relay. With the joint decoding approach the coding in the BC phase might change compared to the compress-and-forward approach, as we do not explicitly aim to decode the correlated variables, i.e. the MAC output representatives, but the transmitted data. An extension along these lines is analyzed in Chapter 5.

4.1.2 Proof of the Coding Theorem

Proof. For the proof of the theorem we can reuse arguments from the proof of Theorem 3.1. The generation of the codebook is the same and so is most of the coding. We will restrict the proof to those things that change: What is replaced is the decoding at the terminal nodes and the choice of the parameter ϵ_q. The decoding sets are changed as well, as they depend on ϵ_q. Furthermore, for some of the parameters of the decoding sets we need a stronger constraint to cope with the new decoding at the receiver. We also have to adapt the error events. All things that are not explicitely stated here can be reused without change form Section 3.1.2.

The first thing to note is, that one of the rate constraints, e.g. $\alpha\left(I(X_1X_2;\hat{Y}_R|Q) - I(Y_R;\hat{Y}_R|Q)\right) + \beta I(X_R;Y_k)$, $k \in \{1,2\}$, might be smaller than 0. We start the proof by assuming probability distributions and a rate pair such that the inequalities (4.1) are strict and therefore we have $I\left(X_1X_2;\hat{Y}_R|Q) - I(Y_R;\hat{Y}_R|Q)\right) + \beta I(X_R;Y_k) > 0$ for $k \in \{1,2\}$. For these rate pairs the proof is similar to the proof of Theorem 3.1 and only the differences will be given in this proof. Finally as for the proof of Theorem 3.1 and used by Corollary 3.2 it turns out, that the rate requirements are not connected directly, i.e. R_2 does not appear in the calculation for receiver 2 while R_1 does not appear in the calculation for receiver 1 and so do the constraints and requirements on the decoding sets related to these rates. Therefore in case that one of the users is idle, i.e. has a rate of 0, the constraints corresponding to that rate are not active and can be disabled. It follows that it is save to set the rate to 0 if $\alpha\left(I(X_1X_2;\hat{Y}_R|Q) - I(Y_R;\hat{Y}_R|Q)\right) + \beta I(X_R;Y_k) < 0$ for some $k \in \{1,2\}$. The achievability of the closure of the rate region is again a consequence of the definition of achievability. The convex hull is achievable by timesharing using up to three codes, one for each of the three different modi operandi.

4.1.2.1 Adapted Random Codebook Generation

For the choice of ϵ_q we now have:

- Let the parameters $\epsilon_q^{(1)}$, $\epsilon_q^{(2)}$,$\epsilon_q^{(3)}$, and $\epsilon_q^{(4)}$, be given as:

$$\epsilon_q^{(1)} := \frac{1}{2\alpha+\beta}\Big(\alpha\Big(I(X_1X_2;\hat{Y}_R|Q) - I(Y_R;\hat{Y}_R|Q)\Big) + \beta I(X_R;Y_2) - R_1\Big),$$

$$\epsilon_q^{(2)} := \frac{1}{2\alpha+\beta}\Big(\alpha\Big(I(X_1X_2;\hat{Y}_R|Q) - I(Y_R;\hat{Y}_R|Q)\Big) + \beta I(X_R;Y_1) - R_2\Big),$$

$$\epsilon_q^{(3)} := \frac{I(X_2;\hat{Y}_R|X_1,Q) - \frac{R_2}{\alpha}}{3},$$

and

$$\epsilon_q^{(4)} := \frac{I(X_1;\hat{Y}_R|X_2,Q) - \frac{R_1}{\alpha}}{3}.$$

Choose $\epsilon_q \in (0, \min\{\epsilon_q^{(1)}, \epsilon_q^{(2)}, \epsilon_q^{(3)}, \epsilon_q^{(4)}\})$.

Using this ϵ_q we can generate the random codebook as it was done in Section 3.1.2.1. Most of the decoding sets could use the same parameters as in Section 3.1.2.2 but with the new ϵ_q. Nevertheless, we will diminish the parameter as ϵ_4 needs a new restriction due to the joint decoding.

4.1.2.2 Adapted Decoding Sets

For the decoding we will use typical set decoding. For a strict definition of the decoding sets we choose parameter for the typical sets as $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon_4 \in (0, \frac{\epsilon_q}{8})$. The missing parameters for the receiver 2 are chosen in an analogous way.

4.1.2.3 Adapted Coding

While in the proof of Theorem 3.1 we decoded $\hat{y}$ first and w_2 thereafter, we do both steps at the same time now, i.e. we have decoding steps iv and v replaced by the new decoding step $\hat{\text{iv}}$:

$\hat{\text{iv}}$ Upon receiving $y_1^{n_2}$ node 1 decides that w_2 was transmitted if $x_2^{n_1}(w_2)$ is the only codeword such that for some $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_Q \rceil}\}$, the sequences $x_2^{n_1}(w_2)$, $\hat{y}_R^{n_1}(i)$, and $x_1^{n_1}(w_1)$ are jointly typical given q^{n_1}, and simultaneously $x_R^{n_2}(i)$ and the received signal $y_1^{n_2}$ are jointly typical, i.e $\exists i \in \{1, 2, \dots, 2^{\lceil \alpha n R_Q \rceil}\} : \left(x_1^{n_1}(w_1), x_2^{n_1}(w_2), \hat{y}_R^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R | q^{n_1})$ and $\left(x_R^{n_2}(i), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, X_1)$.

All other coding steps are identical to the coding described in Section 3.1.2.3.

4.1.2.4 Adapted Error Events

For the bounding of $\mathbb{E}\{\mu_1^{(n)}\}$ we can reuse all the error events defined in Section 3.1.2.4 but E_4. Therefore we replace E_4 by a new error event:

- $\hat{E}_4$: Suppose a codebook is given, $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted, the relay chose some i and $x_R^{n_2}(i)$ is transmitted. $\hat{E}_4$ is the event that $\exists j \neq i, \hat{w}_2 \neq w_2 : \left(x_1^{n_1}(w_1), x_2^{n_1}(\hat{w}_2), \hat{y}_R^{n_1}(j)\right) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R | q^{n_1})$, $\left(x_R^{n_2}(j), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, X_1)$.

Note that the seemingly missing event $j = i$ is already captured by E_6. Now we bound the probability for this error event for receiver 1. The proof for receiver 2 is analogous.

Error event $\hat{E}_4$

$$\mathbb{E}\{Pr[\hat{E}_4]\} \leq \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}) Pr[\hat{E}_{4,1}] Pr[\hat{E}_{4,2}] 2^{\lceil \alpha n R_Q \rceil} 2^{\lfloor n R_2 \rfloor} \tag{4.4}$$

Here $\hat{E}_{4,1}$ is the event that for three sequences $x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}$ drawn independent of each other we have $(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R | q^{n_1})$. $x_1^{n_1}$, $x_2^{n_1}$, and $\hat{y}_R^{n_1}$ are drawn at random according to $p(x_1^{n_1}|q^{n_1})$, $p(x_2^{n_1}|q^{n_1})$, and $p(\hat{y}_R^{n_1}|q^{n_1})$ respectively to capture the averaging over the random codebooks. $\hat{E}_{4,2}$ is the event that for two sequences $x_R^{n_2}, y_1^{n_2}$ drawn independent of each other we have $(x_R^{n_2}, y_1^{n_2}) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(X_R, X_1)$. The factor $2^{\lceil \alpha n R_Q \rceil}$ accounts for the fact that we can use a union bound and the error occurs if at least one $j \neq i$ is found fulfilling the requirements. The factor $2^{\lfloor nR_2 \rfloor}$ accounts for the different possible $\hat{w}_2$.

For sufficiently large n

$$\begin{aligned}
Pr[\hat{E}_{4,1}] &= \sum_{x_1^{n_1} \in \mathcal{X}_1^{n_1}} \sum_{x_2^{n_1} \in \mathcal{X}_2^{n_1}} \sum_{\hat{y}_R^{n_1} \in \hat{\mathcal{Y}}_R^{n_1}} p(x_1^{n_1}|q^{n_1}) p(x_2^{n_1}|q^{n_1}) \chi_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1,X_2,\hat{Y}_R|q^{n_1})}(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) p(\hat{y}_R^{n_1}|q^{n_1}) \\
&\leq |\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R|q^{n_1})| 2^{-n_1\left(H(X_1|Q)-2\epsilon_4\right)} 2^{-n_1\left(H(X_2|Q)-2\epsilon_4\right)} 2^{-n_1\left(H(\hat{Y}_R|Q)-2\epsilon_4\right)}
\end{aligned}$$

due to the properties of the typical set. Furthermore, it follows from these properties that

$$|\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_R|q^{n_1})| \leq 2^{n_1\left(H(X_1,X_2,\hat{Y}_R|Q)+2\epsilon_4\right)}.$$

$Pr[\hat{E}_{4,2}]$ can be bounded in a similar way. Therefore we have

$$\begin{aligned}
\mathbb{E}\{Pr(\hat{E}_4)\} &\leq \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}) 2^{-n_1\left(I(X_1X_2;\hat{Y}_R|Q)-8\epsilon_4\right)} 2^{-n_2\left(I(X_R;Y_1)-6\epsilon_2\right)} 2^{\alpha n R_Q+1} 2^{nR_2} \\
&\leq 2^{-n\left(\alpha\left(I(X_1X_2;\hat{Y}_R|Q)-R_Q-8\epsilon_4\right)+\beta(I(X_R;Y_1)-6\epsilon_2)-R_2\right)+1+I(X_1X_2;\hat{Y}_R|Q)+6\epsilon_2} \\
&= 2^{-n\left(\alpha\left(I(X_1X_2;\hat{Y}_R|Q)-I(Y_R;\hat{Y}_R|Q)\right)+\beta I(X_R;Y_1)-R_2-\tilde{\epsilon}\right)+1+I(X_1X_2;\hat{Y}_R|Q)+6\epsilon_2}
\end{aligned}$$

with

$$\tilde{\epsilon} = \alpha\epsilon_q + \beta 6\epsilon_2 + \alpha 8\epsilon_4 < (2\alpha + \beta)\epsilon_q.$$

This term goes to zero if

$$\alpha\left(I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(X_R; Y_1) - R_2 - (2\alpha + \beta)\epsilon_q > 0$$

By the choice of ϵ_q as

$$\epsilon_q < \frac{1}{(2\alpha + \beta)}\left(\alpha\left(I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(X_R; Y_1) - R_2\right)$$

this is true whenever

$$R_2 < \alpha\left(I(X_1X_2; \hat{Y}_R|Q) - I(Y_R; \hat{Y}_R|Q)\right) + \beta I(X_R; Y_1)$$

as required by the assumption. Therefore for any given ϵ we can find $n^{(4)}$ such that for $\mathbb{E}\{Pr[\hat{E}_4]\} < \frac{\epsilon}{6}$ for $n \geq n^{(4)}$. The rest of the proof is analogous to the proof of Theorem 3.1 and is omitted here. □

4.1.3 A Note on Coding Mechanisms for Joint Decoding

The coding mechanism in the proof of Theorem 4.1 is similar to the coding used in Theorem 3.1. In fact the code used by the relay can be the same for both coding schemes. Only the decoding at the terminal node changes. Instead of using only its one side information as described in Section 3.1.4 the receivers use in addition the knowledge of the other nodes code in the MAC phase for the decoding.

We now describe the decoding procedure at node 1. This node knows its own message w_1 and the code used by node 2 in the MAC phase. Now, node 1 determines the MAC outputs which are jointly typical with the side information w_1 and the message $w_2 = 1$. Furthermore, using the relay's mapping $f(\cdot)$ node 1 can determine the subset $C(w_1, 1)$ of codewords $x_R(i)$, that may occur if w_1 and $w_2 = 1$ are transmitted. This proceeding is repeated for all messages w_2.

The union of these subcodes forms the effective code $C(w_1) = \bigcup_{w_2} C(w_1, w_2)$ used by node 1 for the decoding. Note that the code can be determined offline, as it does not depend on any random variable but only on the statistics of the channels. With high probability the code $C(w_1)$ contains less than $2^{\beta n I(X_R;Y_1)}$ codewords and is therefore decodable by the receiver. The decoding yields both: the message w_2 transmitted by the other node and the index i used for the compression at the relay.

For a practical coding scheme one has to consider the MAC encoding, the quantization at the relay and the BC coding jointly. Except for the statistics of the MAC and the chosen quantization there is no structure in the mechanism, that could be used to create good codes. A simple but potentially suboptimal code for the BC is again a set of codewords, such that all subsets of size $2^{\beta n I(X_R;Y_1)}$ are good codes for the channel to receiver 1, and all subsets of size $2^{\beta n I(X_R;Y_2)}$ are good codes for the channel to receiver 2. Note, that once the MAC statistics, the MAC code and the quantization is known, the index subsets that determine the subcode $C(w_1, w_2)$ for a given w_1, w_2 are known as well. These in turn determine the codes $C(w_1)$ and $C(w_2)$. However, although the code used in the MAC phase should be a good code for the MAC transmission, and enables the decoding knowing $\hat{y}_R^{n_2}$ and the side information, the Example 4.1 showed, that this code can also be used for error correction in the BC. Therefore it is unlikely, that in general good codes can be designed separating MAC and BC phase. On the contrary to the above considerations it may be more promising to design a BC code first, consisting of good interwoven subcodes $C(w_1)$ and $C(w_2)$, and thereafter tuning the quantization and the MAC coding such that theses codes are actually those, which occur at the relay. This is possible as the MAC encoding together with the quantization determine the subcodes $C(w_1, w_2)$, that are used by both nodes in the decoding process.

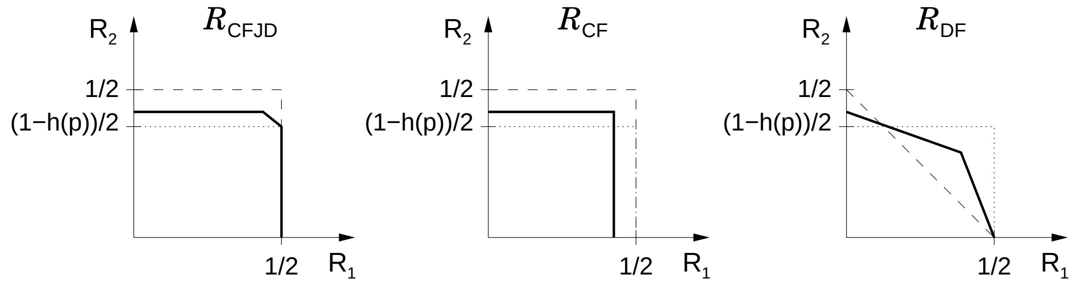

Figure 4.1: The left figure shows the capacity regions of the Channel considered in Section 4.1.4. The achievable rate region is indicated by the solid line. The dashed and dotted lines depict the constraints imposed by the MAC and BC phase respectively. The capacity for the channel is achievable with compress-and-forward and joint decoding. The region displayed in the center is an achievable rate region with compress-and-forward without joint decoding using an identity mapping at the relay and uniform distributed MAC inputs. The figure on the right hand side is the region of rate pairs achievable with decode-and-forward.

4.1.4 Example and Interpretation

As a first example consider a setup similar to the one in Example 4.1. The MAC output is the XOR sum of two binary inputs. The BC channel consists of one lossless channel with binary input to receiver 2. The channel to receiver 1 is a binary symmetric channel with a probability $p_{1,BC}$ that the output bit is inverted. The MAC channel has a maximum sum rate of 1 bit. The channel to receiver 1 can transport $1 - h(p_{1,BC})$ bits per channel use. If we use a uniform input distribution on X_1, X_2 and an identity mapping from y_R to $\hat{y}_R$ we have $I(X_1; Y_R|X_2, Q) = I(X_1; \hat{Y}_R|X_2, Q) = I(\hat{Y}_R; Y_R|Q) = I(X_1X_2; \hat{Y}_R|Q) = 1$ bit. Therefore $I(X_1X_2; \hat{Y}_R|Q) - I(\hat{Y}_R; Y_R|Q) = 0$ and the achievable rate region is given by

$$\begin{aligned} R_1 &\leq \min\{\alpha, (1-\alpha)\} \\ R_2 &\leq \min\{\alpha, (1-\alpha)(1-h(p_{1,BC}))\} \end{aligned}$$

for some $\alpha \in [0, 1]$. Comparison with the outer bound in Lemma 1.1 shows, that this is indeed the capacity of the considered channel.

If we do not use joint decoding the constraints in (3.2) enforce $\alpha < (1-\alpha)(1-h(p_{1,BC}))$ if we use the identity mapping and an uniform input distribution to the MAC[1]. This degrades the achievable rate region. Figure 4.1 shows the capacity region for this example together with the region that is achievable without joint decoding using the sketched strategy. The third region shown is achievable by decode-and-forward.

Note that in a similar setup with a symmetric binary erasure multiple access channel as considered in Section 2.1.5 the same rate pairs are achievable by compress-and-forward with joint decoding. The strategy uses a mapping of the MAC output such that the virtual channel

[1]A non-uniform input distribution achieves some more rate pairs, but the analysis of this is out of the scope of the example. Some remarks on this effect will be given in the discussion of the next example.

$p(\hat{y}_R|x_1, x_2)$ is the channel considered in the example above. For the optimal (uniform) input distribution on X_1, X_2 , we still have $I(X_1X_2; \hat{Y}_R|Q) - I(\hat{Y}_R; Y_R|Q) = 0$ while the individual rate constraints do not change due to that mapping, i.e. we have $I(X_1; Y_R|X_2, Q) = I(X_1; \hat{Y}_R|X_2, Q)$. The resulting rate expressions are that of the cutset outer bound on the capacity region (Lemma 1.1). Therefore we conclude that for the example in Section 2.1.5 capacity can be achieved by compress-and-forward with joint decoding.

Now consider a setup with a noisy MAC. The setup is based on the MAC of the above example, where the output is the XOR sum of two binary inputs. Some binary noise is added to the channel output, that is independent of the channel inputs, i.e. we invert the MAC output with probability p_{MAC}. Using the identity mapping at the relay and a uniform input distribution leads to $I(X_1; \hat{Y}_R|X_2, Q) = 1 - h(p_{MAC})$ and $I(X_1X_2; \hat{Y}_R|Q) - I(\hat{Y}_R; Y_R|Q) = -h(p_{MAC})$. Therefore we get some penalty if we use this strategy. This penalty is caused by the quantization: We spent some bits to describe the MAC output, which contains noise that has no information for the receivers. Whenever $I(X_1X_2; \hat{Y}_R|Q) - I(\hat{Y}_R; Y_R|Q) = -I(\hat{Y}_R; Y_R|X_1, X_2, Q) < 0$ the noise is still included in quantized representative. Therefore some bits are wasted on the noise. We can decrease this penalty be using a less fine quantization. One way[2] of achieving this is to use a quantized variable such that $p(\hat{y}_R|y_R)$ is a binary symmetric channel with crossover probability p_Q. Thereby we degrade the MAC performance and we have for uniform distributed channel inputs $I(X_1X_2; \hat{Y}_R|Q) - I(\hat{Y}_R; Y_R|Q) = h(p_Q) - h(p_{MAC} + p_Q - 2p_{MAC}p_Q)$ and $I(X_1; \hat{Y}_R|X_2, Q) = 1 - h(p_{MAC} + p_Q - 2p_{MAC}p_Q)$. Now suppose we optimize α to achieve a high rate for receiver 1 ignoring the rate of receiver 2. Fixing the strategy as discussed above leads to

$$\alpha = \frac{1 - h(p_{1,BC})}{2 - h(p_Q) - h(p_{1,BC})}$$

and a maximum rate

$$R_2 = \frac{1 - h(p_{1,BC})}{2 - h(p_Q) - h(p_{1,BC})}(1 - h(p_{MAC} + p_Q - 2p_{MAC}p_Q)).$$

Therefore we can calculate the optimal parameter p_Q for this strategy and for the rate R_2. In Figure 4.2 the rate R_2 is plotted over the parameter p_Q for the quantization assuming fixed $p_{MAC} = 0.3$ and $p_{1,BC} = 0.2$ and a corresponding optimal α. It turns out that the optimal p_Q depends on both, p_{MAC} and $p_{1,BC}$. In particular, the optimal parameters p_Q and α will be different if the goal is to maximize R_1. The figure shows, that the degradation of the MAC output can increase the rate in the overall communication.

Similar effects, i.e. the degradation of the performance in one of the transmission steps to increase the overall performance, might be used to increase the rate of R_1 in the first example of the noiseless MAC without joint decoding above: In that example using a non-uniform distri-

[2]We do not argue that this is the optimal way of quantization. But this quantization serves the purpose to show some effects which can occur for the compress-and-forward strategy with joint decoding.

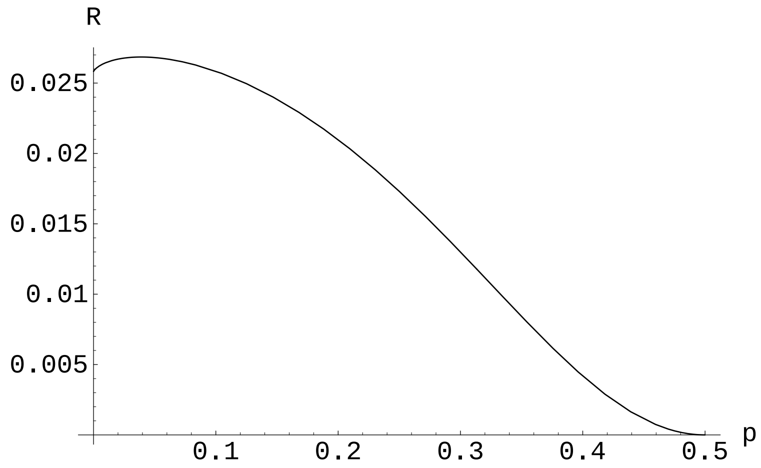

Figure 4.2: The figure the achievable rate $R = R_1$ for receiver 1 over the parameter $p = p_Q$ for the quantization assuming fixed $p_{MAC} = 0.3$ and $p_{1,BC} = 0.2$.

bution for X_2 allows for larger α. This in turn increases the rate R_1. With this strategies for the example at hand the rate region with joint decoding and without joint decoding are the same (up to boundary effects du to the strict inequalities in the constraint (3.2)), but different strategies need to be used. Note that in general it is not possible to use a different input distribution on X_1 without affecting the rates achievable for R_2 in the MAC phase.

Back to the example with a noisy MAC: Suppose for now we choose α and p_Q to maximize R_1. We have $h(p_Q) - h(p_{MAC} + p_Q - 2p_{MAC}p_Q) < 0$ for $p_Q \neq 0.5$. Therefore to achieve the maximum rate R_1 it might be necessary to set $R_2 = 0$ if $p_{1,BC}$ is close to 0.5. If the BC is orthogonal for both receivers, a simple solution for the two problems sketched above is to use a different quantization for the two receivers. This strategy is analyzed in Chapter 5. Note that we cannot use different α, as this parameter determines the timesharing and is common for both receivers.

The example with the noisy MAC output leads the way to a different strategy at the relay: There exists [64, 65] a capacity achieving sequence of linear codes for the binary symmetric channel with parameter p_{MAC}. Now, if both nodes use the same of these linear codes, the XOR sum of the codewords is again a codeword du to the linearity of the code. Therefore the relay will be able to decode the XOR sum of the two messages. As already pointed out in Chapter 2, the relay can use the XOR sum of the messages as an input of coding for the BC with side information at the receivers. Therefore we conclude that for this channel the cutset bound given in Lemma 1.1 is achievable. The strategy in this example is based on the structure of the codes. Because of this structure the relay need to choose one out of only $2^{\alpha n \max\{R_1,R_2\}}$ codewords. Such a reduction of the number of effective codewords the relay has to pick from cannot be ensured with a random coding argument, unless one of the nodes uses all available codewords as in the case if there is no noise in the MAC. Until today the achievable rate region with such structured codes is only known for very few channels.

In [66] a structured code based on nested lattices for certain Gaussian channels was proposed and an achievable rate region was stated. A related topic was introduced as computational coding in [67]. In computational coding the goal is to receive a certain function of several random variables at the receiver of a MAC. Receiving the XOR sum can be seen to be an example of such a computational code. For general channels the achievable rate region for decoding the XOR sum of two messages with structured coding remains unknown. A related result for source coding can be found in [68, 35]. In this references structured codes are used to encode the XOR sum of two dependent random variables. During the work on this thesis we obtained some results for the transmission of correlated binary data via an AWGN MAC, which are related to the design of structured codes. These results were published in [3, 4]. The analysis of computational coding for a MAC is out of the scope of this thesis.

4.2 Partial Decode-and-Forward with Joint Decoding at the Receiver

Similar to the proceeding in Section 3.3.1 the above coding scheme can be superimposed on a decode-and-forward scheme. The resulting coding partially decodes the messages at the relay. The complement information is forwarded by a compress-and-forward approach. At the receiver, a joint decoding is used to decode the compress-and-forward part of the messages.

4.2.1 Coding Theorem

Theorem 4.3. Let $\mathcal{R}_8 \subset \mathbb{R}_+^4$ be the set of all $\left[R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)}\right]$ satisfying

$$
\begin{aligned}
R_1^{(1)} &\leq \min\Big\{\alpha I(U_1; Y_R|U_2, Q), \beta I(V; Y_2)\Big\}\\
R_2^{(1)} &\leq \min\Big\{\alpha I(U_2; Y_R|U_1, Q), \beta I(V; Y_1)\Big\}\\
R_1^{(1)} + R_2^{(1)} &\leq \alpha I(U_1 U_2; Y_R|Q)\\
R_1^{(2)} &\leq \max\Bigg\{\min\Bigg\{ \alpha I(X_1; \hat{Y}_R|X_2, U_1),\\
&\qquad \alpha\Big(I(X_1X_2; \hat{Y}_R|U_1, U_2) - I(Y_R; \hat{Y}_R|U_1, U_2)\Big) + \beta I(X_R; Y_2|V)\Bigg\}, 0\Bigg\}\\
R_2^{(2)} &\leq \max\Bigg\{\min\Bigg\{ \alpha I(X_2; \hat{Y}_R|X_1, U_2),\\
&\qquad \alpha\Big(I(X_1X_2; \hat{Y}_R|U_1, U_2) - I(Y_R; \hat{Y}_R|U_1, U_2)\Big) + \beta I(X_R; Y_1|V)\Bigg\}, 0\Bigg\}
\end{aligned}
\tag{4.5}
$$

for some joint probability distributions $p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)$ and $p(v)p(x_R|v)p_2(y_1, y_2|x_R)$ and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.

An achievable rate region for the two-phase two-way relay channel using a partial decode-and-forward protocol is the set $\mathcal{R}_{\text{PCF-JD}} \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ such that there exists

$\left[R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)}\right] \in \text{ConvexHull}(\mathcal{R}_8)$ with $R_1^{(1)} + R_1^{(2)} = R_1$, $R_2^{(1)} + R_2^{(2)} = R_2$.

Remark 4.6 (Cardinalities of all auxiliary variables)**.** All the cardinalities of all auxiliary variables can be bounded using the Fenchel-Bunt extension of Caratheodory's theorem [62]. The proof of this claim is along the lines of the bounding of the auxiliary variables in the other theorems. These are given in Section 3.2 and in the appendices. The proof for this rate region does not contain any new elements. Therefore it is skipped for brevity. The region $\mathcal{R}_8$ contains the region $\mathcal{R}_{\text{DF}}$ and the region $\mathcal{R}_{\text{CF-JD}}$ as special cases. The region of $\mathcal{R}_{\text{DF}}$ is obtained by choosing $|\hat{\mathcal{Y}}_R| = 1$ $U_1 = X_1$, $U_2 = X_2$, and $V = X_R$; we obtain the regions $\mathcal{R}_{\text{CF-JD}}$ by choosing $U_1 = U_2 = Q$ and $|\mathcal{V}| = 1$.

4.2.2 Proof of the Coding Theorem

Proof. The proof is an extension of the proof of Theorem 3.5 analogous to the proceeding in the proof of Theorem 4.1. We will restrict the proof here to those things that change compared to the proof of Theorem 3.5: What is replaced is the decoding at the terminal nodes and the choice of the parameter ϵ_q. For the choice of ϵ_q we now have:

- Let $\epsilon_q^{(1)} = \frac{1}{2\alpha+\beta}\Big(\beta I(X_R;Y_1|V) - \alpha\Big(I(\hat{Y}_R;X_1,X_2|U_1,U_2) + I(\hat{Y}_R;Y_R|U_1,U_2)\Big) - R_2^{(2)}\Big)$, $\epsilon_q^{(2)} = \frac{1}{2\alpha+\beta}\Big(\beta I(X_R;Y_2|V) - \alpha\Big(I(\hat{Y}_R;X_1,X_2|U_1,U_2) + I(\hat{Y}_R;Y_R|U_1,U_2)\Big) - R_1^{(2)}\Big)$, $\epsilon_q^{(3)} = \frac{I(X_2;\hat{Y}_R|X_1,U_2) - \frac{R_2^{(2)}}{\alpha}}{3}$, and $\epsilon_q^{(4)} = \frac{I(X_1;\hat{Y}_R|X_2,U_2) - \frac{R_1^{(2)}}{\alpha}}{3}$. Choose $\epsilon_q \in (0, \min\{\epsilon_q^{(1)}, \epsilon_q^{(2)}, \epsilon_q^{(3)}, \epsilon_q^{(4)}\})$.

Some parameters for the decoding sets are changed. They are now chosen as $\epsilon_2 = \epsilon_4 = \epsilon_5 = \epsilon_6 \in (0, \frac{\epsilon_q}{8})$. Furthermore we have decoding steps vi and vii replaced by the new decoding step $\hat{\text{iv}}$:

$\hat{\text{vi}}$ Upon receiving $y_1^{n_2}$ and knowing $w_2^{(1)}$, $w_1^{(1)}$ and $w_1^{(2)}$ node 1 decides that $w_2(w_2^{(1)}, w_2^{(2)})$ was transmitted if $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$ is the only codeword such that for some $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_Q \rceil}\}$ the sequences $x_2^{n_1}(w_2^{(2)}|w_2^{(1)})$, $\hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})$, and $x_1^{n_1}(w_1^{(2)}|w_1^{(1)})$ are jointly typical, and simultaneously $x_R^{n_2}(i|w_2^{(1)}, w_1^{(1)})$ and the received signal $y_1^{n_2}$ are jointly typical, i.e. $\exists i \in \{1, 2, \dots, 2^{\lceil \alpha n R_Q \rceil}\}$ such that $\left(x_R^{n_2}(i|w_2^{(1)}, w_1^{(1)}), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_4}^{(n_2)}\left(X_R, Y_1|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)$ and simultaneously $\left(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}), x_2^{n_1}(w_2^{(2)}|w_2^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\right) \in \mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1, X_2, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$. If no or more than one such codewords is found node 1 chooses $w_2(w_2^{(1)}, w_2^{(2)}) = 1$.

We can reuse all the error events but E_8. Therefore we replace E_8 by a new error event:

- $\hat{E}_8$: Suppose a codebook is given, $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted, the relay chose some i and $x_R^{n_2}(i)$ is transmitted. $\hat{E}_8$ is the event that for some $j \neq i$ and $\hat{w}_2^{(2)} \neq w_2^{(2)}$ we have $\left(x_R^{n_2}(i|w_2^{(1)}, w_1^{(1)}), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_4}^{(n_2)}\left(X_R, Y_1|v^{n_2}(w_1^{(1)}, w_2^{(1)})\right)$ and simultaneously $\Big(x_1^{n_1}(w_1^{(2)}|w_1^{(1)}),$ $x_2^{n_1}(\hat{w}_2^{(2)}|w_2^{(1)}), \hat{y}_R^{n_1}(i|w_1^{(1)}, w_2^{(1)})\Big) \in \mathcal{T}_{\epsilon_6}^{(n_1)}\left(X_1, X_2, \hat{Y}_R|u_1^{n_1}(w_1^{(1)}), u_2^{n_1}(w_2^{(1)})\right)$

Note that the seemingly missing event $j = i$ is already captured by E_{10}. Now we bound the probability for the new error event for receiver 1.

Error event $\hat{E}_8$

$$\mathbb{E}\{Pr[\hat{E}_4]\} \leq \sum_{v^{n_2}\in\mathcal{V}^{n_2}} \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1})p(v^{n_2})Pr[\hat{E}_{8,1}]Pr[\hat{E}_{8,2}]2^{\lceil\alpha n R_Q\rceil}2^{\lfloor nR_2^{(2)}\rfloor}$$

Here $\hat{E}_{8,1}$ is the event, that given $u_1^{n_1}, u_2^{n_1}$, and q^{n_1} for three sequences $x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}$ drawn independent of each other we have $(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) \in \mathcal{T}_{\epsilon_6}^{(n_1)}(X_1, X_2, \hat{Y}_R|u_1^{n_1}, u_1^{n_1})$. $x_1^{n_1}, x_2^{n_1}$, and $\hat{y}_R^{n_1}$ are drawn at random according to $p(x_1^{n_1}|u_1^{n_1})$, $p(x_2^{n_1}|u_2^{n_1})$ and $p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1})$ respectively to capture the averaging over the random codebooks and the known side information. $\hat{E}_{8,2}$ is the event, that given v^{n_2} for two sequences $x_R^{n_2}, y_1^{n_2}$ drawn independent of each other according to $p(x_R^{n_2}|v^{n_2})$ and $p(y_1^{n_2}|v^{n_2})$ we have $(x_R^{n_2}, y_1^{n_2}) \in \mathcal{T}_{\epsilon_4}^{(n_2)}(X_R, X_1|v^{n_2})$. The factor $2^{\lceil\alpha n R_Q\rceil}$ accounts for the fact that we can use a union bound and the error occurs if at least one $j \neq i$ is found fulfilling the requirements. The factor $2^{\lfloor nR_2^{(2)}\rfloor}$ accounts for the different possible $\hat{w}_2^{(2)}$.

For sufficiently large n

$$\begin{aligned} Pr[\hat{E}_{8,1}] = &\sum_{x_1^{n_1}\in\mathcal{X}_1^{n_1}} \sum_{x_2^{n_1}\in\mathcal{X}_2^{n_1}} \sum_{\hat{y}_R^{n_1}\in\hat{\mathcal{Y}}_R^{n_1}} p(x_1^{n_1}|u_1^{n_1})p(x_2^{n_1}|u_2^{n_1})p(\hat{y}_R^{n_1}|u_1^{n_1}, u_2^{n_1}) \\ &\qquad\qquad \times \chi_{\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1,X_2,\hat{Y}_R|u_1^{n_1},u_2^{n_1})}(x_1^{n_1}, x_2^{n_1}, \hat{y}_R^{n_1}) \\ &\leq |\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1, X_2, \hat{Y}_R|u_1^{n_1}, u_2^{n_1})| 2^{-n_1\left(H(X_1|U_1)-2\epsilon_6\right)}2^{-n_1\left(H(X_2|U_2)-2\epsilon_6\right)}2^{-n_1\left(H(\hat{Y}_R|U_1,U_2)-2\epsilon_6\right)} \end{aligned}$$

due to the properties of the typical set. Furthermore, it follows from these properties that

$$|\mathcal{T}_{\epsilon_6}^{(n_1)}(X_1, X_2, \hat{Y}_R|u_1^{n_1}, u_2^{n_1})| \leq 2^{n_1\left(H(X_1,X_2,\hat{Y}_R|U_1,U_2)+2\epsilon_6\right)}.$$

$Pr[\hat{E}_{8,2}]$ can be bounded in a similar way. Therefore we have for sufficient large n

$$\begin{aligned} \mathbb{E}\{Pr(\hat{E}_8)\} \leq &\sum_{v^{n_2}\in\mathcal{V}^{n_2}} \sum_{u_1^{n_1}\in\mathcal{U}_1^{n_1}} \sum_{u_2^{n_1}\in\mathcal{U}_2^{n_1}} \sum_{q^{n_1}\in\mathcal{Q}^{n_1}} p(q^{n_1}, u_1^{n_1}, u_2^{n_1})p(v^{n_2})2^{-n_1\left(I(X_1X_2;\hat{Y}_R|U_1,U_2)-8\epsilon_6\right)} \\ &\qquad 2^{-n_2\left(I(X_R;Y_1|V)-6\epsilon_4\right)}2^{\alpha n R_Q+1}2^{nR_2^{(2)}} \\ \leq\ & 2^{-n\left(\alpha\left(I(X_1X_2;\hat{Y}_R|U_1,U_2)-R_Q-8\epsilon_6\right)+\beta\left(I(X_R;Y_1|V)-6\epsilon_4\right)-R_2^{(2)}\right)+1+I(X_1X_2;\hat{Y}_R|U_1,U_2)+6\epsilon_4} \\ &= 2^{-n\left(\alpha\left(I(X_1X_2;\hat{Y}_R|U_1,U_2)-I(Y_R;\hat{Y}_R|U_1,U_2)\right)+\beta I(X_R;Y_1|V)-R_2^{(2)}-\tilde{\epsilon}\right)+1+I(X_1X_2;\hat{Y}_R|U_1,U_2)+6\epsilon_4} \end{aligned}$$

with

$$\tilde{\epsilon} = \alpha\epsilon_q + \beta 6\epsilon_4 + \alpha 8\epsilon_6 < (2\alpha + \beta)\epsilon_q.$$

This term goes to zero if

$$\alpha\left(I(X_1X_2; \hat{Y}_R|U_1, U_2) - I(Y_R; \hat{Y}_R|U_1, U_2)\right) + \beta I(X_R; Y_1|V) - R_2^{(2)} - (2\alpha + \beta)\epsilon_q > 0$$

By the choice of ϵ_q as

$$\epsilon_q < \frac{1}{(2\alpha+\beta)}\left(\alpha\left(I(X_1X_2;\hat{Y}_R|U_1,U_2) - I(Y_R;\hat{Y}_R|U_1,U_2)\right) + \beta I(X_R;Y_1|V) - R_2^{(2)}\right)$$

this is true whenever

$$R_2^{(2)} < \alpha\left(I(X_1X_2;\hat{Y}_R|U_1,U_2) - I(Y_R;\hat{Y}_R|U_1,U_2)\right) + \beta I(X_R;Y_1|V)$$

as required by the assumption. Therefore for any given ϵ we can find $n^{(8)}$ such that $\mathbb{E}\{Pr[\hat{E}_8]\} < \frac{\epsilon}{10}$ for $n \geq n^{(8)}$.

$\square$

4.3 Concluding remarks

In this chapter, the compress-and-forward scheme proposed in Chapter 3 is extended by applying a joint decoding mechanism. An achievable rate region for a two-phase protocol was stated in Theorem 4.1. The gain originates from the observation, that the MAC phase induces some dependencies to the system, which remain unused by the compress-and-forward scheme, namely we can use the joint typicality between both inputs and the output of the MAC to restrict the effective code in the BC phase.

The new scheme has some interesting properties. First it is not required, that the relay can decode the data. Furthermore the terminal node may not be able to decode the relay's transmission without decoding the message of the other node; even more, the receiver is not required to decode the relay's transmission without error. Surprisingly, the relaying of both messages via the relay is possible anyhow.

For certain channels the achievable rate region given in Theorem 4.1 equals the outer bound on the capacity region given by the cutset bound in Lemma 1.1. In general this bound cannot be achieved with the proposed coding technique, as some penalty term caused by the quantization occurs. A promising way to overcome this penalty is sketched in Section 4.1.4. Instead of a quantization the relay decodes the XOR sum of both messages. In fact for the examples where we achieve the capacity the quantization can be interpreted as decoding the XOR sum. Unfortunately for general channels no rate region is known, that is achievable with this strategy. It seems inevitable to use structured codes as in [66, 68, 35, 67] to achieve high rates. Proofs based on random coding argument seem incapable to realize the possible gains. Computational coding [67], i.e. decoding the XOR sum of two messages at the relay, promises larger achievable rate region; beyond the two-phase two-way relay channel, computational coding might be a powerful tool in many network scenarios to increase the achievable rates. Research has just started to focus on this kind of coding. Until today it is unclear what gains can be achieved.

The rate regions given in this chapter still show some ugly boundary effects. Therefore a convexification is needed. In the proof of Theorem 4.1 one can see where these non-continuity stem from. It is the compression at the relay, that adds this undesired property to the overall region. Take a look at (4.4). Note that the factor $2^{\lfloor nR_2 \rfloor}$ is a simplified upper bound. As one codeword is the correct one, $(2^{\lfloor nR_2 \rfloor} - 1)$ is a more precise factor. With this factor, it is obvious that with a change from $R_2 = 0$ to $R_2 > 0$ new conditions need to be satisfied. If $R_2 = 0$, then $\mathbb{E}\{Pr[\hat{E}_4]\} = 0$; but for $R_2 > 0$ the term $(2^{\lfloor nR_2 \rfloor} - 1)$ grows exponentially and needs to be compensated by adequate coding. This indicates that during the quantization an interference like effect occurs: The quantization for the two different flows of information should be performed with a different quality. While the flow with the higher rate needs a finer quantization, the flow that transports less information can be quantized rather coarsely. In the extrema of only one flow of information the quantization needs to focus only on this stream. While the extreme case is already captured by Theorem 4.1, the case of different needed qualities of quantization is not appropriately covered by the coding used in the proof.

In the next chapter we take a closer look at the different flows of information and extend the results given so far by using three flows of information at the relay at the same time: One flow goes to both receivers, and one flow goes to one of the receivers respectively. The resulting region is convex without further timesharing. It includes all the compress-and-forward regions in this thesis.

Chapter 5

Using More than one Representation for Compress-and-Forward

This chapter considers the question, how one can cope with the different information flows that occur in the system. The discussion in the previous chapters shows, that there might be a gain for one node, if the other node stays silent for some time. While in the general coding scheme we transmit a codeword containing information which is crucial for the decoding at both receivers, in this case only one of the receivers needs to get the information. This in turn results in an increase of freedom for the compression at the relay, and therefore may give a larger achievable rate for one of the nodes.

The key feature of the coding approaches is at the same time the crux: In the broadcast only one message is transmitted, which is decoded by both receivers. The feature of this approach is that by the virtue of the decoding we can get rid of interference, as there is no second codeword transmitted. The crux is, that the compression at the relay cannot differentiate between the information flows to the different nodes.

The quantization at the relay needs to be fine enough to capture all information for the information flow from both transmitters to the receivers. The side information available at the receiver restricts the number of codewords in the effective code, as it was discussed in Section 4.1.3. Still, the resulting effective code has in general some overhead, i.e. it contains more codewords then there are messages which could be transmitted. For example for receiver 1 the effective code in the joint decoding scheme contains approximately $2^{n\left(R_2+\alpha\left(I(\hat{Y}_R;Y_R|Q)-I(X_1X_2;\hat{Y}_R|Q)\right)\right)}$ codewords. Now, $I(\hat{Y}_R;Y_R|Q)-I(X_1X_2;\hat{Y}_R|Q)=I(\hat{Y}_R;Y_R|X_1,X_2)$ due to the Markov chain. The factor of this overhead is the same for both receivers, and it depends on the quality of the quantized representation for the MAC output. In Remark 4.3 we already discussed, that in general it is not possible to keep the overhead to zero. This leads to a tradeoff between overhead in the BC and a good representation to achieve a high overall rate. The problem is now, that the optimal tradeoff might be different for the two information flows. The following example outlines the problem that may occur:

Example 5.1. *Consider a setup where the channel in the BC phase to receiver 1 allows only very limited transmission, while the channel to receiver 2 has a very high capacity. At the same time the MAC allows high though limited rates for both nodes. Obviously the R_2 rate is bounded from above by the channel in the BC phase to receiver 1. Therefore this receiver demands only a corse representation of the MAC output. The corse representation keeps the overhead in the effective code small, while at the same time it is sufficient to achieve the best possible throughput in the overall system. At the same time receiver 2 demands a fine representation: it does not care about the overhead in the BC to much, and its rate is maximized by a fine representation. Clearly we need to tradeoff between the demands of both the users. If the representation is chosen such, that receiver 1 has optimal performance, this degrades the performance of receiver 2. In case we choose a fine representation to satisfy the demands of receiver 2 this will degrade the rate R_2. Therefore we have an effect similar to interference. But this does not emerge from the channel, but from the quantization at the relay.*

An approach that follows from the observation outlined in the example above is to separate the different flows of information at the relay and use different representations for the two receivers. Moreover, as some information about the MAC output may be useful for both receivers, in this chapter we consider three different information flows via the relay: one flow is directed to receiver 1 and receiver 2 respectively and a third flow goes to both receivers. As a result we can avoid the interference like effect in the quantization at the relay. This is bought by the need to transmit more then one information flow via the BC, and therefore in general we introduce new interference. The resulting coding scheme contains the coding schemes discussed in Section 4.1 as a special cases. The split of the information flows can only increase freedom in the optimization and therefore the performance of the overall system.

5.1 Extending the Region by using Three Data Streams

For the coding schemes in Chapter 3 and 4 it was necessary to convexify the achievable rate region by timesharing over different codes. As a result, to achieve the boundary of the achievable rate region it could be necessary to use more than one data stream. One information flow is from the relay to both receivers, while one flow is to one of the receivers respectively. The drawback of the convexification is, that the quantization at the relay as well as the BC transmission is performed in an orthogonal fashion for these three flows. This might be suboptimal.

Therefore, in this section we extend the previous results by using three data streams in the coding, that can be in use simultaneously, i.e. non-orthogonal. The technique used is an application of the mechanisms used in the proof of the achievable rate region for the BC [47, 69]. We sketch the proceeding for a simple case without superposition with a decode-and-forward mechanism. Furthermore, the result captures only a first approach in this direction and is therefore not complete. In the analysis we will point out, where further improvements are possible, and what difficulties arise with these more general approaches.

5.1.1 Coding Theorem

Theorem 5.1. An achievable rate region for the two-phase two-way relay channel using a compress-and-forward protocol is the set $\mathcal{R}_{\text{CF-JD-3S}} \subset \mathbb{R}_+^2$ of all rate pairs $[R_1, R_2]$ satisfying

$$\begin{aligned}
R_1 \le\ & \min\Big\{ \alpha I(X_1; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|X_2, Q); \\
& \qquad \alpha\Big(I(X_1X_2; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q) - I(Y_R; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q)\Big) + \beta I(UV_2; Y_2)\Big\} \\
R_2 \le\ & \min\Big\{ \alpha I(X_2; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|X_1, Q); \\
& \qquad \alpha\Big(I(X_1X_2; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q) - I(Y_R; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q)\Big) + \beta I(UV_1; Y_1)\Big\} \\
R_1 + R_2 \le\ & \alpha\Big(I(X_1X_2; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q) + I(X_1X_2; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q) \\
& \quad -I(Y_R; \hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q) - I(Y_R; \hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q)\Big) \\
& \quad +\beta\Big(I(UV_2; Y_2) + I(UV_1; Y_1) - I(V_1; V_2|U)\Big)
\end{aligned} \tag{5.1}$$

for some joint probability distributions $p(q)p(x_1|q)p(x_2|q)p(y_R|x_1, x_2)p(\hat{y}_{R,1+2}|y_R)p(\hat{y}_{R,1}|y_R, \hat{y}_{R,1+2})$ $p(\hat{y}_{R,2}|y_R, \hat{y}_{R,1+2})$ and $p(u, v_1, v_2)p(x_R|u, v_1, v_2)p(y_1, y_2|x_R)$ and some $\alpha, \beta > 0$ with $\alpha + \beta = 1$.

Remark 5.1. For the proof of the theorem we use a joint decoding mechanism. Clearly, an extension of the sequential decoding mechanism form Theorem 3.1 is possible, but it leads to a smaller region.

Remark 5.2 (Convexity of $\mathcal{R}_1$)**.** The region $\mathcal{R}_{\text{CF-JD-3S}}$ is convex. To see that it is convex for fixed α and β note, that one can add Q as a condition to all entropy and mutual information terms without changing the region. If we allow for different timesharing parameters α and β, then we can use arguments analogous to that in Remark 2.11 to prove, that the region is convex.

Remark 5.3 (Cardinality of the auxiliary variables cannot be bounded)**.** Note that the result is incomplete in the sense that we do not give an upper bound on the cardinality of the auxiliary variables V_1 and V_2. While the cardinality of Q, $\hat{Y}_{R,1}$, $\hat{Y}_{R,2}$, $\hat{Y}_{R,1+2}$, and U can be bounded[1] as before using the Fenchel-Bunt extension of Caratheodory's theorem [62], this cannot be done for these variables. The problematic term in the calculation is $I(V_1; V_2|U)$. Applying the method to the variables V_1 or V_2 leads to a bound, which depends on the cardinality of other respective variable. Therefore the cardinality bound has a recursive structure. As a consequence upper bounds cannot be calculated which depend solely on the fixed cardinalities of the input or output alphabets given by the system setup. A similar problem was pointed out in [63] for the broadcast channel with correlated sources.

Remark 5.4 (Possible extension to partial decoding)**.** We could easily extend the proof of Theorem 4.3 to this approach with three data streams. The result is a superposition of a decode-and-forward code and a compress-and-forward code with joint decoding using three data streams.

[1] As the proof does not give any new insights we skip it for brevity.

The proof would be a simple combination of the other proofs given in the thesis and does not give any new insight.

Remark 5.5 (Outlook)**.** The given proof does neither use the dependency between $\hat{Y}_{R,1}$ and $\hat{Y}_{R,2}$, nor the fact that both these variables depend on both codewords transmitted in the MAC. This dependencies could be used as additional side information in the decoding of the BC as they restrict the effective code used by the receiver for the decoding even further. Therefore this may lead to an even larger region. Furthermore, in the theorem the variables $\hat{Y}_{R,1}$ and $\hat{Y}_{R,2}$ are assumed to be independent given $\hat{Y}_{R,1+2}$ and Y_R. This was done to allow a straight forward — namely a separated — quantization of the three data streams at the relay. It is unclear, whether or not a more complex joint quantization could extend the achievable rate region. The coding approaches sketched in this remark are beyond the scope of this thesis and will not be analyzed here.

Proof. In what follows we extend the Proof of the Theorem 4.1. We will focus only on those things that change. In particular we skip the details, if the bounding of the error is a straight forward extension of some bounding in one of the proofs above, i.e. the result can be achieved with exactly the same technique but on slightly different sets.

As in the above proofs we start with assuming strict inequality in (5.1). The achievability of the closure and the case that one of the rates is restricted by 0 can be handled analogous to the procedure in the above proofs. Depending on which of the strict inequalities assumed in the proof are not valid, it may be necessary to adjust parts of the proof, especially the choice of the parameters for some of the the decoding sets or some of the three parameters for the quantization $\epsilon_{q,1+2}$, $\epsilon_{q,1}$, $\epsilon_{q,2}$. The arguments are similar to the effects with idle users and the treatment in the section about boundary effects in the above proofs and will not be handled in detail. By comparison it is easy to see, that the region proved in Theorem 4.1 is a special case of this theorem. Therefore it is also an extension of the region given in Theorem 3.1.

Note, that for some probability distributions the rate constraints in the theorem might be negative for one user. In contrast to the non-continuity effect of the rate region that we had in Theorem 4.1 we do not achieve any more freedom for the probability distribution, if one of the rates is set to 0. This can be seen by observing that in this case we can choose alphabets of cardinality 1 for e.g. $\hat{Y}_{R,1+2}$ and $\hat{Y}_{R,1}$ while all information for receiver 2 is carried by $\hat{Y}_{R,2}$. It turns out that this is the special case with an idle user, which is now included in the theorem. Furthermore, it is obvious that the region is convex, as we can add Q as a condition to all terms without changing the region; the minimum operation as well as the restriction to positive rates does not effect the convexity. Therefore no additional timesharing over codes is needed to convexify the region as it was needed in Corollary 4.2.

We start the proof by assuming that for R_1, R_2, and some positive $R_{V,1}$, $R_{V,1}$ the following inequalities hold for some probability distributions $p(q)p(x_1|q)p(x_2|q)p(y_R|x_1,x_2)p(\hat{y}_{R,1+2}|y_R)$

$p(\hat{y}_{R,1}|y_R,\hat{y}_{R,1+2})p(\hat{y}_{R,2}|y_R,\hat{y}_{R,1+2})$ and $p(u,v_1,v_2)p(x_R|u,v_1,v_2)p(y_1,y_2|x_R)$:

$$\begin{aligned} R_1 &< \alpha I(X_1;\hat{Y}_{R,2}\hat{Y}_{R,1+2}|X_2,Q) && (5.2)\\ R_2 &< \alpha I(X_2;\hat{Y}_{R,1}\hat{Y}_{R,1+2}|X_1,Q) && (5.3)\\ R_1+R_{V,2} &< \alpha\Big(I(X_1X_2;\hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q)-I(Y_R;\hat{Y}_{R,1+2}|Q)\Big)+\beta I(UV_2;Y_2) && (5.4)\\ R_2+R_{V,1} &< \alpha\Big(I(X_1X_2;\hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q)-I(Y_R;\hat{Y}_{R,1+2}|Q)\Big)+\beta I(UV_1;Y_1) && (5.5)\end{aligned}$$

$$\begin{aligned} R_{V,1} &> \alpha I(Y_R;\hat{Y}_{R,1}|\hat{Y}_{R,1+2},Q) && (5.6)\\ R_{V,2} &> \alpha I(Y_R;\hat{Y}_{R,2}|\hat{Y}_{R,1+2},Q) && (5.7)\\ R_{V,1}+R_{V,2} &> \alpha\Big(I(Y_R;\hat{Y}_{R,1}|\hat{Y}_{R,1+2},Q)+I(Y_R;\hat{Y}_{R,2}|\hat{Y}_{R,1+2},Q)\Big)+\beta I(V_1;V_2|U) && (5.8)\end{aligned}$$

Whenever we have strict inequality for R_1, R_2 in (5.1), we can find positive $R_{V,1}$, $R_{V,1}$ fulfilling these inequalities.

5.1.1.1 Random codebook generation

For a given n set $n_1 = \lfloor \alpha n \rfloor$, $n_2 = \lceil \beta n \rceil$.

- Choose one q^{n_1} drawn according to the probability $\prod_{s=1}^{n_1} p(q^{n_1}_{(s)})$.
- Choose $2^{\lfloor nR_1 \rfloor}$ i.i.d. codewords $x_1^{n_1}$ each according to the probability $\prod_{s=1}^{n_1} p(x^{n_1}_{1,(s)}|q^{n_1}_{(s)})$. Label these $x_1^{n_1}(w_1)$, $w_1 \in \{1,2,\ldots,2^{\lfloor nR_1\rfloor}\}$.
- Choose $2^{\lfloor nR_2 \rfloor}$ i.i.d. codewords $x_2^{n_1}$ each according to the probability $\prod_{s=1}^{n_1} p(x^{n_1}_{2,(s)}|q^{n_1}_{(s)})$. Label these $x_2^{n_1}(w_2)$, $w_2 \in \{1,2,\ldots,2^{\lfloor nR_2\rfloor}\}$.
- Let $\epsilon^{(1)}_{q,1+2} := \frac{1}{2\alpha+\beta}\Big(\alpha\Big(I(X_1X_2;\hat{Y}_{R,1}\hat{Y}_{R,1+2}|Q)-I(Y_R;\hat{Y}_{R,1+2}|Q)\Big)+\beta I(UV_1;Y_1)-R_2-R_{V,1}\Big)$, $\epsilon^{(2)}_{q,1+2} := \frac{1}{2\alpha+\beta}\Big(\alpha\Big(I(X_1X_2;\hat{Y}_{R,2}\hat{Y}_{R,1+2}|Q)-I(Y_R;\hat{Y}_{R,1+2}|Q)\Big)+\beta I(UV_2;Y_2)-R_1-R_{V,2}\Big)$, $\epsilon^{(3)}_{q,1+2} := \frac{I(X_2;\hat{Y}_{R,1+2}\hat{Y}_{R,1}|X_1,Q)-\frac{R_2}{\alpha}}{4}$ and $\epsilon^{(4)}_{q,1+2} := \frac{I(X_1;\hat{Y}_{R,1+2}\hat{Y}_{R,2}|X_2,Q)-\frac{R_1}{\alpha}}{4}$. Choose the parameter for the quantization $\epsilon_{q,1+2} \in \Big(0,\min\{\epsilon^{(1)}_{q,1+2},\epsilon^{(2)}_{q,1+2},\epsilon^{(3)}_{q,1+2},\epsilon^{(4)}_{q,1+2}\}\Big)$.
- Choose $\epsilon_{q,1} \in (0,\min\{\epsilon^{(1)}_{q,1},\epsilon^{(2)}_{q,1},\epsilon^{(3)}_{q,1}\})$ where $\epsilon^{(1)}_{q,1} := \frac{R_{V_1}}{\alpha} - I(Y_R;\hat{Y}_{R,1}|\hat{Y}_{R,1+2},Q)$, $\epsilon^{(2)}_{q,1} := \frac{I(X_2;\hat{Y}_{R,1+2}\hat{Y}_{R,1}|X_1,Q)-\frac{R_2}{\alpha}}{4}$, and $\epsilon^{(3)}_{q,1} := \frac{R_{V,1}+R_{V,2}-\alpha\big(I(Y_R;\hat{Y}_{R,1}|\hat{Y}_{R,1+2},Q)+I(Y_R;\hat{Y}_{R,2}|\hat{Y}_{R,1+2},Q)\big)-\beta I(V_1;V_2|U)}{2\alpha}$.
- Choose $\epsilon_{q,2} \in (0,\min\{\epsilon^{(1)}_{q,2},\epsilon^{(2)}_{q,2},\epsilon^{(3)}_{q,2}\})$ where $\epsilon^{(1)}_{q,2} := \frac{R_{V_2}}{\alpha} - I(Y_R;\hat{Y}_{R,2}|\hat{Y}_{R,1+2},Q)$, $\epsilon^{(2)}_{q,2} := \frac{I(X_1;\hat{Y}_{R,1+2}\hat{Y}_{R,2}|X_2,Q)-\frac{R_1}{\alpha}}{4}$, and $\epsilon^{(3)}_{q,2} := \frac{R_{V,1}+R_{V,2}-\alpha\big(I(Y_R;\hat{Y}_{R,1}|\hat{Y}_{R,1+2},Q)+I(Y_R;\hat{Y}_{R,2}|\hat{Y}_{R,1+2},Q)\big)-\beta I(V_1;V_2|U)}{2\alpha}$.
- For each $i \in \{1,2,\ldots,2^{\lceil \alpha n R_{Q,1+2}\rceil}\}$, $R_{Q,1+2} = I(Y_R;\hat{Y}_{R,1+2}|Q)+\epsilon_{q,1+2}$, choose one codeword $\hat{y}^{n_1}_{R,1+2}(i)$ according to $\prod_{s=1}^{n_1} p(\hat{y}^{n_1}_{R,1+2,(s)}|q^{n_1}_{(s)})$ and one codeword $u^{n_2}(i)$ according to $\prod_{s=1}^{n_2} p(u^{n_2}_{(s)})$. The $2^{\lceil \alpha n R_{R,1+2}\rceil}$ codeword pairs are drawn i.i.d..

- For every $u^{n_2}(i)$, $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_{Q,1+2} \rceil}\}$, draw independently $2^{\lfloor n R_{V,1} \rfloor}$ codewords $v_1^{n_2}$ according to $\prod_{s=1}^{n_2} p(v_{1,(s)}^{n_2} 1 | u_{(s)}^{n_2})$. Label these $v_1^{n_2}(j, l|i)$, $l \in \{1, 2, \dots, 2^{\lfloor n R_{V,1} \rfloor}\}$, where j is a bin index defined by $j := l \mod 2^{\lceil \alpha n R_{Q,1} \rceil}$ with $R_{Q,1} = I(Y_R; \hat{Y}_{R,1} | \hat{Y}_{R,1+2}, Q) + \epsilon_{q,1}$.

- For every $u^{n_2}(i)$, $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_{Q,1+2} \rceil}\}$, draw independently $2^{\lfloor n R_{V,2} \rfloor}$ codewords $v_2^{n_2}$ according to $\prod_{s=1}^{n_2} p(v_{2,(s)}^{n_2} 2 | u_{(s)}^{n_2})$. Label these $v_2^{n_2}(k, m|i)$, $m \in \{1, 2, \dots, 2^{\lfloor n R_{V,2} \rfloor}\}$, where m is a bin index defined by $m := k \mod 2^{\lceil \alpha n R_{Q,2} \rceil}$ with $R_{Q,2} = I(Y_R; \hat{Y}_{R,2} | \hat{Y}_{R,1+2}, Q) + \epsilon_{q,2}$.

- For each pair (i, j), $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_{Q,1+2} \rceil}\}$, $j \in \{1, 2, \dots, 2^{\lceil \alpha n R_{Q,1} \rceil}\}$, choose one codeword $\hat{y}_{R,1}^{n_1}(j|i)$ according to $\prod_{s=1}^{n_1} p(\hat{y}_{R,1,(s)}^{n_1} | \hat{y}_{R,1+2,(s)}^{n_1}(i), q_{(s)}^{n_1})$.

- For each pair (i,k), $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_{Q,1+2} \rceil}\}$, $k \in \{1, 2, \dots, 2^{\lceil \alpha n R_{Q,2} \rceil}\}$, choose one codeword $\hat{y}_{R,2}^{n_1}(k|i)$ according to $\prod_{s=1}^{n_1} p(\hat{y}_{R,2,(s)}^{n_1} | \hat{y}_{R,1+2,(s)}^{n_1}(i), q_{(s)}^{n_1})$.

This constitutes a random codebook.

5.1.1.2 Decoding sets

For the decoding we will use typical set decoding. For a strict definition of the decoding sets we choose parameter for the typical sets as $\epsilon_1 = \epsilon_2 = \epsilon_4 = \epsilon_5 = \epsilon_6 \in (0, \min\{\frac{\epsilon_{q,1+2}}{8}; \frac{\epsilon_{q,1}}{8}; \frac{\epsilon_{q,2}}{8}\})$ and $\epsilon_7 < \frac{R_{V,1} + R_{V,2} - \alpha\left(I(Y_R; \hat{Y}_{R,1} | \hat{Y}_{R,1+2}, Q) + I(Y_R; \hat{Y}_{R,2} | \hat{Y}_{R,1+2}, Q) + \epsilon_{q,1} + \epsilon_{q,2}\right) - \beta I(V_1; V_2 | U)}{18\beta}$. The missing parameters for the receiver 2 are chosen in an analogous way.

5.1.1.3 Coding

i To transmit message w_1 node 1 sends $x_1^{n_1}(w_1)$.

ii To transmit message w_2 node 2 sends $x_2^{n_1}(w_2)$.

iii Upon receiving $y_R^{n_1}$ the relay looks for the first i such that $\left(y_R^{n_1}, \hat{y}_{R,1+2}^{n_1}(i)\right) \in \mathcal{T}_{\epsilon_1}^{(n_1)}\left(Y_R, \hat{Y}_{R,1+2} | q^{n_1}\right)$. If no such i is found the relay chooses[2] $i = 1$. Thereafter the relay looks for the first j such that $\left(y_R^{n_1}, \hat{y}_{R,1}^{n_1}(j|i)\right) \in \mathcal{T}_{\epsilon_5}^{(n_1)}\left(Y_R, \hat{Y}_{R,1} | \hat{y}_{R,1+2}^{n_1}(i), q^{n_1}\right)$, and for the first k such that $\left(y_R^{n_1}, \hat{y}_{R,2}^{n_1}(k|i)\right) \in \mathcal{T}_{\epsilon_6}^{(n_1)}\left(Y_R, \hat{Y}_{R,2} | \hat{y}_{R,1+2}^{n_1}(i), q^{n_1}\right)$. If no such j (k) is found the relay chooses $j = 1$ ($k = 1$). This induces a mapping $f : \mathcal{Y}_R^{n_1} \to C_{\hat{y}_{R,1+2}}^{(n)}(q^{n_1}) \times C_{\hat{y}_{R,1}}^{(n)}(\hat{y}_{R,1+2}^{n_1}, q^{n_1}) \times C_{\hat{y}_{R,2}}^{(n)}(\hat{y}_{R,1+2}^{n_1}, q^{n_1})$. Now, the relay looks for the first triple $\left(u^{n_2}(i), v_1^{n_2}(j, l|i), v_2^{n_2}(k, m|i)\right)$ such that $\left(u^{n_2}(i), v_1^{n_2}(j, l|i), v_2^{n_2}(k, m|i)\right) \in \mathcal{T}_{\epsilon_7}^{(n_1)}(U, V_1, V_2)$. If such a triple is found the relay transmits a random $x_R^{n_2}$ drawn according to $p\left(x_R^{n_2} | u^{n_2}(i), v_2^{n_2}(j, l|i), v_2^{n_2}(k, m|i)\right)$. If no such triple is found an arbitrary $x_R^{n_2}$ is transmitted.

iv Upon receiving $y_1^{n_2}$ node 1 decides that w_2 was transmitted if $x_2^{n_1}(w_2)$ is the only codeword such that for some $i \in \{1, 2, \dots, 2^{\lceil \alpha n R_Q \rceil}\}$, some $j \in \{1, 2, \dots, 2^{\lceil \alpha n R_{Q,1} \rceil}\}$, and some

[2] This is done to have a well defined error probability. Equivalently one could declare an error at the relay, but this induces a much more cumbersome notation in the definition of the error probability. Similar arguments apply for the other default choices.

$l \in \{1, 2, \ldots, 2^{\lfloor nR_{V,1} \rfloor}\}$ the sequences $x_2^{n_1}(w_2)$, $\hat{y}_{R,1+2}^{n_1}(i)$, $\hat{y}_{R,1}^{n_1}(j|i)$ and $x_1^{n_1}(w_1)$ are jointly typical given q^{n_1}, and simultaneously $u^{n_2}(i)$, $v_1^{n_2}(j, l|i)$ and the received signal $y_1^{n_2}$ are jointly typical, i.e $\exists(i, j, l)$ with $i \in \{1, 2, \ldots, 2^{\lceil \alpha n R_Q \rceil}\}$, $j \in \{1, 2, \ldots, 2^{\lceil \alpha n R_{Q,1} \rceil}\}$, $l \in \{1, 2, \ldots, 2^{\lfloor nR_{V,1} \rfloor}\}$ such that $\left(x_1^{n_1}(w_1), x_2^{n_1}(w_2), \hat{y}_{R,1+2}^{n_1}(i), \hat{y}_{R,1}^{n_1}(j|i)\right) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_{R,1+2}, \hat{Y}_{R,1}|q^{n_1})$ and simultaneous $\left(u^{n_2}(i), v_1^{n_2}(j, l|i), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(U, V_1, Y_1)$.

v The decoding at node 2 is performed in a analogous way.

5.1.1.4 Error Events

The error events do not change to much even tough the coding seems to have changed dramatically. In fact the event E_1 can be reused without change. This leads to the condition $\epsilon_{q,1+2} > 6\epsilon_1$. For the other two variables $\hat{Y}_{R,1}$ and $\hat{Y}_{R,2}$ a similar new event can be defined and bounded using the same tools, but conditioned on both, Q and $Y_{R,1+2}$. From the definition of the typical set and the factorization constraint of the probability distribution, this is sufficient to proof that we will find a triple (i, j, k) with probability arbitrarily close to 1 for n sufficient large. The bounding uses the assumption that $\epsilon_{q,1} > 6\epsilon_5$ and $\epsilon_{q,2} > 6\epsilon_6$.

E_2 can be changed into events stating that the pair $(u^{n_2}(i), v_1^{n_2}(j, l|i))$ is jointly typical with $y_1^{n_2}$ with high probability, which is obvious. Similar arguments apply for E_3 and E_5 where $\hat{Y}_R$ is replaced by the pair $(\hat{Y}_{R,1+2}, \hat{Y}_{R,1})$ for receiver 1. The proofs uses the assumption that $\epsilon_2 = \epsilon_4$. Furthermore E_5 can be easily proved if we use the assumption $\epsilon_1 = \epsilon_4 = \epsilon_5 = \epsilon_6$.

The replacement of $\hat{Y}_R$ by $(\hat{Y}_{R,1+2}, \hat{Y}_{R,1})$ in the proof can also be used for E_6, now yielding the requirement

$$R_2 < \alpha I(X_2; \hat{Y}_{R,1+2}\hat{Y}_{R,1}|X_1, Q)$$

for receiver 1 and

$$R_1 < \alpha I(X_1; \hat{Y}_{R,1+2}\hat{Y}_{R,2}|X_2, Q)$$

for receiver 2. In the bounding it is used that $\epsilon_1 = \epsilon_4 = \epsilon_5 = \epsilon_6$; $\epsilon_{q,1} < \frac{I(X_2;\hat{Y}_{R,1+2}\hat{Y}_{R,1}|X_1,Q)-\frac{R_2}{\alpha}}{4}$, and $\epsilon_{q,1+2} < \frac{I(X_2;\hat{Y}_{R,1+2}\hat{Y}_{R,1}|X_1,Q)-\frac{R_2}{\alpha}}{4}$.

What needs to be changed is the calculation and definition of error events for the joint decoding, i.e. $\hat{E}_4$ from the proof of Theorem 4.1. Furthermore we now have an additional event E_7, that captures an error in the encoding at the relay, i.e. the event, that there is no jointly typical triple $u^{n_2}(i)$, $v_2^{n_2}(j, l|i)$, $v_2^{n_2}(k, m|i)$ for the given (i, j, k).

Next, we give a definition of the changed error events E_4 and the new event E_7

- E_4: Suppose a codebook is given, $x_1^{n_1}(w_1), x_2^{n_1}(w_2)$ are transmitted, the relay chose some i,j,k and $u^{n_2}(i)$, and some $v_1^{n_2}(j, l|i)$, $v_2^{n_2}(k, m|i)$. An accordant $x_R^{n_2}$ is transmitted. E_4 is the event that there exists a pair $(\hat{i}, \hat{j}) \neq (i, j)$, some $\hat{w}_2 \neq w_2$ and some $\hat{l}$ such that $\left(x_1^{n_1}(w_1), x_2^{n_1}(\hat{w}_2), \hat{y}_{R,1+2}^{n_1}(\hat{i}), \hat{y}_{R,1}^{n_1}(\hat{j}|\hat{i})\right) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_{R,1+2}, \hat{Y}_{R,1}|q^{n_1})$ and simultaneously we have $\left(u^{n_2}(\hat{i}), v_1^{n_2}(\hat{j}, \hat{l}|\hat{i}), y_1^{n_2}\right) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(U, V_1, X_1)$.

- E_7: Suppose the triple (i, j, k) is given. E_7 is the event, that the relay cannot find a triple $u^{n_2}(i), v_1^{n_2}(j, l|i), v_2^{n_2}(k, m|i)$ such that $\exists l, m$ with $(u^{n_2}(i), v_1^{n_2}(j, l|i), v_2^{n_2}(k, m|i)) \in \mathcal{T}_{\epsilon_7}^{(n_1)}(U, V_1, V_2)$.

Now we can bound the two error events defined above.

Error event E_4 The error event E_4 can be bounded as

$$E\{Pr[E_4]\} \leq \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}) Pr[E_{4,1}] Pr[E_{4,2}] 2^{\lceil \alpha n R_{Q,1+2} \rceil} 2^{\lceil \alpha n R_{Q,1} \rceil} 2^{\lfloor n R_2 \rfloor} \lceil 2^{\lfloor n R_{V,1} \rfloor - \lceil \alpha n R_{Q,1} \rceil} \rceil$$

Here $E_{4,1}$ is the event, that for sequences $x_1^{n_1}, x_2^{n_1}, \hat{y}_{R,1+2}^{n_1}, \hat{y}_{R,1}^{n_1}$ we have $(x_1^{n_1}, x_2^{n_1}, \hat{y}_{R,1+2}^{n_1}, \hat{y}_{R,1}^{n_1}) \in \mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_{R,1+2}, \hat{Y}_{R,1}|q^{n_1})$. For this event $x_1^{n_1}$, $x_2^{n_1}, \hat{y}_{R,1+2}^{n_1}$, and $\hat{y}_{R,1}^{n_1}$ are drawn at random according to $p(x_1^{n_1}|q^{n_1})$, $p(x_2^{n_1}|q^{n_1})$, and $p(\hat{y}_{R,1}^{n_1}, \hat{y}_{R,1+2}^{n_1}|q^{n_1})$ respectively to capture the averaging over the random codebooks. $\hat{E}_{4,2}$ is the event, that for sequences $u^{n_2}, v_1^{n_2}, y_1^{n_2}$ we have $(u^{n_2}, v_1^{n_2}, y_1^{n_2}) \in \mathcal{T}_{\epsilon_2}^{(n_2)}(U, V_1, X_1)$. For this error event the sequences are drawn according to $p(u^{n_2}, v_1^{n_2})$ and $p(y_1^{n_2})$. The factor $2^{\lceil \alpha n R_{Q,1+2} \rceil}$ accounts for the fact that we can use a union bound and the error occurs if at least one $\hat{i} \neq i$ is found fulfilling the requirements. The factor $2^{\lceil \alpha n R_{Q,1+2} \rceil}$ accounts for a wrong $\hat{j}$ following the same argument. Furthermore, the event $E_{4,2}$ may happen for any $\hat{l}$; the union bound can be used and therefore leads to the factor $\lceil 2^{\lfloor n R_{V,1} \rfloor - \lceil \alpha n R_{Q,1} \rceil} \rceil$. The factor $2^{\lfloor n R_2 \rfloor}$ accounts for the different possible $\hat{w}_2$.

For sufficiently large n

$$\begin{aligned} Pr[\hat{E}_{4,1}] &= \sum_{x_1^{n_1} \in \mathcal{X}_1^{n_1}} \sum_{x_2^{n_1} \in \mathcal{X}_2^{n_1}} \sum_{\hat{y}_{R,1+2}^{n_1} \in \hat{\mathcal{Y}}_{R,1+2}^{n_1}} \sum_{\hat{y}_{R,1}^{n_1} \in \hat{\mathcal{Y}}_{R,1}^{n_1}} p(x_1^{n_1}|q^{n_1}) p(x_2^{n_1}|q^{n_1}) p(\hat{y}_{R,1+2}^{n_1}, \hat{Y}_{R,1}|q^{n_1}) \\ &\qquad \times \chi_{\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_{R,1+2}, \hat{Y}_{R,1}|q^{n_1})}(x_1^{n_1}, x_2^{n_1}, \hat{y}_{R,1+2}^{n_1}, \hat{y}_{R,1}^{n_1}) \\ &\leq |\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_{R,1+2}, \hat{Y}_{R,1}|q^{n_1})| 2^{-n_1\left(H(X_1|Q) - 2\epsilon_4\right)} 2^{-n_1\left(H(X_2|Q) - 2\epsilon_4\right)} 2^{-n_1\left(H(\hat{Y}_{R,1+2}, \hat{Y}_{R,1}|Q) - 2\epsilon_4\right)} \end{aligned}$$

due to the properties of the typical set. Furthermore, it follows from these properties that

$$|\mathcal{T}_{\epsilon_4}^{(n_1)}(X_1, X_2, \hat{Y}_{R,1+2}, \hat{Y}_{R,1}|q^{n_1})| \leq 2^{n_1\left(H(X_1, X_2, \hat{Y}_{R,1+2}, \hat{Y}_{R,1}|Q) + 2\epsilon_4\right)}.$$

$Pr[\hat{E}_{4,2}]$ can be bounded in a similar way. Therefore we have

$$\begin{aligned} E\{Pr(\hat{E}_4)\} &\leq \sum_{q^{n_1} \in \mathcal{Q}^{n_1}} p(q^{n_1}) 2^{-n_1(I(X_1 X_2; \hat{Y}_{R,1+2} \hat{Y}_{R,1}|Q) - 8\epsilon_4)} 2^{-n_2(I(UV_1;Y_1) - 6\epsilon_2)} 2^{\lceil \alpha n R_{Q,1+2} \rceil} \\ &\qquad 2^{\lceil \alpha n R_{Q,1} \rceil} 2^{\lfloor n R_2 \rfloor} \lceil 2^{\lfloor n R_{V,1} \rfloor - \lceil \alpha n R_{Q,1} \rceil} \rceil \\ &\leq 2^{-n\left(\alpha\left(I(X_1 X_2; \hat{Y}_{R,1+2} \hat{Y}_{R,1}|Q) - R_{Q,1+2} - 8\epsilon_4\right) + \beta(I(UV_1;Y_1) - 6\epsilon_2) - R_2 - R_{V,1}\right) + 2 + I(X_1 X_2; \hat{Y}_{R,1+2} \hat{Y}_{R,1}|Q) + 6\epsilon_2} \\ &= 2^{-n\left(\alpha\left(I(X_1 X_2; \hat{Y}_{R,1+2} \hat{Y}_{R,1}|Q) - I(Y_R; \hat{Y}_{R,1+2}|Q)\right) + \beta I(UV_1;Y_1) - R_2 - R_{V,1} - \tilde{\epsilon}\right) + 2 + I(X_1 X_2; \hat{Y}_{R,1+2} \hat{Y}_{R,1}|Q) + 6\epsilon_2} \end{aligned}$$

with

$$\tilde{\epsilon} = \alpha\epsilon_{q,1+2} + \beta 6\epsilon_2 + \alpha 8\epsilon_4 < (2\alpha + \beta)\epsilon_{q,1+2}.$$

This term goes to zero if

$$\alpha\left(I(X_1X_2; \hat{Y}_{R,1+2}\hat{Y}_{R,1}|Q) - I(Y_R; \hat{Y}_{R,1+2}|Q)\right) + \beta I(UV_1; Y_1) - R_2 - R_{V,1} - (2\alpha + \beta)\epsilon_{q,1+2} > 0.$$

By the choice of $\epsilon_{q,1+2}$ as

$$\epsilon_{q,1+2} < \frac{1}{(2\alpha + \beta)}\left(\alpha\left(I(X_1X_2; \hat{Y}_{R,1+2}\hat{Y}_{R,1}|Q) - I(Y_R; \hat{Y}_{R,1+2}|Q)\right) + \beta I(UV_1; Y_1) - R_2 - R_{V,1}\right)$$

this is true whenever

$$R_2 + R_{V,1} < \alpha\left(I(X_1X_2; \hat{Y}_{R,1+2}\hat{Y}_{R,1}|Q) - I(Y_R; \hat{Y}_{R,1+2}|Q)\right) + \beta I(UV_1; Y_1)$$

as required by the assumption. Therefore for any given ϵ we can find $n^{(4)}$ such that for $E\{Pr[\hat{E}_4]\} < \frac{\epsilon}{7}$ for $n \geq n^{(4)}$.

Error event E_7 The event E_7 can be proved following the lines of the proof of the lemma in [69]. In what follows, we prove that given i and $u^{n_2}(i)$ with high probability there is at least one pair $v_1^{n_2}(j, l|i)$, $v_2^{n_2}(k, m|i)$ for a given (j, k) that is jointly typical given $u^{n_2}(i)$. From the definition of the typical set, it follows that the sequences $(u^{n_2}(i), v_1^{n_2}(j, l|i), v_2^{n_2}(k, m|i)) \in \mathcal{T}_{\epsilon_7}^{(n_1)}(U, V_1, V_2)$.

For a fixed i, j, k and $u^{n_2}(i)$, let $\mathcal{T}_{j,k}(i)$ be the set of codeword pairs in the bin pair (j, k) that are jointly typical sequences given $u^{n_2}(i)$, i.e.

$$\mathcal{T}_{j,k}(i) = \left\{\left(v_1^{n_2}(j, l|i), v_2^{n_2}(k, m|i)\right) : \left(v_1^{n_2}(j, l|i), v_2^{n_2}(k, m|i)\right) \in \mathcal{T}_{\epsilon_7}^{(n_1)}(V_1, V_2|u^{n_2}(i))\right\}.$$

The number of sequences in each bin j is greater than $C_1 = \lfloor 2^{\lfloor nR_{V,1}\rfloor - \lceil \alpha n R_{Q,1}\rceil}\rfloor$. We calculate the error probability for a bin with a small number of sequences. The probability of not finding a pair of sequences for bins with more sequences can only be smaller. Similarly the number of sequences in each bin k is greater than $C_2 = \lfloor 2^{\lfloor nR_{V,2}\rfloor - \lceil \alpha n R_{Q,2}\rceil}\rfloor$.

It is assured that for sufficient large n we have

$$C_1 \geq 2^{nR_{V,1} - \alpha n\left(I(Y_R; \hat{Y}_{R,1}|\hat{Y}_{R,1+2}, Q) + \epsilon_{q,1}\right) - 3} \geq 1$$

as we assumed

$$R_{V,1} > \alpha I(Y_R; \hat{Y}_{R,1}|\hat{Y}_{R,1+2}, Q)$$

and chose

$$\epsilon_{q,1} < \frac{R_{V,1}}{\alpha} - I(Y_R; \hat{Y}_{R,1}|\hat{Y}_{R,1+2}, Q).$$

Similar arguments apply for $C_2 \geq 1$.

As all codewords are drawn at random we assume without loss of generality in what follows $i = j = k = 1$ and set $\mathcal{T} := \mathcal{T}_{1,1}(1)$. Furthermore we use the fact that the bin $j = 1$ contains more than C_1 sequences and the bin $k = 1$ contains more than C_2 sequences (and so does any other bin pair). As more sequences in the bins can only decrease the probability of error in what follows we assume that the considered bins $j = 1$, $k = 1$ contain C_1 and C_2 sequences respectively.

The error probability for event E_7 is given by

$$\mathbb{E}\{Pr[E_7]\} = Pr[|\mathcal{T}| = 0].$$

For $0 < \epsilon^{(7)} < 1$ we have by Chebychev's inequality and using the fact that $\mathbb{E}\{|\mathcal{T}|\} > 0$

$$\begin{aligned}
Pr\left[|\mathcal{T}| = 0\right] &\leq Pr\left[|\mathcal{T}| < (1-\epsilon^{(7)})\mathbb{E}\{|\mathcal{T}|\}\right] \\
&= Pr\left[\mathbb{E}\{|\mathcal{T}|\} - |\mathcal{T}| > \epsilon^{(7)}\mathbb{E}\{|\mathcal{T}|\}\right] \\
&\leq Pr\left[\left||\mathcal{T}| - \mathbb{E}\{|\mathcal{T}|\}\right| > \epsilon^{(7)}\mathbb{E}\{|\mathcal{T}|\}\right] \\
&\leq \frac{\sigma^2(|\mathcal{T}|)}{(\epsilon^{(7)}\mathbb{E}\{|\mathcal{T}|\})^2},
\end{aligned}$$

where $\sigma^2(\cdot) = \mathbb{E}\{(\cdot)^2\} - \mathbb{E}\{\cdot\}^2$ is the variance of the argument.

Now

$$\mathbb{E}\{|\mathcal{T}|\} = C_1 C_2 Pr[E_{7,1}],$$

where $E_{7,1}$ is the event that two sequences drawn according to $p\left(v_1^{n_2}|u^{n_2}(i)\right)$ and $p\left(v_2^{n_2}|u^{n_2}(i)\right)$ are jointly typical, i.e. $(v_1^{n_2}, v_2^{n_2}) \in \mathcal{T}_{\epsilon_7}^{(n_1)}(V_1, V_2|u^{n_2})$. The probability for this event can be lower bounded using the properties of the typical set. Therefore we have

$$\mathbb{E}\{|\mathcal{T}|\} \geq (1-\delta)C_1 C_2 2^{-\beta n(I(V_1;V_2|U)+6\epsilon_7)},$$

where δ can be made arbitrarily small by choosing n large.

It is left to find an upper bound for $\mathbb{E}\{|\mathcal{T}|^2\}$. Now, with the C_1 sequences $v_1^{n_2}(a) = v_1^{n_2}(1, a|1)$

and the C_2 sequences $v_2^{n_2}(b) = v_2^{n_2}(1, b|1)$ in the bins $j = 1, k = 1$ considered in this proof

$$\begin{aligned}
|\mathcal{T}|^2 &= \left(\sum_{\substack{a\in\{1,2,...C_1\}\\ b\in\{1,2,...C_2\}}} \chi_{\mathcal{T}}\left(v_1^{n_2}(a), v_2^{n_2}(b)\right) \right)^2 \\
&= \sum_{\substack{a\in\{1,2,...C_1\}\\ b\in\{1,2,...C_2\}}} \chi_{\mathcal{T}}\left(v_1^{n_2}(a), v_2^{n_2}(b)\right) \\
&\quad + \sum_{\substack{a\neq\tilde{a}\in\{1,2,...C_1\}\\ b\in\{1,2,...C_2\}}} \chi_{\mathcal{T}}\left(v_1^{n_2}(a), v_2^{n_2}(b)\right)\chi_{\mathcal{T}}\left(v_1^{n_2}(\tilde{a}), v_2^{n_2}(b)\right) \\
&\quad + \sum_{\substack{a\in\{1,2,...C_1\}\\ b\neq\tilde{b}\in\{1,2,...C_2\}}} \chi_{\mathcal{T}}\left(v_1^{n_2}(a), v_2^{n_2}(b)\right)\chi_{\mathcal{T}}\left(v_1^{n_2}(a), v_2^{n_2}(\tilde{b})\right) \\
&\quad + \sum_{\substack{a\neq\tilde{a}\in\{1,2,...C_1\}\\ b\neq\tilde{b}\in\{1,2,...C_2\}}} \chi_{\mathcal{T}}\left(v_1^{n_2}(a), v_2^{n_2}(b)\right)\chi_{\mathcal{T}}\left(v_1^{n_2}(\tilde{a}), v_2^{n_2}(\tilde{b})\right)
\end{aligned}$$

where $\chi_{\mathcal{T}}(\cdot)$ is the indicator function on the set $\mathcal{T}$.

After taking expectations we can write

$$\begin{aligned}
\sigma^2\{|\mathcal{T}|\} &= C_1C_2Pr[\left(v_1^{n_2}(a), v_2^{n_2}(b)\right) \in \mathcal{T}\} \\
&\quad +C_2(C_1^2 - C_1)Pr[\left(v_1^{n_2}(a), v_2^{n_2}(b)\right) \in \mathcal{T} \text{ and } \left(v_1^{n_2}(\tilde{a}), v_2^{n_2}(b)\right) \in \mathcal{T}] \\
&\quad +C_1(C_2^2 - C_2)Pr[\left(v_1^{n_2}(a), v_2^{n_2}(b)\right) \in \mathcal{T} \text{ and } \left(v_1^{n_2}(a), v_2^{n_2}(\tilde{b})\right) \in \mathcal{T}] \\
&\quad +(C_1^2 - C_1)(C_2^2 - C_2)Pr[\left(v_1^{n_2}(a), v_2^{n_2}(b)\right) \in \mathcal{T} \text{ and } \left(v_1^{n_2}(\tilde{a}), v_2^{n_2}(\tilde{b})\right) \in \mathcal{T}] \\
&\quad -\left(C_1C_2Pr[(v_1^{n_2}(a), v_2^{n_2}(b)) \in \mathcal{T}]\right)^2.
\end{aligned}$$

Here we use that

$$E\{|\mathcal{T}|\}^2 = \left(C_1C_2Pr[(v_1^{n_2}(a), v_2^{n_2}(b)) \in \mathcal{T}]\right)^2$$

in the calculation of $\sigma^2\{|\mathcal{T}|\}$.

We can bound the probabilities that sequence drawn independent of each other are jointly typical with the technique used in e.g. the bounding of event E_6. Therefore we get for sufficient large n

$$\begin{aligned}
\sigma^2\{|\mathcal{T}|\} &\leq C_1C_2 2^{-\beta n(I(V_1;V_2|U)-6\epsilon_7)} \\
&\quad + C_1(C_2^2 - C_2)2^{-2\beta n(I(V_1;V_2|U)-6\epsilon_7)} \\
&\quad + C_2(C_1^2 - C_1)2^{-2\beta n(I(V_1;V_2|U)-6\epsilon_7)} \\
&\quad + (C_1C_2 - C_1^2C_2 - C_1C_2^2)2^{-2\beta n(I(V_1;V_2|U)-6\epsilon_7)} \\
&= C_1C_2 2^{-\beta n(I(V_1;V_2|U)-6\epsilon_7)} \\
&\quad - C_1C_2 2^{-2\beta n(I(V_1;V_2|U)-6\epsilon_7)} \\
&\leq C_1C_2 2^{-\beta n(I(V_1;V_2|U)-6\epsilon_7)},
\end{aligned}$$

where we used $\left(Pr[(v_1^{n_2}(a), v_2^{n_2}(b)) \in \mathcal{T}]\right)^2 = Pr[(v_1^{n_2}(a), v_2^{n_2}(b)) \in \mathcal{T}$ and $(v_1^{n_2}(\tilde{a}), v_2^{n_2}(\tilde{b})) \in \mathcal{T}]$.

With these bounds on variance and expectation of $|\mathcal{T}|$ we can upper bound $\mathbb{E}\{Pr[E_7]\}$ as

$$\begin{aligned}
E\{Pr[E_7]\} \leq Pr[|\mathcal{T}| = 0] &\leq \frac{C_1C_2 2^{-\beta n(I(V_1;V_2|U)-6\epsilon_7)}}{(\epsilon^{(7)}(1-\delta)C_1C_2 2^{-\beta n(I(V_1;V_2|U)+6\epsilon_7)})^2} \\
&= \frac{1}{(\epsilon^{(7)})^2(1-\delta)^2 C_1C_2 2^{-\beta n(I(V_1;V_2|U)+18\epsilon_7)}}.
\end{aligned}$$

Now for sufficient large n

$$E\{Pr[E_7]\} \leq \frac{1}{(\epsilon^{(7)})^2(1-\delta)^2 2^{-n\left(\alpha R_{Q,1}+\alpha R_{Q,2}-R_{V,1}-R_{V,2}+\beta(I(V_1;V_2|U)+18\epsilon_7)\right)-6}}$$

where we used $C_1 \geq 2^{nR_{V,1}-\alpha nR_{Q,1}-3}$, $C_2 \geq 2^{nR_{V,2}-\alpha nR_{Q,2}-3}$, and where $\delta \to 0$ for $n \to \infty$. This goes to zero if

$$\alpha R_{Q,1} + \alpha R_{Q,2} - R_{V,1} - R_{V,2} + \beta I(V_1; V_2|U) + \beta 18\epsilon_7 < 0$$

which is equal to the condition

$$\alpha\left(I(Y_R; \hat{Y}_{R,1}|\hat{Y}_{R,1+2}, Q) + I(Y_R; \hat{Y}_{R,2}|\hat{Y}_{R,1+2}, Q) + \epsilon_{q,1} + \epsilon_{q,2}\right) + \beta(I(V_1; V_2|U) + 18\epsilon_7) < R_{V,1} + R_{V,2}.$$

This inequality is fulfilled by assumption and by the choice of $\epsilon_{q,1}$, $\epsilon_{q,2}$ and ϵ_7.

The rest of the proof follows immediately using similar arguments as in the proofs given in previous chapters and is not repeated here.

□

5.2 Concluding Remarks

In this chapter we take a first glance on the potential that arises by considering the different information flows, which are present in the two way relay channel. It turns out, that the system design gains some degrees of freedom, if the relay treats the different information flows not jointly, but allows for a different coarseness in the quantization depending on the need of the respective flow. In particular, we use three information flows, which are quantized distinct from each other at the relay. One information flow is form the relay to both respective receivers, while a third information flow is from the receiver to both receivers at the same time. This third stream facilitates a de-facto interference free transmission in the BC, while the individual data streams interfere with each other. Depending on the BC and MAC statistics this interference is accepted, as it is less harmful than the joint quantization at the relay.

The coding uses tools from the proof of the general BC [47, 69] to permit the transmission of different information to the respective receivers. As in the case of compress-and-forward with joint decoding, side information about the dependency of the MAC output from the transmitted codewords is used to restrict the number of codewords in the effective code, which is used for the decoding at the receiver. Due to the inability of the relay to decode the data there is in general still some overhead in the effective code. In contrast to the compress-and-forward scheme with joint decoding, this overhead can now be controlled individually for both receivers. Therefore, the new scheme is promising particularly for systems with non-symmetric rates, more precisely for systems, where the BC channel for one of the transmissions is the bottleneck for the overall system. This feature is bought by additional interference at least for one of the receivers.

Note, that the proposed coding scheme is only a first simple step in the direction of a more general understanding of the two-way relay channel. There are dependencies in the system that are not used so far, namely the dependency of the other receiver's quantized MAC output representation from the transmitted codewords. Furthermore, for the ease of analysis the quantization at the relay is constrained by the assumption that the two respective individual MAC output representations are independent given the channel output and the common representation. A more general treatment and analysis of the system needs to break up these restrictions. This leads to the problem to cope with correlated information transmitted in a multi-user system. Furthermore, the BC coding bases upon the coding for the general BC. It is likely that progress towards a proof of the capacity region of the BC will impact the achievable rate region of the two-way relay channel using more then one data streams.

Chapter 6

Conclusion and Outlook

In many communication scenarios the task of the transmission protocol is to get messages from one node in a network to a distant node. One-hop transmissions over long distances need high transmission power and therefore cause large interference to other links in the network. The same is true if the direct link between transmitter and receiver is weak. In a real communication scenario, this can be caused by shadowing, e.g. if there is no line-of-sight connection between transmitter and receiver possible. Relaying protocols have the potential to circumvent that problem by splitting the distance into several hops. Thereby, these protocols can increase the coverage of cellular systems and enhance the throughput by reducing interference due to a smaller transmission power. For that reason relaying concepts will play a central role in future wireless communication systems.

In this thesis we study the two-way relay channel. In the two-way relay channel the task of the relay is to establish a bi-directional communication between two nodes in the network. Recently, two-way relaying has attracted great interest, as it has the potential to offer gains compared to one-hop communication or one-way transmission protocols. The reason is that for systems with half duplex nodes no additional resourses such as time or frequency have to be allocated compared to the one-way relaying scenario; the transmission in both directions can be performed simultaneously.

The two-way relay channel features an interesting property, that originates from the setup of the system. Both terminal nodes know the message intended for the other respective receiver. This knowledge can be used to eliminate some of the interference in the transmission from the relay to the terminal nodes. In Chapter 2 we state the capacity region for a BC, where the receivers know the message intended for the other respective node. This region shows that a de-facto interference free transmission is possible in this channel. In fact the interference can be canceled by coding and no interference cancellation at the receiver is needed. Both links can be operated as if the other node were not present. The only drawback is the common input distribution to the channel, i.e. one of the links may not achieve the single user capacity, as the transmission needs to fit for both channels at the same time.

We analyze the coding in the broadcast phase in detail and give a practical coding scheme, which uses the mechanisms pointed out by the proof of the capacity region. This scheme uses well developed single user codes to build a joint code for the BC in the two-way relay channel, for the case where the relay decodes the messages. The resulting code achieves the single user performance of the base codes for both users, whenever the marginal channels fulfill a certain symmetry condition. It turns out that some properties of the scheme can be generalized to Gaussian channels if nested lattice codes are used as base codes. As these codes achieve the capacity of the single user Gaussian channel, the resulting joint code will achieve the capacity of the BC in the two-way relay channel with decode-and-forward.

The result of an achievable rate region for the BC phase is used to state an achievable rate region for the two-way relay channel with two phases. For this achievable rate region the relay is assumed to decode the messages, and re-encode both for transmission over the BC. The results are extended to the case that the MAC transmission is replaced by other transmission strategies in order to enable the decoding at the relay. Thereby we allow for transmission protocols that facilitate the transmission via a direct link between the terminal nodes.

Through simple examples one can see that decoding at the relay might be suboptimal. This can be the case if a sum-rate constraint of the MAC restricts the achievable rates in the overall system. Whenever the sum rate of the MAC does not pose an additional constraint on the rate region of the MAC, the decode-and-forward protocol achieves the cutset outer bound on the capacity region and therefore is optimal. If a sum-rate constraint of the MAC is active, knowing the message of one of the nodes could permit the decoding, even though it is impossible without possessing this knowledge. This observation leads to a protocol where the compressed MAC output is forwarded to the terminal nodes which already know one of the transmitted signals. The resulting scheme is a compress-and-forward protocol for the two-way relay channel.

In a first approach the receivers decode the transmission of the relay. Although this transmission seems to be a simple multicast, it turns out that the receivers can use the known message as side information. The reason for this is that the MAC output, and as a consequence also the compressed MAC output, as well as the signal transmitted by the relay depend on the message transmitted by the terminal nodes in the MAC phase. The approach is extended by noting that the MAC output depends also on the message transmitted by the other node. Therefore gains can be achieved if the receiver decodes the transmission of the relay and the message intended for it jointly instead of sequentially. In effect the receiver does not decode the signal transmitted by the relay explicitly, but focuses on the message transmitted by the other respective node. The relay's transmission enables the decoding of this message. The result shows that in networks it can be suboptimal to treat the atoms of the network as a MAC and a BC separately. The gains offered by the protocol can only be achieved if the overall system is considered in the decoding process.

As in the decode-and-forward protocol there is no interference in the BC for the compress-and-forward protocols. Furthermore, the sum-constraint of the MAC vanishes. Due to the compression at the relay, the rate region can still be smaller than the cutset outer bound on the capacity region. If the compression at the relay is such that it is not lossy with respect to the input signals while it eliminates all the information overhead due to noise in the MAC output, then the proposed compress-and-forward scheme can be proven to achieve the capacity of the two-way relay channel.

It turns out that the common compression in the system may hinder the transmission in one way to allow transmission in the other way. Therefore we extend the previous results by allowing an individual compression for all the flows of information that occur in the system. Thereby we can trade off an interference-like effect in the relay's compression for some interference in the BC. The resulting achievable rate region includes all regions proposed in this thesis which are achievable with compress-and-forward.

As the strategies of compress-and-forward and decode-and-forward use different mechanisms to achieve large rates, an hybrid approach can trade off between these effects. For the compress-and-forward protocols we therefore propose a partial decode-and-forward strategy, which is a superposition of both the schemes.

In summary we have three extreme cases: For the decode-and-forward approach, interference is present in the MAC, but not in the BC. In the compress-and-forward approach we suffer from a possibly suboptimal compression and an interference-like effect in the compression. In the extreme case, using an independent compression for the different flows of information we can eliminate this effect, but now we have interference in the BC channel. We conclude that by the superposition of decode-and-forward with the compress-and-forward protocol, which facilitates more than one flow of information, we can balance these negative effects. In effect we can build a system such that the users interfere where it causes the least harm.

6.1 Outlook

Until today the capacity of the two-way relay channel has benn obtained only for very few channels. The achievable rate regions in this thesis are only subsets of the capacity region for most channels. For example the achievable rate region obtained in Chapter 5 could be improved by using more of the statistical dependencies in the system. This includes a more involved compression as well as considerations about how one can further use the side information in the BC to cancel some of the interference. These improvements seem quite demanding though very interesting and important for the understanding of the two-way relay channel.

Further improvements may target the MAC transmission and the "decoding" at the relay. In Chapter 2 we noted that it is sufficient to know the XOR sum of the messages at the relay. This gives rise to the so called computational coding, where the goal is to decode a function of the messages instead of the message itself. A first discussion as well as references can be found in Section 4.1.4. The topic of computational coding has just begun to attract interest in the research community. Therefore some results can be expected in the coming years. It turns out that structured codes are important for this kind of coding, therefore the random coding approaches used in classical information theory will fail.

This thesis focuses on discrete alphabet channels. For channels with continuous alphabets such as Gaussian channels, the decode-and-forward result is known to hold. Most of the other results can probably be adapted to these channels, though the proof is not yet provided. Furthermore for the Gaussian channel, there are several other strategies proposed for the case of one-way communication besides decode-and-forward and compress-and-forward. It is unclear if some of these strategies may improve the rate regions given in this thesis.

Another interesting topic for future research is to drop the assumption that the two-way communication is restricted. This will enable explicit cooperation between the terminal nodes as well as feedback. The most general treatment for the considered setup is the non-restricted two-way relay channel with full-duplex nodes. All the restricted scenarios considered in this thesis can be seen to be special cases of this channel.

Appendix A

Appendix — Bounding of the Cardinalities

A.1 Cardinalities of Auxiliary Random Variables

In this appendix we derive upper bounds for the cardinality of the auxiliary variables in the theorems given in this thesis. The key tool for the bounding is the Fenchel-Bunts extension of Caratheodorys theorem [62] which we restated in Theorem 3.4:

Theorem (Fenchel-Bunts extension of Caratheodorys theorem [62])**.** If $\mathcal{S} \subset \mathbb{R}^n$ has no more than n connected components (in particular, if $\mathcal{S}$ is connected), then any $x \in \text{ConvexHull}(\mathcal{S})$ can be expressed as a convex combination of n elements of $\mathcal{S}$.

A.1.1 The Cardinality of the Auxiliary Variables in Theorem 3.5

A.1.1.1 The cardinality of Q

We can bound the cardinality of Q in Theorem 3.5 similar to the proceeding in Section 3.2 for Theorem 3.1: Define for given channels $p_1(y_R|x_1, x_2)$, $p_2(y_1, y_2|x_R)$, and fixed α, β

$$\mathcal{S} = \bigcup_{p(u_1)p(u_2)p(x_1|u_1)p(x_2|u_2)p(\hat{y}_R|y_R)} \Big\{[\delta_1(p), \delta_2(p), \delta_3(p), \delta_4(p), \delta_5(p), \delta_6(p), \delta_7(p)]$$
$$|p = p(u_1)p(u_2)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)\Big\}$$

where the union is over the compact set of all $p(u_1)p(u_2)p(x_1|u_1)p(x_2|u_2)p(\hat{y}_R|y_R)$ and where

$$\begin{aligned}\delta_1(p) &= \alpha I(X_1;\hat{Y}_R|X_2,U_1)\\ \delta_2(p) &= \alpha I(X_2;\hat{Y}_R|X_1,U_2)\\ \delta_3(p) &= \alpha I(U_1;Y_R|U_2)\\ \delta_4(p) &= \alpha I(U_2;Y_R|U_1)\\ \delta_5(p) &= \alpha I(U_2U_2;Y_R)\\ \delta_6(p) &= \alpha I(\hat{Y}_R;Y_R|X_1,U_2)\\ \delta_7(p) &= \alpha I(\hat{Y}_R;Y_R|X_2,U_1)\end{aligned}$$

Furthermore let $C = \text{ConvexHull}(\mathcal{S})$ and let

$$\hat{\mathcal{S}} = \bigcup_{p(v)p(x_R|v)} \Big\{[\beta I(V,Y_2), \beta I(V,Y_1), \beta I(X_R,Y_1|V), \beta I(X_R,Y_2|V)]\Big\}.$$

The achievable rate region can now be stated as

$$\begin{aligned}\mathcal{R}_4 = \Big\{&[R_1,R_2]\in\mathbb{R}_+^2 : \exists[\delta_1,\delta_2,\delta_3,\delta_4,\delta_5,\delta_6,\delta_7]\in C;[\delta_8,\delta_9,\delta_{10},\delta_{11}]\in\hat{\mathcal{S}} \text{ and}\\ &\exists R_1^{(1)},R_2^{(1)},R_1^{(2)},R_2^{(2)}\geq 0 \text{ with } R_1^{(1)}+R_1^{(2)}=R_1, R_2^{(1)}+R_2^{(2)}=R_2,\\ &R_1^{(2)}\leq\delta_1, R_2^{(2)}\leq\delta_2, R_1^{(1)}\leq\min\{\delta_3,\delta_8\}, R_2^{(1)}\leq\min\{\delta_4,\delta_9\}, R_1^{(1)}+R_2^{(1)}\leq\delta_5, \delta_6<\delta_{10}, \delta_7<\delta_{11}\Big\}.\end{aligned}$$

The set $\mathcal{S}$ is connected, as it is the continuous image of a continuous compact set. Therefore, all points in C can be expressed as a convex combination of at most $\dim\{\mathcal{S}\} = 7$ elements of $\mathcal{S}$. It follows that we can bound the required cardinality of $\mathcal{Q}$ from above by 7.

A.1.1.2 The Cardinality of $\hat{\mathcal{Y}}$

Let $s_1 \in \Delta_{|\mathcal{Y}_R|}$. Define for given channels $p_1(y_R|x_1,x_2)$, $p_2(y_1,y_2|x_R)$, fixed α, β and fixed $p = p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)$

$$\mathcal{S}(p) = \bigcup_{s_1}\{[\delta_1,\delta_2,\delta_6,\delta_7,s_1]\}$$

where the union is over all $s_1 \in \Delta_{|\mathcal{Y}_R|}$, and we have

$$\begin{aligned}\delta_1 = \alpha\Big(&H(X_1|X_2,U_1) + \sum_{i,x_1,x_2,u_1} s_1(i)p(x_1,x_2,u_1|Y_R=i)\\ &\times\Big(\log\Big(\sum_j s_1(j)p(x_2,u_1|Y_R=j)\Big) - \log\Big(\sum_j s_1(j)p(x_1,x_2,u_1|Y_R=j)\Big)\Big)\Big)\end{aligned}$$

$$\delta_2 = \alpha\Big(H(X_2|X_1, U_2) + \sum_{i,x_1,x_2,u_2} s_1(i)p(x_1, x_2, u_2|Y_R = i) \\ \times \Big(\log\Big(\sum_j s_1(j)p(x_1, u_2|Y_R = j)\Big) - \log\Big(\sum_j s_1(j)p(x_1, x_2, u_2|Y_R = j)\Big)\Big)\Big)$$

$$\delta_6 = \alpha\Big(H(Y_R|X_1, U_2) + \sum_{i,x_1,u_2} s_1(i)p(x_1, u_2|Y_R = i) \\ \times \Big(\log\Big(s_1(i)p(x_1, u_2|Y_R = i)\Big) - \log\Big(\sum_j s_1(j)p(x_1, u_2|Y_R = j)\Big)\Big)\Big)$$

$$\delta_7 = \alpha\Big(H(Y_R|X_2, U_1) + \sum_{i,x_2,u_1} s_1(i)p(x_2, u_1|Y_R = i) \\ \times \Big(\log\Big(s_1(i)p(x_2, u_1|Y_R = i)\Big) - \log\Big(\sum_j s_1(j)p(x_2, u_1|Y_R = j)\Big)\Big)\Big)$$

Here we use the common convention $0 \log 0 = 0$ justified by continuity since $x \log x \to 0$ as $x \to 0$. Now, let $C(p) = \text{ConvexHull}(\mathcal{S}(p))$ and let

$$\hat{\mathcal{S}} = \bigcup_{p(v)p(x_R|v)} \Big\{[\beta I(V, Y_2), \beta I(V, Y_1), \beta I(X_R, Y_1|V), \beta I(X_R, Y_2|V)]\Big\}.$$

Furthermore let

$$\bar{C}(p) = \Big\{[\delta_1, \delta_2, \alpha I(U_1; Y_R|U_2, Q), \alpha I(U_2; Y_R|U_1, Q), \alpha I(U_1U_2; Y_R|Q), \delta_6, \delta_7] : \\ [\delta_1, \delta_2, \delta_6, \delta_7, s_1] \in C(p) \text{ and } \forall i \quad s_1(i) = p(Y_R = i)\Big\}.$$

The achievable rate region can be stated as

$$\mathcal{R}_4 = \bigcup_p \Big\{[R_1, R_2] : \exists R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)} \geq 0, \\ \exists[\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7] \in \bar{C}(p) \text{ and } \exists[\delta_8, \delta_9, \delta_{10}, \delta_{11}] \in \hat{\mathcal{S}} \\ \text{with } R_1^{(1)} + R_1^{(2)} = R_1, R_2^{(1)} + R_2^{(2)} = R_2, \\ R_1^{(2)} \leq \delta_1, R_2^{(2)} \leq \delta_2, R_1^{(1)} \leq \min\{\delta_3, \delta_8\}, R_2^{(1)} \leq \min\{\delta_4, \delta_9\}, R_1^{(1)} + R_2^{(1)} \leq \delta_5, \delta_6 < \delta_{10}, \delta_7 < \delta_{11}\Big\}.$$

The set $\mathcal{S}(p)$ is connected, as it is the continuous image of the continuous compact set $\Delta_{|\mathcal{Y}_R|}$. Therefore all points in $C(p)$ can be expressed as a convex combination of at most $\dim\{\mathcal{S}(p)\} = |\mathcal{Y}_R| + 3$ elements of $\mathcal{S}(p)$. Therefore all points in the achievable rate region can be achieved with $|\hat{\mathcal{Y}}_R| \leq |\mathcal{Y}_R| + 3$.

A.1.1.3 The Cardinality of $\mathcal{U}_1$ and $\mathcal{U}_2$

In what follows we bound the cardinality of $\mathcal{U}_1$. The bounding of the cardinality of $\mathcal{U}_2$ follows accordingly.

Let $s_1 \in \Delta_{|\mathcal{X}_1|}$ $s_2 \in \Delta_{|Q|}$ and $s_3 \in \Delta_{|\mathcal{X}_1||Q|}$. Define for given channels $p_1(y_R|x_1, x_2)$, $p_2(y_1, y_2|x_R)$, fixed α, β and fixed $p = p(q, x_1)(u_2|q)p(x_2|u_2)$

$$\mathcal{S}(p) = \bigcup_{s_1, s_2} \{[\gamma_1, \gamma_2, \gamma_3, \gamma_4, s_3]\}$$

where the union is over all possible $s_1 \in \Delta_{|\mathcal{Y}_R|}$, $s_2 \in \Delta_{|Q|}$, and we have

$$\begin{aligned}
\gamma_1 =\ & \alpha\Big(\sum_{i,j,\hat{y}_R,x_2} s_1(i)s_2(j)p(\hat{y}_R, x_2|X_1 = i, Q = j) \\
& \qquad \times\Big(\log(p(\hat{y}_R|X_1 = i, x_2)) - \log\Big(\textstyle\sum_k s_1(k)p(\hat{y}_R|X_1 = k, x_2)\Big)\Big)\Big) \\
\gamma_2 =\ & \alpha\Big(\sum_{i,j,y_R,u_2} s_1(i)s_2(j)p(\hat{y}_R, u_2|X_1 = i, Q = j)\Big(-\log\Big(\sum_k s_1(k)p(y_R|X_1 = k, u_2)\Big)\Big)\Big) \\
\gamma_3 =\ & \alpha\Big(\sum_{i,j,y_R,u_2} s_1(i)s_2(j)p(\hat{y}_R, u_2|X_1 = i, Q = j) \\
& \qquad \times\Big(\log(\textstyle\sum_k s_1(k)p(y_R|x_1 = k, u_2)) - \log\Big(\sum_l s_1(l)p(y_R|X_1 = l, q = j)\Big)\Big)\Big) \\
\gamma_4 =\ & \alpha\Big(\sum_{i,j,\hat{y}_R,x_2} s_1(i)s_2(j)p(\hat{y}_R, x_2|X_1 = i, Q = j)\Big(-\log\Big(\sum_k s_1(k)p(\hat{y}_R|X_1 = k, x_2)\Big)\Big)\Big)
\end{aligned}$$

and $s_3(i, j) = s_1(i)s_2(j)$. In the last equation we use a notation similar to that we use for joint probability distributions to index the elements of the vector s_3. Again we use the common convention $0 \log 0 = 0$ justified by continuity since $x \log x \to 0$ as $x \to 0$. Let $C(p) = \text{ConvexHull}(\mathcal{S}(p))$ and let

$$\hat{\mathcal{S}} = \bigcup_{p(v)p(x_R|v)} \Big\{[\beta I(V, Y_2), \beta I(V, Y_1), \beta I(X_R, Y_1|V), \beta I(X_R, Y_2|V)]\Big\}.$$

Furthermore let

$$\begin{aligned}
\bar{C}(p) = \Big\{\Big[&\gamma_1, \alpha I(X_2; \hat{Y}_R|X_1, U_2), \alpha H(Y_R|U_2, Q) - \gamma_2, \gamma_3, \\
& \alpha H(Y_R|Q) - \gamma_2, \alpha\left(H(\hat{Y}_R|X_1, U_2) - H(\hat{Y}_R|Y_R)\right), \gamma_4 - \alpha H(\hat{Y}_R|Y_R)\Big] \\
& : \exists[\gamma_1, \gamma_2, \gamma_3, \gamma_4, s_3] \in C(p), \text{ with } \forall i, j \quad s_3(i, j) = p(X_1 = i, Q = j)\Big\}
\end{aligned}$$

The achievable rate region can be stated as

$$\mathcal{R}_4 = \bigcup_{p} \Big\{ [R_1, R_2] : \exists R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)} \geq 0;$$
$$\exists [\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7] \in \bar{C}(p); \text{ and } \exists [\delta_8, \delta_9, \delta_{10}, \delta_{11}] \in \hat{\mathcal{S}} \text{ with}$$
$$R_1^{(1)} + R_1^{(2)} = R_1, R_2^{(1)} + R_2^{(2)} = R_2,$$
$$R_1^{(2)} \leq \delta_1, R_2^{(2)} \leq \delta_2, R_1^{(1)} \leq \min\{\delta_3, \delta_8\}, R_2^{(1)} \leq \min\{\delta_4, \delta_9\}, R_1^{(1)} + R_2^{(1)} \leq \delta_5, \delta_6 < \delta_{10}, \delta_7 < \delta_{11} \Big\}.$$

The set $\mathcal{S}(p)$ is connected, as it is the continuous image of the continuous compact set $\Delta_{|\mathcal{X}_1|} \times \Delta_{|Q|}$. Therefore all points in $C(p)$ can be expressed as a convex combination of at most $\dim\{\mathcal{S}(p)\} = |\mathcal{X}_1||Q| + 3$ elements of $\mathcal{S}(p)$. Therefore all points in the achievable rate region can be achieved with $|\mathcal{U}_1| \leq |\mathcal{X}_1||Q| + 3$. The needed cardinality of $\mathcal{U}_2$ can be bounded in an analogous way as $|\mathcal{U}_2| \leq |\mathcal{X}_2||Q| + 3$

A.1.1.4 The Cardinality of $\mathcal{V}$

Define for a given channel $p_1(y_R|x_1, x_2)$, and fixed α, β

$$\mathcal{S} = \bigcup_{p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)p(\hat{y}_R|y_R)} \{[\delta_1(p), \delta_2(p), \delta_3(p), \delta_4(p), \delta_5(p), \delta_6(p), \delta_7(p)]$$
$$|p = p(u_1)p(u_2)p(x_1|u_1)p(x_2|u_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)\}$$

where the union is over the compact set of all $p(q)p(u_1|q)p(u_2|q)p(x_1|u_1)p(x_2|u_2)p(\hat{y}_R|y_R)$ and where

$$\begin{aligned}
\delta_1(p) &= \alpha I(X_1; \hat{Y}_R|X_2, U_1) \\
\delta_2(p) &= \alpha I(X_2; \hat{Y}_R|X_1, U_2) \\
\delta_3(p) &= \alpha I(U_1; Y_R|U_2, Q) \\
\delta_4(p) &= \alpha I(U_2; Y_R|U_1, Q) \\
\delta_5(p) &= \alpha I(U_2 U_2; Y_R|Q) \\
\delta_6(p) &= \alpha I(\hat{Y}_R; Y_R|X_1, U_2) \\
\delta_7(p) &= \alpha I(\hat{Y}_R; Y_R|X_2, U_1)
\end{aligned}$$

Furthermore let $s_1 \in \Delta_{|\mathcal{X}_R|}$ and define for a given channel $p_2(y_1, y_2|x_R)$, fixed α, β and fixed $\hat{p} = p(x_R)$

$$\hat{\mathcal{S}}(\hat{p}) = \bigcup_{s_1 \in \Delta_{|\mathcal{X}_R|}} \{\gamma_1(\hat{p}), \gamma_2(\hat{p}), s_1\}$$

where the union is over all possible $s_1 \in \Delta_{|\mathcal{X}_R|}$, and we have

$$\begin{aligned}\gamma_1(\hat{p}) &= \sum_{y_2,i} s_1(i)p(y_2|X_R = i)\log(\sum_j s_1(j)p(y_2|X_R = j))\\ \gamma_2(\hat{p}) &= \sum_{y_1,i} s_1(i)p(y_1|X_R = i)\log(\sum_j s_1(j)p(y_1|X_R = j))\end{aligned}$$

Now, let $C(\hat{p}) = \mathrm{ConvexHull}(\hat{\mathcal{S}}(\hat{p}))$ and

$$\bar{C} = \bigcup_{\hat{p}} \Big\{[H(Y_2) + \gamma_1, H(Y_1) + \gamma_2, -H(Y_1|X_R) - \gamma_2, -H(Y_2|X_R) - \gamma_1] \\ : \exists[\gamma_1, \gamma_2, s_1] \in C(\hat{p}) \text{ with } \forall i \quad s_1(i) = p(X_R = i)\Big\}.$$

The achievable rate region can now be stated as

$$\begin{aligned}\mathcal{R}_4 = \Big\{&[R_1, R_2] : \exists R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)} \geq 0;\\ &\exists[\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7] \in \mathcal{S} \text{ and } \exists[\delta_8, \delta_9, \delta_{10}, \delta_{11}] \in \bar{C} \text{ with}\\ &R_1^{(1)} + R_1^{(2)} = R_1, R_2^{(1)} + R_2^{(2)} = R_2,\\ &R_1^{(2)} \leq \delta_1, R_2^{(2)} \leq \delta_2, R_1^{(1)} \leq \min\{\delta_3, \delta_8\}, R_2^{(1)} \leq \min\{\delta_4, \delta_9\}, R_1^{(1)} + R_2^{(1)} \leq \delta_5, \delta_6 < \delta_{10}, \delta_7 < \delta_{11}\Big\}.\end{aligned}$$

Now, the set $\hat{\mathcal{S}}$ is connected, as it is the continuous image of a continuous compact set. Therefore all points in C can be expressed as a convex combination of at most $\dim\{\hat{\mathcal{S}}\} = |\mathcal{X}_R| + 1$ elements of $\hat{\mathcal{S}}$. Therefore we can bound the required cardinality of $\mathcal{V}$ from above by $|\mathcal{X}_R| + 1$.

A.1.1.5 The Cardinality of the Auxiliary Random Variables in Corollary 3.6

Applying the above arguments to Corollary 3.6 it follows immediately, that in this case cardinalities $|\mathcal{Q}| \leq 5$, $|\hat{\mathcal{Y}}_R| \leq |\mathcal{Y}_R| + 1$, $|\mathcal{V}| \leq |\mathcal{X}_R| + 1$, $|\mathcal{U}_1| \leq |\mathcal{X}_1||\mathcal{Q}| + 1$, and $|\mathcal{U}_2| \leq |\mathcal{X}_2||\mathcal{Q}| + 3$ are sufficient to achieve all points in the region $\mathcal{R}_5$. For the region $\mathcal{R}_6$ the cardinalities $|\mathcal{Q}| \leq 5$, $|\hat{\mathcal{Y}}_R| \leq |\mathcal{Y}_R| + 1$, $|\mathcal{V}| \leq |\mathcal{X}_R| + 1$, $|\mathcal{U}_1| \leq |\mathcal{X}_1||\mathcal{Q}| + 3$, and $|\mathcal{U}_2| \leq |\mathcal{X}_2||\mathcal{Q}| + 1$ are sufficient.

A.1.2 The Cardinality of Q and $\hat{Y}$ in Theorem 4.1

As mentioned in the remarks for Theorem 4.1 the region stated in the Theorem is not convex, but can be convexified by timesharing over codes for the three regions implicitly defined by the inequalities in the theorem. We can still bound the cardinality of Q and $\hat{Y}$ in Theorem 4.1 similar to the proceeding above:

A.1.2.1 The cardinality of Q

Define for given channels $p_1(y_R|x_1, x_2)$ and fixed α, β

$$\mathcal{S} = \bigcup_{p(x_1)p(x_2)p(\hat{y}_R|y_R)} \Big\{[\delta_1(p), \delta_2(p), \delta_3(p), \delta_4(p)]|p = p(x_1)p(x_2)p_1(y_R|x_1, x_2)p(\hat{y}_R|y_R)\Big\}$$

where the union is over the compact set of all distributions $p(x_1)p(x_2)p(\hat{y}_R|y_R)$ and where

$$\begin{aligned}
\delta_1(p) &= \alpha I(X_1; \hat{Y}_R|X_2) \\
\delta_2(p) &= \alpha I(X_2; \hat{Y}_R|X_1) \\
\delta_3(p) &= \alpha \left(I(X_1X_2; \hat{Y}_R) - I(Y_R; \hat{Y}_R)\right).
\end{aligned} \tag{A.1}$$

Furthermore let $C = \text{ConvexHull}(\mathcal{S})$ and let

$$\hat{\mathcal{S}} = \bigcup_{p(x_R)} \Big\{[\beta I(X_R, Y_1), \beta I(X_R, Y_2)]\Big\}$$

for a given channel $p(y_1, y_2|x_R)$. The achievable rate region can now be stated as

$$\begin{aligned}
\mathcal{R}_7 = \Big\{&[R_1, R_2] : \exists[\delta_1, \delta_2, \delta_3] \in C, [\delta_4, \delta_5] \in \hat{\mathcal{S}} \\
&\text{with } R_1 \leq \max\Big\{0, \min\{\delta_1, \delta_3 + \delta_4\}\Big\}, R_2 \leq \max\Big\{0, \min\{\delta_2, \delta_3 + \delta_5\}\Big\}\Big\}.
\end{aligned}$$

The set $\mathcal{S}$ is connected, as it is the continuous image of a continuous compact set. Therefore all points in C can be expressed as a convex combination of at most $\dim\{\mathcal{S}\} = 3$ elements of $\mathcal{S}$. Therefore we can upper bound the required cardinality of Q by 3.

A.1.2.2 The cardinality of $\hat{\mathcal{Y}}$

Let $s_1 \in \Delta_{|\mathcal{Y}_R|}$. For given channels $p_1(y_R|x_1, x_2)$, $p_2(y_1, y_2|x_R)$, fixed α, β and fixed $p = p(q)$ $p(x_1|q)p(x_2|q)$ let the set $\mathcal{S}(p)$ be given by

$$\mathcal{S}(p) = \bigcup_{s_1} \Big\{[\delta_1, \delta_2, \delta_3, s_1]\Big\}$$

where the union is over all possible $s_1 \in \Delta_{|\mathcal{Y}_R|}$, and we have

$$\begin{aligned}
\delta_1 = \alpha\Big(&H(X_1|X_2, Q) + \sum_{i,x_1,x_2,q} s_1(i)p(x_1, x_2, q|Y_R = i) \\
&\times \left(\log\Big(\sum_j s_1(j)p(x_2, q|Y_R = j)\Big) - \log\Big(\sum_j s_1(j)p(x_1, x_2, q|Y_R = j)\Big)\right)\Big)
\end{aligned}$$

$$\delta_2 = \alpha\Big(H(X_2|X_1, Q) + \sum_{i,x_1,x_2,q} s_1(i)p(x_1, x_2, q|Y_R = i) \\ \times \Big(\log\Big(\sum_j s_1(j)p(x_1, q|Y_R = j)\Big) - \log\Big(\sum_j s_1(j)p(x_1, x_2, q|Y_R = j)\Big)\Big)\Big)$$

$$\delta_3 = -\alpha\Big(H(Y_R|X_1, X_2, Q) + \sum_{i,x_1,q} s_1(i)p(x_1x_2, q|Y_R = i) \\ \times \Big(\log\Big(\sum_j s_1(j)p(x_1, x_2, q|Y_R = j)\Big) - \log\Big(s_1(i)p(x_1, x_2, q|Y_R = i)\Big)\Big)\Big).$$

Here we use the common convention $0 \log 0 = 0$ justified by continuity since $x \log x \to 0$ as $x \to 0$. Let $C(p) = \text{ConvexHull}(\mathcal{S}(p))$ and let

$$\hat{\mathcal{S}} = \bigcup_{p(x_R)} \Big\{[\beta I(X_R, Y_1), \beta I(X_R, Y_2)]\Big\}.$$

Furthermore let $\bar{C}(p) = \Big\{[\delta_1, \delta_2, \delta_3, s_1] \in C(p) | \forall i \quad s_1(i) = p(Y_R = i)\Big\}$. Now the achievable rate region can be stated as

$$\mathcal{R}_7 = \bigcup_p \Big\{[R_1, R_2] : \exists[\delta_1, \delta_2, \delta_3, s_1] \in \bar{C}(p), [\delta_4, \delta_5] \in \hat{\mathcal{S}} \text{ with} \\ 0 \le R_1 \le \max\Big\{0, \min\{\delta_1, \delta_3 + \delta_4\}\Big\}, 0 \le R_2 \le \max\Big\{0, \min\{\delta_2, \delta_3 + \delta_5\}\Big\}\Big\}.$$

The set $\mathcal{S}(p)$ is connected, as it is the continuous image of the continuous compact set $\Delta_{|\mathcal{Y}_R|}$. Therefore all points in $C(p)$ can be expressed as a convex combination of at most $\dim\{\mathcal{S}(p)\} = |\mathcal{Y}_R| + 2$ elements of $\mathcal{S}(p)$. As $\bar{C}(p) \subset C(p)$ all points in $\bar{C}(p)$ can be expressed as a convex combination of at most $\dim\{\mathcal{S}(p)\} = |\mathcal{Y}_R| + 2$ elements of $\mathcal{S}(p)$. Therefore all points in the achievable rate region can be achieved with $|\hat{\mathcal{Y}}_R| \le |\mathcal{Y}_R| + 2$.

Publication List

[1] R. Heras-Evangelio, S. Stanczak, and C. Schnurr, "Energy efficiency of a two-stage relaying scheme in wireless sensor networks," in *Proc. 42nd Annual Conference on Information Sciences and Systems (CISS 2008)*, Princeton University, USA, Mar., pp. 842–847.

[2] ——, "On the optimal power allocation in relay networks based on a two-stage relaying scheme," in *Proc. 7th International ITG Conference on Source and Channel Coding (SCC'08)*, Ulm, Germany, Jan. 2008.

[3] C. Schnurr and S. Stanczak, "On the bit error probability in CDMA channels with correlated binary data: Bounds and optimal sequences," in *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, Nov. 2005, pp. 116–120.

[4] ——, "On correlated binary data: Bounds, optimal sequences, and their dependence on stochastic interactions," in *TURBO-CODING-2006, 4th International Symposium on Turbo Codes and Related Topics, 6th International ITG-Conference on Source and Channel Coding*, Munich, Germany, Apr. 2006.

[5] C. Schnurr, S. Stanczak, and A. Sezgin, "The impact of different mimo strategies on the network-outage performance," in *ITG/IEEE Workshop on Smart Antennas (WSA)*, Vienna, Austria, Feb. 2007.

[6] T. J. Oechtering, C. Schnurr, I. Bjelakovic, and H. Boche, "Achiebable rate region of a two phase bidirectional relay channel," in *Proceedings of the 41st Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2007, pp. 408–413.

[7] ——, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 454–458, Jan. 2008.

[8] I. Bjelakovic, T. J. Oechtering, C. Schnurr, and H. Boche, "On the strong converse for the broadcast capacity region of two-phase bidirectional relaying," in *Proc. IEEE Information Theory Workshop (ITW 2007)*, Bergen, Norway, July 2007.

[9] C. Schnurr, T. J. Oechtering, and H. Boche, "On coding for the broadcast phase in the two-way relay channel," in *Proceedings of the 41st Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2007, pp. 271–276.

[10] R. F. Wyrembelski, T. J. Oechtering, I. Bjelakovic, C. Schnurr, and H. Boche, "Capacity of Gaussian MIMO bidirectional broadcast channels," in *2008 IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, Jul 2008, pp. 584–588.

[11] C. Schnurr, T. J. Oechtering, and S. Stanczak, "Achievable rates for the restricted half-duplex two-way relay channel," in *Asilomar Conference on Signals, Systems, and Computers*, Nov. 2007.

[12] C. Schnurr, S. Stanczak, and T. J. Oechtering, "Achievable rates for the restricted half-duplex two-way relay channel under a partial-decode-and-forward protocol," in *IEEE Information Theory Workshop (ITW 2008)*, May 2008, pp. 134–138.

[13] C. Schnurr, S. Stanczak, and T. Oechtering, "Coding theorems for the restricted half-duplex two-way relay channel with joint decoding," in *2008 IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, Jul 2008, pp. 2688–2692.

References

[14] C. E. Shannon, "Two-way communication channels," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 611–644.

[15] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, Sept. 1971.

[16] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 5, pp. 572–584, Sept. 1979.

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

[18] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of mimo and cooperative mimo techniques in sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.

[19] Y. Lang, D. Wübben, C. Bockelmann, and K.-D. Kammeyer, "A closed power allocation solution for outage restricted distributed mimo multi-hop networks," in *Workshop on Resource Allocation in Wireless Networks (RAWNET 08)*, Berlin, Germany, Apr. 2008.

[20] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *23rd Conference of the IEEE Communications Society (INFOCOM)*, Hong-Kong, Mar. 2004.

[21] S. J. Beak, G. de Veciana, and X. Su, "Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1130–1140, Aug. 2004.

[22] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Asilomar Conference on Signals, Systems, and Computers 2005*, Nov. 2005.

[23] T. J. Oechtering and H. Boche, "Optimal resource allocation for a bidirectional regenerative half-duplex relaying," in *Proc. IEEE International Symposium on Information Theory and its Applications*, Seoul, Korea, 2006, pp. 528–533.

[24] Y. Wu, P. A. Chou, and S. Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," in *Proceedings of the 39th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2005.

[25] C. Fragouli, J.-Y. L. Boudec, and J. Widmer, "Network coding: An instant primer," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, pp. 63–68, 2006.

[26] R. W. Yeung, "Network Coding Theory Part I/II: Single/Multiple Source," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 4–5, pp. 241–381, 2005.

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

[28] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *2006 IEEE International Symposium on Information Theory (ISIT)*, 2006, pp. 1668–1672.

[29] E. Tuncel, "Slepian-wolf coding over broadcast channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. New York: John Wiley & Sons, INC., 1991.

[31] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 3, pp. 277–288, May 1980.

[32] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *The Bell System Technical Journal*, vol. 52, no. 7, pp. 1037–1076, Sept. 1973.

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

[34] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrrily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.

[35] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 396–412, May 1983.

[36] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 626–642, Mar. 2003.

[37] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[38] H. El Gamal, "On the scaling laws of dense wireless sensor networks: The data gathering channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 1229–1234, Mar. 2005.

[39] J. Garcia-Frias, W. Zhong, and Y. Zhao, "Iterative decoding schemes for source and joint source-channel coding of correlated sources," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 2002, pp. 250–256.

[40] A. D. Murugan, P. K. Gopala, and H. El Gamal, "Correlated sources over wireless channels: cooperative source-channel coding," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 988–998, Aug. 2004.

[41] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[42] R. S. Blum, "MIMO capacity with interference," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 793–801, June 2003.

[43] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[44] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symposium on Information Theory*, Tsahkadsor, Armenian, USSR, Sept. 1971, pp. 23–52.

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

[46] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[47] K. Marton, "A coding theorem for the discrete memoryless braodcast channel," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.

[48] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.

[49] R. Knopp, "Two-way radio networks with a star topology," in *2006 International Zurich Seminar on Communications*, Feb. 2006, pp. 154–157.

[50] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[51] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 2, pp. 197–207, Mar. 1973.

[52] R. Ahlswede, "On two-way communication channels and a problem by Zarankiewicz," in *Sixth Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand. Proc.* Publ. House Chechosl. Academy of Sc., Sept. 1971, pp. 23–37.

[53] C. Fragouli, J. Y. Boudec, and J. Widmer, "Network coding: An instant primer," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 1, pp. 63–68, Jan. 2006.

[54] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[55] U. Erez and R. Zamir, "Achieving 1/2log(1+snr) on the awgn channel with lattice encoding and decoding," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[56] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[57] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. Costello, "Nested codes with multiple interpretations," in *2006 40th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2006, pp. 851–856.

[58] J. K. Wolf, "Source coding for a noiseless broadcast channel," in *Proceedings of the 38th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2004, pp. 666–671.

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

[60] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, 1959.

[61] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes," in *IEEE International Conference on Communications (ICC 93)*, vol. 2, Geneva, Switzerland, May 1993, pp. 1064–1070.

[62] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. New York: Springer-Verlag, 2001.

[63] T. S. Han and M. H. M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 5, pp. 641–650, Sept. 1987.

[64] P. Elias, "Coding for noisy channels," in *IRE Convention Record*, vol. 4, 1955, pp. 37–46.

[65] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[66] K. Narayanan, M. P. Wilson, and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," in *Proceedings of 2007 Allerton Conference on Communication, Control and Computing*, Sept. 2007.

[67] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[68] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 2, pp. 219–221, Mar. 1979.

[69] A. El Gamal and E. C. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.