TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Theoretische Informationstechnik

Robust Coding Strategies and Physical Layer Service Integration for Bidirectional Relaying

Rafael Felix Wyrembelski

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. sc. techn. Gerhard Kramer
Prüfer der Dissertation:
1. Univ.-Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
2. Prof. Dr. Vincent Poor (Princeton University, USA)

Die Dissertation wurde am 06.10.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 22.03.2012 angenommen.

Zusammenfassung

Aktuelle Forschungsergebnisse zeigen, dass Relaiskonzepte die Leistungsfähigkeit und Abdeckung drahtloser Kommunikationssysteme deutlich erhöhen können. In dieser Dissertation betrachten wir *bidirektionale Relais-Kommunikation* in einem Netzwerk mit drei Stationen, in dem zwei Stationen mit Hilfe einer Relaisstation kommunizieren.

Im ersten Teil der Dissertation analysieren wir die Auswirkungen ungenauer Kanalkenntnis an Sender und Empfängern und betrachten den dazugehörigen diskreten gedächtnislosen *bidirektionalen Compound Broadcastkanal*. Wir leiten die Kapazitätsregion her und analysieren die Szenarien, in denen zusätzlich entweder der Sender oder die Empfänger perfekte Kanalkenntnis besitzen. Hierbei zeigt sich, dass Kanalkenntnis an den Empfängern keinen Einfluss auf die Kapazitätsregion hat, während Kanalkenntnis am Sender diese vergrößern kann.

Anschließend betrachten wir bidirektionale Relais-Kommunikation in unkoordinierten drahtlosen Netzwerken. In solchen Netzwerken wird die Kommunikation durch unbekannte, sich beliebig ändernde Interferenz gestört. Wir betrachten den diskreten gedächtnislosen *bidirektionalen beliebig variierenden Broadcastkanal* und leiten die Kapazitätsregion für deterministische und randomisierte Kodierungsstrategien her. Weiterhin analysieren wir Empfänger mit Listen-Dekodierern und charakterisieren die zugehörige Kapazitätsregion in Abhängigkeit der erlaubten Listengrößen und der Symmetrisierung des Kanals.

Im letzten Teil der Dissertation betrachten wir die *effiziente Implementierung verschiedener Dienste auf der physikalischen Schicht* für bidirektionale Relais-Netzwerke mit diskreten gedächtnislosen Kanälen sowie für Netzwerke mit mehreren Antennen. Wir betrachten die Implementierung von Multicast Diensten und analysieren den bidirektionalen Broadcastkanal mit gemeinsamen Nachrichten. Wir leiten die entsprechende Kapazitätsregion her. Dabei zeigt sich, dass die kapazitätserreichenden Strategien für die bidirektionalen Broadcastkanäle mit und ohne gemeinsamen Nachrichten eng miteinander verbunden sind. Anschließend betrachten wir die Implementierung von zusätzlichen vertraulichen Nachrichten. Wir nutzen das Kriterium der informationstheoretischen Sicherheit, um die zugehörigen Sicherheitskapazitätsregionen herzuleiten.

Im abschließenden Fazit geben wir einen Ausblick auf offene Probleme und zukünftige Forschungsrichtungen.

Abstract

Recent research developments show that the concept of *bidirectional relaying* significantly improves the performance and coverage in wireless networks. In this thesis we consider bidirectional relaying in a three-node network, where a relay establishes a bidirectional communication between two other nodes using a decode-and-forward protocol.

In the first part we consider the problem of imperfect channel state information and study the discrete memoryless *compound bidirectional broadcast channel*. We derive the capacity region and further discuss the cases where either the transmitter or the receivers have perfect channel state information. It shows that channel knowledge on the receiver side has no influence on the achievable rates while on the transmitter side it increases the capacity.

In the next part we consider bidirectional relaying in uncoordinated wireless networks, where the communication is disturbed by unknown varying interference from other transmitters outside the bidirectional relay network. Accordingly, we study the discrete memoryless *arbitrarily varying bidirectional broadcast channel* and derive the corresponding capacity regions for deterministic and random coding strategies. We further study the influence of list decoding and characterize the list capacity region in terms of the list sizes at the receivers and of the symmetrizability of the channel. Then we impose constraints on the permissible codewords and sequences of channel states and finally analyze the case of unknown varying additive interference.

In the last part we address the problem of *physical layer service integration* in discrete memoryless and multi-antenna Gaussian bidirectional relay networks. First, we consider multicast services and, accordingly, study the bidirectional broadcast channel with common messages. We derive the capacity region and thereby establish a strong connection with the corresponding scenario without common messages. Then, we take into account that there are also services with certain secrecy constraints. We use the concept of information theoretic security to model this requirement and study the additional integration of confidential messages. We derive the capacity-equivocation and secrecy capacity regions.

Finally, we end with a conclusion and give an outlook on open problems and future research directions.

Acknowledgments

During my time at the Technische Universität Berlin and the Technische Universität München I had the opportunity to meet many interesting persons. First and foremost, I want to thank my advisor Prof. Holger Boche for giving me the opportunity to work with him. I am grateful for the scientific freedom, his support, and endless motivation. I would also like to express my sincere gratitude to Prof. Vincent Poor for serving as the second referee of this dissertation. Further, I thank Prof. Gerhard Kramer for acting as the chairman of my thesis committee.

I want to thank all my colleagues for the pleasant social atmosphere and the stimulating discussions. Many thanks go to Tobias Oechtering for collaborating with me and guiding me in my first steps. Further, I thank Igor Bjelaković for collaborating with me and teaching me information theory down to the very last detail. Special thanks go to Ullrich Mönich for sharing the room with me most of the time of our Ph.D. studies.

My heartfelt gratitude goes to my family for their love and unquestioning support. Finally, my special and heartfelt thanks go to Anabel. Without her love and encouragement this work would not have been possible.

Contents

List of Figures

Notation

In this work we denote scalars, vectors, matrices, and sets by lower case letters, bold lower case letters, bold capital letters, and script letters, e.g., x, $\boldsymbol{x}$, $\boldsymbol{X}$, and $\mathcal{X}$. Further, we use:

$\mathbb{N}$	set of positive integers, i.e., $\{1, 2, 3, ...\}$
$\mathbb{R}_+$	set of non-negative real numbers
$\mathbb{C}$	set of complex numbers
∞	infinity
$\emptyset$	empty set
$\mathcal{X}^c$	complement of set $\mathcal{X}$
$\lvert\mathcal{X}\rvert$	cardinality of set $\mathcal{X}$
$\text{co}(\mathcal{X})$	convex hull of set $\mathcal{X}$
$\text{int}(\mathcal{X})$	interior of set $\mathcal{X}$
lhs $:=$ rhs	value of right hand side (rhs) is assigned to left hand side (lhs)
lhs $=:$ rhs	value of left hand side (lhs) is assigned to right hand side (rhs)
$\exists$	there exists
$\forall$	for all
$\lfloor\cdot\rfloor$	floor function maps a real number to largest previous integer
$\lceil\cdot\rceil$	ceiling function maps a real number to smallest following integer
$\lvert\cdot\rvert^+$	abbreviation for $\max\{0, \cdot\}$
$\log$	logarithm to base two
$\ln$	natural logarithm
$\exp$	exponential function
$\boldsymbol{x}^T$, $\boldsymbol{X}^T$	transpose of vector $\boldsymbol{x}$ resp. matrix $\boldsymbol{X}$
$\boldsymbol{x}^H$, $\boldsymbol{X}^H$	Hermitian transpose of vector $\boldsymbol{x}$ resp. matrix $\boldsymbol{X}$
$\boldsymbol{X}^{-1}$	inverse of matrix $\boldsymbol{X}$
$\det(\boldsymbol{X})$	determinant of matrix $\boldsymbol{X}$
$\text{tr}(\boldsymbol{X})$	trace of matrix $\boldsymbol{X}$
$\text{rank}(\boldsymbol{X})$	rank of matrix $\boldsymbol{X}$
$\boldsymbol{X} \succeq \boldsymbol{0}$	matrix $\boldsymbol{X}$ is positive semidefinite
$\boldsymbol{I}_N$	identity matrix of dimension $N \times N$
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$	inner product between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$
$\lVert\boldsymbol{x}\rVert$	Euclidean norm of vector $\boldsymbol{x}$
$o(g(x))$	$f(x)$ is little-o of $g(x)$ if $\lim_{x\to\infty} \frac{g(x)}{f(x)} = 0$

We denote random variables by non-italic capital letters and their realizations and ranges by lower case italic letters and script letters, e.g., X, x, and $\mathcal{X}$, respectively. The notation X^n stands for the sequence $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ of length n.

Let $\mathcal{P}(\mathcal{X})$ be the set of all probability distributions on $\mathcal{X}$. We denote (arbitrary) probability distributions by lower case letters, e.g., $p_{\mathrm{X}} \in \mathcal{P}(\mathcal{X})$ is the probability distribution associated with the random variable X, while capital letters are devoted to *types*, e.g., $P_{\mathrm{X}} \in \mathcal{P}_0(n, \mathcal{X})$ where $\mathcal{P}_0(n, \mathcal{X})$ is the set of all types on $\mathcal{X}$ of length n. We refer to Appendix B for further details.

Let $p_{\mathrm{X}} \in \mathcal{P}(\mathcal{X})$ be a probability distribution and $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ be a stochastic matrix. Then, we denote the entropy of the random variable X by $H(\mathrm{X})$. To emphasize the dependency of the entropy on the probability distribution $p_{\mathrm{X}} \in \mathcal{P}(\mathcal{X})$, we also write $H(\mathrm{X}) = H(p_{\mathrm{X}})$ interchangeably. This extends to joint entropy $H(\mathrm{X}, \mathrm{Y}) = H(p_{\mathrm{XY}})$ and conditional entropy $H(\mathrm{Y}|\mathrm{X}) = H(W|p_{\mathrm{X}})$ in a natural way. We denote the mutual information between the random variables X and Y by $I(\mathrm{X}; \mathrm{Y})$. Accordingly, we also write $I(\mathrm{X}; \mathrm{Y}) = I(p_{\mathrm{X}}, W)$ to emphasize the dependency on the input distribution $p_{\mathrm{X}} \in \mathcal{P}(\mathcal{X})$ and the channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$. For further details we refer to standard text books on (multi-user) information theory as for example by Gallager [Gal68], Wolfowitz [Wol78], Csiszár and Körner [CK81], Cover and Thomas [CT06], Kramer [Kra08], or El Gamal and Kim [EK11].

Further, we use the following notation:

$W^{\otimes n}$	n-th memoryless extension of the stochastic matrix W
$\mathrm{X} - \mathrm{Y} - \mathrm{Z}$	random variables X, Y, and Z form a Markov chain in this order
$\mathcal{P}(\mathcal{X})$	set of all probability distributions on $\mathcal{X}$
$\mathcal{P}_0(n, \mathcal{X})$	set of all types on $\mathcal{X}$ of length n
$H(\mathrm{X}) = H(p_{\mathrm{X}})$	traditional entropy of discrete random variable X
$h(\mathrm{X})$	differential entropy of continuous random variable X
$H_2(\cdot)$	binary entropy function
$I(\mathrm{X}; \mathrm{Y}) = I(p, W)$	mutual information between X and Y
$D(p_{\mathrm{X}} \Vert p_{\mathrm{Y}})$	(Kullback-Leibler) information divergence between p_{X} and p_{Y}
$\mathrm{X} \sim \mathcal{N}(m, \sigma^2)$	X is Gaussian distributed with mean m and variance σ^2
$\mathrm{X} \sim \mathcal{CN}(m, \sigma^2)$	X is complex Gaussian distributed with mean m and variance σ^2
$\mathrm{var}[\mathrm{X}]$	variance of X
$\mathbb{E}_{\mathrm{X}}[\cdot]$	expectation with respect to X
$\mathbb{P}\{\cdot\}$	probability
$\mathcal{T}_{\mathrm{X}}^{(n)}$	set of sequences of type P_{X}
$\mathcal{T}_{p_{\mathrm{X}},\epsilon}^{(n)}$	set of (strongly) ϵ-typical sequences with respect to p_{X}
$\mathcal{A}_{\epsilon}^{(n)}(\mathrm{X})$	set of (weakly) ϵ-typical sequences with respect to p_{X}

Abbreviations

3GPP	3rd Generation Partnership Project
arg	argument
AVBBC	arbitrarily varying bidirectional broadcast channel
AVC	arbitrarily varying channel
AVGBC	arbitrarily varying general broadcast channel
AVMAC	arbitrarily varying multiple access channel
BBC	bidirectional broadcast channel
BC	broadcast channel
CC	compound channel
CSI	channel state information
CSIT	channel state information at the transmitter
CSIR	channel state information at the receiver
det	deterministic
DMC	discrete memoryless channel
iid	independent and identical distributed
inf	infimum
KKT	Karush-Kuhn-Tucker
lim	limes
limsup	limes superior
liminf	limes inferior
LTE	Long-Term Evolution
MAC	multiple access channel
max	maximum
min	minimum
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
OFDM	orthogonal frequency division multiplex
ran	random
SIMO	single-input multiple-output
SVD	singular value decomposition
sup	supremum
TDMA	time division multiple access

1 Introduction

1.1 Motivation

Almost all technological advances would not be possible without the immense progress in integrated circuit design. In 1965 Moore observed that the number of transistors on an integrated circuit or chip doubles approximately every two years while the production costs remain constant [Moo65]. This is widely known as *Moore's law*. Although this prediction was proposed almost 50 years ago, it is astonishing that it is still valid and it seems to continue for the next years. However, sooner or later this growth will come to an end due to limits of miniaturization at atomic levels. By then at the latest we have to find other concepts to keep the technological progress alive.

With the rapid development of integrated circuits the capabilities of electronic devices, such as processing speed or memory capacity, increase exponentially. This made the development of wireless communication systems possible which are nowadays omnipresent. It started with cellular networks for voice communication only and continued to high speed data services such as mobile Internet or video streaming. However, the available network bandwidth is the most defining bottleneck for wireless communication systems, since favorable frequency bands are scarce. The ongoing technological development will extend the usable frequencies to higher regions but with an increasing frequency the radio propagation conditions will be more and more susceptible. Thus, the problem of coverage in wireless systems will be even more intricate especially when the direct link does not have the desired quality due to distance or shadowing. At a first glance, an increase in transmit power seems to be an obvious option but this will result in higher interference for other users and a higher energy consumption. Clearly, operators and especially manufacturers of mobile devices are interested in a low energy consumption since this results in longer operating times of the devices. Furthermore, current cellular systems are usually interference limited so that a simple increase in transmit power is no option. A possible solution for conventional network structures would be to increase the number of base stations, but this would lead to significant higher costs in infrastructure.

Besides networks with wired infrastructure such as conventional cellular networks, there are also networks without any fixed infrastructure which are called ad-hoc networks. Such networks are becoming more and more attractive since they exploit the broadcast nature

of the wireless medium so that in principle each node is connected to each other node in the network. In practice, due to radio propagation conditions each node can only transmit reliably to nodes in a certain neighborhood. Therefore, information in such networks is usually not exchanged directly but by multi-hop communication. This means that a source-destination pair usually cooperates with other nodes so that the information is relayed by one or more intermediate nodes. A survey about cooperative communication can be found for example in [KMY06]. Because of this relaying feature, ad-hoc networks have further favorable properties. For example each additional node that comes into the network will increase the overall connectivity in the network. Further, the routing tasks of a certain node can be taken over by other nodes if this node should disappear from the network due to bad channel conditions or low battery capacity. Ad-hoc networks do not need a wired backbone which makes them easy to deploy. It is clear that for certain applications such as sensor networks or car-to-car communication, a flexible network topology is more favorable than a fixed cellular topology.

The advantages of multi-hop communication make it worth to integrate relaying techniques also in conventional cellular networks to improve the performance and especially the coverage and connectivity. Not surprisingly, this is intensively discussed at the moment by the Third Generation Partnership Project's Long Term Evolution-Advanced (LTE-Advanced) group, cf. for example [PPTH09]. Instead of a direct connection between the base station and a mobile device, relays can be used to support the exchange of information between them. These relays can be other mobile terminals or can be fixed and placed over the cells. Since relays are less complex and do not require a connection to the wired backbone, the coverage and connectivity of cellular networks can be improved with lower infrastructure costs compared to installing additional base stations. Moreover, the use of relays splits up the distance between the base station and the mobile device so that the transmission is performed in several steps. Since the received signal power falls off super-linear with the distance [Rap02], the power that is needed for the transmission can be further reduced.

The discussion shows that wireless networks benefit from multi-hop and relay communication. If a relay is used within a cellular network for range extension, the information flow is usually bidirectional due to the underlying communication scenario. Therefore, it is not surprisingly that the first multi-user communication problem considered in the literature is Shannon's two-way channel [Sha61]. Here, two users simultaneously want to exchange information as effectively as possible. In [Sha61] Shannon derived the capacity region for the restricted two-way channel, which means that feedback between the encoders is not allowed. Until now, the general case is still unsolved.

Another important basis for multi-hop communication is the relay channel. It was introduced by van der Meulen in [van71] and considers a three-node scenario where a relay supports the communication from a source to a destination. In [CE79] Cover and El Gamal established the capacities for the degraded relay channel, the reversely degraded relay channel, and the relay

channel with feedback from both receivers to the source and relay. More recently, Kramer, Gastpar, and Gupta discussed several relay scenarios under wireless communication aspects in [KGG05]. But unfortunately, similarly to the two-way channel the capacity of the general relay channel remains unknown.

If relays are used within wireless networks, further implementational difficulties appear. Since it is technically almost impossible to sufficiently isolate a received signal from a transmitted signal within the same frequency band, a relay has to allocate orthogonal resources for transmission and reception. Therefore, relays usually operate in half-duplex modus which leads to an inherent loss in spectral efficiency. The consequence is that the use of half-duplex relays is only beneficial if we find techniques which compensate the loss in spectral efficiency. Recent research developments have shown that it is promising to exploit the bidirectional property of the communication to compensate the loss in spectral efficiency [RW07, LJS05, WCK05, Kno06].

All this makes the concept of bidirectional relaying attractive. It applies to three-node networks, where a half-duplex relay node establishes a bidirectional communication between two other nodes. There exist several strategies for bidirectional relaying which are classified by the processing at the relay node. First works mostly consider amplify-and-forward strategies [RW07, ZLCC09, RH10, NQS09, LSPL10, CHK09] or decode-and-forward strategies [RW07, OSBB08, KMT08, LK09, LTXW09]. Other schemes are compress-and-forward [SOS07, GTN08] or compute-and-forward [WNPS10, NCL10, BC07, NG11, OKJ10] approaches, where the relay decodes a certain function of both individual messages.

In this thesis we consider a decode-and-forward protocol, where the relay decodes both messages it received in the initial multiple access phase. It then re-encodes and transmits both messages in such a way that both receiving nodes can decode their intended message using their own message from the previous phase as side information. It is shown in [OSBB08, KMT08, Xie07, KS07] that capacity is achieved by a single data stream that combines both messages based on the network coding idea [ACLY00, YLCZ05, FS07]. This concept breaks with the common model to regard information flows as *"fluids"* [ACLY00] and constitutes a paradigm shift.

Currently, there are many ongoing research activities on bidirectional relaying and its extensions. For example, [PKA09] provides a survey of different processing strategies. The efficient integration of bidirectional relaying in a cellular downlink is presented in [ODS10]. Bidirectional relaying for multiple pairs of nodes is analyzed in [CY09, YZGK10, SAKH11]. Bidirectional relaying with an additional private message for the relay in the MAC phase is addressed in [HGS09]. Beamforming strategies for multi-antenna bidirectional relaying with analog network coding is presented in [ZLCC09]. A deterministic approach that characterizes the capacity of the full-duplex bidirectional relay channel within a constant gap is given in [AST10]. A four-node network with bidirectional communication is discussed in [SWS09], while [ILH10] addresses the problem of joint network and channel coding for

multi-way relaying. A source coding counterpart, i.e., source coding with complementary side information, is studied in [TGK11].

1.2 Contribution and Outline of the Thesis

In Chapter 2 we introduce the concept of bidirectional relaying in a three-node network. We briefly review the multiple access (MAC) and bidirectional broadcast (BBC) phase of the decode-and-forward protocol and summarize the corresponding capacity regions for discrete memoryless and MIMO Gaussian channels. Then we briefly discuss the bidirectional achievable rate region that results if fixed or optimal time division between both phases is applied. Throughout this chapter we assume perfect channel state information at all nodes. Parts of the overview are published in [WOB08b].

In Chapter 3 we analyze bidirectional relaying under channel uncertainty. We use the concept of *compound channels* to model the imperfect channel state information at the nodes. We briefly review the compound multiple access channel and then establish the capacity region of the compound bidirectional broadcast channel without any CSI at all nodes and for partial CSI where either the transmitter or the receivers have perfect CSI. Finally, we give a numerical example and a game-theoretic interpretation. Parts of the results are published in [WBOB09, WBOB10].

In Chapter 4 we extensively examine bidirectional relaying in uncoordinated wireless networks. We use the concept of *arbitrarily varying channels* to model the unknown interference from other transmitting nodes outside the bidirectional relay network. We briefly review the arbitrarily varying multiple access channel and then analyze the arbitrarily varying bidirectional broadcast channel (AVBBC) in detail.

- In Section 4.3 we consider the case where the transmitter and the receivers have access to a common randomness so that they can coordinate their choice of encoder and decoders. We establish the corresponding random code capacity region of the AVBBC using results from the compound BBC, cf. Chapter 3, and Ahlswede's robustification technique [Ahl80b, Ahl86]. Parts of the results are published in [WBB09a] and should be published in [WBB11].

- In Section 4.4 we assume that common randomness is not available so that transmitter and receivers have to use a deterministic coding strategy. We derive the deterministic code capacity region of the AVBBC using Ahlswede's elimination technique [Ahl78] and establish a dichotomy behavior of the deterministic code capacity region. Parts of the results are published in [WBB09a, WBB09b].

- In Section 4.5 we consider list decoding at the receivers and characterize the list capacity region of the AVBBC using the concept of symmetrizability based on [Hug97]. Parts of the results are published in [WBB10b] and should be published in [WBB11].
- In Section 4.6 we impose constraints on the input and state sequences. We establish the random code and deterministic code capacity region of the AVBBC under input and state constraints. Parts of the results are published in [WBB10c] and should be published in [WBB12].
- In Section 4.7 we analyze the AVBBC with unknown varying additive interference. We discuss the impact of coordination and establish the corresponding random code and deterministic code capacity regions. Parts of the results are published [WBB10a].

In Chapter 5 we analyze physical layer service integration in bidirectional relay networks. Here, the relay establishes not only a bidirectional communication between the two other nodes, but also integrates additional common and confidential services.

- In Section 5.2 we discuss the scenario where the relay transmits an additional common message in the bidirectional broadcast phase and derive the corresponding capacity regions for discrete memoryless and MIMO Gaussian channels. We further analyze the transmit covariance optimization problem and therewith establish a strong connection between the BBC with and without common messages. Parts of the results are published in [WOB10, WOB11].
- In Section 5.3 we analyze the case where the relay transmits an additional confidential message that is intended for one node and should be kept secret from the other, non-legitimate node. We establish the capacity-equivocation and secrecy capacity regions of the BBC with confidential messages using similar techniques as in [CK78] for the classical broadcast channel with common and confidential messages. Parts of the results are published in [WB11b] and will be published in [WB12a].
- In Section 5.4 the relay transmits additional common and confidential messages. We present the corresponding capacity-equivocation and secrecy capacity regions that unify the previous results. Parts of the results are published in [WB11a] and should be published in [WB11c].
- In Section 5.5 we derive similar results for MIMO Gaussian bidirectional relay networks. We establish the secrecy capacity region of the MIMO Gaussian BBC with common and confidential messages using channel enhancement arguments as in [LLL10] for the classical MIMO Gaussian broadcast channel with common and confidential messages. Parts of the results are published in [WB11d, WB11e] and should be published in [WB11c].

Finally, in Chapter 6 we conclude the thesis and give an outlook on future research directions and open problems.

Further Results which are not Part of this Thesis

During my time at the Technische Universität Berlin and the Technische Universität München we obtained further interesting results which are not part of this thesis.

- In coauthored works with Tobias Oechtering and Eduard Jorswieck we study optimal transmit strategies for the multi-antenna bidirectional broadcast channel. In [OWB08b, OWB09a] we analyze the MISO Gaussian BBC and show that it is always optimal to transmit into the subspace spanned by the channels. Further, there exists always an optimal single-beam transmit strategy whose transmit covariance matrix is of rank one. This reflects the single stream processing based on the network coding idea. In [OWB09b, OJWB09] we study optimal transmit strategies for the MIMO Gaussian BBC and show that in general there exist different equivalent transmit strategies with different ranks. But for the special case where the ranks of the channels are equal to the number of antennas at the relay node and a full-rank transmission is optimal, the optimal transmit covariance matrix can be obtained. The same is true for the case of parallel channels which is in particular a relevant scenario since it immediately provides also solutions for the power allocation problem of a single-antenna OFDM system.
- In a coauthored work with Tobias Oechtering [OWB08a] we study the optimal time-division for bidirectional relaying where the relay is equipped with multiple antennas while the other two nodes are equipped with a single antenna only. We characterize the bidirectional achievable rate region for this scenario.
- In a work with Aydin Sezgin [WSB11] we study the bidirectional broadcast wiretap channel where the bidirectional communication itself has to be secure from possible eavesdroppers outside the bidirectional relay network. We derive inner and outer bounds on the capacity region of the bidirectional broadcast wiretap channel. This model differs from the BBC with confidential messages which addresses the problem of realizing confidential communication within such a network, cf. Section 5.3.

A complete list of all publications can be found in the appendix.

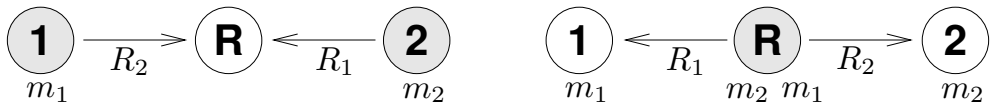
Copyright Information

2 Decode-and-Forward Bidirectional Relaying

In this work we consider a three-node network where a relay node establishes a bidirectional communication between two other nodes. Since it is difficult for the nodes to isolate simultaneously transmitted and received wireless signals within the same frequency band, we assume half-duplex nodes and therefore allocate orthogonal resources in time for orthogonal transmission and reception. Accordingly, the whole transmission is separated into two phases which causes an inherent loss in spectral efficiency for unidirectional protocols. The loss in spectral efficiency can be significantly reduced by the concept of *bidirectional relaying* which advantageously exploits the property of bidirectional communication [RW07, LJS05, WCK05, Kno06]. This is also known as two-way relaying.

In the first phase of a decode-and-forward protocol the two nodes transmit their messages to the relay node. Since the relay is assumed to decode both messages, the first phase corresponds to the classical multiple access channel (MAC). In the succeeding phase it remains for the relay to re-encode and broadcast the messages in such a way that both receiving nodes can decode their intended messages. Since both nodes can use their own messages from the previous phase as side information for decoding, this channel differs from the classical broadcast channel is therefore called *bidirectional broadcast channel (BBC)*.

We do not allow any feedback in the decode-and-forward protocol or any other cooperation between the encoders. Thus, both phases decouple and this setup is known as a *restricted* bidirectional relay channel. Throughout this chapter we assume perfect channel state information (CSI) at all nodes.



(a) MAC phase (b) BBC phase

Figure 2.1: Decode-and-Forward bidirectional relaying in a three-node network.

2.1 Multiple Access Phase

In the initial MAC phase nodes 1 and 2 transmit their messages m_1 and m_2 with rates R_2 and R_1 to the relay node. Since the relay has to decode both messages, this is the classical multiple access channel.

Discrete Memoryless Multiple Access Channel

Here, we briefly restate the capacity region of the *discrete memoryless multiple access channel* which was independently established by Ahlswede [Ahl71] and Liao [Lia72]. Nowadays it is part of any standard book on (multi-user) information theory [Wol78, CK81, CT06, Kra08, EK11].

For the multiple access phase let $\mathcal{X}_i$, $i = 1, 2$, and $\mathcal{Y}$ be finite input and output sets. Then, for input and output sequences $x_i^n \in \mathcal{X}_i^n$, $i = 1, 2$, and $y^n \in \mathcal{Y}^n$ of length n, let $V^{\otimes n}(y^n|x_1^n, x_2^n) \coloneqq \prod_{k=1}^n V(y_k|x_{1,k}, x_{2,k})$.

Definition 2.1. *The* discrete memoryless multiple access channel *is defined by*

$$\left\{V^{\otimes n} : \mathcal{X}_1^n \times \mathcal{X}_2^n \to \mathcal{P}(\mathcal{Y}^n)\right\}_{n\in\mathbb{N}}$$

which we simply denote by V with a slight abuse of notation.

Theorem 2.2 ([Ahl71, Lia72])**.** *The capacity region $\mathcal{R}(V)$ of the multiple access channel V is the set of all rate pairs $(R_2, R_1) \in \mathbb{R}_+^2$ that satisfy*[1]

$$\begin{aligned} R_2 &\leq I(\mathrm{X}_1; \mathrm{Y}|\mathrm{X}_2, \mathrm{U}) \\ R_1 &\leq I(\mathrm{X}_2; \mathrm{Y}|\mathrm{X}_1, \mathrm{U}) \\ R_2 + R_1 &\leq I(\mathrm{X}_1, \mathrm{X}_2; \mathrm{Y}|\mathrm{U}) \end{aligned}$$

for random variables $\mathrm{U} - (\mathrm{X}_1, \mathrm{X}_2) - \mathrm{Y}$ with joint probability distribution $p_{\mathrm{U}}(u)p_{\mathrm{X}_1|\mathrm{U}}(x_1|u)p_{\mathrm{X}_2|\mathrm{U}}(x_2|u)V(y|x_1, x_2)$. Thereby, U is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of U can be bounded by $|\mathcal{U}| \leq 2$.

[1] Recall that the MAC is considered within the two-phase decode-and-forward protocol. Therefore, the individual rates look "swapped". Here, R_2 and R_1 denote the rates from node 1 to the relay node and from node 2 to the relay node, respectively, cf. also Figure 2.1.

MIMO Gaussian Multiple Access Channel

Next, we briefly restate the capacity region of the *MIMO Gaussian multiple access channel*. In principle, the capacity region for MIMO Gaussian channels follows from the corresponding region for discrete memoryless channels, cf. [Ahl71, Lia72] and Theorem 2.2. For further details we refer for example to [VG97, GJJV03, BCC$^+$07] and references therein.

For the multiple access phase we assume N_R antennas at the relay node and N_i antennas at node i, $i = 1, 2$. Then, the discrete-time complex-valued input-output relation between nodes 1 and 2 and the relay node is given by

$$\boldsymbol{y} = \boldsymbol{H}_1\boldsymbol{x}_1 + \boldsymbol{H}_2\boldsymbol{x}_2 + \boldsymbol{n}$$

where $\boldsymbol{y} \in \mathbb{C}^{N_R \times 1}$ denotes the output at the relay node, $\boldsymbol{H}_i \in \mathbb{C}^{N_R \times N_i}$ the multiplicative channel matrix, $\boldsymbol{x}_i \in \mathbb{C}^{N_i \times 1}$ the input of node i, and $\boldsymbol{n} \in \mathbb{C}^{N_R \times 1}$ the independent additive noise according to a circular symmetric complex Gaussian distribution $\mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{N_R})$. We assume an average transmit power constraint $\mathrm{tr}(\boldsymbol{Q}_i) \leq P_i$ with $\boldsymbol{Q}_i = \mathbb{E}\{\boldsymbol{x}_i\boldsymbol{x}_i^H\}$ at node i, $i = 1, 2$.

For the capacity region of the MIMO Gaussian MAC we need the following region. For given covariance matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ let $\mathcal{R}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ be the set of all rate tuples $(R_2, R_1) \in \mathbb{R}_+^2$ that satisfy

$$\begin{aligned} R_2 &\leq \log\det\left(\boldsymbol{I}_{N_R} + \tfrac{1}{\sigma^2}\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H\right) \\ R_1 &\leq \log\det\left(\boldsymbol{I}_{N_R} + \tfrac{1}{\sigma^2}\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right) \\ R_2 + R_1 &\leq \log\det\left(\boldsymbol{I}_{N_R} + \tfrac{1}{\sigma^2}\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^H + \tfrac{1}{\sigma^2}\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^H\right). \end{aligned}$$

Theorem 2.3. *The capacity region $\mathcal{R}(\boldsymbol{H}_1, \boldsymbol{H}_2|P_1, P_2)$ of the MIMO Gaussian MAC under average power constraints P_1 and P_2 is given by*

$$\mathcal{R}(\boldsymbol{H}_1, \boldsymbol{H}_2|P_1, P_2) = co\left(\bigcup_{tr(\boldsymbol{Q}_i)\leq P_i,\, \boldsymbol{Q}_i\succeq\boldsymbol{0}, i=1,2} \mathcal{R}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)\right).$$

2.2 Bidirectional Broadcast Phase

It is reasonable to assume that the relay has successfully decoded the messages both nodes have sent in the previous MAC phase, if the rates are chosen within the corresponding MAC capacity region. Therefore, we assume that the relay has perfect knowledge about messages m_1 and m_2. Now, the relay re-encodes and broadcasts both messages in such a way that nodes 1 and 2 can decode m_2 and m_1 respectively using their own messages m_1 and m_2 as side information.

Discrete Memoryless Bidirectional Broadcast Channel

Here, we briefly restate the capacity region of the *discrete memoryless bidirectional broadcast channel*. Capacity-achieving strategies can be found, for instance, in [OSBB08, KMT08, KS07, Xie07] and can be further deduced from [Tun06].

For the bidirectional broadcast phase let $\mathcal{X}$ and $\mathcal{Y}_i$, $i = 1, 2$, be finite input and output sets. Then, for input and output sequences $x^n \in \mathcal{X}^n$ and $y_i^n \in \mathcal{Y}_i^n$, $i = 1, 2$, of length n, let $W^{\otimes n}(y_1^n, y_2^n|x^n) \coloneqq \prod_{k=1}^n W^{\otimes n}(y_{1,k}, y_{2,k}|x_k)$.

Definition 2.4. *The* discrete memoryless broadcast channel *is defined by*

$$\left\{W^{\otimes n} : \mathcal{X}^n \to \mathcal{P}(\mathcal{Y}_1^n \times \mathcal{Y}_2^n)\right\}_{n\in\mathbb{N}}$$

which we simply denote by W with a slight abuse of notation.

Since we do not allow any cooperation between the receiving nodes, it is sufficient to consider the marginal transition probabilities $W_i^{\otimes n}(y_i^n|x^n) = \prod_{k=1}^n W(y_{i,k}|x_k)$, $i = 1, 2$, only.

Interestingly, the optimal coding strategy for the BBC is based on the idea of network coding [ACLY00, YLCZ05, FS07]. The philosophy of network coding is to convey as much information to the receiving nodes which allows them to conclude on the intended message using their own (side) information. This concept breaks with the common model which regards information flows as "*fluids*" [ACLY00] and constititutes a paradigm shift.

Theorem 2.5 ([OSBB08, KMT08, KS07, Xie07])**.** *The capacity region $\mathcal{R}(W)$ of the discrete memoryless bidirectional broadcast channel W is the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy*

$$R_1 \leq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) \tag{2.1a}$$
$$R_2 \leq I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) \tag{2.1b}$$

for random variables $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ with joint probability distribution $p_{\mathrm{U}}(u)p_{\mathrm{X}|\mathrm{U}}(x|u)W(y_1, y_2|x)$. Thereby, U is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of U can be bounded by $|\mathcal{U}| \leq 2$.

Remark 2.6. *Following [KS07, Theorem 1] it is further possible to get rid of the time-sharing random variable U in (2.1) so that we end up with $R_i \leq I(\mathrm{X};\mathrm{Y}_i)$, $i = 1, 2$.*

MIMO Gaussian Bidirectional Broadcast Channel

Next, we briefly restate the capacity region of the *MIMO Gaussian bidirectional broadcast channel* which was established in [WOB+08a] by extending the corresponding result for discrete memoryless channels to MIMO Gaussian channels.

For the bidirectional broadcast phase we assume N_R antennas at the relay node and N_i antennas at node i, $i = 1, 2$. Then, the discrete-time complex-valued input-output relation between the relay node and node i, $i = 1, 2$, is given by

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x} + \boldsymbol{n}_i$$

where $\boldsymbol{y}_i \in \mathbb{C}^{N_i \times 1}$ denotes the output at node i, $\boldsymbol{H}_i \in \mathbb{C}^{N_i \times N_R}$ the multiplicative channel matrix, $\boldsymbol{x} \in \mathbb{C}^{N_R \times 1}$ the input of the relay node, and $\boldsymbol{n}_i \in \mathbb{C}^{N_i \times 1}$ the independent additive noise according to a circular symmetric complex Gaussian distribution $\mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{N_i})$. We assume an average transmit power constraint $\mathrm{tr}(\boldsymbol{Q}) \leq P$ with $\boldsymbol{Q} = \mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^H\}$ at the relay node.

For the capacity region of the MIMO Gaussian bidirectional broadcast channel we need the following region. For a given covariance matrix $\boldsymbol{Q}$ let $\mathcal{R}(\boldsymbol{Q})$ be the set of all rate tuples $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy

$$\begin{aligned} R_1 &\leq \log\det\left(\boldsymbol{I}_{N_1} + \tfrac{1}{\sigma^2}\boldsymbol{H}_1 \boldsymbol{Q} \boldsymbol{H}_1^H\right) \\ R_2 &\leq \log\det\left(\boldsymbol{I}_{N_2} + \tfrac{1}{\sigma^2}\boldsymbol{H}_2 \boldsymbol{Q} \boldsymbol{H}_2^H\right). \end{aligned}$$

Theorem 2.7 ([WOB+08a])**.** *The capacity region $\mathcal{R}(\boldsymbol{H}_1, \boldsymbol{H}_2|P)$ of the MIMO Gaussian BBC under average power constraint P is given by*

$$\mathcal{R}(\boldsymbol{H}_1, \boldsymbol{H}_2|P) = \bigcup_{tr(\boldsymbol{Q}) \leq P,\, \boldsymbol{Q} \succeq \boldsymbol{0}} \mathcal{R}(\boldsymbol{Q}).$$

Similar to the discrete case, cf. Theorem 2.5, it is optimal to transmit only one data stream that carries all the information based on the network coding idea. Theorem 2.7 shows that it is further optimal to let the input $\boldsymbol{x}$ be Gaussian distributed, but it does not specify the optimal covariance matrix $\boldsymbol{Q}$. This is analyzed in detail in [OWB09a, OJWB09] and, interestingly, it shows that the philosophy of network coding carries over to the signal processing part as well. For example, in [OWB09a] it is shown that for the MISO Gaussian BBC there exists always an optimal single-beam transmit strategy which reflects the single stream processing based on the network coding idea. This manifests the paradigm shift to consider information flows not as "*fluids*" [ACLY00].

2.3 Bidirectional Achievable Rate Region

It is clear that for a successful bidirectional exchange of both messages, i.e., the transmission of message m_1 from node 1 to node 2 with rate R_1 and message m_2 from node 2 to node 1 with rate R_2, the rate pair (R_1, R_2) has to be achievable in the MAC phase as well as in the BBC phase. This underlines the intensive studies of both phases each for its own, since they constitute the basis for the following analysis of the bidirectional achievable rate region.

In the following we exemplarily discuss the analysis for discrete memoryless channels, but of course the same argumentation also holds for MIMO Gaussian channels.

Fixed Time Division

For the time division we define the parameter $\alpha \in [0, 1]$ so that we have in the MAC phase the scaled rate region $\alpha\mathcal{R}(V)$ and in the BBC phase $(1-\alpha)\mathcal{R}(W)$. Since the rate pair has to be achievable in both phases, the bidirectional achievable rate region for given fixed time division parameter α is given by the intersection of the scaled versions of the MAC and BBC rate regions, i.e.,

$$\mathcal{R}_{\text{BR}}(\alpha) \coloneqq \alpha\mathcal{R}(V) \cap (1-\alpha)\mathcal{R}(W).$$

A detailed discussion about the bidirectional achievable rate region with fixed time division can be found in [OB06].

Optimal Time Division

A fixed time division between the two phases can be suboptimal especially if one scaled rate region of a phase is much smaller than the other scaled rate region. Consequently, the next step is to optimize the time division between the two phases. Obviously, the bidirectional achievable rate region is given by the union over all possible time division parameters α as follows

$$\mathcal{R}_{\text{BRopt}} \coloneqq \bigcup_{\alpha\in[0,1]} \mathcal{R}_{\text{BR}}(\alpha) = \bigcup_{\alpha\in[0,1]} \big(\alpha\mathcal{R}(V) \cap (1-\alpha)\mathcal{R}(W)\big).$$

The rate region for optimal time division contains the achievable rate regions for all time division parameters α so that each rate pair on the boundary can be achieved by a certain time division parameter α. The bidirectional achievable rate region with optimal time division using superposition encoding in the BBC phase with an average power constraint can be found in [OB08c] and with an average energy constraint in [OB08a]. The optimal division for multi-antenna bidirectional relaying can be found in [ZKWB08, OWB08a].

3 Bidirectional Relaying Under Channel Uncertainty

To date, bidirectional relaying has been analyzed under the assumption of perfect channel state information (CSI) at all nodes. However, due to the nature of the wireless channel, uncertainty in the channel state information is a ubiquitous phenomenon. It is clear that this should be taken into account for the design of practical systems to make them robust against such impairments. A bidirectional relay network with channel uncertainty at all nodes is visualized in Figure 3.1.

The traditional and most popular approach to mitigate the channel uncertainty is based on channel estimation. But this is only one specific approach, and it is natural to address the problem of reliable communication under channel uncertainty from a more general point of view in order to gain insights to the best possible approach for this setting. A well accepted model for channel uncertainty is to assume that the exact channel realization is not known to the nodes; rather, it is assumed that all nodes only know that the realization belongs to a pre-specified set of channels. If this channel remains fixed during the whole transmission of a codeword, this corresponds to the concept of the *compound channel* which was independently introduced and analyzed by Blackwell, Breiman, and Thomasian [BBT59] and Wolfowitz [Wol60, Wol78].

It seems worthwhile to study the compound channel as a model of channel uncertainty from an optimal coding perspective to gain an understanding of how robust coding strategies should be designed. This is especially important to know for wireless systems, which make strict demands on the quality of service. As an example, one can think of wireless control applications where certain rates have to be guaranteed regardless of the current channel realization. For such applications most performance measures, for example the ergodic capacity, are not appropriate, since they characterize rates which are only achievable on average. Rather, a performance measure is needed which characterizes the guaranteed rates. The concept of the compound channel is an attractive model which allows treating such problems from a general point of view and obtaining bounds on maximal achievable rates. Moreover, this concept allows us to assess the resulting gain based on an improvement in the channel state information. This makes it possible to consider the trade-off between the contribution of CSI and the effort which would be needed to improve it, e.g., by using better or longer training sequences.

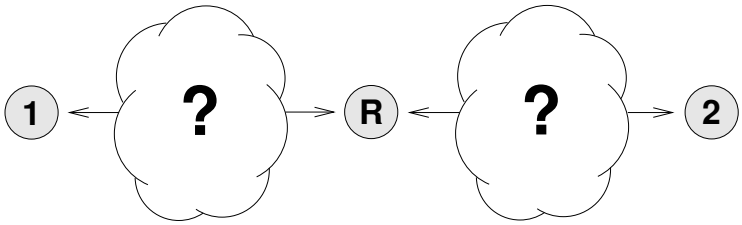


Figure 3.1: Bidirectional relaying under channel uncertainty.

The discrete memoryless compound channel was first analyzed in [BBT59, Wol60, Wol78] and later extended to Gaussian channels in [RV68]. Recently, the concept of the compound channel has gained a lot of attention. For example the compound channel where the transmitter has certain non-causal side information is studied in [MDT06], while [SP09] discusses the case with feedback. The single-user MIMO Gaussian compound channel is analyzed under several aspects in [PCL03, WES05, WES07, LC08]. There are further extensions to multi-user settings. The compound MAC was first introduced in [Ahl74] and further analyzed in [Han98, Han03] using the information-spectrum approach. The case where the encoders and decoders can (partially) cooperate is considered in [MYK05, WBBJ11] and [SGP$^+$09], respectively. The compound broadcast channel is discussed in [WSK07, WLS$^+$09], while [SEP08, RPV09] addresses the compound interference channel. There is further some work related to physical layer secrecy, cf. also Section 5.3. The compound wiretap is analyzed in [LPV08, PDT09, LKPS09, EU10b, BBS11a]. Interference alignment for the compound wiretap channel is discussed in [XU10, Khi11].

In this chapter we analyze bidirectional relaying for compound channels. Therefore, we briefly summarize in Section 3.1 the well understood compound multiple access channel for the first phase of the decode-and-forward protocol. To capture the second phase, we introduce the *compound bidirectional broadcast channel* in Section 3.2 and present a universal coding strategy that overcomes the absent channel state information at all nodes. This establishes the capacity region for the compound BBC with no CSI. Then, in Section 3.4 we discuss the cases where either the receivers or the transmitter have perfect CSI. Interestingly, we show that CSIR does not influence the capacity region, while CSIT can advantageously be exploited to enlarge the capacity region. Section 3.5 presents a numerical example and a game-theoretic interpretation using the game against nature framework. Finally, a concluding discussion is given in Section 3.6.

The analysis of bidirectional relaying for compound channels is not only relevant in itself, since it yields results for bidirectional relaying in common communication scenarios such as flat fading channels, but also since these results constitute the basis for further analysis of more complex uncertainty models such as *arbitrarily varying channels*. Here the channel may vary during the transmission from symbol to symbol in an unknown and arbitrary

manner, as occurs for example, in fast fading channels. Bidirectional relaying is extended to this model of uncertainty in Chapter 4. Research in this area is also the basis for the analysis of multi-user settings in uncoordinated wireless networks where the receiving nodes are confronted with unknown varying interference.

3.1 Compound Multiple Access Channel

In this section we briefly restate the capacity region of the *compound multiple access channel* which models the first phase of the decode-and-forward bidirectional relaying protocol under channel uncertainty. The compound MAC was first analyzed by Ahlswede [Ahl74].

Let $\mathcal{X}_i$, $i = 1, 2$, and $\mathcal{Y}$ be finite input and output sets. Then, for a fixed channel realization $s \in \mathcal{S}$ and for input and output sequences $x_i^n \in \mathcal{X}_i^n$, $i = 1, 2$, and $y^n \in \mathcal{Y}^n$ of length n, the discrete memoryless multiple access channel is given by $V_s^{\otimes n}(y^n|x_1^n, x_2^n) \coloneqq \prod_{k=1}^n V_s(y_k|x_{1,k}, x_{2,k})$.

Definition 3.1. *The discrete memoryless* compound multiple access channel $\mathfrak{V}$ *is defined by a family*

$$\mathfrak{V} \coloneqq \left\{V_s^{\otimes n} : \mathcal{X}_1^n \times \mathcal{X}_2^n \to \mathcal{P}(\mathcal{Y}^n)\right\}_{n\in\mathbb{N}, s\in\mathcal{S}}.$$

Theorem 3.2 ([Ahl74])**.** *The capacity region $\mathcal{R}(\mathfrak{V})$ of the compound multiple access channel $\mathfrak{V}$ is the set of all rate pairs $(R_2, R_1) \in \mathbb{R}_+^2$ that satisfy*[1]

$$\begin{aligned} R_2 &\leq \inf_{s\in\mathcal{S}} I(\mathrm{X}_1; \mathrm{Y}_s|\mathrm{X}_2, \mathrm{U}) \\ R_1 &\leq \inf_{s\in\mathcal{S}} I(\mathrm{X}_2; \mathrm{Y}_s|\mathrm{X}_1, \mathrm{U}) \\ R_2 + R_1 &\leq \inf_{s\in\mathcal{S}} I(\mathrm{X}_1, \mathrm{X}_2; \mathrm{Y}_s|\mathrm{U}) \end{aligned}$$

for random variables[2] $\mathrm{U} - (\mathrm{X}_1, \mathrm{X}_2) - \mathrm{Y}_s$ *with joint probability distributions* $\{p_{\mathrm{U}}(u)p_{\mathrm{X}_1|\mathrm{U}}(x_1|u)p_{\mathrm{X}_2|\mathrm{U}}(x_2|u)V_s(y|x_1, x_2)\}_{s\in\mathcal{S}}$*. Thereby,* U *is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of* U *can be bounded by* $|\mathcal{U}| \leq 2$*.*

[1] Recall that the compound MAC is considered within the two-phase decode-and-forward protocol. Therefore, the individual rates look "swapped", cf. Chapter 2 and especially Figure 2.1.

[2] With a slight abuse of notation we write Y_s to indicate that the output depends on the specific channel realization $s \in \mathcal{S}$.

3.2 Compound Bidirectional Broadcast Channel

Here we turn to the bidirectional broadcast phase of the decode-and-forward bidirectional relaying protocol. Again we assume that the transmission takes place over a channel which is unknown to the transmitter and the receivers. It is only known to the nodes that the realization is from a pre-specified set of channels $\mathcal{S}$ and that it remains fixed during the whole transmission of a codeword. No restrictions are imposed on the set $\mathcal{S}$. Since $\mathcal{S}$ can be arbitrary, it includes, in particular, the case of infinitely many channels. Accordingly, this constitutes an appropriate model for common communication scenarios such as flat fading channels.

Let $\mathcal{X}$ and $\mathcal{Y}_i$, $i = 1, 2$, be finite input and output sets. Then for a fixed $s \in \mathcal{S}$ and for input and output sequences $x^n \in \mathcal{X}^n$ and $y_i^n \in \mathcal{Y}_i^n$, $i = 1, 2$, of length n, the discrete memoryless broadcast channel is given by $W_s^{\otimes n}(y_1^n, y_2^n|x^n) \coloneqq \prod_{k=1}^n W_s(y_{1,k}, y_{2,k}|x_k)$.

Definition 3.3. *The discrete memoryless* compound broadcast channel $\mathfrak{W}$ *is defined by a family*

$$\mathfrak{W} \coloneqq \left\{W_s^{\otimes n} : \mathcal{X}^n \to \mathcal{P}(\mathcal{Y}_1^n \times \mathcal{Y}_2^n)\right\}_{n\in\mathbb{N}, s\in\mathcal{S}}.$$

Since we do not allow any cooperation between the receiving nodes, it is sufficient to consider the marginal transition probabilities $W_{i,s}^{\otimes n}(y_i^n|x^n) = \prod_{k=1}^n W_{i,s}(y_{i,k}|x_k)$, $i = 1, 2$, only. Thereby, $W_{i,s}$ denotes the channel between the relay and node i for channel realization $s \in \mathcal{S}$.

We consider the standard model with a block code of arbitrary but fixed length n. Let $\mathcal{M}_i \coloneqq \{1, 2, ..., M_i^{(n)}\}$ be the message set of node i, $i = 1, 2$, which is also known at the relay node. Further, we use the abbreviation $\mathcal{M} \coloneqq \mathcal{M}_1 \times \mathcal{M}_2$.

Definition 3.4. *A deterministic*[3] $(n, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{\text{det}}(\mathfrak{W})$ *for the compound BBC* $\mathfrak{W}$ *consists of universal codewords*

$$x_m^n \in \mathcal{X}^n,$$

one for each message $m = (m_1, m_2) \in \mathcal{M}$*, and mutually disjoint decoding sets at nodes 1 and 2*

$$\mathcal{D}_{m_2|m_1}^{(1)} \subseteq \mathcal{Y}_1^n \qquad \textit{and} \qquad \mathcal{D}_{m_1|m_2}^{(2)} \subseteq \mathcal{Y}_2^n$$

for all $m_1 \in \mathcal{M}_1$ *and* $m_2 \in \mathcal{M}_2$.

[3] For the compound BBC we will only consider deterministic codes so that we will often suppress the word *deterministic* in the following. This is in contrast to the arbitrarily varying bidirectional broadcast channel, where we have to carefully distinguish between random and deterministic codes, cf. Chapter 4.

We require disjoint decoding sets in the sense that for given fixed $m_1 \in \mathcal{M}_1$ at node 1 the decoding sets have to be disjoint, i.e., $\mathcal{D}^{(1)}_{m_2|m_1} \cap \mathcal{D}^{(1)}_{\hat{m}_2|m_1} = \emptyset$ for $\hat{m}_2 \neq m_2$. Clearly, for different $m_1, \hat{m}_1 \in \mathcal{M}_1$, $m_1 \neq \hat{m}_1$, at node 1, the decoding sets need not be disjoint. Accordingly for node 2, for given fixed $m_2 \in \mathcal{M}_2$ at node 2 the decoding sets have to be disjoint, i.e., $\mathcal{D}^{(2)}_{m_1|m_2} \cap \mathcal{D}^{(2)}_{\hat{m}_1|m_2} = \emptyset$ for $\hat{m}_1 \neq m_1$.

Note that neither the codewords at the transmitter nor the decoding sets at the receivers depend on the actual channel realization. This reflects the fact that all nodes do not know the exact channel realization and therefore have to chose their codewords and decoding sets universally such that they work for the whole set of channels.

When channel realization $s \in \mathcal{S}$ governs the transmission and the relay node has sent the codeword $x^n_m \in \mathcal{X}^n$ for message $m = (m_1, m_2)$ according to codebook $\mathcal{C}_{\text{det}}(\mathfrak{W})$, and nodes 1 and 2 have received $y^n_1 \in \mathcal{Y}^n_1$ and $y^n_2 \in \mathcal{Y}^n_2$, the decoder at node 1 is in error if y^n_1 is not in $\mathcal{D}^{(1)}_{m_2|m_1}$. Accordingly, the decoder at node 2 is in error if y^n_2 is not in $\mathcal{D}^{(2)}_{m_1|m_2}$. This allows us to define the probabilities of error for given message $m = (m_1, m_2)$ and given channel realization $s \in \mathcal{S}$ as

$$e_1(m, s|\mathcal{C}_{\text{det}}(\mathfrak{W})) \coloneqq W^{\otimes n}_{1,s}\big((\mathcal{D}^{(1)}_{m_2|m_1})^c|x^n_m\big) = \sum_{y^n_1 \notin \mathcal{D}^{(1)}_{m_2|m_1}} W^{\otimes n}_{1,s}(y^n_1|x^n_m)$$

$$e_2(m, s|\mathcal{C}_{\text{det}}(\mathfrak{W})) \coloneqq W^{\otimes n}_{2,s}\big((\mathcal{D}^{(2)}_{m_1|m_2})^c|x^n_m\big) = \sum_{y^n_2 \notin \mathcal{D}^{(2)}_{m_1|m_2}} W^{\otimes n}_{2,s}(y^n_2|x^n_m).$$

Thus, the average probability of error for channel realization $s \in \mathcal{S}$ at node i, $i = 1, 2$, is[4]

$$\bar{e}_i(s) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_i(m, s).$$

Finally, we define the supremum of all average probabilities of error as $\mu^{(n)}_i \coloneqq \sup_{s \in \mathcal{S}} \bar{e}_i(s)$, $i = 1, 2$.

Definition 3.5. *A rate pair $(R_1, R_2) \in \mathbb{R}^2_+$ is said to be* achievable *for the compound BBC $\mathfrak{W}$ if for any $\delta > 0$ there is an $n(\delta) \in \mathbb{N}$ and a sequence $\{\mathcal{C}^{(n)}_{det}(\mathfrak{W})\}_{n \in \mathbb{N}}$ of $(n, M^{(n)}_1, M^{(n)}_2)$-codes such that for all $n \geq n(\delta)$ we have*

$$\frac{1}{n} \log M^{(n)}_2 \geq R_1 - \delta \qquad \textit{and} \qquad \frac{1}{n} \log M^{(n)}_1 \geq R_2 - \delta$$

[4] We will suppress the dependency on the used codebook in the definition of the probability of error if it is clear from the context which codebook is used. Since for the compound BBC we always consider deterministic codes, we simply write $e_i(m, s)$ instead of $e_i(m, s|\mathcal{C}_{\text{det}}(\mathfrak{W}))$ for notational convenience, $i = 1, 2$.

while

$$\mu_i^{(n)} = \sup_{s\in\mathcal{S}} \bar{e}_i(s) \to 0,$$

$i = 1, 2$*, as* $n \to \infty$*. The set of all achievable rate pairs is the (deterministic code)* capacity region *of the compound BBC* $\mathfrak{W}$ *and is denoted by* $\mathcal{R}_{det}(\mathfrak{W})$.

Remark 3.6. *The definitions require that we have to find codes such that* $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ *as* $n \to \infty$ *for all channels in the set* $\mathcal{S}$ *simultaneously. This means the codes are universal with respect to the channel realization.*

3.3 Universal Strategy and Capacity Region

Now we are in a position to present the universal strategy for the BBC phase of the decode-and-forward protocol which overcomes the channel uncertainty at the transmitter and the receivers. But first, we prove an outer bound of the capacity region which gives us an intuition of what is at best possible for the compound BBC.

Lemma 3.7. *Any given sequence* $\{\mathcal{C}_{det}^{(n)}(\mathfrak{W})\}_{n\in\mathbb{N}}$ *of* $(n, M_1^{(n)}, M_2^{(n)})$*-codes for the compound BBC* $\mathfrak{W}$ *with* $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ *must satisfy*

$$\frac{1}{n}\log M_2^{(n)} \leq \inf_{s\in\mathcal{S}} I(\mathrm{X}; \mathrm{Y}_{1,s}|\mathrm{U}) + o(n^0) \tag{3.1a}$$

$$\frac{1}{n}\log M_1^{(n)} \leq \inf_{s\in\mathcal{S}} I(\mathrm{X}; \mathrm{Y}_{2,s}|\mathrm{U}) + o(n^0) \tag{3.1b}$$

for random variables $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_{1,s}, \mathrm{Y}_{2,s})$ *with joint probability distributions* $\{p_{\mathrm{U}}(u)p_{\mathrm{X}|\mathrm{U}}(x|u)W_s(y_1, y_2|x)\}_{s\in\mathcal{S}}$.

Proof. From [OSBB08] we know that for a specific channel realization $s \in \mathcal{S}$ the rates are bounded from above by $\frac{1}{n}H(M_2) \leq I(\mathrm{X}; \mathrm{Y}_{1,s}|\mathrm{U}) + \epsilon_1^{(n)}$ and $\frac{1}{n}H(M_1) \leq I(\mathrm{X}; \mathrm{Y}_{2,s}|\mathrm{U}) + \epsilon_2^{(n)}$ where $\epsilon_1^{(n)}, \epsilon_2^{(n)} \to 0$ as $n \to \infty$. Since the rates have to be achievable for all $s \in \mathcal{S}$ simultaneously, it follows immediately that for the compound BBC with channel uncertainty at the transmitter and the receivers the rates are bounded from above by the infimum of the mutual information terms as stated in (3.1). This proves the lemma. □

Next, we start with the case where the compound channel has finitely many elements and derive the corresponding capacity region. With this result we then are able to solve the general case of an arbitrary, not necessarily finite, set of channels, which, of course, is the more relevant case since it covers communication scenarios as for example flat fading channels.

3.3.1 Finite Compound Channel

In this subsection we restrict the set of channels $\mathcal{S}$ to be finite and present a universal strategy which actually achieves the rates under this condition as stated in Lemma 3.7, cf. (3.1). For this we need the following lemma which shows the existence of such a strategy whose probability of error is arbitrarily small.

Lemma 3.8. *Let the finite compound BBC $\mathfrak{W}$ be given by a finite index set $\mathcal{S} = \{1, ..., S\}$. For any block length $n \in \mathbb{N}$, input distribution $p \in \mathcal{P}(\mathcal{X})$, and*[5]

$$R_1 \leq \min_{i=1,...,S} I(p, W_{1,i}) - \frac{\tau}{2} \qquad \text{and} \qquad R_2 \leq \min_{i=1,...,S} I(p, W_{2,i}) - \frac{\tau}{2},$$

$\tau > 0$, there is a $(n, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{det}(\mathfrak{W})$ where the probability of error $\mu_i^{(n)}$ at node i, $i = 1, 2$, averaged over all codebooks is bounded from above by

$$\mathbb{E}_{\mathrm{X}^n}[\mu_i^{(n)}] \leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_i|} 2^{-nc\epsilon^2} + S^2 \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1 - (n+1)^{|\mathcal{X}|} 2^{-nc\epsilon^2}} 2^{-n\frac{\tau}{4}} \tag{3.2}$$

with $c = \frac{1}{2 \ln 2}$.

Proof. For given distribution $p^{\otimes n}(x^n)$ and $\epsilon > 0$ let $\mathcal{T}_{p,\epsilon}^{(n)}$ denote the set of (strongly) typical sequences on $\mathcal{X}^n$, cf. also Appendix B.2.1. We restrict the possible inputs to this subset and define the new input distribution

$$p'(x^n) := \begin{cases} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}_{p,\epsilon}^{(n)})} & \text{if } x^n \in \mathcal{T}_{p,\epsilon}^{(n)} \\ 0 & \text{else.} \end{cases} \tag{3.3}$$

Let $\tau > 0$ and set the rates $R_1 := \min_{i=1,...,S} I(p, W_{1,i}) - \frac{\tau}{2}$ and $R_2 := \min_{i=1,...,S} I(p, W_{2,i}) - \frac{\tau}{2}$. Then we generate $|\mathcal{M}| = |\mathcal{M}_1||\mathcal{M}_2|$ independent codewords X_m^n, one for each message $m = (m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, of length n with $|\mathcal{M}_1| := \lfloor 2^{nR_2} \rfloor$ and $|\mathcal{M}_2| := \lfloor 2^{nR_1} \rfloor$ according to p'. This implies that all generated random codewords $\mathrm{X}_m^n \in \mathcal{T}_{p,\epsilon}^{(n)}$ almost surely.

Next, we specify the decoding sets of nodes 1 and 2 in detail. They are given by

$$\mathcal{D}_{m_2|m_1}^{(1)}(\mathrm{X}^n) := \Big(\bigcup_{i=1}^{S} \mathcal{T}_{W_{1,i},\epsilon}^{(n)}(\mathrm{X}_m^n) \Big) \cap \Big(\bigcup_{\substack{\hat{m}_2 \in \mathcal{M}_2 \\ \hat{m}_2 \neq m_2}} \bigcup_{i=1}^{S} \mathcal{T}_{W_{1,i},\epsilon}^{(n)}(\mathrm{X}_{m_1 \hat{m}_2}^n) \Big)^c \tag{3.4a}$$

$$\mathcal{D}_{m_1|m_2}^{(2)}(\mathrm{X}^n) := \Big(\bigcup_{i=1}^{S} \mathcal{T}_{W_{2,i},\epsilon}^{(n)}(\mathrm{X}_m^n) \Big) \cap \Big(\bigcup_{\substack{\hat{m}_1 \in \mathcal{M}_1 \\ \hat{m}_1 \neq m_1}} \bigcup_{i=1}^{S} \mathcal{T}_{W_{2,i},\epsilon}^{(n)}(\mathrm{X}_{\hat{m}_1 m_2}^n) \Big)^c \tag{3.4b}$$

[5] Without loss of generality we can assume that $R_i > 0$, $i = 1, 2$, since the rate $R_i = 0$ is always achievable.

where $\mathrm{X}^n \coloneqq \{\mathrm{X}^n_{m_1 m_2}\}_{m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2}$ so that the decoding sets depend on all randomly generated codewords. The definition of the decoding sets is motivated as follows. The first part in (3.4a) ensures that for given $m = (m_1, m_2)$ the decoding sets for node 1 are mostly defined by all output sequences that are $W_{1,s}$-typical under the input $\mathrm{X}^n_{m_1 m_2}$ for all channel realizations $s \in \mathcal{S}$. The second part excludes all such output sequences that are further $W_{1,s}$-typical to another input $X^n_{m_1 \hat{m}_2}$ with $\hat{m}_2 \neq m_2$ so that the decoding sets are unambiguously defined and therewith mutually disjoint. Clearly, the decoding sets in (3.4b) are motivated accordingly.

When x^n_m with $m = (m_1, m_2)$ has been sent, and y^n_1 and y^n_2 have been received at nodes 1 and 2, the decoder at node 1 is in error if either y^n_1 is not in $\bigcup_{i=1}^{S} \mathcal{T}^{(n)}_{W_{1,i},\epsilon}(x^n_m)$ or if y^n_1 is in $\bigcup_{i=1}^{S} \mathcal{T}^{(n)}_{W_{1,i},\epsilon}(x^n_{m_1 \hat{m}_2})$ for some $\hat{m}_2 \neq m_2$, cf. (3.4a). The error events at node 2 are defined in an analogous way.

In the following we present the analysis of the probability of error for node 1, the analysis for node 2 follows accordingly using the same arguments. For a given channel realization $s \in \mathcal{S}$ the union bound[6] yields for average probability of error $\bar{e}_1(s) \leq E_1(s) + E_2(s)$ with

$$E_1(s) \coloneqq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W_{1,s}^{\otimes n}\Big(\big(\bigcup_{i=1}^{S} \mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_m) \big)^c \big| \mathrm{X}^n_m \Big) \tag{3.5a}$$

$$E_2(s) \coloneqq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W_{1,s}^{\otimes n}\Big(\bigcup_{\substack{\hat{m}_2 \in \mathcal{M}_2 \\ \hat{m}_2 \neq m_2}} \bigcup_{i=1}^{S} \mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1 \hat{m}_2}) \big| \mathrm{X}^n_{m_1 m_2} \Big). \tag{3.5b}$$

Next, we average over all codebooks and show that $\mathbb{E}_{\mathrm{X}^n}[\bar{e}_1(s)] \leq \mathbb{E}_{\mathrm{X}^n}[E_1(s) + E_2(s)]$ can be bounded uniformly in s from above by a term which decreases exponentially fast for increasing block length n. For fixed $s \in \mathcal{S}$ we get for the first error event (3.5a)

$$\begin{aligned}
\mathbb{E}_{\mathrm{X}^n}[E_1(s)] &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathbb{E}_{\mathrm{X}^n}\Big[W_{1,s}^{\otimes n}\Big(\big(\bigcup_{i=1}^{S} \mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_m) \big)^c \big| \mathrm{X}^n_m \Big) \Big] \\
&= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathbb{E}_{\mathrm{X}^n}\Big[W_{1,s}^{\otimes n}\Big(\bigcap_{i=1}^{S} \big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_m) \big)^c \big| \mathrm{X}^n_m \Big) \Big] \\
&\leq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathbb{E}_{\mathrm{X}^n}\Big[W_{1,s}^{\otimes n}\Big(\big(\mathcal{T}^{(n)}_{W_{1,s},\epsilon}(\mathrm{X}^n_m) \big)^c \big| \mathrm{X}^n_m \Big) \Big] \\
&\leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|} 2^{-nc\epsilon^2}
\end{aligned} \tag{3.6}$$

[6]The probability of the union of two events E_1 and E_2 is bounded from above by $\mathbb{P}\{E_1 \cup E_2\} \leq \mathbb{P}\{E_1\} + \mathbb{P}\{E_2\}$, which is known as *union bound*, cf. for example [Pro00, Sec. 5.2].

with $c = \frac{1}{2\ln 2}$ where the first inequality follows from the monotonicity of the probability and the last one from Lemma B.11, cf. (B.6) in Appendix B.2.1. For the second error event (3.5b) we have

$$\begin{aligned}
\mathbb{E}_{\mathrm{X}^n}[E_2(s)] &= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathbb{E}_{\mathrm{X}^n}\Big[W_{1,s}^{\otimes n}\Big(\bigcup_{\substack{\hat{m}_2\in\mathcal{M}_2\\ \hat{m}_2\neq m_2}}\bigcup_{i=1}^{S}\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})|\mathrm{X}^n_{m_1m_2}\Big)\Big] \\
&\leq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}\sum_{\substack{\hat{m}_2\in\mathcal{M}_2\\ \hat{m}_2\neq m_2}}\sum_{i=1}^{S}\mathbb{E}_{\mathrm{X}^n}\Big[W_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})|\mathrm{X}^n_{m_1m_2}\Big)\Big] \\
&= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}\sum_{\substack{\hat{m}_2\in\mathcal{M}_2\\ \hat{m}_2\neq m_2}}\sum_{i=1}^{S}\mathbb{E}_{\mathrm{X}^n_{m_1\hat{m}_2}}\mathbb{E}_{\mathrm{X}^n_{m_1m_2}}\Big[W_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})|\mathrm{X}^n_{m_1m_2}\Big)\Big] \qquad (3.7)
\end{aligned}$$

where the last equality follows from the fact that X^n is an iid sequence. Next, we compute the expectations. For the inner expectation we get

$$\begin{aligned}
&\mathbb{E}_{\mathrm{X}^n_{m_1m_2}}\Big[W_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})|\mathrm{X}^n_{m_1m_2}\Big)\Big] \\
&\quad= \sum_{x^n\in\mathcal{X}^n} p'(x^n)W_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})|x^n\Big) \\
&\quad= \sum_{x^n\in\mathcal{T}^{(n)}_{p,\epsilon}} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}^{(n)}_{p,\epsilon})}W_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})|x^n\Big) \\
&\quad\leq \sum_{x^n\in\mathcal{X}^n} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}^{(n)}_{p,\epsilon})}W_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})|x^n\Big) \\
&\quad= \frac{1}{p^{\otimes n}(\mathcal{T}^{(n)}_{p,\epsilon})}q_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})\Big) \\
&\quad\leq \frac{1}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}}q_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})\Big) \qquad (3.8)
\end{aligned}$$

where $c = \frac{1}{2\ln 2}$ and $q_{1,s} \in \mathcal{P}(\mathcal{Y}_1)$ denotes the output distribution generated by p and $W_{1,s}$. The second equality follows from (3.3) and the last inequality follows from Lemma B.10, cf. (B.5). Since $\mathrm{X}^n_{m_1\hat{m}_2} \in \mathcal{T}^{(n)}_{p,\epsilon}$ almost surely, we can apply Lemma B.12 more precisely (B.10b) and obtain for the second expectation

$$\mathbb{E}_{\mathrm{X}^n_{m_1\hat{m}_2}}\Big[q_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{m_1\hat{m}_2})\Big)\Big] \leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-n(I(p,W_{1,i})-\varphi(\epsilon)-\psi(\epsilon))} \qquad (3.9)$$

for $\epsilon \in (0, \frac{1}{4|\mathcal{X}||\mathcal{Y}_1|})$. From (3.7)–(3.9) we get

$$\mathbb{E}_{\mathrm{X}^n}[E_2(s)] \leq (|\mathcal{M}_2|-1)\sum_{i=1}^{S}\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}}2^{-n(I(p,W_{1,i})-\varphi(\epsilon)-\psi(\epsilon))}.$$

Since $|\mathcal{M}_2| = \lfloor 2^{nR_1} \rfloor$ and $R_1 = \min_{i=1,\dots,S} I(p, W_{1,i}) - \frac{\tau}{2}$ we have $\frac{\tau}{2} \leq I(p, W_{1,i}) - R_1$ for all $i = 1, ..., S$ so that we obtain

$$\mathbb{E}_{\mathrm{X}^n}[E_2(s)] \leq S \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}} 2^{-n(\frac{\tau}{2}-\varphi(\epsilon)-\psi(\epsilon))}.$$

Next, we set $\epsilon \in (0, \frac{1}{4|\mathcal{X}||\mathcal{Y}_1|})$ small enough to ensure that $\frac{\tau}{4} \leq \frac{\tau}{2} - \varphi(\epsilon) - \psi(\epsilon)$ so that with (3.6) we get

$$\begin{aligned}\mathbb{E}_{\mathrm{X}^n}[\bar{e}_1(s)] &\leq \mathbb{E}_{\mathrm{X}^n}[E_1(s) + E_2(s)] \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-nc\epsilon^2} + S\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}} 2^{-n\frac{\tau}{4}}\end{aligned}$$

with $c = \frac{1}{2\ln 2}$. This gives us a bound on the average probability of error $\bar{e}_1(s)$ for one specific channel realization $s \in \mathcal{S}$. To obtain an upper bound for $\mu_1^{(n)}$, i.e., the maximum of all channel realizations, we need the average BBC

$$\overline{W}_1 := \frac{1}{S}\sum_{s\in\mathcal{S}} W_{1,s}.$$

From the definition it is clear that $\overline{W}_1 \geq \frac{1}{S} W_{1,s}$ holds for all $s \in \mathcal{S}$ which implies that

$$\mu_1^{(n)} = \max_{s\in\mathcal{S}} \bar{e}_1(s) \leq S\bar{e}_1(\overline{W}_1)$$

where $\bar{e}_1(\overline{W}_1)$ is the average probability of error with respect to the average BBC $\overline{W}_1$. Therefore, we have $\mathbb{E}_{\mathrm{X}^n}[\bar{e}_1(\overline{W}_1)] = \frac{1}{S}\sum_{s\in\mathcal{S}} \mathbb{E}_{\mathrm{X}^n}[\bar{e}_1(s)]$ due to the linearity of the expectation. Finally, we obtain for the average probability of error at node 1

$$\begin{aligned}\mathbb{E}_{\mathrm{X}^n}[\mu_1^{(n)}] &\leq S\mathbb{E}_{\mathrm{X}^n}[\bar{e}_1(\overline{W}_1)] \\ &\leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-nc\epsilon^2} + S^2\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}} 2^{-n\frac{\tau}{4}}\end{aligned}$$

with $c = \frac{1}{2\ln 2}$ as stated in (3.2). Similar reasoning leads for the probability of error at node 2 to $\mathbb{E}_{\mathrm{X}^n}[\mu_2^{(n)}] \leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_2|}2^{-nc\epsilon^2} + S^2\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_2|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}} 2^{-n\frac{\tau}{4}}$ which proves the lemma. □

The crucial point of this lemma is that the concept of typical sequences from Csiszár and Körner [CK81] allows us to establish bounds on the probability of error that decrease exponentially fast for increasing block length. This property will be important for the extension to the general case of an arbitrary set of channels $\mathcal{S}$. However, this lemma together with the outer bound in Lemma 3.7 immediately yields the capacity region of the finite compound BBC $\mathfrak{W}$, which is stated in the following corollary.

Corollary 3.9. *The capacity region $\mathcal{R}_{det}(\mathfrak{W})$ of the finite compound BBC $\mathfrak{W}$ given by the finite index set $\mathcal{S} = \{1, ..., S\}$ is the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy*

$$R_1 \leq \min_{i=1,..,S} I(\mathrm{X}; \mathrm{Y}_{1,i}|\mathrm{U}) \tag{3.10a}$$

$$R_2 \leq \min_{i=1,..,S} I(\mathrm{X}; \mathrm{Y}_{2,i}|\mathrm{U}) \tag{3.10b}$$

for random variables $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_{1,i}, \mathrm{Y}_{2,i})$ with joint probability distributions $\{p_{\mathrm{U}}(u)p_{\mathrm{X}|\mathrm{U}}(x|u)W_i(y_1, y_2|x)\}_{i=1,...,S}$. Thereby, U is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of U can be bounded by $|\mathcal{U}| \leq 2$.

Proof. The achievability follows immediately from Lemma 3.8 which states that all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying $R_1 \leq \min_{i=1,..,S} I(p, W_{1,i}) = \min_{i=1,..,S} I(\mathrm{X}; \mathrm{Y}_{1,i})$ and $R_2 \leq \min_{i=1,..,S} I(p, W_{2,i}) = \min_{i=1,..,S} I(\mathrm{X}; \mathrm{Y}_{2,i})$ are achievable with $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. The desired region is determined by establishing the convex hull by first introducing an auxiliary random variable U and applying standard arguments. Similarly to [OSBB08] it follows from Fenchel-Bunt's extension of Carathéodory's theorem [HUL01] that any rate pair is achievable by time-sharing between two rate pairs, i.e., $|\mathcal{U}| = 2$ is enough.

The weak converse follows from Lemma 3.7. Since the strategy from Lemma 3.8 already achieves these rate pairs, the capacity region of the finite compound BBC $\mathfrak{W}$ is determined by the corollary. □

Remark 3.10. *Similarly as in [KS07, Theorem 1] it is further possible to get rid of the time-sharing random variable U in (3.10) so that we end up with $R_1 \leq \min_{i=1,...,S} I(\mathrm{X}; \mathrm{Y}_{1,i})$ and $R_2 \leq \min_{i=1,...,S} I(\mathrm{X}; \mathrm{Y}_{2,i})$.*

3.3.2 Arbitrary Compound Channel

With the previous result we are able to establish the capacity region for the compound BBC $\mathfrak{W}$ with an arbitrary, possibly infinite, set of channels $\mathcal{S}$. Therefore we need the following two lemmas which are slightly adapted from [BBT59] to our scenario.

Lemma 3.11. *Let $\mathcal{X}$, $\mathcal{Y}_i$, $i = 1, 2$, be given. For every integer $N \geq 2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2$ there is a compound broadcast channel $\widetilde{\mathfrak{W}}$ where the index set $\mathcal{S}_N$ has at most $(N+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}$ elements such that for any W_s from $\mathcal{S}$ there is a channel $\widetilde{W}_s$ from $\mathcal{S}_N$ such that*

(a) $\left|W_s(y_1, y_2|x) - \widetilde{W}_s(y_1, y_2|x)\right| \leq \frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{N}$ *for all* x, y_1, y_2

(b) $W_s(y_1, y_2|x) \leq 2^{\frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{N}} \widetilde{W}_s(y_1, y_2|x)$ *for all* x, y_1, y_2

(c) *For any* $p \in \mathcal{P}(\mathcal{X})$ *it holds* $|I(p, W_{i,s}) - I(p, \widetilde{W}_{i,s})| \leq 2|\mathcal{Y}_1||\mathcal{Y}_2|(\frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{N})^{\frac{1}{2}}$, $i = 1, 2$.

Proof. The proof is almost identical to [BBT59, Lemma 4] and is therefore omitted. □

This lemma shows that we can approximate any given set of channels $\mathcal{S}$ by a finite set of channels $\mathcal{S}_N$ such that any channel $s \in \mathcal{S}$ is close in several senses to one of the new constructed channels in $\mathcal{S}_N$. Further, from the next lemma we see that if there is a "good" code for a channel, the same code can be used for all channels in a certain neighborhood of this channel.

Lemma 3.12. *Let* W_s *and* $\widetilde{W}_s$ *be two channels and* $A \in \mathbb{R}_+$ *a non-negative number such that* $W_s(y_1, y_2|x) \leq 2^A \widetilde{W}_s(y_1, y_2|x)$ *for all* x, y_1, y_2. *Then any* $(n, M_1^{(n)}, M_2^{(n)})$*-code for* $\widetilde{W}_s$ *is also a* $(n, M_1^{(n)}, M_2^{(n)})$*-code for* W_s *with* $\mu_i^{(n)} \leq 2^{nA}\tilde{\mu}_i^{(n)}$, $i = 1, 2$.

Proof. The proof is almost identical to [BBT59, Lemma 5] and is therefore omitted. □

With these two lemmas and the result for the finite compound BBC, we are able to prove our main result which is the capacity region of the compound BBC with an arbitrary set of channels.

Theorem 3.13. *The capacity region* $\mathcal{R}_{det}(\mathfrak{W})$ *of the compound BBC* $\mathfrak{W}$*, where the index set* $\mathcal{S}$ *can be arbitrary, is the set of all rate pairs* $(R_1, R_2) \in \mathbb{R}_+^2$ *that satisfy*

$$R_1 \leq \inf_{s \in \mathcal{S}} I(\mathrm{X}; \mathrm{Y}_{1,s}|\mathrm{U}) \tag{3.11a}$$

$$R_2 \leq \inf_{s \in \mathcal{S}} I(\mathrm{X}; \mathrm{Y}_{2,s}|\mathrm{U}) \tag{3.11b}$$

for random variables $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_{1,s}, \mathrm{Y}_{2,s})$ *with joint probability distributions* $\{p_{\mathrm{U}}(u)p_{\mathrm{X}|\mathrm{U}}(x|u)W_s(y_1, y_2|x)\}_{s \in \mathcal{S}}$. *Thereby,* U *is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of* U *can be bounded by* $|\mathcal{U}| \leq 2$.

Proof. We start with an approximation of the arbitrary set of channels. Therefore we choose $N \geq \max\{\frac{|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\epsilon^2}, \frac{8|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\tau}\}$ and large enough to ensure that $\frac{\tau}{2} > 2|\mathcal{Y}_1||\mathcal{Y}_2|(\frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{N})^{\frac{1}{2}}$. For each W_s from $\mathcal{S}$ we select a $\widetilde{W}_s$ according to Lemma 3.11 and denote the set of approximated channels by $\mathcal{S}_N$. Since $\mathcal{S}_N$ has at most $(N+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}$ elements, we know from Lemma 3.8 that if we choose $R_1 \leq \min_{i \in \mathcal{S}_N} I(p, \widetilde{W}_{1,i}) - \frac{\tau}{2}$ and $R_2 \leq \min_{i \in \mathcal{S}_N} I(p, \widetilde{W}_{2,i}) - \frac{\tau}{2}$, $\tau > 0$, then there exists a $(n, M_1^{(n)}, M_2^{(n)})$-code with $|\mathcal{M}_1| = \lfloor 2^{nR_2} \rfloor$ and $|\mathcal{M}_2| = \lfloor 2^{nR_1} \rfloor$ for $\mathcal{S}_N$ with probability of error for node i, $i = 1, 2$

$$\tilde{\mu}_i^{(n)} \leq S_N(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}2^{-nc\epsilon^2} + S_N^2 \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1 - (n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}} 2^{-n\frac{\tau}{4}}$$

with $S_N = (N+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}$ and $c = \frac{1}{2\ln 2}$. For each W_s from $\mathcal{S}$ there exists a $\widetilde{W}_s$ from $\mathcal{S}_N$ such that $W_s(y_1, y_2|x) \leq 2^{\frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{N}} \widetilde{W}_s(y_1, y_2|x)$ for all x, y_1, y_2 so that from Lemma 3.12 the code for $\mathcal{S}_N$ is also a code for $\mathcal{S}$ with

$$\begin{aligned}\mu_i^{(n)} \leq\ & S_N(n+1)^{|\mathcal{X}||\mathcal{Y}_k|} 2^{-n(c\epsilon^2 - \frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{N})} \\ & + S_N^2 \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}} 2^{-n(\frac{\tau}{4} - \frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{N})}.\end{aligned} \tag{3.12}$$

Since $N \geq \max\{\frac{|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\epsilon^2}, \frac{8|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\tau}\}$, we have $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. This means that the code constructed for the approximated channel is also a good code for the original channel. It remains to show that the code achieves rates arbitrarily close to the desired rates. From Lemma 3.11 we know that $|I(p, W_{i,s}) - I(p, \widetilde{W}_{i,s})| \leq 2|\mathcal{Y}_1||\mathcal{Y}_2|(\frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{N})^{\frac{1}{2}} \leq \frac{\tau}{2}$, $i = 1, 2$ so that

$$\inf_{s\in\mathcal{S}} I(p, W_{i,s}) - \tau \leq \min_{s\in\mathcal{S}_N} I(p, \widetilde{W}_{i,s}) - \frac{\tau}{2}$$

which proves the achievability of the rates given in (3.11). The optimality of the strategy follows similarly to Corollary 3.9 from Lemma 3.7 which finally proves the theorem. □

Remark 3.14. *Similarly as in [KS07, Theorem 1] it is further possible to get rid of the time-sharing random variable* U *in (3.11) so that we end up with* $R_1 \leq \inf_{s\in\mathcal{S}} I(\mathrm{X}; \mathrm{Y}_{1,s})$ *and* $R_2 \leq \inf_{s\in\mathcal{S}} I(\mathrm{X}; \mathrm{Y}_{2,s})$.

3.4 Partial Channel State Information at Transmitter or Receivers

In this section we discuss the scenarios where either the receivers or the transmitter have perfect channel state information while the other part still has no channel knowledge and only knows the set of channels.

3.4.1 CSI at the Receivers

We start with the scenario where the receivers have perfect CSI so that they can adapt their decoders to the specific channel realization. Consequently, we now have a whole family of decoders at nodes 1 and 2, one for each channel realization $s \in \mathcal{S}$. Note that we still only have one universal encoder at the relay due to the channel uncertainty at the transmitter. The definition of a code slightly changes as follows.

Definition 3.15. *A deterministic* $(n, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{\text{det}}^{\text{CSIR}}(\mathfrak{W})$ *for the compound BBC* $\mathfrak{W}$ *with CSIR consists of universal codewords*

$$x_m^n \in \mathcal{X}^n,$$

one for each message $m = (m_1, m_2) \in \mathcal{M}$*, and families of mutually disjoint decoding sets at nodes 1 and 2*

$$\mathcal{D}_{s,m_2|m_1}^{(1)} \subseteq \mathcal{Y}_1^n \qquad \textit{and} \qquad \mathcal{D}_{s,m_1|m_2}^{(2)} \subseteq \mathcal{Y}_2^n$$

for all $m_1 \in \mathcal{M}_1$ *and* $m_2 \in \mathcal{M}_2$*, one family for each channel realization* $s \in \mathcal{S}$*.*

The following theorem shows that CSIR does not lead to an improved capacity region as long as the transmitter merely knows the set of channels.

Theorem 3.16. *The capacity region* $\mathcal{R}_{det}^{CSIR}(\mathfrak{W})$ *of the compound BBC* $\mathfrak{W}$ *with CSIR is equal to the capacity region of the compound BBC* $\mathfrak{W}$ *with channel uncertainty at all nodes, i.e.,*

$$\mathcal{R}_{det}^{CSIR}(\mathfrak{W}) = \mathcal{R}_{det}(\mathfrak{W}).$$

Proof. If we apply the coding strategy for channel uncertainty at all nodes, cf. Theorem 3.13, it is clear that we can achieve the same rate pairs as if we have perfect CSI at the receivers. Consequently, it remains to show that this strategy is already optimal which means that, even with CSIR, no higher rates are achievable. The reasoning is as follows. In our communication scenario we have the following Markov chains $(\mathrm{M}_1, \mathrm{M}_2) - \mathrm{X} - \mathrm{Y}_{1,s} - \hat{\mathrm{M}}_2$ and $(\mathrm{M}_1, \mathrm{M}_2) - \mathrm{X} - \mathrm{Y}_{2,s} - \hat{\mathrm{M}}_1$ for node 1 and 2, respectively, where $\hat{\mathrm{M}}_i$, $i = 1, 2$ denotes the decoded message. From the data processing inequality [CK81, Lemma 3.11] follows immediately that $I(\mathrm{M}_1, \mathrm{M}_2; \hat{\mathrm{M}}_1) \leq I(\mathrm{X}; \mathrm{Y}_{2,s})$ and $I(\mathrm{M}_1, \mathrm{M}_2; \hat{\mathrm{M}}_2) \leq I(\mathrm{X}; \mathrm{Y}_{1,s})$ which shows that the decoder does not effect the achievable rate. This permits the proof of the optimality of the universal strategy similarly to the case of channel uncertainty at all nodes, cf. Theorem 3.13. □

Remark 3.17. *An intuitive explanation of why CSIR does not lead to an improved capacity region is already indicated in [Wol78]. Even if the channel used for the transmission is not known to the receivers, it can be estimated with arbitrary accuracy by the receivers. For sufficiently large block length* n *the part "wasted" for the estimation is a negligible part of* n*, and goes to zero as* $n \to \infty$*.*

3.4.2 CSI at the Transmitter

Here, the transmitter has perfect CSI so that it can adapt its encoder to the specific channel realization. Consequently, we now have a whole family of encoders at the relay node, one for each channel realization, while we still have universal decoding sets.

Definition 3.18. *A deterministic* $(n, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{\text{det}}^{\text{CSIT}}(\mathfrak{W})$ *for the compound BBC* $\mathfrak{W}$ *with CSIT consists of families of codewords*

$$x_{s,m}^n \in \mathcal{X}^n$$

for all $m = (m_1, m_2) \in \mathcal{M}$*, one family for each channel realizations* $s \in \mathcal{S}$*, and mutually disjoint decoding sets at nodes 1 and 2*

$$\mathcal{D}_{m_2|m_1}^{(1)} \subseteq \mathcal{Y}_1^n \qquad \text{and} \qquad \mathcal{D}_{m_1|m_2}^{(2)} \subseteq \mathcal{Y}_2^n$$

for all $m_1 \in \mathcal{M}_1$ *and* $m_2 \in \mathcal{M}_2$*.*

In the following we show that CSIT leads to an improved capacity region of the compound BBC, which is in contrast to the previous discussed case of CSIR. The introduced concept of types and typical sequences from Csiszár and Körner [CK81] permits a proof of the capacity region which is quite similar to the case of channel uncertainty at all nodes. Thus, we concentrate on the crucial points where the reasoning differs from the derivation in Section 3.3. Similarly, we start with an outer bound on the capacity region to get an intuition what is at best possible with CSIT.

Lemma 3.19. *Any given sequence of* $(n, M_1^{(n)}, M_2^{(n)})$*-codes with* $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ *must satisfy*

$$\frac{1}{n} \log M_2^{(n)} \leq \inf_{s \in \mathcal{S}} I(\mathrm{X}_s; \mathrm{Y}_{1,s}|\mathrm{U}) + o(n^0)$$
$$\frac{1}{n} \log M_1^{(n)} \leq \inf_{s \in \mathcal{S}} I(\mathrm{X}_s; \mathrm{Y}_{2,s}|\mathrm{U}) + o(n^0)$$

for random variables $\mathrm{U} - \mathrm{X}_s - (\mathrm{Y}_{1,s}, \mathrm{Y}_{2,s})$ *with joint probability distributions* $\{p_{\mathrm{U}}(u) p_{\mathrm{X}_s|\mathrm{U}}(x|u) W_s(y_1, y_2|x)\}_{s \in \mathcal{S}}$*.*

Proof. The proof is similar to the proof of Lemma 3.7 and therefore omitted for brevity. □

Next, we present a universal strategy which actually achieves the rates stated in the previous lemma. The crucial point is to establish an upper bound on the probability of error for the case of a finite set $\mathcal{S}$ similar to the one given in Lemma 3.8. Then the rest of the proof of the capacity region follows accordingly.

Lemma 3.20. *Let the index set* $\mathcal{S} = \{1, ..., S\}$ *denote a finite compound BBC* $\mathfrak{W}$ *with CSIT. For any block length* $n \in \mathbb{N}$*, input distributions* p_i*,* $i = 1, ..., S$*, and* $R_1 \leq \min_{i=1,..,S} I(p_i, W_{1,i}) - \frac{\tau}{2}$*,* $R_2 \leq \min_{i=1,..,S} I(p_i, W_{2,i}) - \frac{\tau}{2}$*,* $\tau > 0$*, there is a*

$(n, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{det}^{CSIT}(\mathfrak{W})$ where the probability of error $\mu_i^{(n)}$ at node i, $i = 1, 2$, averaged over all codebooks is bounded from above by

$$\mathbb{E}_{\mathrm{X}^n}[\mu_i^{(n)}] \leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}2^{-nc\epsilon^2} + S^2\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}}2^{-n\frac{\tau}{4}} \tag{3.13}$$

with $c = \frac{1}{2\ln 2}$.

Proof. Since the transmitter can adapt its encoder to the specific channel realization, we have a family of input distributions p_i, one for each channel realization $i \in \{1, ..., S\}$. For each distribution $p_i^{\otimes n}(x^n)$, $i = 1, ..., S$, and $\epsilon > 0$ let $\mathcal{T}_{p_i,\epsilon}^{(n)}$ denote the corresponding set of typical sequences on $\mathcal{X}^n$. We restrict the possible inputs to these subsets and define the new input distributions

$$p_i'(x^n) \coloneqq \begin{cases} \frac{p_i^{\otimes n}(x^n)}{p_i^{\otimes n}(\mathcal{T}_{p_i,\epsilon}^{(n)})} & \text{if } x^n \in \mathcal{T}_{p_i,\epsilon}^{(n)} \\ 0 & \text{else.} \end{cases}$$

Let $\tau > 0$ and set the rates $R_1 \coloneqq \min_{i=1,...,S} I(p_i, W_{1,i}) - \frac{\tau}{2}$ and $R_2 \coloneqq \min_{i=1,...,S} I(p_i, W_{2,i}) - \frac{\tau}{2}$. For each $i \in \{1, ..., S\}$ we generate $|\mathcal{M}| = |\mathcal{M}_1||\mathcal{M}_2|$ independent codewords $\mathrm{X}_{i,m}^n$, one for each $m = (m_1, m_2)$, of length n with $|\mathcal{M}_1| \coloneqq \lfloor 2^{nR_1} \rfloor$ and $|\mathcal{M}_2| \coloneqq \lfloor 2^{nR_1} \rfloor$ according to p_i'. This implies that all generated random codewords $\mathrm{X}_{i,m}^n \in \mathcal{T}_{p_i,\epsilon}^{(n)}$ almost surely, $i = 1, ..., S$.

Since for each channel realization $s \in \mathcal{S}$, the used random codewords differ, the definitions of the decoding sets of nodes 1 and 2 slightly change as follows

$$\mathcal{D}_{m_2|m_1}^{(1)}(\mathrm{X}^n) \coloneqq \Big(\bigcup_{i=1}^{S} \mathcal{T}_{W_{1,i},\epsilon}^{(n)}(\mathrm{X}_{i,m}^n)\Big) \cap \Big(\bigcup_{\substack{\hat{m}_2 \in \mathcal{M}_2 \\ \hat{m}_2 \neq m_2}} \bigcup_{i=1}^{S} \mathcal{T}_{W_{1,i},\epsilon}^{(n)}(\mathrm{X}_{i,m_1\hat{m}_2}^n)\Big)^c$$

$$\mathcal{D}_{m_1|m_2}^{(2)}(\mathrm{X}^n) \coloneqq \Big(\bigcup_{i=1}^{S} \mathcal{T}_{W_{2,i},\epsilon}^{(n)}(\mathrm{X}_{i,m}^n)\Big) \cap \Big(\bigcup_{\substack{\hat{m}_1 \in \mathcal{M}_1 \\ \hat{m}_1 \neq m_1}} \bigcup_{i=1}^{S} \mathcal{T}_{W_{2,i},\epsilon}^{(n)}(\mathrm{X}_{i,\hat{m}_1 m_2}^n)\Big)^c$$

where $\mathrm{X}^n \coloneqq \{\mathrm{X}_{i,m_1m_2}^n\}_{i\in\mathcal{S},m_1\in\mathcal{M}_1,m_2\in\mathcal{M}_2}$ so that the decoding sets depend on all generated codewords for all channel realizations. Consequently, the corresponding error events $E_1(s)$ and $E_2(s)$ at node 1 are now given by

$$E_1(s) \coloneqq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W_{1,s}^{\otimes n}\Big((\bigcup_{i=1}^{S} \mathcal{T}_{W_{1,i},\epsilon}^{(n)}(\mathrm{X}_{i,m}^n))^c | \mathrm{X}_{s,m}^n\Big)$$

$$E_2(s) \coloneqq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W_{1,s}^{\otimes n}\Big(\bigcup_{\substack{\hat{m}_2 \in \mathcal{M}_2 \\ \hat{m}_2 \neq m_2}} \bigcup_{i=1}^{S} \mathcal{T}_{W_{1,i},\epsilon}^{(n)}(\mathrm{X}_{i,m_1\hat{m}_2}^n) | \mathrm{X}_{s,m_1m_2}^n\Big).$$

We continue with the analysis of the probability of error for node 1. Again the analysis for node 2 follows accordingly using the same arguments. As in Lemma 3.8 we average over all codebooks and show that $\mathbb{E}_{\mathrm{X}^n}[\bar{e}_1(s)] \leq \mathbb{E}_{\mathrm{X}^n}[E_1(s) + E_2(s)]$ can be bounded uniformly in s from above by a term which decreases exponentially fast for increasing block length n. The derivation for $\mathbb{E}_{\mathrm{X}^n}[E_1(s)]$ proceeds exactly as in Lemma 3.8 and leads to the same upper bound

$$\begin{aligned}\mathbb{E}_{\mathrm{X}^n}[E_1(s)] &= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathbb{E}_{\mathrm{X}^n}\Big[W_{1,s}^{\otimes n}\Big((\bigcup_{i=1}^{S}\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{i,m}))^c|\mathrm{X}^n_{s,m}\Big)\Big] \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-nc\epsilon^2}\end{aligned}$$

with $c = \frac{1}{2\ln 2}$, cf. also (3.6). The first steps of the derivation for $\mathbb{E}_{\mathrm{X}^n}[E_2(s)]$ are similar to Lemma 3.8 up to

$$\mathbb{E}_{\mathrm{X}^n}[E_2(s)] \leq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}\sum_{\substack{\hat{m}_2\in\mathcal{M}_2\\ \hat{m}_2\neq m_2}}\sum_{i=1}^{S}\frac{\mathbb{E}_{\mathrm{X}^n_{i,m_1\hat{m}_2}}\Big[q_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{i,m_1\hat{m}_2})\Big)\Big]}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}}$$

where $q_{1,s} \in \mathcal{P}(\mathcal{Y}_1)$ denotes the output distribution generated by p_s and $W_{1,s}$, cf. also (3.8). The crucial point is that since for each $i \in \{1,...,S\}$ we have a different input distribution p_i, the input distribution and channel may not "coincide" with the output distribution. But nevertheless we can apply Lemma B.12 since $\mathrm{X}^n_{i,m_1\hat{m}_2} \in \mathcal{T}^{(n)}_{p_i,\epsilon}$ almost surely. Further this lemma is also applicable if the distributions do not match, cf. (B.11) in Appendix B.2.1. We obtain similarly to Lemma 3.8

$$\mathbb{E}_{\mathrm{X}^n_{i,m_1\hat{m}_2}}\Big[q_{1,s}^{\otimes n}\Big(\mathcal{T}^{(n)}_{W_{1,i},\epsilon}(\mathrm{X}^n_{i,m_1\hat{m}_2})\Big)\Big] \leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-n(I(p_i,W_{1,i})-\varphi(\epsilon)-\psi(\epsilon))}$$

for $\epsilon \in (0, \frac{1}{4|\mathcal{X}||\mathcal{Y}_1|})$. The rest of the proof proceeds exactly as in Lemma 3.8 so that we end up with the following upper bound for the average probability of error at node i, $i = 1, 2$,

$$\mathbb{E}_{\mathrm{X}^n}[\mu_i^{(n)}] \leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}2^{-nc\epsilon^2} + S^2\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}}2^{-n\frac{\tau}{4}}$$

with $c = \frac{1}{2\ln 2}$ as stated in (3.13) which finally proves the lemma. □

With the previous Lemmas 3.19 and 3.20 we are now able to establish the capacity region of the compound BBC $\mathfrak{W}$ with CSIT for finite and arbitrary sets of channels. The proofs proceed exactly as in Section 3.3 so that we omit them for brevity.

Corollary 3.21. *The capacity region $\mathcal{R}_{det}^{CSIT}(\mathfrak{W})$ of the finite compound BBC $\mathfrak{W}$ with CSIT is the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy*

$$R_1 \leq \min_{i=1,..,S} I(\mathrm{X}_i; \mathrm{Y}_{1,i}|\mathrm{U}) \tag{3.14a}$$

$$R_2 \leq \min_{i=1,..,S} I(\mathrm{X}_i, \mathrm{Y}_{2,i}|\mathrm{U}) \tag{3.14b}$$

for random variables $\mathrm{U} - \mathrm{X}_i - (\mathrm{Y}_{1,i}, \mathrm{Y}_{2,i})$ with joint probability distributions $\{p_{\mathrm{U}}(u)p_{\mathrm{X}_i|\mathrm{U}}(x|u)W_i(y_1, y_2|x)\}_{i=1,...,S}$. The cardinality U can be bounded by $|\mathcal{U}| \leq 2$.

Theorem 3.22. *The capacity region $\mathcal{R}_{det}^{CSIT}(\mathfrak{W})$ of the compound BBC $\mathfrak{W}$, where the set of channels $\mathcal{S}$ can be arbitrary, is the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy*

$$R_1 \leq \inf_{s\in\mathcal{S}} I(\mathrm{X}_s, \mathrm{Y}_{1,s}|\mathrm{U}) \tag{3.15a}$$

$$R_2 \leq \inf_{s\in\mathcal{S}} I(\mathrm{X}_s, \mathrm{Y}_{2,s}|\mathrm{U}) \tag{3.15b}$$

for random variables $\mathrm{U} - \mathrm{X}_s - (\mathrm{Y}_{1,s}, \mathrm{Y}_{2,s})$ with joint probability distributions $\{p_{\mathrm{U}}(u)p_{\mathrm{X}_s|\mathrm{U}}(x|u)W_s(y_1, y_2|x)\}_{s\in\mathcal{S}}$. The cardinality U can be bounded by $|\mathcal{U}| \leq 2$.

Remark 3.23. *Similarly as in [KS07, Theorem 1] it is further possible to get rid of the time-sharing random variable U in (3.14) and (3.15) so that we end up with the corresponding expressions without U.*

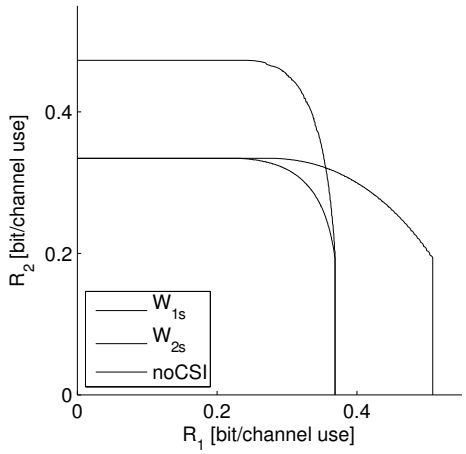
3.5 Numerical Example and Game-Theoretic Interpretation

In this section we give a numerical example which illustrates how CSIT improves the capacity region of the compound BBC. Therefore, let $|\mathcal{X}| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and consider a particular set of channels $\mathcal{S} = \{s_1, s_2\}$ with two possible states. Then the compound BBC $\mathfrak{W}$ is specified by marginal channels that are given by the following transition probability matrices

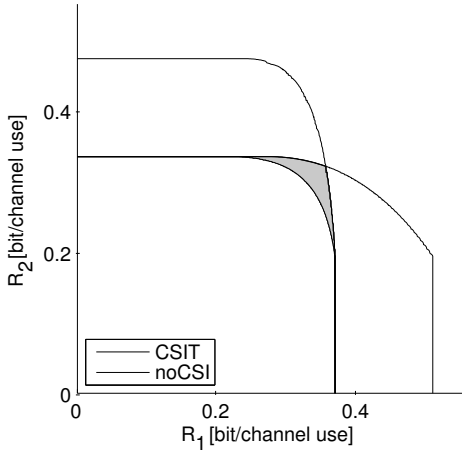
$$W_{1,s_1} \coloneqq \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0 & 0.7 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} \qquad W_{2,s_1} \coloneqq \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

$$W_{1,s_2} \coloneqq \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0 & 1 & 0 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} \qquad W_{2,s_2} \coloneqq \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$$

Figure 3.2 depicts the capacity regions $\mathcal{R}_{\text{det}}(\mathfrak{W})$ and $\mathcal{R}_{\text{det}}^{\text{CSIT}}(\mathfrak{W})$ of this particular compound BBC $\mathfrak{W}$ for channel uncertainty at all nodes and CSIT, respectively. How CSIT affects the



(a) Capacity region $\mathcal{R}_{\text{det}}(\mathfrak{W})$ of the compound BBC $\mathfrak{W}$ with channel uncertainty at all nodes.



(b) Capacity region $\mathcal{R}_{\text{det}}^{\text{CSIT}}(\mathfrak{W})$ of the compound BBC $\mathfrak{W}$ with CSIT.

Figure 3.2: Capacity regions of the compound BBC $\mathfrak{W}$ with channel uncertainty at all nodes and CSIT, respectively. The shaded area in Fig. 3.2(b) illustrates the gain in the capacity region based on the available CSIT.

maximal achievable rates is shown in Figure 3.2(b), where the shaded area illustrates the gain in the capacity region due to the available CSIT. The rate regions for channel realizations s_1 (dashed line) and s_2 (dashed-dotted line) are included for convenience.

Similar to the single-user compound channel it is possible to analyze the compound BBC from a game-theoretic perspective. Therefore, we assume that the nodes and *nature* play a two-player zero-sum game [AH94, BO98] with the mutual information I as the payoff function as depicted in Figure 3.3. This is called a *game against nature* [Mil51].

In this game, the set of channels $\mathcal{S}$ corresponds to nature's action space. Nature's aim is to establish the worst communication condition by selecting $s \in \mathcal{S}$ such that the mutual information is minimized. The set of input distributions $\mathcal{P}(\mathcal{X})$ corresponds to the action space of the player. Clearly, the player wants to maximize the mutual information and therefore tries to choose the best input distribution $p \in \mathcal{P}(\mathcal{X})$. Then for given $p \in \mathcal{P}(\mathcal{X})$ and $s \in \mathcal{S}$ the outcome of the game is given by the following achievable rate region

$$\begin{aligned}\mathcal{R}(p,s) = \big\{(R_1,R_2) \in \mathbb{R}_+^2 : R_1 \leq I(p,W_{1,s}),\\ R_2 \leq I(p,W_{2,s})\big\}.\end{aligned} \tag{3.16}$$

Within this game against nature framework the game can be played in two different ways. First, the player and nature moves simultaneously without knowing the other's choice. And

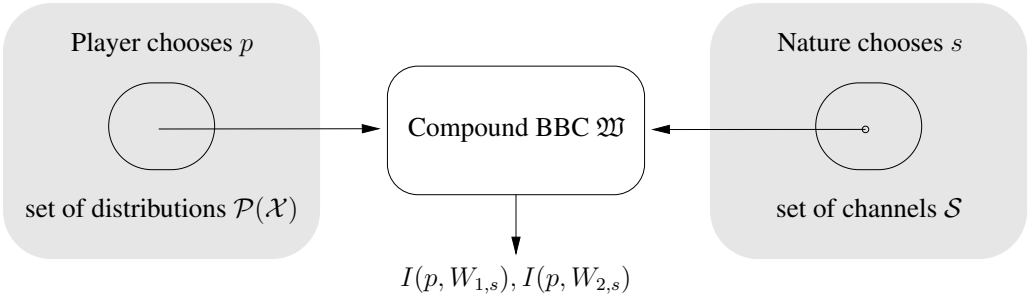


Figure 3.3: Transmission over the compound BBC $\mathfrak{W}$ as a game against nature.

second, nature moves first so that the player is aware of nature's choice. In the single-user scenario, these two types of the game lead to well-known $\max\min$ and $\min\max$ formulations for the outcome of the game, cf. for example [PCL03]. Since we have a multi-user scenario that deals with rate regions, cf. (3.16), we have a vector-valued problem and the $\max$ and $\min$ expressions extend to the union and intersection. The outcomes of the game are then

$$\mathcal{R}_{\text{det}}(\mathfrak{W}) = \text{co}\Big(\bigcup_{p\in\mathcal{P}(\mathcal{X})} \bigcap_{s\in\mathcal{S}} \mathcal{R}(p,s) \Big) \tag{3.17}$$

and

$$\mathcal{R}_{\text{det}}^{\text{CSIT}}(\mathfrak{W}) = \text{co}\Big(\bigcap_{s\in\mathcal{S}} \bigcup_{p\in\mathcal{P}(\mathcal{X})} \mathcal{R}(p,s) \Big). \tag{3.18}$$

We see that (3.17) and (3.18) are equivalent to the capacity regions given in Theorems 3.13 and 3.22. Accordingly, they correspond to the cases of channel uncertainty at all nodes and CSIT, respectively. Note that in the theorems the convex hull is established by the time-sharing variable U. Moreover, it follows immediately from (3.17) and (3.18) that $\mathcal{R}_{\text{det}}(\mathfrak{W}) \subseteq \mathcal{R}_{\text{det}}^{\text{CSIT}}(\mathfrak{W})$ which agrees with the intuition and the previous results that CSIT improves the capacity region.

3.6 Discussion

In practical wireless communication systems channel uncertainty is a ubiquitous phenomenon. The question must be asked if it is advantageous to improve the available channel state information at the nodes or if the nodes should be left with the uncertainty. The concept of the compound channel allows us to derive robust coding strategies that are appropriate for wireless applications where certain rates have to be guaranteed even in the case of channel uncertainty. Further, the analysis shows the best possible rates that are achievable under

channel uncertainty. This allows us to assess if it is worthwhile to improve the channel state information at the nodes, e.g., by using longer training sequences or feedback. In particular, this is important to know for the design of wireless networks.

In this chapter we addressed the bidirectional broadcast channel and presented robust coding strategies which guarantee certain rates regardless of the current channel realization. These immediately lead to a characterization of the capacity region of the compound BBC which is an useful result since it constitutes the basis for further analysis of multi-user settings under more complex models of channel uncertainty and in uncoordinated wireless networks. In particular, our results provide valuable insights since for the general broadcast channel with discrete channels and finite alphabets the capacity region for compound channels is not known and, consequently, similar results are not available. To date, only some special cases are treated as for example the case with degraded MIMO Gaussian channels [WLS$^+$09] which is a quite different setting to the one we considered here.

Furthermore, the analysis shows that CSIR does not improve the capacity region if the transmitter merely knows the set of channels, which is at first counter-intuitive. But at second glance this becomes clear if one realizes that transmitted symbols "wasted" for channel estimation are negligible for large block lengths. Further, we show that CSIT can advantageously be used to improve the capacity region since the transmitter can adapt its encoder to the specific channel realization. Adaptive bidirectional relaying with quantized CSIT is analyzed in [KP11]. The game-theoretic interpretation of the compound BBC reveals interesting generalizations, which keep the characteristics of the single-user compound channel but includes now multi-user effects. This is a nice property of the compound BBC, which is not self-evident for multi-user scenarios.

4 Bidirectional Relaying in Uncoordinated Networks

The ongoing research progress reveals a paradigm shift from coordinated to uncoordinated wireless systems. While most current systems such as conventional cellular systems are usually coordinated in a centralized way, several future systems will act in an uncoordinated and self-organizing way, e.g., ad-hoc or sensor networks. The main issue that comes along with this development is that interference becomes an ubiquitous phenomenon and will be one of the main impairments in future wireless networks. Since the resulting interference cannot be longer coordinated in a centralized way, new concepts are needed especially for the frequency usage.

In the previous chapter we studied an isolated bidirectional relay network under channel uncertainty. The next step is to consider bidirectional relaying within an (uncoordinated) wireless network. Although uncertainty in the channel state information is ubiquitous and should not be disregarded, the most defining impairment now is the fact that the communication is disturbed by interference from other transmitting nodes as illustrated in Figure 4.1. If there is no a priori knowledge about applied transmit strategies such as coding or modulation schemes of all other transmitting nodes, there is no knowledge about the induced interference. Thus, a reasonable model is to assume that the channel may vary from symbol to symbol in an unknown and arbitrary manner. This is the concept of *arbitrarily varying channels (AVC)*.

The point-to-point AVC was first introduced by Blackwell, Breiman, and Thomasian [BBT60] who established its random code capacity. Interestingly, for this channel the random code and deterministic code capacity need not be equal. In more detail, Ahlswede showed in his famous work [Ahl78] that the AVC displays a dichotomy behavior: the deterministic code capacity either equals the random code capacity or else is zero. Unfortunately, he missed a characterization for the deterministic code capacity to be non-zero. Finally, Ericson [Eri85] and Csiszár and Narayan [CN88b] established non-symmetrizability as a necessary and sufficient condition for the AVC to have a non-zero deterministic code capacity. Roughly speaking, a symmetrizable AVC leads to a zero deterministic code capacity, since such a channel can *emulate* a valid input which makes it impossible for the decoder to decide on the correct codeword.

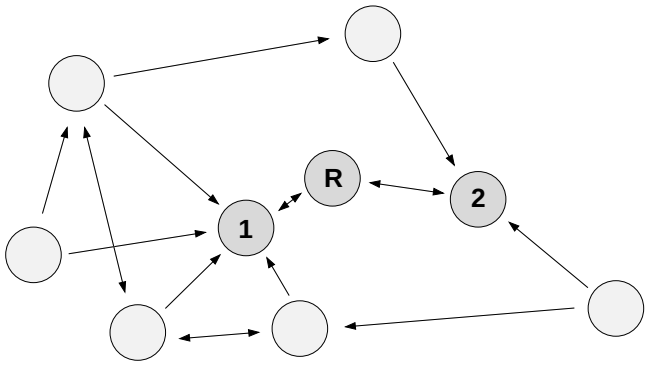


Figure 4.1: Bidirectional relaying in (uncoordinated) wireless networks.

While decoding rules of deterministic codes decide on exactly one transmitted codeword or message, the concept of list decoding allows the decoder to narrow the transmitted codeword to a list of certain alternatives. List decoding was first considered by Elias [Eli57] and Wozencraft [Woz58] but it is shown that it does not change the capacity of ordinary channels. The situation changes for AVCs where it might help to dissolve the ambiguity of codewords caused by symmetrizable channels. The corresponding list code capacity was established and analyzed in detail independently by Blinovsky, Narayan, and Pinsker [BNP95] and Hughes [Hug97].

Further effects occur, if constraints on the permissible codewords and sequences of channel states are imposed. This assumption is motivated by the fact that in real communication systems the transmitter as well as possible interferers are usually limited in their transmit power. The single-user AVC under input and state constraints was analyzed in detail by Csiszár and Narayan [CN88a, CN88b]. There, it is shown that due to the imposed constraints the deterministic code capacity may be positive even for symmetrizable channels, but may be less than its random code capacity.

Besides the point-to-point case there are important extensions to multi-user settings as well. The arbitrarily varying multiple access channel (AVMAC) is analyzed in [Jah81, Gub90, AC99, Nit10], where the random code and deterministic code capacity regions are established. It is shown that the latter may have an empty interior which is completely characterized and analyzed in terms of an appropriate concept of symmetrizability [Gub90, AC99]. The AVMAC with constraints on input and states is considered in [Gub91, GH95], where it is shown that the random code capacity region is non-convex in general [GH95].

While the AVMAC is well understood, there are only partial results known until now for the arbitrarily varying general broadcast channel (AVGBC). An achievable deterministic code rate region for the AVGBC is established in [Jah81] but it is not further analyzed when its interior is non-empty. On the other hand, [HB06] analyzes an achievable deterministic code

rate region of the AVGBC in terms of symmetrizability but imposes further the assumption of degraded message sets. Further, in [Jah81, HB06] only achievable rate regions are presented but no converse results or outer bounds on the capacity regions are given.

In this chapter we analyze bidirectional relaying for arbitrarily varying channels. Therefore, we briefly summarize in Section 4.1 the well understood AVMAC for the first phase of the decode-and-forward protocol. To capture the second phase, we introduce the *arbitrarily varying bidirectional broadcast channel (AVBBC)* in Section 4.2. As a first step, in Section 4.3 we establish the random code capacity region presenting the proof of achievability and the weak converse. Similar to Ahlswede's dichotomy result for the point-to-point AVC, we show in Section 4.4 that the deterministic list capacity region of the AVBBC either equals its random code capacity region or else has an empty interior. Then, in Section 4.5 we use an appropriate concept of symmetrizability to establish non-symmetrizability as a necessary and sufficient condition for the list capacity region to have a non-empty interior. Furthermore, we present a weak converse that completely establishes the list capacity region. Then, Section 4.6 discusses the case where constraints are imposed on input and state sequences. The corresponding random code and deterministic code capacity regions are given. In Section 4.7 we briefly address the scenario where the transmission is disturbed by unknown varying additive interference and end with a discussion in Section 4.8.

4.1 Arbitrarily Varying Multiple Access Channel

In this section we briefly restate the *arbitrarily varying multiple access channel (AVMAC)* which models the first phase of the decode-and-forward bidirectional relaying protocol.

We introduce a finite state set $\mathcal{S}$. Further, let $\mathcal{X}_i$, $i = 1, 2$, and $\mathcal{Y}$ be finite input and output sets. Then, for fixed state sequence $s^n \in \mathcal{S}^n$ of length n and input and output sequences $x_i^n \in \mathcal{X}_i^n$, $i = 1, 2$, and $y^n \in \mathcal{Y}^n$, the discrete memoryless multiple access channel is given by $V^{\otimes n}(y^n|x_1^n, x_2^n, s^n) \coloneqq \prod_{k=1}^n V(y_k|x_{1,k}, x_{2,k}, s_k)$.

Definition 4.1. *The discrete memoryless* arbitrarily varying multiple access channel $\mathfrak{V}^n$ *is the family*

$$\mathfrak{V}^n \coloneqq \left\{V^{\otimes n} : \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{S}^n \to \mathcal{P}(\mathcal{Y}^n)\right\}_{n\in\mathbb{N}, s^n\in\mathcal{S}^n}.$$

Further, for any probability distribution $q \in \mathcal{P}(\mathcal{S})$ we denote the averaged multiple access channel by

$$\overline{V}_q(y|x_1, x_2) = \sum_{s\in\mathcal{S}} V(y|x_1, x_2, s)q(s).$$

Random Code Capacity Region

In contrast to the compound MAC there is a difference for the AVMAC if random or deterministic codes are used. The random code capacity was characterized by Jahn [Jah81].

Theorem 4.2 ([Jah81])**.** *The random code capacity region $\mathcal{R}_{ran}(\mathfrak{V}^n)$ of the AVMAC $\mathfrak{V}^n$ is the set of all rate pairs $(R_2, R_1) \in \mathbb{R}_+^2$ that satisfy*[1]

$$\begin{aligned} R_2 &\leq \inf_{q\in\mathcal{P}(\mathcal{S})} I(\mathrm{X}_1;\overline{\mathrm{Y}}_q|\mathrm{X}_2,\mathrm{U}) \\ R_1 &\leq \inf_{q\in\mathcal{P}(\mathcal{S})} I(\mathrm{X}_2;\overline{\mathrm{Y}}_q|\mathrm{X}_1,\mathrm{U}) \\ R_2+R_1 &\leq \inf_{q\in\mathcal{P}(\mathcal{S})} I(\mathrm{X}_1,\mathrm{X}_2;\overline{\mathrm{Y}}_q|\mathrm{U}) \end{aligned}$$

for random variables $\mathrm{U} - (\mathrm{X}_1,\mathrm{X}_2) - \overline{\mathrm{Y}}_q$ and joint probability distributions $\{p_{\mathrm{U}}(u)p_{\mathrm{X}_1|\mathrm{U}}(x_1|u)p_{\mathrm{X}_2|\mathrm{U}}(x_2|u)\overline{V}_q(y|x_1,x_2)\}_{q\in\mathcal{P}(\mathcal{S})}$. Thereby, U is an auxiliary random variable and describes a possible time-sharing operation.

Deterministic Code Capacity Region

The first step of the derivation of $\mathcal{R}_{\text{det}}(\mathfrak{V}^n)$ was done by Jahn [Jah81]. He showed that the deterministic code capacity region of the AVMAC $\mathfrak{V}^n$ displays the following behavior

$$\mathcal{R}_{\text{det}}(\mathfrak{V}^n) = \mathcal{R}_{\text{ran}}(\mathfrak{V}^n) \qquad \text{if } \operatorname{int}(\mathcal{R}_{\text{det}}(\mathfrak{V}^n)) \neq \emptyset \tag{4.1}$$

by extending Ahlswede's robustification technique [Ahl80b] and elimination technique [Ahl78] for the single-user channel to the multiple access channel. Unfortunately, in [Jah81] he missed a characterization when the interior of $\mathcal{R}_{\text{det}}(\mathfrak{V}^n)$ is non-empty.

To characterize when $\operatorname{int}(\mathcal{R}_{\text{det}}(\mathfrak{V}^n)) \neq \emptyset$, Gubner introduced in [Gub90] a natural extension of symmetrizability for the AVMAC:

Definition 4.3. i*) An AVMAC $\mathfrak{V}^n$ is* $(\mathcal{X}_1,\mathcal{X}_2)$-symmetrizable *if for some channel $U : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{P}(\mathcal{S})$ the following*

$$\sum_{s\in\mathcal{S}} V(y|x_1,x_2,s)U(s|x_1',x_2') = \sum_{s\in\mathcal{S}} V(y|x_1',x_2',s)U(s|x_1,x_2)$$

holds for every $x_1, x_1' \in \mathcal{X}_1$, $x_2, x_2' \in \mathcal{X}_2$, and $y \in \mathcal{Y}$.

[1] Recall that the compound MAC is considered within the two-phase decode-and-forward protocol. Therefore, the individual rates look "swapped", cf. Chapter 2 and especially Figure 2.1.

ii) *An AVMAC $\mathfrak{V}^n$ is* $\mathcal{X}_1$-symmetrizable *if for some channel $U_1 : \mathcal{X}_1 \to \mathcal{P}(\mathcal{S})$ the following*

$$\sum_{s\in\mathcal{S}} V(y|x_1, x_2, s)U_1(s|x_1') = \sum_{s\in\mathcal{S}} V(y|x_1', x_2, s)U_1(s|x_1)$$

holds for every $x_1, x_1' \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, and $y \in \mathcal{Y}$.

iii) *An AVMAC $\mathfrak{V}^n$ is* $\mathcal{X}_2$-symmetrizable *if for some channel $U_2 : \mathcal{X}_2 \to \mathcal{P}(\mathcal{S})$ the following*

$$\sum_{s\in\mathcal{S}} V(y|x_1, x_2, s)U_2(s|x_2') = \sum_{s\in\mathcal{S}} V(y|x_1, x_2', s)U_2(s|x_2)$$

holds for every $x_1 \in \mathcal{X}_1$, $x_2, x_2' \in \mathcal{X}_2$, and $y \in \mathcal{Y}$.

Using this definition of symmetrizability Ahlswede and Cai [AC99] were able to show that the AVMAC has a capacity region whose interior is non-empty if and only if the AVMAC is non-$(\mathcal{X}_1, \mathcal{X}_2)$-symmetrizable, non-$\mathcal{X}_1$-symmetrizable, and non-$\mathcal{X}_2$-symmetrizable. Together with the result of Jahn [Jah81] this finally establishes the deterministic code capacity region $\mathcal{R}_{\text{det}}(\mathfrak{V}^n)$ of the AVMAC.

Theorem 4.4 ([Jah81, AC99])**.** *For a non-$(\mathcal{X}_1, \mathcal{X}_2)$-symmetrizable, non-$\mathcal{X}_1$-symmetrizable, and non-$\mathcal{X}_2$-symmetrizable AVMAC $\mathfrak{V}^n$ the deterministic code capacity region $\mathcal{R}_{det}(\mathfrak{V}^n)$ under the average error criterion is*

$$\mathcal{R}_{det}(\mathfrak{V}^n) = \mathcal{R}_{ran}(\mathfrak{V}^n).$$

We have interior$(\mathcal{R}_{det}(\mathfrak{V}^n)) = \emptyset$, if and only if the AVMAC $\mathfrak{V}^n$ is $(\mathcal{X}_1, \mathcal{X}_2)$-symmetrizable, $\mathcal{X}_1$-symmetrizable, or $\mathcal{X}_2$-symmetrizable.

List Decoding

Recently, it was Nitinawarat [Nit10] who extended the idea of list decoding for the single-user AVC [BNP95, Hug97] to the AVMAC. This is based on a concept of symmetrizability which distinguishes among different degrees of symmetry:

Definition 4.5. *For a positive integer t, an AVMAC $\mathfrak{V}^n$ is t-symmetrizable if either of the following holds.*

i) *There exists a conditional distribution $U(s|x_1', x_1'', ..., x_t', x_t'')$, $s \in \mathcal{S}$, $(x_1', x_1''), ..., (x_t', x_t'') \in \mathcal{X}_1 \times \mathcal{X}_2$ such that for any $x_0', x_1', ..., x_t' \in \mathcal{X}_1$, $x_0'', x_1'', ..., x_t'' \in \mathcal{X}_2$, $y \in \mathcal{Y}$ and any permutation π on $\{0, ..., t\}$*

$$\sum_{s\in\mathcal{S}} V(y|x_0'x_0'', s)U(s|x_1', x_1'', ..., x_t', x_t'') = \sum_{s\in\mathcal{S}} V(y|x'_{\pi(0)}, x''_{\pi(0)}, s)U(s|x'_{\pi(1)}, x''_{\pi(1)}, ..., x'_{\pi(t)}, x''_{\pi(t)}).$$

ii) *There exists a conditional distribution* $U(s|x'_1,...,x'_a,x''_1,...,x''_b)$, $s \in \mathcal{S}$, $x'_1,...,x'_a \in \mathcal{X}_1$, $x''_1,...,x''_b \in \mathcal{X}_2$ *for some* a, b *satisfying* $(a+1)(b+1) \geq t+1$ *such that for any* $x'_0,...,x'_a \in \mathcal{X}_1$, $x''_0,...,x''_b \in \mathcal{X}_2$, $s \in \mathcal{S}$, $y \in \mathcal{Y}$, *and any permutations* π *on* $(0,...,a)$ *and* σ *on* $(0,...,b)$,

$$\sum_{s\in\mathcal{S}} V(y|x'_0,x''_0,s)U(s|x'_1,...,x'_a,x''_1,...,x''_b) = \sum_{s\in\mathcal{S}} V(y|x'_{\pi(0)},x''_{\sigma(0)},s)U(s|x'_{\pi(1)},...,x'_{\pi(a)},x''_{\sigma(1)},...,x''_{\sigma(b)}).$$

The symmetrizability of the AVMAC $\mathfrak{V}^n$ *is given by the largest integer* T *such that the AVMAC* $\mathfrak{V}^n$ *is* T*-symmetrizable.*

Using this definition of symmetrizability for the case that the decoder maps the received signal into a list of size L, the list capacity region $\mathcal{R}_{\text{list}}(\mathfrak{V}^n|L)$ of the AVMAC $\mathfrak{V}^n$ is characterized by the following results [Nit10].

Theorem 4.6 ([Nit10])**.** *For an AVMAC* $\mathfrak{V}^n$ *with symmetrizability* T*, the list capacity region* $\mathcal{R}_{list}(\mathfrak{V}^n|L)$ *has an empty interior for every list size* $L \leq T$.

Theorem 4.7 ([Nit10])**.** *For an AVMAC* $\mathfrak{V}^n$ *with symmetrizability* T*, the list capacity region is given by*

$$\mathcal{R}_{list}(\mathfrak{V}^n|L) = \mathcal{R}_{ran}(\mathfrak{V}^n) \qquad \text{if } L \geq (T+1)^2(T+2)-1.$$

Constraints on Input and States

The AVMAC is further analyzed for the case where constraints are imposed on the inputs and states. Therefore, cost functions $g_i(x_i)$ on $\mathcal{X}_i$, $i=1,2$, for the inputs and $l(s)$ on $\mathcal{S}$ for the states are defined as

$$g_i(x_i^n) \coloneqq \frac{1}{n}\sum_{k=1}^{n} g_i(x_{i,k}), \quad i=1,2$$
$$l(s^n) \coloneqq \frac{1}{n}\sum_{k=1}^{n} l(s_k).$$

As in [GH95] we define for given auxiliary distribution $p_{\mathrm{U}} \in \mathcal{P}(\mathcal{U})$ the set of all input probability distributions that satisfy the input constraint Γ_i, $i=1,2$, as

$$\mathcal{P}(\mathcal{X}_i,\Gamma_i|p_{\mathrm{U}}) \coloneqq \left\{p_{\mathrm{X}_i|\mathrm{U}} \in \mathcal{P}(\mathcal{X}_i|\mathcal{U}) : \mathbb{E}_{p_{\mathrm{X}_i|\mathrm{U}}}[g_i(p_{\mathrm{X}_i|\mathrm{U}})] \leq \Gamma_i\right\}.$$

Further note that

$$\mathbb{E}_q[l(q)] = \sum_{u\in\mathcal{U}} p_{\mathrm{U}}(u)\sum_{s\in\mathcal{S}} q(s|u)l(s)$$

depends only on p_{U} and q. Therefore we define the set of all probability distributions $q \in \mathcal{P}(\mathcal{S}|p_{\mathrm{U}})$ that satisfy $\mathbb{E}_q[l(q)] \leq \Lambda$ by

$$\mathcal{P}(\mathcal{S}, \Lambda|p_{\mathrm{U}}) \coloneqq \big\{q : q \in \mathcal{P}(\mathcal{S}|p_{\mathrm{U}}), \mathbb{E}_q[l(q)] \leq \Lambda\big\}.$$

The random code capacity region of the AVMAC was determined by Gubner and Hughes [GH95].

Theorem 4.8 ([GH95]). *The random code capacity region $\mathcal{R}_{ran}(\mathfrak{V}^n|\Gamma_1, \Gamma_2, \Lambda)$ of the AVMAC $\mathfrak{V}^n$ under input constraints Γ_i, $i = 1, 2$, and state constraint Λ is*

$$\mathcal{R}_{ran}(\mathfrak{V}^n|\Gamma_1, \Gamma_2, \Lambda) \coloneqq \bigcup_{\substack{p_{\mathrm{U}} \in \mathcal{P}(\mathcal{U}), |\mathcal{U}| \leq \infty \\ p_{\mathrm{X}_i|\mathrm{U}} \in \mathcal{P}(\mathcal{X}_i, \Gamma_i|p_{\mathrm{U}}), i=1,2}} \mathcal{R}(p_{\mathrm{U}}, p_{\mathrm{X}_1|\mathrm{U}}, p_{\mathrm{X}_2|\mathrm{U}})$$

with

$$\begin{aligned}\mathcal{R}(p_{\mathrm{U}}, p_{\mathrm{X}_1|\mathrm{U}}, p_{\mathrm{X}_2|\mathrm{U}}) \coloneqq \big\{(R_2, R_1) \in \mathbb{R}_+^2 : & R_2 \leq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda|p_{\mathrm{U}})} I(\mathrm{X}_1; \mathrm{Y}|\mathrm{X}_2, \mathrm{U}) \\ & R_1 \leq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda|p_{\mathrm{U}})} I(\mathrm{X}_2; \mathrm{Y}|\mathrm{X}_1, \mathrm{U}) \\ & R_1 + R_2 \leq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda|p_{\mathrm{U}})} I(\mathrm{X}_1, \mathrm{X}_2; \mathrm{Y}|\mathrm{U})\big\}\end{aligned}$$

for $\{p_{\mathrm{U}}(u)p_{\mathrm{X}_1|\mathrm{U}}(x_1|u)p_{\mathrm{X}_2|\mathrm{U}}(x_2|u)q(s|u)V(y|x_1, x_2, s)\}_{q \in \mathcal{P}(\mathcal{S}, \Lambda|p_{\mathrm{U}})}$.

Unfortunately, due to the state constraint the random code capacity region is *not* convex in general. This is in contrast to the case without any state constraints where the corresponding region is indeed convex [Jah81].

For the deterministic code capacity region of the AVMAC with constraints on input and states there are only partial results available [Gub91, Gub92] and it remains unsolved in general.

4.2 Arbitrarily Varying Bidirectional Broadcast Channel

The transmission is affected by arbitrarily varying channels, which is modeled with the help of a finite state set $\mathcal{S}$. Further, let $\mathcal{X}$ and $\mathcal{Y}_i$, $i = 1, 2$, be finite input and output sets. Then, for a fixed state sequence $s^n \in \mathcal{S}^n$ of length n and input and output sequences $x^n \in \mathcal{X}^n$ and $y_i^n \in \mathcal{Y}_i^n$, $i = 1, 2$, the discrete memoryless broadcast channel is given by $W^{\otimes n}(y_1^n, y_2^n|x^n, s^n) \coloneqq \prod_{k=1}^n W(y_{1,k}, y_{2,k}|x_k, s_k)$.

Definition 4.9. *The discrete memoryless* arbitrarily varying broadcast channel $\mathfrak{W}^n$ *is defined by a family*

$$\mathfrak{W}^n := \left\{W^{\otimes n} : \mathcal{X}^n \times \mathcal{S}^n \to \mathcal{P}(\mathcal{Y}_1^n \times \mathcal{Y}_2^n)\right\}_{n\in\mathbb{N}, s^n\in\mathcal{S}^n}.$$

Since we do not allow any cooperation between the receiving nodes, it is sufficient to consider only the marginal transition probabilities $W_i^{\otimes n}(y_i^n|x^n, s^n)$, $i = 1, 2$. Further, for any probability distribution $q \in \mathcal{P}(\mathcal{S})$ we denote the averaged broadcast channel by

$$\overline{W}_q(y_1, y_2|x) := \sum_{s\in\mathcal{S}} W(y_1, y_2|x, s)q(s) \tag{4.2}$$

and the corresponding averaged marginal channels by $\overline{W}_{1,q}(y_1|x)$ and $\overline{W}_{2,q}(y_2|x)$.

For the following analysis we need a concept of symmetrizability which distinguishes among different degrees of symmetry. In more detail, we say a channel $\widetilde{W}_i(y_i|x_1, ..., x_t)$ with input alphabet $\mathcal{X}^t$ and output alphabet $\mathcal{Y}_i$ is symmetric in $x_1, ..., x_t$ if the channel is invariant under all permutations of the inputs $x_1, ..., x_t$ for all $y_i, x_1, ..., x_t$. This leads to the following definition.

Definition 4.10. *For any $t_i \geq 1$, $i = 1, 2$, an arbitrarily varying broadcast channel is* $(\mathcal{Y}_i, t_i)$-symmetrizable *if there is a channel $U_i : \mathcal{X}^{t_i} \to \mathcal{P}(\mathcal{S})$ such that*

$$\widetilde{W}_i(y_i|x_0, x_1, ..., x_{t_i}) := \sum_{s\in\mathcal{S}} W_i(y_i|x_0, s)U_i(s|x_1, ..., x_{t_i}) \tag{4.3}$$

is symmetric in $x_0, x_1, ..., x_{t_i}$ for all $x_0, x_1, ..., x_{t_i} \in \mathcal{X}$ and $y_i \in \mathcal{Y}_i$. For convenience, we take all arbitrarily varying broadcast channels to be $(\mathcal{Y}_i, 0)$-symmetrizable, $i = 1, 2$.

Intuitively, a $(\mathcal{Y}_i, t_i)$-symmetrizable channel can be interpreted as a channel where the state sequence can emulate t_i replicas of the channel input. Further, from the definition it is clear that if an arbitrarily varying broadcast channel is $(\mathcal{Y}_i, t_i)$-symmetrizable, then it is also $(\mathcal{Y}_i, t_i')$-symmetrizable for all $0 \leq t_i' \leq t_i$, $i = 1, 2$.

Definition 4.11. *The* symmetrizability *of an arbitrarily varying broadcast channel is defined by the largest integers t_1 and t_2 such that the channel is $(\mathcal{Y}_1, t_1)$-symmetrizable and $(\mathcal{Y}_2, t_2)$-symmetrizable. This pair of largest integers is denoted by (T_1, T_2).*

Remark 4.12. *The concept of symmetrizability for the arbitrarily varying broadcast channel introduced above is a natural extension of the one proposed for the single-user AVC under list decoding in [BNP95, Hug97]. Additionally, we call a $(\mathcal{Y}_i, 1)$-symmetrizable channel in the sense of Definition 4.10 simply a* $\mathcal{Y}_i$-symmetrizable *channel according to the terminology*

used for the single-user AVC in [Eri85, CN88b], which does not distinguish among different degrees of symmetry. In this case, condition (4.3) can be written as

$$\sum_{s\in\mathcal{S}} W_i(y_i|x,s)U_i(s|x') = \sum_{s\in\mathcal{S}} W_i(y_i|x',s)U_i(s|x) \tag{4.4}$$

which means that the channel $\widetilde{W}_i(y_i|x,x')$ is symmetric in x, x' for all $x, x' \in \mathcal{X}$ and $y_i \in \mathcal{Y}_i$, $i = 1, 2$.

We consider the standard model with a block code of arbitrary but fixed length n. Let $\mathcal{M}_i := \{1, ..., M_i^{(n)}\}$ be the message set at node i, $i = 1, 2$, which is also known at the relay node. Further, we make use of the abbreviation $\mathcal{M} := \mathcal{M}_1 \times \mathcal{M}_2$.

Definition 4.13. *A deterministic $(n, M_1^{(n)}, M_2^{(n)}, L_1, L_2)$-*list code *$\mathcal{C}_{list}(\mathfrak{W}^n)$ of length n with list sizes (L_1, L_2) for the arbitrarily varying bidirectional broadcast channel (AVBBC) $\mathfrak{W}^n$ consists of codewords*

$$x_m^n \in \mathcal{X}^n,$$

one for each message $m = (m_1, m_2) \in \mathcal{M}$, and list decoders at nodes 1 and 2

$$\mathcal{L}^{(1)} : \mathcal{Y}_1^n \times \mathcal{M}_1 \to \hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$$
$$\mathcal{L}^{(2)} : \mathcal{Y}_2^n \times \mathcal{M}_2 \to \hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$$

where $\hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$ is the set of all subsets of $\mathcal{M}_2$ with cardinality at most L_1 and, similarly, $\hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$ is the set of all subsets of $\mathcal{M}_1$ with cardinality at most L_2.

When x_m^n with $m = (m_1, m_2)$ has been sent, and y_1^n and y_2^n have been received at nodes 1 and 2, the list decoder at node 1 is in error if m_2 is not in $\mathcal{L}^{(1)}(y_1^n, m_1)$. Accordingly, the list decoder at node 2 is in error if m_1 is not in $\mathcal{L}^{(2)}(y_2^n, m_2)$. This allows us to define the probabilities of error for given message $m = (m_1, m_2)$ and state sequence $s^n \in \mathcal{S}^n$ as

$$e(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) := \sum_{\substack{(y_1^n, y_2^n): m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1) \\ \vee m_1 \notin \mathcal{L}^{(2)}(y_2^n, m_2)}} W^{\otimes n}\big(y_1^n, y_2^n|x_m^n, s^n\big) \tag{4.5}$$

and the corresponding marginal probabilities of error at nodes 1 and 2 by $e_1(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) := \sum_{y_1^n: m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} W_1^{\otimes n}\big(y_1^n|x_m^n, s^n\big)$ and $e_2(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) := \sum_{y_2^n: m_1 \notin \mathcal{L}^{(2)}(y_2^n, m_2)} W_2^{\otimes n}\big(y_2^n|x_m^n, s^n\big)$, respectively. Thus, the *average probability of error* for state sequence $s^n \in \mathcal{S}^n$ is given by

$$\bar{e}(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) := \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} e(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \tag{4.6}$$

and the corresponding marginal average probability of error at node i by $\bar{e}_i(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \coloneqq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} e_i(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))$, $i = 1, 2$. Clearly, we always have $\bar{e}(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \leq \bar{e}_1(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) + \bar{e}_2(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))$.

For given $0 < \lambda^{(n)} < 1$, $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ is called a $(n, M_1^{(n)}, M_2^{(n)}, L_1, L_2, \lambda^{(n)})$-list code (with average probability of error $\lambda^{(n)}$) for the AVBBC $\mathfrak{W}^n$ if

$$\bar{e}(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \leq \lambda^{(n)} \qquad \text{for all } s^n \in \mathcal{S}^n.$$

Definition 4.14. *A rate pair $(R_1, R_2) \in \mathbb{R}_+^2$ is said to be* list achievable *for the AVBBC $\mathfrak{W}^n$ if for any $\delta > 0$ there exists an $n(\delta) \in \mathbb{N}$ and a sequence $\{\mathcal{C}_{list}^{(n)}(\mathfrak{W}^n)\}_{n\in\mathbb{N}}$ of deterministic $(n, M_1^{(n)}, M_2^{(n)}, L_1, L_2, \lambda^{(n)})$-list codes such that for all $n \geq n(\delta)$ we have*

$$\frac{1}{n}\log\left(\frac{M_1^{(n)}}{L_2}\right) \geq R_2 - \delta \quad \textit{and} \quad \frac{1}{n}\log\left(\frac{M_2^{(n)}}{L_1}\right) \geq R_1 - \delta$$

while

$$\max_{s^n\in\mathcal{S}^n} \bar{e}(s^n|\mathcal{C}_{list}(\mathfrak{W}^n)) \leq \lambda^{(n)}$$

with $\lambda^{(n)} \to 0$ as $n \to \infty$. The set of all achievable rate pairs with list sizes (L_1, L_2) is the list capacity region *of the AVBBC $\mathfrak{W}^n$ and is denoted by $\mathcal{R}_{list}(\mathfrak{W}^n|L_1, L_2)$.*

Remark 4.15. *The definitions above require that we have to find codes such that the average probability of error goes to zero as the block length tends to infinity for all possible state sequences simultaneously. This means that the codes are universal with respect to the state sequence.*

Remark 4.16. *For list sizes $L_1 = L_2 = 1$ the list code $\mathcal{C}_{list}(\mathfrak{W}^n)$ as given in Definition 4.13 reduces to a usual deterministic $(n, M_1^{(n)}, M_2^{(n)})$-*code *$\mathcal{C}_{det}(\mathfrak{W}^n)$ where each decoder maps its received sequence into exactly one message. The definitions of a* deterministically achievable *rate pair and the* deterministic code capacity region *$\mathcal{R}_{det}(\mathfrak{W}^n)$ follow accordingly.*

Now we are in the position to state the list capacity region of the AVBBC $\mathfrak{W}^n$. For this purpose we define the region

$$\mathcal{R}(\overline{\mathfrak{W}}) \coloneqq \bigcup_{P_{\mathrm{X}}} \Big\{(R_1, R_2) \in \mathbb{R}_+^2 : R_1 \leq \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{1,q}) \tag{4.7a}$$

$$R_2 \leq \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{2,q})\Big\} \tag{4.7b}$$

for joint probability distributions $\{P_{\mathrm{X}}(x)\overline{W}_q(y_1, y_2|x)\}_{q\in\mathcal{P}(\mathcal{S})}$.

Theorem 4.17. *The list capacity region $\mathcal{R}_{list}(\mathfrak{W}^n|L_1, L_2)$ with list sizes (L_1, L_2) of the AVBBC $\mathfrak{W}^n$ with symmetrizability (T_1, T_2) is*

$$\mathcal{R}_{list}(\mathfrak{W}^n|L_1, L_2) = \mathcal{R}(\overline{\mathfrak{W}}) \qquad \text{if } L_1 > T_1 \text{ and } L_2 > T_2.$$

We have $int(\mathcal{R}_{list}(\mathfrak{W}^n|L_1, L_2)) = \emptyset$ if and only if $L_1 \leq T_1$ or $L_2 \leq T_2$.

The theorem shows that every AVBBC has a characteristic pair of minimum list sizes $(T_1 + 1, T_2 + 1)$ that enables bidirectional communication at all rate pairs $(R_1, R_2) \in \mathcal{R}(\overline{\mathfrak{W}})$. On the other hand, if $L_1 \leq T_1$ or $L_2 \leq T_2$, then there is no reliable communication possible, not even at very low rates.

In addition, from Theorem 4.17 we immediately obtain the deterministic code capacity region $\mathcal{R}_{\text{det}}(\mathfrak{W}^n)$ if we restrict both list sizes to one, i.e., $L_1 = L_2 = 1$.

Corollary 4.18. *For a non-$\mathcal{Y}_1$-symmetrizable[2] and non-$\mathcal{Y}_2$-symmetrizable AVBBC $\mathfrak{W}^n$ the deterministic code capacity region $\mathcal{R}_{det}(\mathfrak{W}^n)$ is given by*

$$\mathcal{R}_{det}(\mathfrak{W}^n) = \mathcal{R}(\overline{\mathfrak{W}}).$$

We have $int(\mathcal{R}_{det}(\mathfrak{W}^n)) = \emptyset$ if and only if the AVBBC $\mathfrak{W}^n$ is $\mathcal{Y}_1$-symmetrizable or $\mathcal{Y}_2$-symmetrizable.

In the following we prove Theorem 4.17. Although the goal is to establish the list capacity region, we first prove the random code capacity region, where we allow the relay and the receivers to coordinate their choice of encoder and decoders. It will be convenient to use this result to establish the desired list capacity region.

4.3 Random Code Construction

In this section, we restrict the list sizes at the receiving nodes to one and derive the optimal random coding strategy for the AVBBC. Thereby, the word "random" refers to the fact that the encoder and decoders are chosen according to a common random experiment whose outcome has to be known at all nodes in advance. This leads directly to the following definition.

[2]Note that according to Remark 4.12, we call a $(\mathcal{Y}_i, 0)$-symmetrizable channel in the sense of Definition 4.10 also a non-$\mathcal{Y}_i$-symmetrizable channel, $i = 1, 2$.

Definition 4.19. *A* random $(n, M_1^{(n)}, M_2^{(n)}, \mathrm{Z})$-code $\mathcal{C}_{ran}(\mathfrak{W}^n)$ *of length* n *for the AVBBC* $\mathfrak{W}^n$ *is given by a family* $\mathcal{C}_{ran}(\mathfrak{W}^n) \coloneqq \{\mathcal{C}(z) : z \in \mathcal{Z}\}$ *of deterministic* $(n, M_1^{(n)}, M_2^{(n)})$*-codes*

$$\mathcal{C}(z) \coloneqq \big\{\big(x_m^n(z), \mathcal{D}_{m_2|m_1}^{(1)}(z), \mathcal{D}_{m_1|m_2}^{(2)}(z)\big) : m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2\big\}$$

together with a random variable $\mathrm{Z} \in \mathcal{Z}$ *distributed according to* $p_{\mathrm{Z}} \in \mathcal{P}(\mathcal{Z})$.

Here, it will be convenient to use the notion of decoding sets to specify the decoding rule as also done for the compound BBC in Chapter 3, cf. Definition 3.4. This means that the decoding sets at nodes 1 and 2 of one deterministic code $\mathcal{C}(z)$, $z \in \mathcal{Z}$, are given by $\mathcal{D}_{m_2|m_1}^{(1)}(z) \subseteq \mathcal{Y}_1^n$ and $\mathcal{D}_{m_1|m_2}^{(2)}(z) \subseteq \mathcal{Y}_2^n$ for all $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$. Since $\mathcal{C}(z)$ is a deterministic code (with list sizes one), the decoding sets must be disjoint. In more detail, for given m_1 at node 1 the decoding sets must satisfy $\mathcal{D}_{m_2|m_1}^{(1)}(z) \cap D_{\hat{m}_2|m_1}^{(1)}(z) = \emptyset$ for $\hat{m}_2 \neq m_2$, and similarly for given m_2 at node 2 the decoding sets must satisfy $\mathcal{D}_{m_1|m_2}^{(2)}(z) \cap D_{\hat{m}_1|m_2}^{(2)}(z) = \emptyset$ for $\hat{m}_1 \neq m_1$.

The average probability of error of the deterministic code $\mathcal{C}(z)$, $z \in \mathcal{Z}$, for state sequence $s^n \in \mathcal{S}^n$ can be written as

$$\bar{e}(s^n|\mathcal{C}(z)) \coloneqq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W^{\otimes n}\big((\mathcal{D}_{m_2|m_1}^{(1)}(z) \times \mathcal{D}_{m_1|m_2}^{(2)}(z))^c | x_m^n(z), s^n\big).$$

Then, the average probability of error of the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for state sequence $s^n \in \mathcal{S}^n$ is given by

$$\bar{e}(s^n|\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)) \coloneqq \mathbb{E}_{\mathrm{Z}}[\bar{e}(s^n|\mathcal{C}(\mathrm{Z}))]$$

and, accordingly, the corresponding marginal average probability of error at node i by $\bar{e}_i(s^n|\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)) \coloneqq \mathbb{E}_{\mathrm{Z}}[\bar{e}_i(s^n|\mathcal{C}(\mathrm{Z}))]$, $i = 1, 2$. For given $0 < \lambda^{(n)} < 1$, $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ is called a $(n, M_1^{(n)}, M_2^{(n)}, \mathrm{Z}, \lambda^{(n)})$-code (with average probability of error $\lambda^{(n)}$) for the AVBBC $\mathfrak{W}^n$ if

$$\bar{e}(s^n|\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)) \leq \lambda^{(n)} \qquad \text{for all } s^n \in \mathcal{S}^n.$$

Then the definitions of a *randomly achievable* rate pair and the *random code capacity region* $\mathcal{R}_{\text{ran}}(\mathfrak{W}^n)$ follow accordingly.

Theorem 4.20. *The random code capacity region* $\mathcal{R}_{ran}(\mathfrak{W}^n)$ *of the AVBBC* $\mathfrak{W}^n$ *is given by*

$$\mathcal{R}_{ran}(\mathfrak{W}^n) = \mathcal{R}(\overline{\mathfrak{W}}),$$

cf. also (4.7).

Remark 4.21. *From the definitions of the codes it is clear that the deterministic code $\mathcal{C}_{det}(\mathfrak{W}^n)$ is a special or degenerated case of the random code $\mathcal{C}_{ran}(\mathfrak{W}^n)$. More precisely, $\mathcal{C}_{det}(\mathfrak{W}^n)$ can be interpreted as a random code that consists of only one deterministic code. Consequently, the deterministic code capacity region $\mathcal{R}_{det}(\mathfrak{W}^n)$ must be contained in the random code capacity region $\mathcal{R}_{ran}(\mathfrak{W}^n)$, i.e., $\mathcal{R}_{det}(\mathfrak{W}^n) \subseteq \mathcal{R}_{ran}(\mathfrak{W}^n)$.*

In the following subsections we give the proof of the random code capacity region which is mainly based on Ahlswede's *robustification technique* [Ahl80b, Ahl86].

4.3.1 Compound Bidirectional Broadcast Channel

The first key idea is to exploit results from the compound BBC, cf. Chapter 3. Therefore, we construct a suitable compound broadcast channel by defining the convex hull of all averaged broadcast channels, cf. (4.2), as

$$\left\{\overline{W}_q(y_1, y_2|x)\right\}_{q\in\mathcal{P}(\mathcal{S})}.$$

We observe that this already corresponds to a compound broadcast channel where each probability distribution $q \in \mathcal{P}(\mathcal{S})$ parametrizes one element of the compound channel which we denote by $\overline{\mathfrak{W}}$. The capacity region of this compound BBC $\overline{\mathfrak{W}}$ is known from previous studies, cf. Theorem 3.13. There, it is shown that the deterministic code capacity region $\mathcal{R}_{\text{det}}(\overline{\mathfrak{W}})$ of the compound BBC $\overline{\mathfrak{W}}$ is given by

$$\mathcal{R}_{\text{det}}(\overline{\mathfrak{W}}) = \mathcal{R}(\overline{\mathfrak{W}}),$$

cf. also (4.7).

The achievability of the rates specified by $\mathcal{R}(\overline{\mathfrak{W}})$, cf. (4.7a) and (4.7b), is proved by showing the existence of a deterministic $(n, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ with arbitrarily small average probability of error. In more detail, in Section 3.3 it is shown that the average probability of error of $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ can be bounded from above by

$$\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W_q^{\otimes n}\big((\mathcal{D}^{(1)}_{m_2|m_1}\times\mathcal{D}^{(2)}_{m_1|m_2})^c|x_m^n\big) \le \lambda^{(n)}_{\overline{\mathfrak{W}}} \qquad \text{for all } q\in\mathcal{P}(\mathcal{S})$$

with $\lambda^{(n)}_{\overline{\mathfrak{W}}} = \lambda^{(n)}_{\overline{\mathfrak{W}},1} + \lambda^{(n)}_{\overline{\mathfrak{W}},2}$ where $\lambda^{(n)}_{\overline{\mathfrak{W}},i}$ is an upper bound on the marginal average probability of error at node i, $i = 1, 2$. More precisely, for n large enough $\lambda^{(n)}_{\overline{\mathfrak{W}},i}$ is given by

$$\lambda^{(n)}_{\overline{\mathfrak{W}},i} = (n+1)^{|\mathcal{X}||\mathcal{Y}_i|}2^{-n\frac{c\epsilon^2}{2}} + \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}}2^{-n\frac{\tau}{8}}$$

which decreases exponentially fast for increasing block length n. Thereby, ϵ, τ, and c are positive constants, cf. also (3.12).

Together with the definition of the averaged broadcast channel (4.2) this immediately implies that for $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ the probability of a successful transmission over the compound BBC $\overline{\mathfrak{W}}$ is bounded from below by

$$\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W_q^{\otimes n}\big(\mathcal{D}^{(1)}_{m_2|m_1}\times\mathcal{D}^{(2)}_{m_1|m_2}|x_m^n\big) > 1-\lambda^{(n)}_{\overline{\mathfrak{W}}}$$

or equivalently by

$$\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}\sum_{s^n\in\mathcal{S}^n} W^{\otimes n}\big(\mathcal{D}^{(1)}_{m_2|m_1}\times\mathcal{D}^{(2)}_{m_1|m_2}|x_m^n,s^n\big)q^{\otimes n}(s^n) > 1-\lambda^{(n)}_{\overline{\mathfrak{W}}} \tag{4.8}$$

for all $q^{\otimes n}=\prod_{k=1}^n q$ and $q\in\mathcal{P}(\mathcal{S})$.

4.3.2 Robustification

Next, we follow [Ahl80b, Ahl86] and use the deterministic code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ to construct a random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$.

Let Π_n be the group of permutations acting on $(1,2,...,n)$. For given sequence $s^n=(s_1,...,s_n)\in\mathcal{S}^n$ and permutation $\pi\in\Pi_n:\mathcal{S}^n\to\mathcal{S}^n$ we denote the permuted sequence $(s_{\pi(1)},...,s_{\pi(n)})\in\mathcal{S}^n$ by $\pi(s^n)$. Further, we denote the inverse permutation by π^{-1} so that $\pi^{-1}\big(\pi(s^n)\big)=s^n$.

Theorem 4.22 (Robustification technique [Ahl86])**.** *Let $f:\mathcal{S}^n\to[0,1]$ be a function such that for some $\alpha\in(0,1)$ the inequality*

$$\sum_{s^n\in\mathcal{S}^n} f(s^n)q^{\otimes n}(s^n) > 1-\alpha \qquad \textit{for all } q\in\mathcal{P}_0(n,\mathcal{S}) \tag{4.9}$$

is satisfied. Then the inequality

$$\frac{1}{n!}\sum_{\pi\in\Pi_n} f\big(\pi(s^n)\big) > 1-(n+1)^{|\mathcal{S}|}\alpha \qquad \textit{for all } s^n\in\mathcal{S}^n$$

is also satisfied.

Since (4.9) is fulfilled with $\alpha=\lambda^{(n)}_{\overline{\mathfrak{W}}}$ by (4.8), from the robustification technique and

$$f\big(\pi(s^n)\big)=\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W^{\otimes n}\big(\mathcal{D}^{(1)}_{m_2|m_1}\times\mathcal{D}^{(2)}_{m_1|m_2}|x_m^n,\pi(s^n)\big)$$

we immediately obtain a random $(n, M_1^{(n)}, M_2^{(n)}, \Pi_n)$-code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ given by the family

$$\mathcal{C}_{\text{ran}}(\mathfrak{W}^n) = \big\{(\pi^{-1}(x_m^n), \pi^{-1}(\mathcal{D}^{(1)}_{m_2|m_1}), \pi^{-1}(\mathcal{D}^{(2)}_{m_1|m_2})) : \\ m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2, \pi \in \Pi_n\big\} \tag{4.10}$$

where the permutations π are uniformly distributed on Π_n and

$$\pi^{-1}(\mathcal{D}^{(1)}_{m_2|m_1}) = \bigcup_{y_1^n \in \mathcal{D}^{(1)}_{m_2|m_1}} \pi^{-1}(y_1^n) \quad \text{and} \quad \pi^{-1}(\mathcal{D}^{(2)}_{m_1|m_2}) = \bigcup_{y_2^n \in \mathcal{D}^{(2)}_{m_1|m_2}} \pi^{-1}(y_2^n).$$

Since Π_n is the group of permutations of size n, the cardinality of Π_n is $n!$ so that the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ consists of $n!$ deterministic $(n, M_1^{(n)}, M_2^{(n)})$-codes.

From the robustification technique follows that the average probability of error of $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ is bounded from above by

$$\bar{e}(s^n|\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)) \leq (n+1)^{|\mathcal{S}|}\lambda^{(n)}_{\overline{\mathfrak{W}}} =: \lambda^{(n)}_{\mathfrak{W},\text{ran}} \qquad \text{for all } s^n \in \mathcal{S}^n. \tag{4.11}$$

The way how we constructed the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ from the deterministic code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ has the following consequence. All rate pairs achievable for the compound BBC $\overline{\mathfrak{W}}$ using the deterministic code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ are also achievable for the AVBBC $\mathfrak{W}^n$ using the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$. Consequently, the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ actually achieves all rate pairs satisfying (4.7a) and (4.7b) as stated in Theorem 4.20, which proves the achievability. □

4.3.3 Converse

It remains to show that the presented random coding strategy actually achieves all possible rate pairs so that no other rate pairs are achievable.

As a first step, it is easy to show that the average probability of error of the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ equals the average probability of error of the random code for the compound BBC $\overline{\mathfrak{W}}$. Hence, it is clear that we cannot achieve higher rates as for the constructed compound BBC $\overline{\mathfrak{W}}$ with random codes. The deterministic rates of the compound channel are given in Theorem 3.13. As in [AW69] for the single-user compound channel, it can easily be shown that for the compound BBC $\overline{\mathfrak{W}}$ the achievable rates for deterministic and random codes are equal. Since the constructed random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ already achieves these rates, the converse is proved.

This finishes the proof of Theorem 4.20 and therewith establishes the random code capacity region $\mathcal{R}_{\text{ran}}(\mathfrak{W}^n)$ of the AVBBC $\mathfrak{W}^n$. □

4.4 Deterministic Code Construction

A random coding strategy as constructed in the previous section requires *common randomness* between all nodes, since the encoder and decoders depend all on the same random permutation, cf. (4.10), which has to be known to all nodes in advance. If this kind of resource is not available, we are interested in deterministic strategies.

One way to ensure that the decoders are chosen according to the same random permutation as the encoder, is to inform the receivers which one is used by the encoder. Consequently, the transmitter has to communicate first the chosen permutation to the receivers and then to transmit the message according to the randomly selected code. If the number of all possible codes could be kept small enough, the transmission of those additional information would not cause an essential loss in rate. Inspired by this idea we establish the following behavior of the list capacity region which is similar to Ahlswede's dichotomy result for the single-user AVC [Ahl78].

Lemma 4.23. *The list capacity region $\mathcal{R}_{list}(\mathfrak{W}^n|L_1, L_2)$ for the AVBBC $\mathfrak{W}^n$ displays the following behavior:*

$$\mathcal{R}_{list}(\mathfrak{W}^n|L_1, L_2) = \mathcal{R}_{ran}(\mathfrak{W}^n) \qquad \text{if } int(\mathcal{R}_{list}(\mathfrak{W}^n|L_1, L_2)) \neq \emptyset. \tag{4.12}$$

In the following two subsections we prove the lemma using Ahlswede's *elimination technique* [Ahl78]. Therefore, we start with a random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ and construct a list code $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ which achieves the same rate pairs as the random code.

4.4.1 Random Code Reduction

The first step of the elimination technique [Ahl78] is the *random code reduction*. Here, we construct a new random code by selecting a relatively small number of deterministic codes from the original random code using the following lemma suitable for the BBC.

Lemma 4.24 (Random Code Reduction)**.** *As given in (4.10) let $\mathcal{C}_{ran}(\mathfrak{W}^n)$ be a random code for the AVBBC $\mathfrak{W}^n$ and let $\lambda^{(n)}_{\mathfrak{W}^n,ran}$ be an upper bound on the average probability of error of this code as specified in (4.11). For any ϵ and K^2 that satisfy*

$$\epsilon > 2\lambda^{(n)}_{\mathfrak{W}^n,ran} \qquad \text{and} \qquad K^2 > \frac{2}{\epsilon}\log(|\mathcal{S}|^n) \tag{4.13}$$

there exist K^2 deterministic codes $\mathcal{C}_{i,j}$, $i = 1, ..., K$, $j = 1, ..., K$ such that

$$\frac{1}{K^2}\sum_{i,j}\bar{e}(s^n|\mathcal{C}_{i,j}) < \epsilon \qquad \text{for all } s^n \in \mathcal{S}^n. \tag{4.14}$$

Proof. A random code reduction for the single-user AVC was first proposed in [Ahl78]. Our proof for the AVBBC is inspired by [CK81, Lemma 6.8] where a similar result for the single-user AVC in terms of maximal probability of error is established.

First, from the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ we select K^2 independent permutations $\pi_{i,j} \in \Pi_n$, $i = 1, ..., K$, $j = 1, ..., K$ according to the uniform distribution. Each such permutation $\pi_{i,j}$ specifies one deterministic code which is denoted by $\mathcal{C}_{i,j}$ in the following. Then, for given state sequence $s^n \in \mathcal{S}^n$, we have

$$\begin{aligned}\mathbb{P}\Big\{\frac{1}{K^2}\sum_{i,j}\bar{e}(s^n|\mathcal{C}_{i,j}) \geq \epsilon\Big\} &= \mathbb{P}\Big\{\exp\big(\sum_{i,j}\bar{e}(s^n|\mathcal{C}_{i,j})\big) \geq \exp\big(K^2\epsilon\big)\Big\} \\ &\leq \exp\big(-K^2\epsilon\big)\mathbb{E}\Big[\exp\big(\sum_{i,j}\bar{e}(s^n|\mathcal{C}_{i,j})\big)\Big] \qquad (4.15)\end{aligned}$$

where the last step follows from Markov's inequality. Since the random variables $\mathcal{C}_{i,j}$, $i = 1, ..., K$, $j = 1, ..., K$ are independent and identically distributed, we get for the expectation

$$\begin{aligned}\mathbb{E}\Big[\exp\big(\sum_{i,j}\bar{e}(s^n|\mathcal{C}_{i,j})\big)\Big] &= \mathbb{E}\Big[\exp\big(\bar{e}(s^n|\mathcal{C}_{1,1})\big)\Big]^{K^2} \\ &\leq \big(1 + \mathbb{E}\big[\bar{e}(s^n|\mathcal{C}_{1,1})\big]\big)^{K^2} \\ &\leq \big(1 + \lambda^{(n)}_{\mathfrak{W}^n,\text{ran}}\big)^{K^2}\end{aligned}$$

where we further used the inequality $\exp(x) \leq 1 + x$, $0 \leq x \leq 1$ (recall that exp is to the basis 2). This and (4.15) yield

$$\begin{aligned}\mathbb{P}\Big\{\frac{1}{K^2}\sum_{i,j}\bar{e}(s^n|\mathcal{C}_{i,j}) \geq \epsilon\Big\} &\leq \exp\big(-K^2\epsilon\big)\big(1 + \lambda^{(n)}_{\mathfrak{W}^n,\text{ran}}\big)^{K^2} \\ &\leq \exp\big(-K^2(\epsilon - \lambda^{(n)}_{\mathfrak{W}^n,\text{ran}})\big).\end{aligned}$$

Finally, with (4.13) this implies

$$\mathbb{P}\Big\{\frac{1}{K^2}\sum_{i,j}\bar{e}(s^n|\mathcal{C}_{i,j}) \geq \epsilon \quad \text{for all } s^n \in \mathcal{S}^n\Big\} \leq |\mathcal{S}|^n \exp\big(-K^2(\epsilon - \lambda^{(n)}_{\mathfrak{W}^n,\text{ran}})\big). \qquad (4.16)$$

This means that there exists a collection of deterministic codes $\mathcal{C}_{i,j}$, $i = 1, ..., K$, $j = 1, ..., K$ which satisfy (4.14) proving the lemma.

Note that we split up the K^2 chosen deterministic codes into two groups to be conform with the definition of a deterministic code for the AVBBC, cf. Definition 4.13 and especially Remark 4.16. □

The random code reduction shows that for any random code there exists another "reduced" random code which is uniformly distributed over K^2 deterministic codes with an average probability of error less than ϵ under the assumption that (4.13) holds.

A direct consequence of Lemma 4.24 is that for any random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ which achieves the random code capacity of the AVBBC $\mathfrak{W}^n$, there exists another "reduced" random code $\widetilde{\mathcal{C}}_{\text{ran}}(\mathfrak{W}^n)$ which does likewise. Furthermore, from [Ahl78] we know that it is sufficient to select no more than $K^2 = n^2$ deterministic codes to obtain $\widetilde{\mathcal{C}}_{\text{ran}}(\mathfrak{W}^n)$ with the desired properties. This can easily be seen in (4.16) where the choice $K^2 = n^2$ leads to a code whose probability of error exceeds ϵ with a super exponentially small probability since $|\mathcal{S}|^n$ grows exponentially in n only.

In more detail, for any $\epsilon > 0$ and sufficiently large n there exist n^2 deterministic codes

$$\mathcal{C}_{i,j} \coloneqq \left\{ \left(\pi_{i,j}^{-1}(x_m^n), \pi_{i,j}^{-1}(\mathcal{D}_{m_2|m_1}^{(1)}), \pi_{i,j}^{-1}(\mathcal{D}_{m_1|m_2}^{(2)}) \right) : m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2 \right\} \in \mathcal{C}_{\text{ran}}(\mathfrak{W}^n),$$

$i = 1, ..., n$, $j = 1, ..., n$, such that

$$\bar{e}(s^n | \mathcal{C}_{\text{ran}}(\mathfrak{W}^n)) = \frac{1}{n^2} \sum_{i,j} \bar{e}(s^n | \mathcal{C}_{i,j}) < \epsilon =: \tilde{\lambda}^{(n)}_{\mathfrak{W}^n,\text{ran}} \qquad \text{for all } s^n \in \mathcal{S}^n. \tag{4.17}$$

The "reduced" random code $\widetilde{\mathcal{C}}_{\text{ran}}(\mathfrak{W}^n)$ with "exponentially few" elements is given by

$$\widetilde{\mathcal{C}}_{\text{ran}}(\mathfrak{W}^n) \coloneqq \left\{ \mathcal{C}_{i,j} : i = 1, ..., n; j = 1, ..., n \right\}$$

where the indices i, j are drawn according to the uniform distribution on $\{1, ..., n\} \times \{1, ..., n\}$. Clearly, the "reduced" random code $\widetilde{\mathcal{C}}_{\text{ran}}(\mathfrak{W}^n)$ also achieves the random code capacity of the AVBBC $\mathfrak{W}^n$.

4.4.2 Elimination of Randomness

Up to now we have constructed a random code with "exponentially few" elements that achieves the random code capacity of the AVBBC $\mathfrak{W}^n$. The next step of the elimination technique [Ahl78] is the *elimination of randomness*. This means that we convert the "reduced" random code into a list code by adding short prefixes to the original codewords to inform the decoders which of the n^2 deterministic codes is actually used [Ahl78, CK81].

Clearly, this is only possible, if the list capacity region $\mathcal{R}_{\text{list}}(\mathfrak{W}^n | L_1, L_2)$ fulfills $\text{int}(\mathcal{R}_{\text{list}}(\mathfrak{W}^n | L_1, L_2)) \neq \emptyset$, which means that transmission at positive rates is possible in both directions. Then, there exists for sufficiently large n a sequence of list codes $\mathcal{C}_{\text{pre}}$ with sequences

$$x_{i,j}^{l_n} \in \mathcal{X}^{l_n}$$

of length l_n and list decoders at nodes 1 and 2

$$\mathcal{L}^{(1)} : \mathcal{Y}_1^{l_n} \times \{1, ..., n\} \to \hat{\mathfrak{P}}_{L_1}(\{1, ..., n\})$$
$$\mathcal{L}^{(2)} : \mathcal{Y}_2^{l_n} \times \{1, ..., n\} \to \hat{\mathfrak{P}}_{L_2}(\{1, ..., n\})$$

for all $i, j = 1, ..., n$, cf. also Definition 4.13. Further, the average probability of error may be

$$\bar{e}(s^n|\mathcal{C}_{\text{pre}}) = \frac{1}{n^2} \sum_{i,j} \sum_{\substack{(y_1^{l_n}, y_2^{l_n}): j \notin \mathcal{L}^{(1)}(y_1^{l_n}, i) \\ \vee i \notin \mathcal{L}^{(2)}(y_2^{l_n}, j)}} W^{\otimes l_n}\left(y_1^{l_n}, y_2^{l_n} | x_{i,j}^{l_n}, s^{l_n}\right)$$
$$\leq \epsilon =: \lambda_{\text{pre}}^{(l_n)} \qquad \text{for all } s^{l_n} \in \mathcal{S}^{l_n} \tag{4.18}$$

where $\frac{l_n}{n} \to 0$ as $n \to \infty$. This code is used to specify which code will be used in the following.

Next, we define the final list code $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ by concatenating $\mathcal{C}_{\text{pre}}$ with $\widetilde{\mathcal{C}}_{\text{ran}}(\mathfrak{W}^n)$. The new code has a block length $l_n + n$ and the message set $\{1, ..., n\} \times \{1, ..., n\} \times \mathcal{M}_1 \times \mathcal{M}_2$. The transmit sequence is a juxtaposition of the prefix codeword $x_{i,j}^{l_n}$ and the codeword $\pi_{i,j}^{-1}(x_m^n)$ where the former determines the code $\mathcal{C}_{i,j}$ used for the following message.

From (4.17) and (4.18) follows that the average probability of error of $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ for given state sequence $s^{l_n+n} \in \mathcal{S}^{l_n} \times \mathcal{S}^n$ can be bounded from above as

$$\bar{e}(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) < \lambda_{\text{pre}}^{(l_n)} + \tilde{\lambda}_{\mathfrak{W}^n,\text{ran}}^{(n)} = 2\epsilon.$$

Moreover, the rate of the final list code $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ from the relay to node 1 is given by

$$\frac{1}{l_n + n} \log\left(\frac{nM_2^{(n)}}{L_1}\right) = \frac{1}{l_n + n}\left(\log M_2^{(n)} + \log\left(\frac{n}{L_1}\right)\right)$$
$$= \frac{1}{\frac{l_n}{n} + 1} \frac{1}{n} \log M_2^{(n)} + \frac{1}{1 + \frac{n}{l_n}} \frac{1}{l_n} \log\left(\frac{n}{L_1}\right) \underset{n\to\infty}{\longrightarrow} R_1$$

since $\frac{1}{\frac{l_n}{n}+1} \to 1$ and $\frac{1}{1+\frac{n}{l_n}} \to 0$ as $n \to \infty$ and $\frac{1}{n} \log M_2^{(n)} = R_2$. Similarly, we get $\frac{1}{l_n+n} \log(\frac{nM_1^{(n)}}{L_2}) \to R_2$ for the rate from the relay to node 2. This shows that the overall rate of the final, concatenated list code is only negligibly affected by the addition of the prefixes.

Consequently, all rate pairs achievable with the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ are also achievable with the list code $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ with arbitrarily small average probability of error if $\text{int}(\mathcal{R}_{\text{list}}(\mathfrak{W}^n|L_1, L_2)) \neq \emptyset$ as stated in (4.12) which proves Lemma 4.23. □

Remark 4.25. *Due to the concatenated structure of $\mathcal{C}_{list}(\mathfrak{W}^n)$ this code is a special case of a list code for the AVBBC $\mathfrak{W}^n$, cf. Definition 4.13, and consequently, $\mathcal{C}_{list}(\mathfrak{W}^n)$ might not achieve the maximal achievable rates. But the converse in Section 4.5.2 shows that $\mathcal{C}_{list}(\mathfrak{W}^n)$ actually achieves all possible rate pairs so that this concatenated structure is already optimal.*

4.5 List Decoding

Although Lemma 4.23 characterizes the general behavior of the list capacity region of the AVBBC, it does not specify in detail, when the list capacity region has an empty interior. Therefore we fill this hiatus in the following.

4.5.1 Symmetrizability

Already Blackwell, Breiman, and Thomasian observed that under certain conditions the deterministic code capacity of the single-user AVC is zero [BBT60]. Based on an idea of Ericson [Eri85], Csiszár and Narayan showed that non-symmetrizability is a necessary condition for the single-user AVC to have a non-zero capacity [CN88b]. Independently, Blinovsky, Narayan, and Pinsker [BNP95] and Hughes [Hug97] extended this idea to the case of list decoding.

Here, we want to establish similar results for the AVBBC. For this purpose we use the concept of symmetrizability as introduced in Section 4.2 and define the maximum single-user rates as

$$R_{i,\max} \coloneqq \max_{P_{\mathrm{X}}} \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{i,q}), \qquad i = 1, 2.$$

The following result relates the symmetrizability and the maximum single-user rates.

Theorem 4.26. *If $R_{i,max} = 0$, then the AVBBC $\mathfrak{W}^n$ is $(\mathcal{Y}_i, t_i)$-symmetrizable for all $t_i \geq 1$, $i = 1, 2$. If $R_{i,max} > 0$, then any $(\mathcal{Y}_i, t_i)$-symmetrizable AVBBC $\mathfrak{W}^n$ satisfies*

$$t_i \leq \frac{\log(\min\{|\mathcal{Y}_i|, |\mathcal{S}|\})}{R_{i,max}}. \tag{4.19}$$

Proof. The proof can be found in Appendix A.1. □

From Theorem 4.26 follows that for any AVBBC, whose random code capacity region has a non-empty interior, the symmetrizability is always finite. The next lemma presents a lower bound on the average probability of error in a similar way as in [Hug97] for the single-user case.

Lemma 4.27. *Let (T_1, T_2) be the symmetrizability of an AVBBC $\mathfrak{W}^n$. Then any list code $\mathcal{C}_{list}(\mathfrak{W}^n)$ of block length n with $M_1^{(n)} M_2^{(n)}$ messages and $L_1 \leq T_1$ satisfies*

$$\max_{s^n \in \mathcal{S}^n} \bar{e}_1(s^n | \mathcal{C}_{list}(\mathfrak{W}^n)) \geq \left(1 - \frac{L_1}{K_1 + 1}\right) \left(\frac{M_2^{(n)} - K_1}{M_2^{(n)}}\right)$$

where $K_1 = \min\{M_2^{(n)} - 1, T_1\}$. Similarly, any list code $\mathcal{C}_{list}(\mathfrak{W}^n)$ of block length n with $M_1^{(n)} M_2^{(n)}$ messages and $L_2 \leq T_2$ satisfies

$$\max_{s^n \in \mathcal{S}^n} \bar{e}_2(s^n | \mathcal{C}_{list}(\mathfrak{W}^n)) \geq \left(1 - \frac{L_2}{K_2 + 1}\right) \left(\frac{M_1^{(n)} - K_2}{M_1^{(n)}}\right)$$

where $K_2 = \min\{M_1^{(n)} - 1, T_2\}$.

Proof. The proof can be found in Appendix A.2. □

The lemma indicates when the interior of the list capacity region of the AVBBC $\mathfrak{W}^n$ will be empty. In more detail, if $L_i \leq T_i$, $i = 1, 2$, then $\max_{s^n \in \mathcal{S}^n} \bar{e}_i(s^n | \mathcal{C}_{\text{list}}(\mathfrak{W}^n)) > 0$, $i = 1, 2$, which results in $\text{int}(\mathcal{R}_{\text{list}}(\mathfrak{W}^n | L_1, L_2)) = \emptyset$. Consequently, $L_i > T_i$, $i = 1, 2$ is a necessary condition for $\mathcal{R}_{\text{list}}(\mathfrak{W}^n | L_1, L_2) = \mathcal{R}_{\text{ran}}(\mathfrak{W}^n)$. In other words, for fixed list sizes (L_1, L_2), non-$(\mathcal{Y}_1, L_1)$-symmetrizability and non-$(\mathcal{Y}_2, L_2)$-symmetrizability is necessary for $\mathcal{R}_{\text{list}}(\mathfrak{W}^n | L_1, L_2) = \mathcal{R}_{\text{ran}}(\mathfrak{W}^n)$.

4.5.2 Achieving Positive Rates

In this subsection, we present a coding strategy that achieves the desired rates as specified in Theorem 4.17 if $L_i > T_i$, $i = 1, 2$. Moreover, this immediately shows that $L_i > T_i$, $i = 1, 2$, is also a sufficient condition for $\text{int}(\mathcal{R}_{\text{list}}(\mathfrak{W}^n | L_1, L_2)) \neq \emptyset$. The coding strategy in the following is based on [Hug97] where a similar strategy is presented for the single-user case.

Coding Strategy

To achieve positive rates we need a suitable set of codewords $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, with properties as stated in the following lemma.

Lemma 4.28. *For any $L_1 \geq 1, L_2 \geq 1$, $\epsilon > 0$, $n \geq \max\{n_0(\epsilon, L_1), n_0(\epsilon, L_2)\}$, $M_1^{(n)} \geq L_2 \exp(n\epsilon)$, $M_2^{(n)} \geq L_1 \exp(n\epsilon)$, and given type P_{X}, there exist codewords $x^n_{m_1,m_2} \in \mathcal{X}^n$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, each of type P_{X}, such that for every $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n$, and every joint type $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}}$ with $\mathrm{X}^{L_1} = (\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_{L_1})$ we have for each $m_1 \in \mathcal{M}_1$*

$$\big|\big\{\hat{m}_2 : (x^n, x^n_{m_1,\hat{m}_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}\big\}\big| \leq \exp\big(n(|R_1 - I(\mathrm{X}_k; \mathrm{X}, \mathrm{S})|^+ + \epsilon)\big) \tag{4.20a}$$

$$\frac{1}{M_2^{(n)}}\big|\big\{m_2 : (x^n_{m_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{S}}\big\}\big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big) \qquad \text{if } I(\mathrm{X};\mathrm{S}) \geq \epsilon \tag{4.20b}$$

$$\frac{1}{M_2^{(n)}}\big|\big\{m_2 : (x^n_{m_1,m_2}, x^n_{m_1,\hat{m}_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}} \text{ for some } \hat{m}_2 \neq m_2\big\}\big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big)$$
$$\text{if } I(\mathrm{X}; \mathrm{X}_k, \mathrm{S}) - |R_1 - I(\mathrm{X}_k;\mathrm{S})|^+ \geq \epsilon \tag{4.20c}$$

for $k = 1, ..., L_1$. Moreover, if $R_1 < \min_k I(\mathrm{X}_k;\mathrm{S})$, then $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, further satisfy

$$\big|\big\{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2) : (x^n, x^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}}\big\}\big| \leq \exp(n\epsilon) \tag{4.20d}$$

$$\frac{1}{M_2^{(n)}}\big|\big\{m_2 : (x^n_{m_1,m_2}, x^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}} \text{ for some}$$
$$\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{m_2\})\big\}\big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big) \qquad \text{if } I(\mathrm{X}; \mathrm{X}^{L_1}, \mathrm{S}) \geq \epsilon \tag{4.20e}$$

with $\mathcal{J} = \{j_1, ..., j_{L_1}\} \in \mathfrak{P}_{L_1}(\mathcal{M}_2)$ and $x^n_{m_1,\mathcal{J}}$ denotes the ordered L_1-tuple $(x^n_{m_1,j_1}, x^n_{m_1,j_2}, ..., x^n_{m_1,j_{L_1}})$ where the indices are ordered as $j_1 < j_2 < ... < j_{L_1}$. Similarly, for every $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n$, and every joint type $P_{\mathrm{X}\mathrm{X}^{L_2}\mathrm{S}}$ we have for each $m_2 \in \mathcal{M}_2$

$$\big|\big\{\hat{m}_1 : (x^n, x^n_{\hat{m}_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}\big\}\big| \leq \exp\big(n(|R_2 - I(\mathrm{X}_k; \mathrm{X}, \mathrm{S})|^+ + \epsilon)\big) \tag{4.20f}$$

$$\frac{1}{M_1^{(n)}}\big|\big\{m_1 : (x^n_{m_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{S}}\big\}\big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big) \quad \text{if } I(\mathrm{X};\mathrm{S}) \geq \epsilon \tag{4.20g}$$

$$\frac{1}{M_1^{(n)}}\big|\big\{m_1 : (x^n_{m_1,m_2}, x^n_{\hat{m}_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}} \text{ for some } \hat{m}_1 \neq m_1\big\}\big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big)$$
$$\text{if } I(\mathrm{X}; \mathrm{X}_k, \mathrm{S}) - |R_2 - I(\mathrm{X}_k;\mathrm{S})|^+ \geq \epsilon \tag{4.20h}$$

for $k = 1, ..., L_2$. Moreover, if $R_2 < \min_k I(\mathrm{X}_k;\mathrm{S})$, then $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, further satisfy

$$\big|\big\{\mathcal{J}' \in \mathfrak{P}_{L_2}(\mathcal{M}_1) : (x^n, x^n_{\mathcal{J}',m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_2}\mathrm{S}}\big\}\big| \leq \exp(n\epsilon) \tag{4.20i}$$

$$\frac{1}{M_1^{(n)}}\big|\big\{m_1 : (x^n_{m_1,m_2}, x^n_{\mathcal{J}',m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_2}\mathrm{S}} \text{ for some}$$

$$\mathcal{J}' \in \mathfrak{P}_{L_2}(\mathcal{M}_1 \backslash \{m_1\})\} \big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big) \qquad \text{if } I(\mathrm{X};\mathrm{X}^{L_2},\mathrm{S}) \geq \epsilon \tag{4.20j}$$

with $\mathcal{J}' = \{j'_1, ..., j'_{L_2}\} \in \mathfrak{P}_{L_2}(\mathcal{M}_1)$ *and* $x^n_{\mathcal{J}',m_2}$ *denotes the ordered* L_2*-tuple* $(x^n_{j'_1,m_2}, x^n_{j'_2,m_2}, ..., x^n_{j'_{L_2},m_2})$ *where the indices are ordered as* $j'_1 < j'_2 < ... < j'_{L_2}$.

Proof. The proof can be found in Appendix A.3. □

The proof of the lemma shows that good codewords are obtained by randomly selecting codewords from the set of sequences of a fixed type. All such codewords will possess the desired properties with probability arbitrarily close to 1.

Decoding Strategy

A crucial part is to define suitable decoding rules at the receiving nodes 1 and 2. For the single-user AVC with list size one Csiszár and Narayan use in [CN88b] a generalized divergence typicality decoder based on an idea of Dobrushin and Stambler [DS75] which decides on the basis of a joint typicality test together with a threshold test using empirical mutual information quantities. Blinovsky et al. [BNP95] and Hughes [Hug97] use a generalization of the above mentioned decoder that is modified in such a way that it also applies to greater list sizes. We follow their approach and define for this purpose a family of joint distributions P_{XSY_i} of random variables X, S, and Y_i with values in $\mathcal{X}$, $\mathcal{S}$, and $\mathcal{Y}_i$, respectively, by

$$\mathcal{D}_{\eta_i} \coloneqq \big\{P_{\mathrm{XSY}_i} : D(P_{\mathrm{XSY}_i} \| P_{\mathrm{X}} \otimes P_{\mathrm{S}} \otimes W_i) \leq \eta_i\big\}, \qquad i = 1, 2,$$

with $\eta_i \geq 0$ and where $P_{\mathrm{X}} \otimes P_{\mathrm{S}} \otimes W_i$ denotes a joint distribution on $\mathcal{X} \times \mathcal{S} \times \mathcal{Y}_i$ with probability mass function $P_{\mathrm{X}}(x)P_{\mathrm{S}}(s)W_i(y_i|x,s)$. In particular, we have $P_{\mathrm{XSY}_i} \in \mathcal{D}_0$ if and only if

$$P_{\mathrm{XSY}_i}(x,s,y_i) = P_{\mathrm{X}}(x)P_{\mathrm{S}}(s)W_i(y_i|x,s).$$

Therewith we are able to define the decoding rule at node 1 for list size L_1 as follows.

Definition 4.29. *For given codewords* $x^n_{m_1,m_2} \in \mathcal{T}^{(n)}_X$, $m_1 = 1, ..., M^{(n)}_1$, $m_2 = 1, ..., M^{(n)}_2$, *and (small)* $\eta_1 > 0$ *the decoding rule* $\mathcal{L}^{(1)} : \mathcal{Y}^n_1 \times \mathcal{M}_1 \to \hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$ *at node 1 is defined as follows: we have* $m_2 \in \mathcal{L}^{(1)}(y^n_1, m_1)$ *if and only if*

i) there exists an $s^n \in \mathcal{S}^n$ *such that*

$$P_{x^n_{m_1,m_2},s^n,y^n_1} \in \mathcal{D}_{\eta_1}$$

ii) for each choice of L_1 other distinct codewords $x^n_{m_1,j_1}, ..., x^n_{m_1,j_{L_1}}$, where each satisfies

$$P_{x^n_{m_1,j_i}, s^n_i, y^n_1} \in \mathcal{D}_{\eta_1} \qquad 1 \leq i \leq L_1$$

for some $s^n_i \in \mathcal{S}^n$, we have

$$I(\mathrm{X}, \mathrm{Y}_1; \mathrm{X}^{L_1} | \mathrm{S}) \leq \eta_1$$

where $\mathrm{X}^{L_1} = (\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_{L_1})$ and $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}$ is the joint type of $(x^n_{m_1,m_2}, x^n_{m_1,j_1}, ..., x^n_{m_1,j_{L_1}}, s^n, y^n_1)$.

The decoding rule $\mathcal{L}^{(2)} : \mathcal{Y}_2^n \times \mathcal{M}_2 \to \hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$ at node 2 with list size L_2 is defined accordingly with (small) constant $\eta_2 > 0$. To establish the list capacity region for $L_i > T_i$, $i = 1, 2$ (cf. Theorem 4.17), we have to ensure that the decoding rule as specified in Definition 4.29 is well defined. This means that the decoding rule satisfies the given constraints on the list sizes, i.e., $|\mathcal{L}^{(1)}(y^n_1, m_1)| \leq L_1$ for all $m_1 \in \mathcal{M}_1$ and $|\mathcal{L}^{(2)}(y^n_2, m_2)| \leq L_2$ for all $m_2 \in \mathcal{M}_2$. We show that the decoding rule already satisfies $|\mathcal{L}^{(i)}(y^n_i, m_i)| \leq T_i + 1$ for all $y^n_i \in \mathcal{Y}^n_i$ and $m_i \in \mathcal{M}_i$, $i = 1, 2$, which is clearly sufficient. Here is where the symmetrizability conditions come in.

Lemma 4.30. *Let $\beta > 0$, then for a sufficiently small η_i, $i = 1, 2$, no ensemble $(\mathrm{X}^{T_i+2}, \mathrm{S}^{T_i+2}, \mathrm{Y}_i)$ can simultaneously satisfy*

$$\min_x P_{\mathrm{X}}(x) \geq \beta$$

and

$$\begin{aligned} &P_{\mathrm{X}_k} = P, \qquad P_{\mathrm{X}_k \mathrm{S}_k \mathrm{Y}_i} \in \mathcal{D}_{\eta_i} \\ &I(\mathrm{X}_k, \mathrm{Y}_i; \mathrm{X}_k^{T_i+2} | \mathrm{S}_k) \leq \eta_i \qquad 1 \leq k \leq T_i + 2 \end{aligned} \tag{4.21}$$

with $\mathrm{X}_k^{T_i+2} = (\mathrm{X}_1, ..., \mathrm{X}_{k-1}, \mathrm{X}_{k+1}, ..., \mathrm{X}_{T_i+2})$.

Proof. The proof can be found in Appendix A.4. □

Positive Rates

So far we defined coding and decoding rules. Next, we show that this strategy is sufficient to achieve the desired rates if the list sizes are great enough, i.e., $L_i > T_i$, $i = 1, 2$. Clearly, it suffices to show this for $L_k = T_k + 1$, since any rate pair achievable with these list sizes is also achievable with greater list sizes.

Lemma 4.31. *Let $L_i = T_i + 1$, $i = 1, 2$, and $\beta > 0$, $\delta > 0$. For any type P_{X} satisfying $\min_x P_{\mathrm{X}}(x) \geq \beta$, there exists a list code $\mathcal{C}_{list}(\mathfrak{W}^n)$ of block length $n \geq n_2$ with list sizes (L_1, L_2) and codewords $x^n_{m_1,m_2} \in \mathcal{T}^{(n)}_{\mathrm{X}}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, such that*

$$\frac{1}{n}\log\left(\frac{M_2^{(n)}}{L_1}\right) > \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{1,q}) - \delta, \quad \frac{1}{n}\log\left(\frac{M_1^{(n)}}{L_2}\right) > \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{2,q}) - \delta$$

while

$$\max_{s^n\in\mathcal{S}^n} \bar{e}_i(s^n|\mathcal{C}_{list}(\mathfrak{W}^n)) < 2^{-n\gamma_i}, \qquad i = 1, 2 \tag{4.22}$$

where n_2 and $\gamma_i > 0$ depend only on β, δ, and the AVBBC $\mathfrak{W}^n$.

Proof. The proof follows [Hug97, Lemma 3] where a similar result is shown for the single-user AVC.

Let $x^n_{m_1,m_2} \in \mathcal{T}^{(n)}_{\mathrm{X}}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, be codewords with properties as specified in Lemma 4.28 (ϵ will be chosen later) and $R_1 = \frac{1}{n}\log(\frac{M_2^{(n)}}{L_1})$ and $R_2 = \frac{1}{n}\log(\frac{M_1^{(n)}}{L_2})$ satisfying

$$\inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{1,q}) - \delta < R_1 < \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{1,q}) - \frac{2}{3}\delta \tag{4.23a}$$

$$\inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{2,q}) - \delta < R_2 < \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{2,q}) - \frac{2}{3}\delta. \tag{4.23b}$$

Let the list decoders $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ be as given in Definition 4.29. By Lemma 4.30 we can choose η_1 and η_2 small enough to ensure that $|\mathcal{L}^{(i)}(y_i^n, m_i)| \leq T_i + 1$ for all $y_i^n \in \mathcal{Y}_i^n$ and $m_i \in \mathcal{M}_i$, $i = 1, 2$.

Furthermore, $I(\mathrm{X}; \mathrm{Y}_i)$ is uniformly continuous in P_{XY_i} and divergence dominates the variational distance [CK81, p. 58] so that we can choose η_i small enough to ensure that $P_{\mathrm{XSY}_i} \in \mathcal{D}_{\eta_i}$, $i = 1, 2$, which implies

$$I(\mathrm{X}; \mathrm{Y}_i) \geq \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{i,q}) - \frac{\delta}{3}. \tag{4.24}$$

In the following we carry out the analysis for the probability of error at node 1. Then the analysis for node 2 follows accordingly using the same arguments. We establish an exponentially decreasing upper bound on the probability of error as postulated in (4.22) for node 1 for a fixed state sequence $s^n \in \mathcal{S}^n$.

For each $m_1 \in \mathcal{M}_1$ we first observe from Definition 4.29 that y_1^n is erroneously decoded when message $m = (m_1, m_2)$ is sent and $m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)$. This means that decoding

rule i) or decoding rule ii) must be violated. Consequently, we have $P_{x_m^n,s^n,y_1^n} \notin \mathcal{D}_{\eta_1}$ or there exists a joint type $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}$ with $(x_{m_1,m_2}^n, x_{m_1,\mathcal{J}}^n, s^n, y_1^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}$ for some $\mathcal{J} \in \mathfrak{P}_{L_1}(M_2^{(n)}\backslash\{m_2\})$ such that a) $P_{\mathrm{S}\mathrm{Y}_1} \in \mathcal{D}_{\eta_1}$, b) $P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_1} \in \mathcal{D}_{\eta_1}$ for some S_k, $1 \leq k \leq L_1$, and c) $I(\mathrm{X}\mathrm{Y}_1;\mathrm{X}^{L_1}|\mathrm{S}) > \eta_1$. Let $\mathcal{E}_{\eta_1}$ denote the set of all types $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}$ that satisfy the conditions a)–c). Consequently, we can bound the probability of error for message $m = (m_1, m_2)$ and state sequence $s^n \in \mathcal{S}^n$ as follows

$$\begin{aligned} e_1(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) &= \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} W_1^{\otimes n}(y_1^n|x_m^n, s^n) \\ &\leq \sum_{y_1^n : P_{x_m^n,s^n,y_1^n} \notin \mathcal{D}_{\eta_1}} W_1^{\otimes n}(y_1^n|x_m^n, s^n) + \sum_{P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1} \in \mathcal{E}_{\eta_1}} e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \end{aligned} \tag{4.25}$$

with

$$e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \coloneqq \sum_{\substack{y_1^n : (x_{m_1,m_2}^n, x_{m_1,\mathcal{J}}^n, s^n, y_1^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1} \\ \text{for some } \mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2\backslash\{m_2\})}} W_1^{\otimes n}(y_1^n|x_{m_1,m_2}^n, s^n). \tag{4.26}$$

Next, we define the set

$$\mathcal{A}_{m_1} \coloneqq \left\{ m_2 : I(\mathrm{X};\mathrm{S}) < \epsilon \text{ where } P_{\mathrm{X}\mathrm{S}} = P_{x_{m_1,m_2}^n, s^n} \right\}$$

and use the trivial bound $e_1((m_1, m_2), s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \leq 1$ for all $m_2 \in (\mathcal{A}_{m_1})^c$. With this and (4.25) we get for the average probability of error

$$\begin{aligned} \bar{e}_1(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \leq{}& \frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} |(\mathcal{A}_{m_1})^c| \\ &+ \frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} \sum_{m_2 \in \mathcal{A}_{m_1}} \sum_{y_1^n : P_{x_m^n,s^n,y_1^n} \notin \mathcal{D}_{\eta_1}} W_1^{\otimes n}(y_1^n|x_m^n, s^n) \\ &+ \frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} \sum_{m_2 \in \mathcal{A}_{m_1}} \sum_{P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1} \in \mathcal{E}_{\eta_1}} e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)). \end{aligned} \tag{4.27}$$

Property (4.20b) of Lemma 4.28 and Lemma B.5, cf. Appendix B.1, imply

$$\frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} |(\mathcal{A}_{m_1})^c| \leq (n+1)^{|\mathcal{X}||\mathcal{S}|} \exp\Big(- n\frac{\epsilon}{2}\Big) \leq \exp\Big(- n\frac{\epsilon}{3}\Big) \tag{4.28}$$

for the first term, where the last inequality holds for sufficiently large n.

To bound the second term we observe that for any $m_2 \in \mathcal{A}_{m_1}$

$$\begin{aligned}\sum_{y_1^n : P_{x_m^n, s^n, y_1^n} \notin \mathcal{D}_{\eta_1}} W_1^{\otimes n}(y_1^n | x_m^n, s^n) &\leq \sum_{P_{\mathrm{XSY}_1} \notin \mathcal{D}_{\eta_1}} W_1^{\otimes n}(\mathcal{T}^{(n)}_{\mathrm{Y}_1|\mathrm{XS}}(x_m^n, s^n) | x_m^n, s^n) \\ &\leq \sum_{P_{\mathrm{XSY}_1} \notin \mathcal{D}_{\eta_1}} \exp\big(- nD(P_{\mathrm{XSY}_1} \| P_{\mathrm{XS}} \otimes W_1)\big) \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{S}||\mathcal{Y}_1|} \exp\big(- n(\eta_1 - \epsilon)\big) \\ &\leq \exp\big(- n(\eta_1 - 2\epsilon)\big) \end{aligned} \tag{4.29}$$

where the second inequality follows from Lemma B.7 and the third inequality from Lemma B.5 and

$$\begin{aligned} D(P_{\mathrm{XSY}_1} \| P_{\mathrm{XS}} \otimes W_1) &= D(P_{\mathrm{XSY}_1} \| P_{\mathrm{X}} \otimes P_{\mathrm{S}} \otimes W_1) - I(\mathrm{X};\mathrm{S}) \\ &> \eta_1 - \epsilon. \end{aligned}$$

It remains to bound for $P_{\mathrm{XX}^{L_1}\mathrm{SY}_1} \in \mathcal{E}_{\eta_1}$ the term

$$\frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} \sum_{m_2 \in \mathcal{A}_{m_1}} e_{\mathrm{XX}^{L_1}\mathrm{SY}_1}(m, s^n | \mathcal{C}_{\text{list}}(\mathfrak{W}^n)). \tag{4.30}$$

To this end, we consider two cases, i.e., $R_1 < \min_k I(\mathrm{X}_k;\mathrm{S})$ and $R_1 \geq \min_k I(\mathrm{X}_k;\mathrm{S})$.

Case 1: $R_1 < \min_k I(\mathrm{X}_k;\mathrm{S})$. We note that it suffices to bound $e_{\mathrm{XX}^{L_1}\mathrm{SY}_1}(m, s^n | \mathcal{C}_{\text{list}}(\mathfrak{W}^n))$ for $P_{\mathrm{XX}^{L_1}\mathrm{SY}_1}$ satisfying

$$I(\mathrm{X}; \mathrm{X}^{L_1}\mathrm{S}) < \epsilon \tag{4.31}$$

since otherwise property (4.20e) shows that (4.30) is bounded by

$$\frac{1}{M_2^{(n)}} \big| \big\{ m_2 : (x^n_{m_1,m_2}, x^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XX}^{L_1}\mathrm{S}} \text{ for some } \mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{m_2\}) \big\} \big| < \exp\Big(- n \frac{\epsilon}{2} \Big). \tag{4.32}$$

We see from (4.26) that we can bound

$$\begin{aligned} & e_{\mathrm{XX}^{L_1}\mathrm{SY}_1}(m, s^n | \mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \\ &\quad \leq \sum_{\substack{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{m_2\}): \\ (x^n_{m_1,m_2}, x^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XX}^{L_1}\mathrm{S}}}} W_1^{\otimes n}(\mathcal{T}^{(n)}_{\mathrm{Y}|\mathrm{XX}^{L_1}\mathrm{S}}(x^n_{m_1,m_2}, x^n_{m_1,\mathcal{J}}, s^n) | x^n_{m_1,m_2}, s^n) \\ &\quad \leq \sum_{\substack{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{m_2\}): \\ (x^n_{m_1,m_2}, x^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XX}^{L_1}\mathrm{S}}}} \exp\big(- nI(\mathrm{Y}_1; \mathrm{X}^{L_1} | \mathrm{X}, \mathrm{S})\big) \\ &\quad \leq \exp\big(- n(I(\mathrm{Y}_1; \mathrm{X}^{L_1} | \mathrm{X}, \mathrm{S}) - \epsilon)\big) \end{aligned}$$

where the second inequality follows from Lemma B.7 and the last inequality from the property (4.20d) of the codewords. From (4.31) we obtain

$$I(\mathrm{X};\mathrm{X}^{L_1}|\mathrm{S}) \leq I(\mathrm{X};\mathrm{X}^{L_1},\mathrm{S}) < \epsilon$$

and hence

$$I(\mathrm{Y}_1;\mathrm{X}^{L_1}|\mathrm{X},\mathrm{S}) = I(\mathrm{X},\mathrm{Y}_1;\mathrm{X}^{L_1}|\mathrm{S}) - I(\mathrm{X};\mathrm{X}^{L_1}|\mathrm{S}) > \eta_1 - \epsilon$$

from which we conclude that

$$e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \leq \exp\big(-n(\eta_1 - 2\epsilon)\big). \tag{4.33}$$

Case 2: $R_1 \geq \min_k I(\mathrm{X}_k;\mathrm{S})$. For the second case choose any k such that $R_1 \geq I(\mathrm{X}_k;\mathrm{S})$. Then, from (4.26) follows that

$$e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \leq \sum_{\substack{y_1^n:(x^n_{m_1,m_2},x^n_{m_1,\hat{m}_2},s^n,y_1^n)\in\mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}\mathrm{Y}_1}\\ \text{for some } \hat{m}_2\neq m_2}} W_1^{\otimes n}(y_1^n|x^n_{m_1,m_2}, s^n). \tag{4.34}$$

It is sufficient to bound $e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))$ for $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}$ satisfying

$$I(\mathrm{X};\mathrm{X}_k,\mathrm{S}) < |R_1 - I(\mathrm{X}_k;\mathrm{S})|^+ + \epsilon \tag{4.35}$$

since otherwise the property (4.20c) of the codewords and (4.34) show that (4.30) is bounded by

$$\frac{1}{M_2^{(n)}}\big|\{m_2 : (x^n_{m_1,m_2}, x^n_{m_1,\hat{m}_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}} \text{ for some } m_2 \neq \hat{m}_2\}\big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big). \tag{4.36}$$

Moreover, we can assume in the following that $P_{\mathrm{X}_k} = P_{\mathrm{X}}$, since otherwise we have $e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) = 0$. Therefore, from (4.34) we can bound

$$\begin{aligned}
&e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m, s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))\\
&\quad\leq \sum_{\substack{m_2\neq\hat{m}_2:\\ (x^n_{m_1,m_2},x^n_{m_1,\hat{m}_2},s^n)\in\mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}}} W_1^{\otimes n}(\mathcal{T}^{(n)}_{\mathrm{Y}_1|\mathrm{X}\mathrm{X}_k\mathrm{S}}(x^n_{m_1,m_2}, x^n_{m_1,\hat{m}_2}, s^n)|x^n_{m_1,m_2}, s^n)\\
&\quad\leq \sum_{\substack{m_2\neq\hat{m}_2:\\ (x^n_{m_1,m_2},x^n_{m_1,\hat{m}_2},s^n)\in\mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}}} \exp\big(-nI(\mathrm{Y}_1;\mathrm{X}_k|\mathrm{X},\mathrm{S})\big)\\
&\quad\leq \exp\big(-n(I(\mathrm{Y}_1;\mathrm{X}_k|\mathrm{X},\mathrm{S}) - |R_1 - I(\mathrm{X}_k;\mathrm{X},\mathrm{S})|^+ - \epsilon)\big)
\end{aligned}$$

where the second inequality follows from Lemma B.7 and the last inequality from property (4.20a) of the codewords. Since $R_1 \geq I(\mathrm{X}_k;\mathrm{S})$, we get from (4.35) that

$$\begin{aligned} R_1 &> I(\mathrm{X};\mathrm{X}_k,\mathrm{S}) + I(\mathrm{X}_k;\mathrm{S}) - \epsilon \\ &\geq I(\mathrm{X};\mathrm{X}_k|\mathrm{S}) + I(\mathrm{X}_k;\mathrm{S}) - \epsilon \\ &= I(\mathrm{X}_k;\mathrm{X},\mathrm{S}) - \epsilon. \end{aligned}$$

Thus, we get for the probability of error

$$\begin{aligned} e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m,s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) &\leq \exp\big(-n(I(\mathrm{Y}_1;\mathrm{X}_k|\mathrm{X},\mathrm{S}) + I(\mathrm{X}_k;\mathrm{X},\mathrm{S}) - R_1 - 2\epsilon)\big) \\ &= \exp\big(-n(I(\mathrm{X}_k;\mathrm{X},\mathrm{S},\mathrm{Y}_1) - R_1 - 2\epsilon)\big) \\ &\leq \exp\big(-n(I(\mathrm{X}_k;\mathrm{Y}_1) - R_1 - 2\epsilon)\big). \end{aligned}$$

Since $P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_1} \in \mathcal{C}_{\eta_i}$ for some S_k and P_{X_k}, it follows from (4.23a) and (4.24) that

$$I(\mathrm{X}_k;\mathrm{Y}_1) - R_1 \geq \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}},\overline{W}_{1,q}) - R_1 - \frac{\delta}{3} > \frac{\delta}{3}$$

so that

$$e_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}\mathrm{Y}_1}(m,s^n) \leq \exp\left(-n(\frac{\delta}{3} - 2\epsilon)\right). \tag{4.37}$$

Now, we choose $\epsilon < \min\{\frac{\delta}{6}, \frac{\eta_1}{2}\}$ so that (4.28), (4.29), (4.32), (4.33), (4.36), and (4.37) imply that the average probability of error decreases exponentially fast for sufficiently large n. Since the derived bound holds uniformly for all $s^n \in \mathcal{S}^n$, the first part of the proof is complete. Similarly, we can bound the probability of error at node 2 using the same arguments. □

Converse

To complete the proof of Theorem 4.17 it remains to show that the presented strategy actually achieves all possible rate pairs so that no other rate pairs are achievable. For $L_i \leq T_i$ the converse part is already established by Lemma 4.27, since it shows that for $L_i \leq T_i$ no positive rates are achievable. Consequently, it remains to consider the case $L_i > T_i$, $i = 1, 2$. To avoid trivialities we further assume $L_1 \leq M_2^{(n)}$ and $L_2 \leq M_1^{(n)}$ in the following.

We have to show that any given sequence of $(n, M_1^{(n)}, M_2^{(n)}, L_1, L_2, \lambda^{(n)})$-list codes with list sizes $L_i > T_i$, $i = 1, 2$, and $\lambda^{(n)} \to 0$ must satisfy

$$R_2 = \frac{1}{n}\log\left(\frac{M_1^{(n)}}{L_2}\right) \leq \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}},\overline{W}_{2,q}) + o(n^0) \tag{4.38a}$$

$$R_1 = \frac{1}{n}\log\left(\frac{M_2^{(n)}}{L_1}\right) \leq \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}},\overline{W}_{1,q}) + o(n^0) \tag{4.38b}$$

for joint probability distributions $\{P_{\mathrm{X}}(x)\overline{W}_q(y_1, y_2|x)\}_{q\in\mathcal{P}(\mathcal{S})}$.

As a first step it is easy to show that any list code that is a good code for an AVBBC is also a good code for an appropriately constructed compound BBC. In more detail, let $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ be a $(n, M_1^{(n)}, M_2^{(n)}, L_1, L_2, \lambda^{(n)})$-list code for an AVBBC $\mathfrak{W}^n$ with average probability of error at node i, $i = 1, 2$,

$$\bar{e}_i(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \leq \lambda^{(n)} \qquad \text{for all } s^n \in \mathcal{S}^n. \tag{4.39}$$

Since (4.39) holds for all $s^n \in \mathcal{S}^n$, it immediately follows that the same is also true for any affine combination, i.e.,

$$\sum_{s^n\in\mathcal{S}^n} \bar{e}_i(s^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))q^{\otimes n}(s^n) \leq \lambda^{(n)} \qquad \text{for all } q \in \mathcal{P}(\mathcal{S}). \tag{4.40}$$

With the definition of the probability of error, cf. (4.5) and (4.6), for receiving node 1 Equation (4.40) reads as

$$\sum_{s^n\in\mathcal{S}^n} \frac{1}{|\mathcal{M}|} \sum_{y_1^n : m_2\notin\mathcal{L}^{(1)}(y_1^n, m_1)} W_1^{\otimes n}(y_1^n|x_m^n, s^n)q^{\otimes n}(s^n) \leq \lambda^{(n)} \qquad \text{for all } q \in \mathcal{P}(\mathcal{S})$$

or equivalently with the definition of an averaged broadcast channel, cf. (4.2), as

$$\frac{1}{|\mathcal{M}|} \sum_{y_1^n : m_2\notin\mathcal{L}^{(1)}(y_1^n, m_1)} \overline{W}_{1,q}^{\otimes n}(y_1^n|x_m^n) \leq \lambda^{(n)} \qquad \text{for all } q \in \mathcal{P}(\mathcal{S}). \tag{4.41}$$

The same arguments yield for receiving node 2 $\frac{1}{|\mathcal{M}|}\sum_{y_2^n : m_1\notin\mathcal{L}^{(2)}(y_2^n, m_2)} \overline{W}_{2,q}^{\otimes n}(y_2^n|x_m^n) \leq \lambda^{(n)}$ for all $q \in \mathcal{P}(\mathcal{S})$. Since (4.41) holds for all $q \in \mathcal{P}(\mathcal{S})$, the list code $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ is also a good code for the compound BBC $\overline{\mathfrak{W}} = \{\overline{W}_q(y_1, y_2|x)\}_{q\in\mathcal{P}(\mathcal{S})}$. Consequently for the AVBBC $\mathfrak{W}^n$ we cannot achieve higher rates as for the constructed compound BBC $\overline{\mathfrak{W}}$. Therefore, to establish the converse result for the AVBBC $\mathfrak{W}^n$, it remains to show that the rates for the compound BBC $\overline{\mathfrak{W}}$ are already bounded from above by (4.38).

Furthermore, it is sufficient to show that for a specific $q \in \mathcal{P}(\mathcal{S})$ the rates are bounded by

$$R_1 \leq I(P_{\mathrm{X}}, \overline{W}_{1,q}) + o(n^0) \quad \text{and} \quad R_2 \leq I(P_{\mathrm{X}}, \overline{W}_{2,q}) + o(n^0). \tag{4.42}$$

Since for the compound BBC $\overline{\mathfrak{W}}$ the rates have to satisfy (4.42) for all possible $q \in \mathcal{P}(\mathcal{S})$, the rates are immediately bounded by the corresponding infima $\inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{1,q}) + o(n^0)$ and $\inf_{q\in\mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{2,q}) + o(n^0)$. To prove (4.42) we need a version of Fano's lemma suitable for list decoding.

Lemma 4.32. *Let* U *be a random variable with values in* $\mathcal{M} \coloneqq \{1, ..., M\}$ *and* V *a random variable with values in* $\hat{\mathfrak{P}}_L(\mathcal{M})$*, i.e., the set of all subsets of* $\mathcal{M}$ *that contains at most* L *elements. Then*

$$H(\mathrm{U}|\mathrm{V}) \leq H_2\big(\mathbb{P}\{\mathrm{U} \notin \mathrm{V}\}\big) + \log L + \mathbb{P}\{\mathrm{U} \notin \mathrm{V}\} \log\Big(\frac{M}{L} - 1\Big)$$

with $H_2(\cdot)$ *the binary entropy.*

Proof. The proof can be found in [AGK76]. □

Now we are in the position to prove (4.42). Therefore, let $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ be any $(n, M_1^{(n)}, M_2^{(n)}, L_1, L_2, \lambda^{(n)})$-list code with

$$\frac{1}{|\mathcal{M}_1|} \sum_{m_1 \in \mathcal{M}_1} \frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1,m_2}^n) \leq \lambda_1^{(n)}$$
$$\frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \frac{1}{|\mathcal{M}_1|} \sum_{m_1 \in \mathcal{M}_1} \sum_{y_2^n : m_1 \notin \mathcal{L}^{(2)}(y_2^n, m_2)} \overline{W}_{2,q}^{\otimes n}(y_2^n | x_{m_1,m_2}^n) \leq \lambda_2^{(n)}.$$

Let us consider random variables U_i, X^n, Y_i^n, $i = 1, 2$, with values in $\mathcal{M}_i$, $\mathcal{X}^n$, $\mathcal{Y}_i^n$, $i = 1, 2$, respectively, and with

$$\begin{aligned}
&\mathbb{P}\{\mathrm{U}_1 = m_1, \mathrm{U}_2 = m_2, \mathrm{X}^n = x^n, \mathrm{Y}_1^n = y_1^n, \mathrm{Y}_2^n = y_2^n\} \\
&\qquad = \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} p(x_{m_1,m_2}^n | m_1, m_2) \overline{W}_q^{\otimes n}(y_1^n, y_2^n | x_{m_1,m_2}^n) \\
&\qquad = \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} p(x_{m_1,m_2}^n | m_1, m_2) \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1,m_2}^n) \overline{W}_{2,q}^{\otimes n}(y_2^n | x_{m_1,m_2}^n)
\end{aligned}$$

where $p(x_{m_1,m_2}^n | m_1, m_2) = 1$ if x_{m_1,m_2}^n is the codeword corresponding to $m = (m_1, m_2)$ or is equal to 0 else. Further, let V_1 and V_2 be random variables with values in $\hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$ and $\hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$, respectively, and set $\mathrm{V}_1 \coloneqq \mathcal{L}^{(1)}(\mathrm{Y}_1^n, \mathrm{U}_1)$ and $\mathrm{V}_2 \coloneqq \mathcal{L}^{(2)}(\mathrm{Y}_2^n, \mathrm{U}_2)$.

By definition

$$\begin{aligned}\mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\} &= \sum_{m_2 \in \mathcal{M}_2} \mathbb{P}\{\mathrm{U}_2 = m_2, m_2 \notin \mathcal{L}^{(1)}(\mathrm{Y}_1^n, m_1)\} \\ &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{x_{m_1,m_2}^n \in \mathcal{X}^n} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} p(x_{m_1,m_2}^n | m_1, m_2) \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1,m_2}^n) \\ &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1,m_2}^n) \\ &= \frac{1}{|\mathcal{M}_1|} \sum_{m_1 \in \mathcal{M}_1} \frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1,m_2}^n) \leq \lambda_1^{(n)}.\end{aligned}$$

By averaging separately over all $m_1 \in \mathcal{M}_1$ and all $m_2 \in \mathcal{M}_2$ we see that there exists for each block length n a fixed $m_1^* \in \mathcal{M}_1$ such that

$$\frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1^*)} \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1^*,m_2}^n) \leq \lambda_1^{(n)}$$

is fulfilled. Then $\mathrm{U}_2 - \mathrm{X}^n - \mathrm{Y}_1^n - \mathrm{V}_1(m_1^*, q)$ forms a Markov chain and we are in the same position as in the single-user case with a single-user list code. We want to emphasize that the random variable $\mathrm{V}_1(m_1^*, q)$ clearly depends on the specific $m_1^* \in \mathcal{M}_1$ and $q \in \mathcal{P}(\mathcal{S})$ but for the sake of brevity we write V_1 for $\mathrm{V}_1(m_1^*, q)$ in the following. We get

$$\begin{aligned}\log M_2^{(n)} &= H(\mathrm{U}_2) = H(\mathrm{U}_2|\mathrm{V}_1) + I(\mathrm{U}_2; \mathrm{V}_1) \\ &\leq H(\mathrm{U}_2|\mathrm{V}_1) + I(\mathrm{X}^n; \mathrm{Y}_1^n) \\ &\leq H_2\big(\mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\}\big) + \log L_1 + \mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\} \log\Big(\frac{M_2^{(n)}}{L_1}\Big) + I(\mathrm{X}^n; \mathrm{Y}_1^n) \\ &\leq H_2\big(\mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\}\big) + \log L_1 + \mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\} \log\Big(\frac{M_2^{(n)}}{L_1}\Big) + \sum_{k=1}^n I(\mathrm{X}_k; \mathrm{Y}_{1,k}) \quad (4.43)\end{aligned}$$

where the equality and inequalities follow from the definition of mutual information, the data processing inequality, Lemma 4.32 with $\mathcal{L}^{(1)} : \mathcal{Y}_1^n \times \mathcal{M}_1 \to \hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$, and the memoryless property of the channel. We can rewrite the mutual information term on the right hand side of (4.43), cf. Appendix B.1, as

$$\begin{aligned}\log M_2^{(n)} &\leq H_2\big(\mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\}\big) + \log L_1 + \mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\} \log\Big(\frac{M_2^{(n)}}{L_1}\Big) + \sum_{k=1}^n I(P_{\mathrm{X}_k}, \overline{W}_{1,q}) \\ &\leq H_2\big(\mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\}\big) + \log L_1 + \mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\} \log\Big(\frac{M_2^{(n)}}{L_1}\Big) + nI(P_{\mathrm{X}}, \overline{W}_{1,q}) \quad (4.44)\end{aligned}$$

where the last inequality follows from the concavity of mutual information and $P_{\mathrm{X}} \coloneqq \frac{1}{n}\sum_{k=1}^{n} P_{\mathrm{X}_k}$. We note that due to the continuity of the mutual information and the compactness of the set of probability distributions, the dependency of P_{X} on the block length n vanishes asymptotically.

Rearranging the terms in (4.43) and dividing by n leads to

$$\big(1-\mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\}\big)\frac{1}{n}\log\Big(\frac{M_2^{(n)}}{L_1}\Big) \leq \frac{1}{n}H_2\big(\mathbb{P}\{\mathrm{U}_2 \notin \mathrm{V}_1\}\big) + I(P_{\mathrm{X}}, \overline{W}_{1,q})$$

or

$$\big(1-\lambda_1^{(n)}\big)\frac{1}{n}\log\Big(\frac{M_2^{(n)}}{L_1}\Big) \leq \frac{1}{n}H_2\big(\lambda_1^{(n)}\big) + I(P_{\mathrm{X}}, \overline{W}_{1,q}). \tag{4.45}$$

The same arguments yield for receiving node 2 $(1-\lambda_2^{(n)})\frac{1}{n}\log(\frac{M_1^{(n)}}{L_2}) \leq \frac{1}{n}H_2(\lambda_2^{(n)}) + I(P_{\mathrm{X}}, \overline{W}_{2,q})$. Finally, (4.42) and therewith the converse follows from this and (4.45). □

4.6 Input and State Constraints

Next, we analyze the case where constraints on the input and state sequences are imposed. Thereby, we consider only the case with list sizes one, i.e., we assume $L_1 = L_2 = 1$ throughout the rest of this section. We follow [CN88b] and define cost functions $g(x)$ and $l(s)$ on $\mathcal{X}$ and $\mathcal{S}$, respectively. For convenience, we assume that $\min_{x\in\mathcal{X}} g(x) = \min_{s\in\mathcal{S}} l(s) = 0$ and define the maximum values as $g_{\max} \coloneqq \max_{x\in\mathcal{X}} g(x)$ and $l_{\max} \coloneqq \max_{s\in\mathcal{S}} l(s)$. For given sequences $x^n = (x_1, ..., x_n)$ and $s^n = (s_1, ..., s_n)$ we set

$$g(x^n) \coloneqq \frac{1}{n}\sum_{k=1}^{n} g(x_k) \tag{4.46a}$$

$$l(s^n) \coloneqq \frac{1}{n}\sum_{k=1}^{n} l(s_k). \tag{4.46b}$$

Further, for notational convenience we define the costs caused by given probability distributions $p \in \mathcal{P}(\mathcal{X})$ and $q \in \mathcal{P}(\mathcal{S})$ as

$$g(p) = \sum_{x\in\mathcal{X}} p(x)g(x) \qquad \text{and} \qquad l(q) = \sum_{s\in\mathcal{S}} q(s)l(s)$$

and observe that, if we consider types, these definitions immediately yield

$$g(x^n) = g(P_{x^n}) \qquad \text{and} \qquad l(s^n) = l(P_{s^n})$$

for every $x^n \in \mathcal{X}^n$ and every $s^n \in \mathcal{S}^n$, respectively, cf. also [CN88b].

This allows us to define the set of all state sequences of length n that satisfy a given state constraint Λ by

$$\mathcal{S}_\Lambda^n \coloneqq \Big\{ s^n \in \mathcal{S}^n : \frac{1}{n} \sum_{k=1}^{n} l(s_k) = \mathbb{E}_{P_{s^n}}[l(s^n)] \leq \Lambda \Big\}.$$

Furthermore, the set of all probability distributions $q \in \mathcal{P}(\mathcal{S})$ that satisfy $\mathbb{E}_q[l(q)] \leq \Lambda$ is given by

$$\mathcal{P}(\mathcal{S}, \Lambda) \coloneqq \big\{ q : q \in \mathcal{P}(\mathcal{S}), \mathbb{E}_q[l(q)] \leq \Lambda \big\}.$$

In Corollary 4.18 we have shown that an AVBBC $\mathfrak{W}^n$ (without state constraint) has a deterministic code capacity region whose interior is empty if $\mathfrak{W}^n$ is $\mathcal{Y}_1$-symmetrizable or $\mathcal{Y}_2$-symmetrizable. If we impose a state constraint, the situation changes significantly. Now, it is possible that the interior of the deterministic code capacity region is non-empty even if the $\mathfrak{W}^n$ is $\mathcal{Y}_i$-symmetrizable in the sense of Definition 4.10 and Remark 4.12, respectively. Rather, $\mathcal{Y}_i$-symmetrizability enters the picture via

$$\Lambda_i(P_{\mathrm{X}}) = \begin{cases} \min\limits_{U_i \in \mathcal{U}_i} \sum\limits_{x \in \mathcal{X}} \sum\limits_{s \in \mathcal{S}} P_{\mathrm{X}}(x) U_i(s|x) l(s) & \text{if } \mathcal{U}_i \neq \emptyset \\ \infty & \text{if } \mathcal{U}_i = \emptyset \end{cases} \tag{4.47}$$

$i = 1, 2$, which indicates whether the symmetrization violates the imposed state constraint or not. Thereby, $\mathcal{U}_i$ is the set of all channels $U_i : \mathcal{X} \to \mathcal{P}(\mathcal{S})$ which satisfy (4.4). For given type P_{X} the quantity $\Lambda_i(P_{\mathrm{X}})$ is called *symmetrizability costs* and can be interpreted as the minimum costs which are needed to symmetrize the AVBBC $\mathfrak{W}^n$. Clearly, if $\mathfrak{W}^n$ is $\mathcal{Y}_i$-symmetrizable, then $\mathcal{U}_i \neq \emptyset$ and $\Lambda_i(P_{\mathrm{X}})$ is finite. Further, if $\mathfrak{W}^n$ is non-$\mathcal{Y}_i$-symmetrizable, then $\mathcal{U}_i = \emptyset$, and we set the symmetrizability costs $\Lambda_i(P_{\mathrm{X}}) = \infty$ for convenience.

A deterministic code for the AVBBC $\mathfrak{W}^n$ under input and state constraints is defined in a similar way as for the case without constraints, cf. Definition 4.13. The only difference is that we additionally require all valid codewords to satisfy the input constraint Γ.

Definition 4.33. *A* deterministic $(n, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{det}(\mathfrak{W}^n)$ *of length* n *for the AVBBC* $\mathfrak{W}^n$ *under input constraint* Γ *and state constraint* Λ *is a family*

$$\mathcal{C}_{det}(\mathfrak{W}^n) \coloneqq \big\{ (x_m^n, \mathcal{D}^{(1)}_{m_2|m_1}, \mathcal{D}^{(2)}_{m_1|m_2}) : m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2 \big\}$$

with codewords

$$x_m^n \in \mathcal{X}^n \qquad \text{with } g(x_m^n) \leq \Gamma,$$

one for each message $m = (m_1, m_2)$*, satisfying the input constraint* Γ*, and decoding sets at nodes 1 and 2*

$$\mathcal{D}^{(1)}_{m_2|m_1} \subseteq \mathcal{Y}_1^n \quad \text{and} \quad \mathcal{D}^{(2)}_{m_1|m_2} \subseteq \mathcal{Y}_2^n$$

for all $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$. For given m_1 at node 1 the decoding sets must be disjoint, i.e., $\mathcal{D}^{(1)}_{m_2|m_1} \cap D^{(1)}_{\hat{m}_2|m_1} = \emptyset$ for $\hat{m}_2 \neq m_2$, and similarly for given m_2 at node 2 the decoding sets must satisfy $\mathcal{D}^{(2)}_{m_1|m_2} \cap D^{(2)}_{\hat{m}_1|m_2} = \emptyset$ for $\hat{m}_1 \neq m_1$.

The definition of the probability of error follows accordingly with the restriction that we only have to consider state sequences that satisfy the state constraint Λ. Thus, for given $0 < \lambda^{(n)} < 1$, the code $\mathcal{C}_{\text{det}}(\mathfrak{W}^n)$ is called a $(n, M_1^{(n)}, M_2^{(n)}, \lambda^{(n)})$-code (with average probability of error $\lambda^{(n)}$) for $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ if

$$\bar{e}(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \leq \lambda^{(n)} \qquad \text{for all } s^n \in \mathcal{S}^n_\Lambda.$$

Definition 4.34. *A rate pair $(R_1, R_2) \in \mathbb{R}^2_+$ is said to be* deterministically achievable *for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ if for any $\delta > 0$ there exists an $n(\delta) \in \mathbb{N}$ and a sequence of deterministic $(n, M_1^{(n)}, M_2^{(n)}, \lambda^{(n)})$-codes $\{\mathcal{C}^{(n)}_{det}(\mathfrak{W}^n)\}_{n\in\mathbb{N}}$ with codewords $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, each satisfying $g(x^n_{m_1,m_2}) \leq \Gamma$, such that for all $n \geq n(\delta)$ we have*

$$\frac{1}{n} \log M_1^{(n)} \geq R_2 - \delta \quad \textit{and} \quad \frac{1}{n} \log M_2^{(n)} \geq R_1 - \delta$$

while

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}(s^n|\mathcal{C}_{det}(\mathfrak{W}^n)) \leq \lambda^{(n)}$$

with $\lambda^{(n)} \to 0$ as $n \to \infty$, $k = 1, 2$. The set of all achievable rate pairs is the deterministic code capacity region *of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ and is denoted by $\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma, \Lambda)$.*

If $\Gamma \geq g_{\max}$ or $\Lambda \geq l_{\max}$, the input or state sequences are not restricted by its corresponding constraint, respectively. Consequently, we denote the capacity region with state constraint and no input constraint by $\mathcal{R}_{\text{det}}(\mathfrak{W}^n|g_{\max}, \Lambda)$ and the capacity region with input constraint and no state constraint by $\mathcal{R}_{\text{det}}(\mathfrak{W}^n|\Gamma, l_{\max})$.

Finally, the definition of a *random* $(n, M_1^{(n)}, M_2^{(n)}, \text{Z})$-*code* $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ follows accordingly from the one without constraints, cf. Definition 4.19, where we simply additionally require that each deterministic code of the random code satisfies the constraints individually. Clearly, the definitions of a *randomly achievable* rate pair under input and state constraints and the *random code capacity region* under input and state constraints follow accordingly.

4.6.1 Random Code Capacity Region

Here, we derive the random code capacity region of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ. For this purpose we define for given type P_{X} the region

$$\begin{aligned}\mathcal{R}(P_{\mathrm{X}}|\Lambda) \coloneqq \big\{(R_1,R_2)\in\mathbb{R}_+^2 : R_1 \leq \inf_{q\in\mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}},\overline{W}_{1,q}) \\ R_2 \leq \inf_{q\in\mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}},\overline{W}_{2,q})\big\}\end{aligned} \tag{4.48}$$

for joint probability distributions $\{P_{\mathrm{X}}(x)\overline{W}_q(y_1,y_2|x)\}_{q\in\mathcal{P}(\mathcal{S},\Lambda)}$.

Theorem 4.35. *The random code capacity region $\mathcal{R}_{ran}(\mathfrak{W}^n|\Gamma,\Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ is*

$$\mathcal{R}_{ran}(\mathfrak{W}^n|\Gamma,\Lambda) = \bigcup_{P_{\mathrm{X}}:g(P_{\mathrm{X}})\leq\Gamma} \mathcal{R}(P_{\mathrm{X}}|\Lambda).$$

The proof of the random code capacity region under input and state constraints is quite similar to the corresponding case without any constraints. To incorporate the constraints slight adaptations are needed, primarily in the robustification technique. For completeness we present the proof in the following.

Compound Bidirectional Broadcast Channel

As in Section 4.3.1 for the AVBBC without constraints on input and states we start with a construction of a suitable compound BBC, where the key idea is to restrict it in an appropriate way. Having the state constraint Λ in mind, it is reasonable to restrict our attention to all probability distributions $q \in \mathcal{P}(\mathcal{S},\Lambda)$. Let us consider the family of averaged broadcast channels, cf. (4.2),

$$\big\{\overline{W}_q(y_1,y_2|x)\big\}_{q\in\mathcal{P}(\mathcal{S},\Lambda)} \tag{4.49}$$

and observe that this already corresponds to a compound BBC where each permissible probability distribution $q\in\mathcal{P}(\mathcal{S},\Lambda)$ parametrizes one element of the compound channel which we denote by $\overline{\mathfrak{W}}$ in the following. The capacity region of the compound BBC is given in Chapter 3. It is shown that for given input distribution P_{X} all rate pairs (R_1,R_2) satisfying $(R_1,R_2)\in\mathcal{R}(P_{\mathrm{X}}|\Lambda)$, cf. (4.48), are deterministically achievable. In particular, this is valid for a input distribution P_{X} that satisfies the input constraint $g(P_{\mathrm{X}})\leq\Gamma$.

In more detail, it is shown that there exists a deterministic code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ such that all rate pairs $(R_1,R_2)\in\mathcal{R}(P_{\mathrm{X}}|\Lambda)$ are achievable for input type P_{X} while

the average probability of error can be bounded from above by

$$\bar{e}(q|\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})) \coloneqq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \overline{W}_q^{\otimes n}\big((\mathcal{D}^{(1)}_{m_2|m_1}\times\mathcal{D}^{(2)}_{m_1|m_2})^c|x_m^n\big) \leq \lambda^{(n)}_{\overline{\mathfrak{W}}} \quad \text{for all } q\in\mathcal{P}(\mathcal{S},\Lambda)$$

with $\lambda^{(n)}_{\overline{\mathfrak{W}}} = \lambda^{(n)}_{\overline{\mathfrak{W}},1} + \lambda^{(n)}_{\overline{\mathfrak{W}},2}$ where $\lambda^{(n)}_{\overline{\mathfrak{W}},i}$ is the average probability of error at node i, $i=1,2$. Moreover, for n large enough, we have

$$\lambda^{(n)}_{\overline{\mathfrak{W}},i} = (n+1)^{|\mathcal{X}||\mathcal{Y}_i|}2^{-n\frac{c\epsilon^2}{2}} + \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2}}2^{-n\frac{\tau}{8}}$$

which decreases exponentially fast for increasing block length n. Thereby, $\epsilon>0$, $\tau>0$, and $c>0$ are constants, cf. also (3.12).

Together with (4.2) this immediately implies that for $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ the average probability of a successful transmission over the compound BBC $\overline{\mathfrak{W}}$ is bounded from below by

$$\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \overline{W}_q^{\otimes n}(\mathcal{D}^{(1)}_{m_2|m_1}\times\mathcal{D}^{(2)}_{m_1|m_2}|x_m^n) > 1-\lambda^{(n)}_{\overline{\mathfrak{W}}}$$

or equivalently by

$$\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}\sum_{s^n\in\mathcal{S}^n} W^{\otimes n}(\mathcal{D}^{(1)}_{m_2|m_1}\times\mathcal{D}^{(2)}_{m_1|m_2}|x_m^n,s^n)q^{\otimes n}(s^n) > 1-\lambda^{(n)}_{\overline{\mathfrak{W}}}$$

for all $q^{\otimes n}=\prod_{k=1}^n q$ and $q\in\mathcal{P}(\mathcal{S},\Lambda)$.

Robustification

As in Section 4.3.2 for the AVBBC without state constraints we use the deterministic code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ to construct a random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ. Therefore, we need a version of the robustification technique, cf. Theorem 4.22, that incorporates the state constraint.

Theorem 4.36 (Robustification technique)**.** *Let $f:\mathcal{S}^n\to[0,1]$ be a function such that for some $\alpha\in(0,1)$ the inequality*

$$\sum_{s^n\in\mathcal{S}^n} f(s^n)q^{\otimes n}(s^n) > 1-\alpha \qquad \text{for all } q\in\mathcal{P}_0(n,\mathcal{S},\Lambda) \tag{4.50}$$

holds where $\mathcal{P}_0(n,\mathcal{S},\Lambda)\coloneqq\{q\in\mathcal{P}_0(n,\mathcal{S}):\mathbb{E}_q[l(q)]\leq\Lambda\}$. Then it also holds

$$\frac{1}{n!}\sum_{\pi\in\Pi_n} f\big(\pi(s^n)\big) > 1-(n+1)^{|\mathcal{S}|}\alpha \qquad \text{for all } s^n\in\mathcal{S}^n_\Lambda.$$

Proof. The proof can be found in Appendix A.5. □

With the robustification technique and

$$f\big(\pi(s^n)\big) = \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} W^{\otimes n}\big(\mathcal{D}^{(1)}_{m_2|m_1} \times \mathcal{D}^{(2)}_{m_1|m_2} | x^n_m, \pi(s^n)\big)$$

we immediately obtain a random $(n, M_1^{(n)}, M_2^{(n)}, \Pi_n)$-code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ which is given by the family

$$\begin{aligned}\mathcal{C}_{\text{ran}}(\mathfrak{W}^n) = \big\{(\pi^{-1}(x^n_m), \pi^{-1}(\mathcal{D}^{(1)}_{m_2|m_1}), \pi^{-1}(\mathcal{D}^{(2)}_{m_1|m_2})) : \\ m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2, \pi \in \Pi_n\big\}\end{aligned} \tag{4.51}$$

where the permutations π are uniformly distributed on Π_n and

$$\pi^{-1}(\mathcal{D}^{(1)}_{m_2|m_1}) = \bigcup_{y_1^n \in \mathcal{D}^{(1)}_{m_2|m_1}} \pi^{-1}(y_1^n) \quad \text{and} \quad \pi^{-1}(\mathcal{D}^{(2)}_{m_1|m_2}) = \bigcup_{y_2^n \in \mathcal{D}^{(2)}_{m_1|m_2}} \pi^{-1}(y_2^n).$$

From the robustification technique follows that the average probability of error of $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ is bounded from above by

$$\bar{e}(s^n|\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)) \leq (n+1)^{|\mathcal{S}|}\lambda^{(n)}_{\overline{\mathfrak{W}}} =: \lambda^{(n)}_{\mathfrak{W},\text{ran}} \qquad \text{for all } s^n \in \mathcal{S}^n_\Lambda. \tag{4.52}$$

Moreover, because of the construction it is clear that for given input P_{X}, the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ achieves for the AVBBC $\mathfrak{W}^n$ the same rate pairs as $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ as specified in (4.48). Finally, taking the union over all input distributions P_{X} that satisfy the input constraint $g(P_{\text{X}}) \leq \Gamma$ establishes the achievability of the random code capacity $\mathcal{R}_{\text{ran}}(\mathfrak{W}^n|\Gamma,\Lambda)$ as stated in Theorem 4.35.

Converse

As a first step, it is easy to show that the average probability of error for the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ equals the average probability of error for the random code for the compound BBC $\overline{\mathfrak{W}}$. Hence, it is clear that we cannot achieve higher rates as for the constructed compound BBC $\overline{\mathfrak{W}}$ with random codes. The deterministic rates of the compound channel can be found in Chapter 3. Additionally, as in [AW69] for the single-user compound channel, it can easily be shown that for the compound BBC the achievable rates for deterministic and random codes are equal. Since the constructed random code for the AVBBC $\mathfrak{W}^n$ already achieves these rates, the converse is established.

This finishes the proof of Theorem 4.35 and therewith the random code capacity region $\mathcal{R}_{\text{ran}}(\mathfrak{W}^n|\Gamma,\Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ. □

4.6.2 Deterministic Code Capacity Region

A random coding strategy as constructed in the previous section requires *common randomness* between all nodes, since the encoder and the decoders depend all on the same random permutation which has to be known at all nodes in advance. If this kind of resource is not available, we are interested in deterministic strategies. In this section, we derive the deterministic code capacity region of the AVBBC with constraints on input and states.

Theorem 4.37. *If* $\max_{P_{\mathrm{X}}:g(P_{\mathrm{X}})\leq\Gamma}\Lambda_i(P_{\mathrm{X}}) > \Lambda$, $i=1,2$, *the deterministic code capacity region* $\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma,\Lambda)$ *of the AVBBC* $\mathfrak{W}^n$ *under input constraint* Γ *and state constraint* Λ *is*

$$\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma,\Lambda) = \bigcup_{\substack{P_{\mathrm{X}}:\, g(P_{\mathrm{X}})\leq\Gamma \\ \Lambda_i(P_{\mathrm{X}})>\Lambda,\, i=1,2}} \mathcal{R}(P_{\mathrm{X}}|\Lambda).$$

If $\max_{P_{\mathrm{X}}:g(P_{\mathrm{X}})\leq\Gamma}\Lambda_1(P_{\mathrm{X}}) < \Lambda$ *or* $\max_{P_{\mathrm{X}}:g(P_{\mathrm{X}})\leq\Gamma}\Lambda_2(P_{\mathrm{X}}) < \Lambda$, *we have* $int(\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma,\Lambda)) = \emptyset$.

From Theorem 4.37 we immediately obtain the deterministic code capacity region of the AVBBC $\mathfrak{W}^n$ with state constraint Λ and no input constraint, i.e., $\mathcal{R}_{\mathrm{det}}(\mathfrak{W}^n|g_{\max},\Lambda)$.

Corollary 4.38. *If* $\max_{P_{\mathrm{X}}}\Lambda_i(P_{\mathrm{X}}) > \Lambda$, $i=1,2$, *the deterministic code capacity region* $\mathcal{R}_{det}(\mathfrak{W}^n|g_{max},\Lambda)$ *of the AVBBC* $\mathfrak{W}^n$ *with state constraint* Λ *and no input constraint is given by*

$$\mathcal{R}_{det}(\mathfrak{W}^n|g_{max},\Lambda) = \bigcup_{P_{\mathrm{X}}:\,\Lambda_i(P_{\mathrm{X}})>\Lambda, i=1,2} \mathcal{R}(P_{\mathrm{X}}|\Lambda)$$

If $\max_{P_{\mathrm{X}}}\Lambda_1(P_{\mathrm{X}}) < \Lambda$ *or* $\max_{P_{\mathrm{X}}}\Lambda_2(P_{\mathrm{X}}) < \Lambda$, *we have* $int(\mathcal{R}_{det}(\mathfrak{W}^n|g_{max},\Lambda)) = \emptyset$.

We observe that the deterministic code capacity region $\mathcal{R}_{\mathrm{det}}(\mathfrak{W}^n|\Gamma,\Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ displays a dichotomy behavior similarly as in the unconstrained case: it either equals a non-empty region or else has an empty interior. Unfortunately, this knowledge cannot be exploited to prove the corresponding deterministic code capacity region, since, as already observed in [CN88b] for the single-user AVC, *Ahlswede's elimination technique* [Ahl78] does not work anymore if constraints are imposed on the permissible codewords and sequences of states. Consequently, to prove Theorem 4.37 we need a proof idea which does not rely on this technique. In the following we present the proof which is therefore mainly based on an extension of [CN88b].

Symmetrizability

The following result shows that under state constraint Λ no code with codewords of type P_{X} satisfying $\Lambda_1(P_{\mathrm{X}}) < \Lambda$ or $\Lambda_2(P_{\mathrm{X}}) < \Lambda$ can be good.

Lemma 4.39. *For a $\mathcal{Y}_1$-symmetrizable AVBBC $\mathfrak{W}^n$ any deterministic code $\mathcal{C}_{det}(\mathfrak{W}^n)$ of block length n with codewords $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, each of type P_{X} with $\Lambda_1(P_{\mathrm{X}}) < \Lambda$, and $M_2^{(n)} \geq 2$ has*

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_1(s^n | \mathcal{C}_{det}(\mathfrak{W}^n)) \geq \frac{M_2^{(n)} - 1}{2M_2^{(n)}} - \frac{1}{n} \frac{l_{max}^2}{(\Lambda - \Lambda_1(P_{\mathrm{X}}))^2}.$$

Similarly, for a $\mathcal{Y}_2$-symmetrizable AVBBC $\mathfrak{W}^n$ any deterministic code $\mathcal{C}_{det}(\mathfrak{W}^n)$ of block length n with codewords $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, each of type P_{X} with $\Lambda_2(P_{\mathrm{X}}) < \Lambda$, and $M_1^{(n)} \geq 2$ has

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_2(s^n | \mathcal{C}_{det}(\mathfrak{W}^n)) \geq \frac{M_1^{(n)} - 1}{2M_1^{(n)}} - \frac{1}{n} \frac{l_{max}^2}{(\Lambda - \Lambda_2(P_{\mathrm{X}}))^2}.$$

Proof. The proof can be found in Appendix A.6. □

Remark 4.40. *The lemma indicates that for a successful transmission using codewords of type P_{X} the symmetrizability costs $\Lambda_i(P_{\mathrm{X}})$, $i = 1, 2$, have to exceed the permissible (or available) costs Λ since otherwise the AVBBC $\mathfrak{W}^n$ can be symmetrized which prohibits any reliable or error-free communication. This already establishes the second part of Theorem 4.37 and therewith characterizes when int$(\mathcal{R}_{det}(\mathfrak{W}^n | \Gamma, \Lambda)) = \emptyset$.*

Achievability

Next, we present a coding strategy with codewords of type P_{X} that achieves the desired rates as specified in Theorem 4.37 if the symmetrizability costs exceed the permissible costs, i.e., $\Lambda_1(P_{\mathrm{X}}) > \Lambda$ and $\Lambda_2(P_{\mathrm{X}}) > \Lambda$. Fortunately, we are in the same position as in the single-user AVC [CN88b]: the coding strategy for the AVBBC without constraints must only slightly be modified to apply also to the AVBBC with constraints.

We need codewords $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$ with the following properties.

Lemma 4.41. *For any $\epsilon > 0$, $n \geq n_0(\epsilon)$, $M_i^{(n)} \geq \exp(n\epsilon)$, $i = 1, 2$, and given type P_{X}, there exist codewords $x^n_{m_1,m_2} \in \mathcal{T}_{\mathrm{X}}^{(n)}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$ such that for every $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}_\Lambda^n$, and every joint type $P_{\mathrm{XX'S}}$, with $R_1 = \frac{1}{n}\log M_2^{(n)}$ and $R_2 = \frac{1}{n}\log M_1^{(n)}$, we have for each fixed $m_1 \in \mathcal{M}_1$ the following properties*

$$\big|\{\hat{m}_2 : (x^n, x^n_{m_1,\hat{m}_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XX'S}}\}\big| \leq \exp\big(n(|R_1 - I(\mathrm{X}';\mathrm{X},\mathrm{S})|^+ + \epsilon)\big) \tag{4.53a}$$

$$\frac{1}{M_2^{(n)}}\big|\{m_2 : (x^n_{m_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XS}}\}\big| \leq \exp\big(-n\frac{\epsilon}{2}\big) \qquad \text{if } I(\mathrm{X};\mathrm{S}) > \epsilon \tag{4.53b}$$

$$\frac{1}{M_2^{(n)}}\big|\{m_2 : (x^n_{m_1,m_2}, x^n_{m_1,\hat{m}_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XX'S}} \text{ for some } \hat{m}_2 \neq m_2\}\big| \leq \exp\big(-n\frac{\epsilon}{2}\big)$$
$$\text{if } I(\mathrm{X};\mathrm{X}',\mathrm{S}) - |R_1 - I(\mathrm{X}';\mathrm{S})|^+ > \epsilon \tag{4.53c}$$

and further for each fixed $m_2 \in \mathcal{M}_2$

$$\big|\{\hat{m}_1 : (x^n, x^n_{\hat{m}_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XX'S}}\}\big| \leq \exp\big(n(|R_2 - I(\mathrm{X}';\mathrm{X},\mathrm{S})|^+ + \epsilon)\big) \tag{4.53d}$$

$$\frac{1}{M_1^{(n)}}\big|\{m_1 : (x^n_{m_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XS}}\}\big| \leq \exp\big(-n\frac{\epsilon}{2}\big) \qquad \text{if } I(\mathrm{X};\mathrm{S}) > \epsilon \tag{4.53e}$$

$$\frac{1}{M_1^{(n)}}\big|\{m_1 : (x^n_{m_1,m_2}, x^n_{\hat{m}_1,m_2}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{XX'S}} \text{ for some } \hat{m}_1 \neq m_1\}\big| \leq \exp\Big(-n\frac{\epsilon}{2}\Big)$$
$$\text{if } I(\mathrm{X};\mathrm{X}',\mathrm{S}) - |R_2 - I(\mathrm{X}';\mathrm{S})|^+ > \epsilon. \tag{4.53f}$$

Proof. The properties can be deduced from the corresponding result for list decoding, cf. Lemma 4.28, by setting the list sizes to one, i.e., $L_1 = L_2 = 1$, and taking the input constraint into account. Therefore we omit the details for brevity. □

We follow [CN88b] and define the decoding sets similarly as for the single-user AVC under input and state constraints. Therefore, we define the set

$$\mathcal{D}_{\eta_i}(\Lambda) = \big\{P_{\mathrm{XSY}_i} : D(P_{\mathrm{XSY}_i} \| P_{\mathrm{X}} \otimes P_{\mathrm{S}} \otimes W_i) \leq \eta_i, l(P_{\mathrm{S}}) \leq \Lambda\big\}, \qquad i = 1, 2.$$

Then, the decoding sets at node 1 are specified as follows.

Definition 4.42. *For given codewords $x^n_{m_1,m_2} \in \mathcal{T}_{\mathrm{X}}^{(n)}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, and $\eta_1 > 0$ we have $y_1^n \in \mathcal{D}^{(1)}_{m_2|m_1}$ if and only if*

i) there exists an $s^n \in \mathcal{S}_\Lambda^n$ such that $P_{x^n_{m_1,m_2},s^n,y_1^n} \in \mathcal{D}_{\eta_1}(\Lambda)$

ii) for each codeword $x^n_{m_1,\hat{m}_2}$ with $\hat{m}_2 \neq m_2$ which satisfies $P_{x^n_{m_1,\hat{m}_2},s'^n,y_1^n} \in \mathcal{D}_{\eta_1}(\Lambda)$ for some $s'^n \in \mathcal{S}_\Lambda^n$, we have $I(\mathrm{X},\mathrm{Y}_1;\mathrm{X}'|\mathrm{S}) \leq \eta_1$ where $\mathrm{X}, \mathrm{X}', \mathrm{S}, \mathrm{Y}_1$ are dummy random variables such that $P_{\mathrm{XX'SY}_1}$ equals the joint type of $(x^n_{m_1,m_2}, x^n_{m_1,\hat{m}_2}, s^n, y_1^n)$.

The decoding sets at node 2 are defined accordingly with $\eta_2 > 0$. A key part is now to ensure that these decoding sets are unambiguously defined. This means that they are disjoint for small enough η_1 and η_2 which can be shown analogously to the single-user case [CN88b]. Here is where the conditions on the symmetrizability costs, $\Lambda_i(P_{\mathrm{X}}) > \Lambda$, $i = 1, 2$, come in.

Lemma 4.43. *Let $\alpha > 0$ and $\beta > 0$, then for a sufficiently small η_i, $i = 1, 2$, no quintuple of random variables $(\mathrm{X}, \mathrm{X}', \mathrm{S}, \mathrm{S}', \mathrm{Y}_i)$ can simultaneously satisfy $P_{\mathrm{X}} = P_{\mathrm{X}'}$ with*

$$\Lambda_i(P_{\mathrm{X}}) \geq \Lambda + \alpha \qquad \textit{and} \qquad \min_{x \in \mathcal{X}} P_{\mathrm{X}}(x) \geq \beta$$

and

$$P_{\mathrm{XSY}_i} \in \mathcal{D}_{\eta_i}(\Lambda), \qquad P_{\mathrm{X'S'Y}_i} \in \mathcal{D}_{\eta_i}(\Lambda), \tag{4.54a}$$

$$I(\mathrm{X}, \mathrm{Y}_i; \mathrm{X}'|\mathrm{S}) \leq \eta_i, \qquad I(\mathrm{X}', \mathrm{Y}_i; \mathrm{X}|\mathrm{S}') \leq \eta_i. \tag{4.54b}$$

Proof. The proof can be found in Appendix A.7. □

So far we defined coding and decoding rules. Next, we show that codewords of type P_{X} with properties as given in Lemma 4.41 and decoding sets as given in Definition 4.42 suffices to achieve all rate pairs as specified by the region $\mathcal{R}(P_{\mathrm{X}}|\Lambda)$, cf. (4.48).

Lemma 4.44. *Given $\Lambda > 0$ and arbitrarily small $\alpha > 0$, $\beta > 0$, and $\delta > 0$, for any type P_{X} satisfying*

$$\Lambda_i(P_{\mathrm{X}}) \geq \Lambda + \alpha, \ i = 1, 2, \qquad \min_{x \in \mathcal{X}} P_{\mathrm{X}}(x) \geq \beta,$$

there exist a code $\mathcal{C}_{det}(\mathfrak{W}^n)$ of block length $n \geq n_0$ with codewords $x^n_{m_1,m_2} \in \mathcal{T}^{(n)}_{\mathrm{X}}$, $m_1 = 1, ..., M^{(n)}_1$, $m_2 = 1, ..., M^{(n)}_2$, such that

$$\frac{1}{n} \log M^{(n)}_1 > \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_{\mathrm{X}}, \overline{W}_{2,q}) - \delta, \qquad \frac{1}{n} \log M^{(n)}_2 > \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) - \delta$$

while

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_i(s^n | \mathcal{C}_{det}(\mathfrak{W}^n)) \leq \exp(-n\gamma_i), \qquad i = 1, 2, \tag{4.55}$$

where n_0 and $\gamma_i > 0$ depend only on α, β, δ, and the AVBBC $\mathfrak{W}^n$.

Proof. The proof follows [CN88b, Lemma 5] and is similar to the proof for list decoding, cf. Lemma 4.31, if we restrict the list sizes to one, i.e., $L_1 = L_2 = 1$, and take the constraints into account. Therefore we omit the details for brevity. □

Converse

It remains to show that there are no other rate pairs achievable than these rate pairs which are already characterized by Theorem 4.37. If $\Lambda_i(P_{\mathrm{X}}) < \Lambda$, $i = 1, 2$, the converse is already established by Lemma 4.39. Consequently, we only need to consider the case where $\Lambda_i(P_{\mathrm{X}}) > \Lambda$, $i = 1, 2$, in the following.

Lemma 4.45. *For any $\Lambda > 0$, $\delta > 0$, and $\epsilon < 1$, there exists n_0 such that for any deterministic code $\mathcal{C}_{det}(\mathfrak{W}^n)$ of block length $n \geq n_0$ with $M_1^{(n)} M_2^{(n)}$ codewords, each of type P_{X}, satisfying*

$$\frac{1}{n} \log M_2^{(n)} \geq \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) + \delta$$

implies

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_1(s^n | \mathcal{C}_{det}(\mathfrak{W}^n)) > \epsilon.$$

And similarly, if the codewords satisfy $\frac{1}{n} \log M_1^{(n)} \geq \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{2,q}) + \delta$, we have $\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_2(s^n | \mathcal{C}_{det}(\mathfrak{W}^n)) > \epsilon$.

Proof. The proof follows [CN88b, Lemma 2] where a similar converse result is shown for the single-user case. We carry out the analysis for receiving node 1, then the result for receiving node 2 follows accordingly using the same argumentation.

Let us consider a joint probability distribution

$$P_{\mathrm{X}\mathrm{S}\mathrm{Y}_1}(x, s, y_1) = P_{\mathrm{X}}(x) q(s) W_1(y_1 | x, s). \tag{4.56}$$

If some probability distribution $q \in \mathcal{P}(\mathcal{S}, \Lambda)$ satisfies

$$\mathbb{E}_q[l(q)] \leq \Lambda(1 - \eta), \tag{4.57}$$

for some $\eta > 0$ which depends on δ but not on P_{X}, we have

$$I(\mathrm{X}; \mathrm{Y}_1) \leq \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) + \frac{\delta}{2}. \tag{4.58}$$

To prove (4.58) let $q^* \in \mathcal{P}(\mathcal{S}, \Lambda)$ be a probability distribution which achieves the infimum in $\inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q})$ so that we have $I(\mathrm{X}; \mathrm{Y}_1^*) = \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q})$ for $P_{\mathrm{X}\mathrm{S}^*\mathrm{Y}_1^*}$ as given in (4.56) with $\mathbb{E}_{q^*}[l(q^*)] \leq \Lambda$. Next, we use q^* to construct a new probability distribution with slightly smaller costs than Λ as required in (4.57). Therefore, let $s_0 \in \mathcal{S}$ with $l(s_0) = 0$ and define

$$q(s) := \begin{cases} (1 - \eta) q^*(s) & \text{if } s \neq s_0 \\ \eta + (1 - \eta) q^*(s) & \text{if } s = s_0. \end{cases}$$

Clearly, $q(s)$ satisfies (4.57), and therefore (4.58) holds for sufficiently small η, since $I(\mathrm{X};\mathrm{Y}_1)$ is a uniformly continuous in (P_X, q) if P_{XSY_1} is given as in (4.56).

Similarly as in [CN88b, Lemma 2], we now consider any deterministic code $\mathcal{C}_{\text{det}}(\mathfrak{W}^n)$ with codewords $x^n_{m_1,m_2}$, $m_1 = 1,...,M_1^{(n)}$, $m_2 = 1,...,M_2^{(n)}$, and decoding sets $\mathcal{D}^{(1)}_{m_2|m_1}$ and $\mathcal{D}^{(2)}_{m_1|m_2}$ for all $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$, cf. Definition 4.33. Further, let $\mathrm{S}^n = (\mathrm{S}_1,...,\mathrm{S}_n) \in \mathcal{S}^n$ be a sequence, where each element is independent and identically distributed according to q as constructed above. Then for receiving node 1 we get for each fixed $m_1 \in \mathcal{M}_1$ for the probability of error

$$\begin{aligned}
\mathbb{E}_q[\bar{e}_1(\mathrm{S}^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))] &= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathbb{E}_q[e_1((m_1,m_2),\mathrm{S}^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))] \\
&= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \sum_{y_1^n\notin\mathcal{D}^{(1)}_{m_2|m_1}} \mathbb{E}_q[W_1^n(y_1^n|x^n_{m_1,m_2},\mathrm{S}^n)] \\
&= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \sum_{y_1^n\notin\mathcal{D}^{(1)}_{m_2|m_1}} \prod_{k=1}^{n} \mathbb{E}_q[W_1(y_{1,k}|x_{m_1,m_2,k},\mathrm{S}_k)]. \qquad (4.59)
\end{aligned}$$

Next, we set

$$\overline{W}_{1,q}(y_1|x) = \mathbb{E}_q[W_1(y_1|x,s)] \qquad (4.60)$$

which is, in fact, a discrete memoryless channel (DMC). For each $m_1 \in \mathcal{M}_1$, (4.59) yields that $\mathbb{E}_q[\bar{e}_1(\mathrm{S}^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))] = \bar{e}_1(\overline{W}_{1,q}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))$ where $\bar{e}_1(\overline{W}_{1,q}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))$ is the average probability of error when the deterministic code $\mathcal{C}_{\text{det}}(\mathfrak{W}^n)$ is used on the DMC $\overline{W}_{1,q}$. Next, we observe that

$$\begin{aligned}
\mathbb{P}\big\{l(\mathrm{S}^n) > \Lambda\big\} &= \mathbb{P}\bigg\{\frac{1}{n}\sum_{k=1}^{n} l(\mathrm{S}_k) > \mathbb{E}_q[l(q)] + \eta\Lambda\bigg\} \\
&\le \frac{\big(\text{var}[l(q)]\big)^2}{n(\eta\Lambda)^2} \\
&\le \frac{l_{\text{max}}^2}{n\eta^2\Lambda^2}
\end{aligned}$$

which follows from (4.57), (4.46b), and Chebyshev's inequality so that we get

$$\begin{aligned}
\max_{s^n:l(s^n)\le\Lambda} \bar{e}_1(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) &\ge \mathbb{E}_q[\bar{e}_1(\mathrm{S}^n)] - \mathbb{P}\big\{l(\mathrm{S}^n) > \Lambda\big\} \\
&\ge \bar{e}_1(\overline{W}_{1,q}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) - \frac{l_{\text{max}}^2}{n\eta^2\Lambda^2}. \qquad (4.61)
\end{aligned}$$

Now, we are almost done. We observe that the definition of P_{XSY_1} as given in (4.58) implies that Y_1 is connected with X by the channel $\overline{W}_{1,q}$ as defined in (4.60). For such a DMC a strong converse in terms of maximal error can be found in [CK81] which immediately yields also a strong converse for the DMC in terms of average probability of error as needed here. In more detail, (4.59) implies, by the strong converse for a DMC with codewords of type P_{X} that if all codewords $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, each of type P_{X}, then, for each $m_1 \in \mathcal{M}_1$, the average probability of error $\bar{e}_1(\overline{W}_{1,q}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))$ is arbitrarily close to 1 if $\frac{1}{n}\log M_2^{(n)} \geq \inf_{q\in\mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) + \delta$ and n sufficiently large enough. Finally, this together with (4.61) completes the first part of the proof.

The result for receiving node 2 follows accordingly using the same argumentation which completes the proof of the lemma. □

Capacity Region

Summarizing the results obtained so far, we see that for given input distribution P_{X} the achievable rates for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ are given by $\mathcal{R}(P_{\mathrm{X}}|\Lambda)$ if $\Lambda_i(P_{\mathrm{X}}) > \Lambda$, $i = 1, 2$. Taking the union over all such valid inputs we finally obtain

$$\mathcal{R}_{\text{det}}(\mathfrak{W}^n|\Gamma,\Lambda) = \bigcup_{\substack{P_{\mathrm{X}}:\, g(P_{\mathrm{X}})\leq\Gamma,\\ \Lambda_i(P_{\mathrm{X}})>\Lambda, i=1,2}} \mathcal{R}(P_{\mathrm{X}}|\Lambda).$$

On the other hand, we have $\text{int}(\mathcal{R}_{\text{det}}(\mathfrak{W}^n|\Gamma,\Lambda)) = \emptyset$ if $\max_{P_{\mathrm{X}}:g(P_{\mathrm{X}})\leq\Gamma} \Lambda_1(P_{\mathrm{X}}) < \Lambda$ or $\max_{P_{\mathrm{X}}:g(P_{\mathrm{X}})\leq\Gamma} \Lambda_2(P_{\mathrm{X}}) < \Lambda$ which follows immediately from Lemma 4.39. This, indeed, establishes the main result of this work which is the deterministic code capacity region $\mathcal{R}_{\text{det}}(\mathfrak{W}^n|\Gamma,\Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ as stated in Theorem 4.37.

Remark 4.46. *The case where $\Lambda_i(P_{\mathrm{X}}) = \Lambda$, $i \in \{1, 2\}$, remains unsolved in a similar way as for the single-user AVC [CN88b]. Likewise, we expect that int$(\mathcal{R}(P_{\mathrm{X}}|\Lambda)) = \emptyset$ in that case.*

4.7 Unknown Varying Additive Interference

In this section we analyze the case where the transmission in the bidirectional broadcast phase is corrupted by unknown varying additive interference. Therefore, we also call this a *BBC with unknown varying interference*. Clearly, the interference at both receivers may differ so that we introduce two artificial interferers or *jammers*, one for each receiver, to model this scenario. Then the BBC with unknown varying interference $\mathfrak{W}^n$ is specified by

the flat fading input-output relation between the relay node and node i, $i = 1, 2$, which is given by

$$y_i = x + j_i + n_i.$$

Here, $y_i \in \mathbb{R}$ denotes the output at node i, $x \in \mathbb{R}$ the input, $j_i \in \mathbb{R}$ the additive interference, and $n_i \in \mathbb{R}$ the Gaussian noise of the channel distributed according to $\mathcal{N}(0, \sigma^2)$.

The transmit powers of the relay and of the artificial jammers are restricted by average power constraints Γ and Λ_i, $i = 1, 2$, respectively. This means, all permissible input sequences $x^n = (x_1, x_2, ..., x_n)$ of length n must satisfy

$$\frac{1}{n} \sum_{k=1}^{n} x_k^2 \leq \Gamma \tag{4.62}$$

and all permissible jamming sequences $j_i^n = (j_{i,1}, j_{i,2}, ..., j_{i,n})$, $i = 1, 2$, of length n must satisfy

$$\frac{1}{n} \sum_{k=1}^{n} j_{i,k}^2 \leq \Lambda_i. \tag{4.63}$$

From conditions (4.62) and (4.63) follow that all permissible codewords and interfering sequences lie on or within an n-dimensional sphere of radius $\sqrt{n\Gamma}$ or $\sqrt{n\Lambda_i}$, $i = 1, 2$, respectively.

Similarly as for the discrete memoryless AVBBC it makes a difference for the BBC with unknown varying interference, if we consider deterministic or random coding strategies. Hence, we want specify their different impact on the transmission in the following.

4.7.1 Traditional Interference Coordination

The traditional interference coordination is in general based on a system design which ensures that the interference at the receivers does not exceed a certain threshold. For example in current cellular networks this is realized by separating cells in space which operate at the same frequency.

Theorem 4.47. *The deterministic code capacity region $\mathcal{R}_{det}(\mathfrak{W}^n)$ of the BBC with unknown varying interference $\mathfrak{W}^n$ with input constraint Γ and jamming constraints Λ_1 and Λ_2 is the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy*

$$R_i \leq \begin{cases} \frac{1}{2} \log\left(1 + \frac{\Gamma}{\Lambda_i + \sigma^2}\right) & \text{if } \Gamma > \Lambda_i \\ 0 & \text{if } \Gamma \leq \Lambda_i \end{cases} \tag{4.64}$$

$i = 1, 2$. This means int$(\mathcal{R}_{det}(\mathfrak{W}^n)) \neq \emptyset$ if and only if $\Gamma > \Lambda_1$ and $\Gamma > \Lambda_2$.

First, we consider the case when $\Gamma \leq \Lambda_1$ or $\Gamma \leq \Lambda_2$. Let $x^n_{m_1,m_2} \in \mathbb{R}^n$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$ with $M_1^{(n)} \geq 2$ and $M_2^{(n)} \geq 2$ be arbitrary codewords satisfying the input constraint (4.62). For $\Gamma \leq \Lambda_1$ we can consider the jamming sequences $j^n_{1,m_1,m_2} = x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$. Then for each $m_1 \in \mathcal{M}_1$ at node 1 the following holds. For each pair $(k, l) \in \mathcal{M}_2 \times \mathcal{M}_2$ with $k \neq l$ we have for the probability of error at node 1

$$\begin{aligned}
&\mathbb{E}\big[e_1((m_1,k), j_l^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))\big] + \mathbb{E}\big[e_1((m_1,l), j_k^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))\big] \\
&\quad = \mathbb{P}\big\{x^n_{m_1,k} + j^n_{1,m_1,l} + n_1^n \notin \mathcal{D}^{(1)}_{k|m_1}\big\} + \mathbb{P}\big\{x^n_{m_1,l} + j^n_{1,m_1,k} + n_1^n \notin \mathcal{D}^{(1)}_{l|m_1}\big\} \\
&\quad = \mathbb{P}\big\{x^n_{m_1,k} + j^n_{1,m_1,l} + n_1^n \in (\mathcal{D}^{(1)}_{k|m_1})^c\big\} + \mathbb{P}\big\{x^n_{m_1,k} + j^n_{1,m_1,l} + n_1^n \notin \mathcal{D}^{(1)}_{l|m_1}\big\} \\
&\quad \geq \mathbb{P}\big\{x^n_{m_1,k} + j^n_{1,m_1,l} + n_1^n \in (\mathcal{D}^{(1)}_{k|m_1})^c\big\} + \mathbb{P}\big\{x^n_{m_1,k} + j^n_{1,m_1,l} + n_1^n \in \mathcal{D}^{(1)}_{k|m_1}\big\} \\
&\quad = \mathbb{P}\{x^n_{m_1,k} + j^n_{1,m_1,l} + n_1^n \in (\mathcal{D}^{(1)}_{k|m_1})^c \cup \mathcal{D}^{(1)}_{k|m_1}\} = 1.
\end{aligned}$$

Hence, for a fixed $m_1 \in \mathcal{M}_1$ this leads for the average probability of error to

$$\begin{aligned}
&\frac{1}{M_2^{(n)}} \sum_{k=1}^{M_2^{(n)}} \bar{e}_1(j^n_{1,m_1,k}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \\
&= \frac{1}{M_2^{(n)}} \frac{1}{M_1^{(n)} M_2^{(n)}} \sum_{k=1}^{M_2^{(n)}} \sum_{m_1'=1}^{M_1^{(n)}} \sum_{m_2'=1}^{M_2^{(n)}} e_1((m_1', m_2'), j^n_{1,m_1,k}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \\
&\geq \frac{1}{M_1^{(n)}(M_2^{(n)})^2} \sum_{m_1'}^{M_1^{(n)}} \frac{M_2^{(n)}(M_2^{(n)} - 1)}{2} \\
&= \frac{M_1^{(n)} M_2^{(n)}(M_2^{(n)} - 1)}{2M_1^{(n)}(M_2^{(n)})^2} = \frac{M_2^{(n)} - 1}{2M_2^{(n)}} \geq \frac{1}{4}.
\end{aligned}$$

This implies that $\bar{e}_1(j^n_{1,m_1,m_2}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \geq \frac{1}{4}$ for at least one $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$. Since the average probability of error is bounded from below by a positive constant, a reliable transmission from the relay to node 1 is not possible so that we end up with $R_1 = 0$. The case $\Gamma \leq \Lambda_2$ similarly leads to $R_2 = 0$.

With the following this becomes intuitively clear. Since we have $\Gamma \leq \Lambda_i$, it can happen that the interfering sequence looks like another valid codeword. Node i now receives a superposition of two codewords and cannot distinguish which of the codewords was transmitted by the relay and which was the interfering sequence. Thus, reliable communication can no longer be guaranteed.

Remark 4.48. *Interestingly, Theorem 4.47 shows that the existence of positive rates only depends on the interference and is completely independent of the noise. Consequently, the goal of the traditional interference coordination is to ensure that the received interference will be small enough. Otherwise, there is no communication possible, not even at very low rates.*

Now, we turn to the case when $\Gamma > \Lambda_1$ and $\Gamma > \Lambda_2$. To show that the rates given in (4.64) are actually achievable, we follow [CN91] where a similar result is proved for the corresponding single-user scenario. The strategy is outlined in the following.

Without loss of generality we assume that $\Gamma = 1$ and further $0 < \Lambda_i < 1$, $i = 1, 2$. Then it suffices to show that for every small $\delta > 0$ and sufficiently large n there exist $M_1^{(n)} M_2^{(n)}$ codewords $x^n_{m_1,m_2}$ (on the unit sphere) with $M_1^{(n)} = \exp(nR_2)$ and $M_2^{(n)} = \exp(nR_1)$ and $C_i - 2\delta < R_i < C_i - \delta$ with $C_i := \frac{1}{2}\log(1 + \frac{1}{\Lambda_i + \sigma^2})$, $i = 1, 2$, cf. (4.64), such that the average probability is arbitrarily small for all j_i^n satisfying (4.63). To ensure that the probability of error gets arbitrarily small, the codewords must possess certain properties which are guaranteed by the following lemma. This is a straightforward extension of the single-user case [CN91, Lemma 1] to the BBC with unknown varying interference.

Lemma 4.49. *For every $\epsilon > 0$, $8\sqrt{\epsilon} < \eta < 1$, $K > 2\epsilon$, and $M_1^{(n)} = \exp(nR_2)$, $M_2^{(n)} = \exp(nR_1)$ with $2\epsilon \le R_i \le K$, $i = 1, 2$, for $n \ge n_0(\epsilon, \eta, K)$ there exist unit vectors $x^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$ such that for every unit vector u^n and constants α, β in $[0, 1]$, we have for each $m_1 \in \mathcal{M}_1$*

$$\left|\left\{m_2 : \left\langle x^n_{m_1,m_2}, u^n \right\rangle \ge \alpha\right\}\right| \le \exp\left(n(|R_1 + \tfrac{1}{2}\log(1 - \alpha^2)|^+ + \epsilon)\right)$$

and, if $\alpha \ge \eta$, $\alpha^2 + \beta^2 > 1 + \eta - \exp(-2R_1)$

$$\frac{1}{M_2^{(n)}}\Big|\big\{\hat{m}_2 : |\left\langle x^n_{m_1,m_2}, x^n_{m_1,\hat{m}_2}\right\rangle| \ge \alpha, |\left\langle x^n_{m_1,m_2}, u^n\right\rangle| \ge \beta, \text{ for some } m_2 \ne \hat{m}_2\big\}\Big| \le \exp(-n\epsilon)$$

and similarly for each $m_2 \in \mathcal{M}_2$

$$\left|\left\{m_1 : \left\langle x^n_{m_1,m_2}, u^n \right\rangle \ge \alpha\right\}\right| \le \exp\left(n(|R_2 + \tfrac{1}{2}\log(1 - \alpha^2)|^+ + \epsilon)\right)$$

and, if $\alpha \ge \eta$, $\alpha^2 + \beta^2 > 1 + \eta - \exp(-2R_2)$

$$\frac{1}{M_1^{(n)}}\Big|\big\{\hat{m}_1 : |\left\langle x^n_{m_1,m_2}, x^n_{\hat{m}_1,m_2}\right\rangle| \ge \alpha, |\left\langle x^n_{m_1,m_2}, u^n\right\rangle| \ge \beta, \text{ for some } m_1 \ne \hat{m}_1\big\}\Big| \le \exp(-n\epsilon).$$

At the receiving nodes it suffices to use a minimum-distance decoder. Then for each $m_1 \in \mathcal{M}_1$ the decoding sets at node 1 and for each $m_2 \in \mathcal{M}_2$ at node 2 are given by

$$\mathcal{D}^{(1)}_{m_2|m_1} \coloneqq \{y_1^n : \|y_1^n - x^n_{m_1,m_2}\|^2 < \|y_1^n - x^n_{m_1,\hat{m}_2}\|^2 \text{ for all } \ m_2 \neq \hat{m}_2\} \tag{4.65a}$$

$$\mathcal{D}^{(2)}_{m_1|m_2} \coloneqq \{y_2^n : \|y_2^n - x^n_{m_1,m_2}\|^2 < \|y_2^n - x^n_{\hat{m}_1,m_2}\|^2 \text{ for all } \ m_1 \neq \hat{m}_1\}. \tag{4.65b}$$

With the presented coding and decoding rule, the probability of error gets arbitrarily small for increasing block length, which can be shown analogously to [CN91]. The details are rather technical and therefore omitted for brevity.

It remains to show that the described strategy is optimal, which means that no other rate pairs are achievable. From the previous discussions, cf. especially Remark 4.21, we already know that the capacity region of the deterministic code capacity region is included in the capacity region of the random code capacity region. In the next subsection, from Theorem 4.50 we see that for $\Gamma > \Lambda_i$, $i = 1, 2$, the maximal achievable rates for both strategies are equal. Since the described strategy already achieves these rates, the optimality is proved.

4.7.2 Relay-to-Receivers Coordination

Next, we study a strategy with a different degree of coordination. We assume that the relay and the receivers are synchronized in such a manner that they can coordinate their choice of the encoder and decoders based on an access to a common resource which is independent of the current message.

This can be realized by using a random code. If we transmit at rates R_1 and R_2 with exponentially many messages, i.e., $\exp(nR_1)$ and $\exp(nR_2)$, we know from [Ahl78] that it suffices to use a random code which consists of n^2 pairs of encoder and decoders and a uniformly distributed random variable whose value indicates which of the pair all nodes have to use. The access to the common random variable can be realized by an external source, e.g., a satellite signal, or a preamble prior to the transmission. Clearly, for sufficiently large block length the (polynomial) costs for the coordination are negligible. We call this *relay-to-receivers coordination*. Due to the more involved coordination we expect an improvement in the performance compared to the traditional coordination approach, especially for high interference.

Theorem 4.50. *The random code capacity region $\mathcal{R}_{ran}(\mathfrak{W}^n)$ of the BBC with unknown varying interference $\mathfrak{W}^n$ with input constraint Γ and jamming constraints Λ_1 and Λ_2 is the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy*

$$R_i \leq \frac{1}{2}\log\left(1 + \frac{\Gamma}{\Lambda_i + \sigma^2}\right), \quad i = 1, 2. \tag{4.66}$$

The theorem can be proved analogously to [HN87] where a similar result is proved for the single-user case. The random strategy which achieves the rates given in (4.66) is outlined in the following.

The codewords $x^n_{m_1,m_2}$ are uniformly distributed on the n-sphere of radius $\sqrt{n\Gamma}$. Similar to the traditional approach, a minimum-distance decoder as given in (4.65) at the receiving nodes is sufficient. It remains to show that for all rate pairs satisfying (4.66) the probability of error gets arbitrarily small for increasing block length. This can be done similarly to [HN87].

The optimality of the presented random strategy, which means that no other rate pairs are achievable, follows immediately from [HN87] and can be shown by standard arguments.

Remark 4.51. *The capacity region $\mathcal{R}_{ran}(\mathfrak{W}^n)$ is identical to the one if the interfering sequences would consist of iid Gaussian symbols distributed according to $\mathcal{N}(0, \Lambda_i)$, $i = 1, 2$. This means, the arbitrary, possibly non-Gaussian, unknown interference do not affect the achievable rates more than Gaussian noise of the same power.*

4.8 Discussion

In the previous Chapter 3 it has been shown that for compound channels communication in bidirectional relay networks is still possible, but at reduced rates compared to the case of perfect CSI. In this chapter it has been substantiated that for arbitrarily varying channels the impact is much more dramatic. More precisely, based on Ahlswede's elimination technique [Ahl78] we revealed the following dichotomy of the deterministic code capacity region of an AVBBC: it either equals its random code capacity region or else has an empty interior. Unfortunately, many channels of practical interest are symmetrizable which results in an ambiguity of the codewords at the receivers. These channels prohibit any reliable communication and therewith fall in the latter category.

As in [BNP95, Hug97] for the point-to-point AVC, the concept of list decoding is an adequate technique to dissolve the ambiguity caused by the symmetric channels and to enable reliable communication in bidirectional relay networks. The key idea for the analysis was to introduce a generalized notion of symmetrizability which distinguishes among different degrees of symmetry. This allowed us to reveal a connection between the degree of the symmetry of a channel and the needed list size at the decoder. It is shown that if a list size is greater than the symmetrizability of the channel, the decoder is able to successfully dissolve a possible ambiguity of the codewords. Fortunately, the symmetrizability of an AVBBC is always finite so that there are always finite list sizes at the receivers which permit reliable communication is scenarios, where usual deterministic decoding techniques fail.

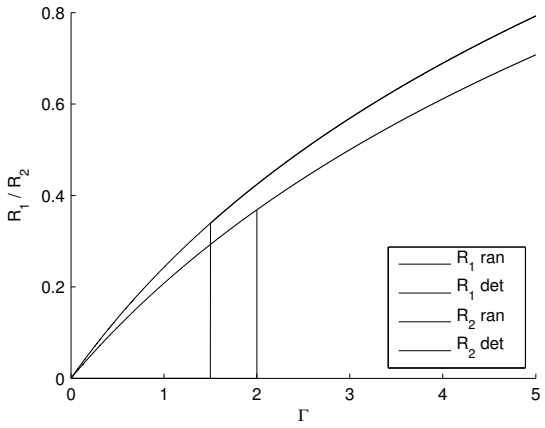


Figure 4.2: Achievable rates for the BBC with unknown varying interference for the traditional interference coordination (deterministic code) and relay-to-receivers coordination (random code) with $\Lambda_1 = 1.5$, $\Lambda_2 = 2$, and $\sigma^2 = 1$.

Imposing constraints on the permissible sequences of channel states reveals further phenomena. Now, even when the channel is symmetrizable, the deterministic code capacity region of the AVBBC under input and state constraints may be non-empty but less than its random code capacity region. Thereby, we observed that the constraints on the state sequences may reduce the deterministic code capacity region so that it is in general strictly smaller than the corresponding random code capacity region, but they preserve the general dichotomy behavior of the deterministic code capacity region: it still equals either a non-empty region or else has an empty interior. Although the deterministic code capacity region displays a dichotomy behavior, it cannot be exploited to prove the corresponding capacity region since Ahlswede's elimination technique [Ahl78] does not work anymore in the presence of constraints on input and states, cf. also [LN98]. This necessitated a proof technique which does not rely on the dichotomy behavior and is based on an idea of Csiszár and Narayan [CN88b].

Figure 4.2 depicts the maximal achievable rates for the case where the transmission is corrupted by unknown varying additive interference. If the power of the interference is greater than the transmit power of the relay, the interference can look like other valid codewords and the receivers cannot reliably distinguish between the intended signal and the interference anymore. Consequently, a traditional interference coordination based on a deterministic coding strategy is only reasonable if the interference can be made small enough, since it treats the interference as some kind of additional Gaussian noise. Thus, especially in the case of high interference, a more sophisticated coordination based on a random coding strategy is needed for reliable communication. It is shown that a coordination of the encoder and decoders based on a common resource is sufficient to handle the interference even if it is stronger than the desired signal.

5 Physical Layer Service Integration in Bidirectional Relay Networks

Recently, significant progress has been made in improving the performance of next generation cellular networks. Proposed techniques such as multiuser MIMO, channel adaptive scheduling, cooperative multi-point transmission, or relaying can increase the spectral efficiency.

An additional research area that is gaining importance is the efficient implementation of certain services at the physical layer. For example, in current cellular systems, operators offer not only traditional services such as (bidirectional) voice communication, but also further multicast services or confidential services that are subject to certain secrecy constraints. Nowadays, the integration of multiple services is realized by policies that allocate different services on different logical channels and by applying secrecy techniques at higher levels. In general, this is quite inefficient, and there is a trend to merge multiple coexisting services efficiently from an information theoretic point of view so that they work on the same wireless resources. This is referred to as *physical layer service integration* and has the potential to significantly increase the spectral efficiency for next generation wireless networks and, especially, 5G cellular networks. Accordingly, this is being intensively discussed at the moment by the 3rd Generation Partnership Projects Long-Term Evolution Advanced (3GPP LTE-Advanced) group.

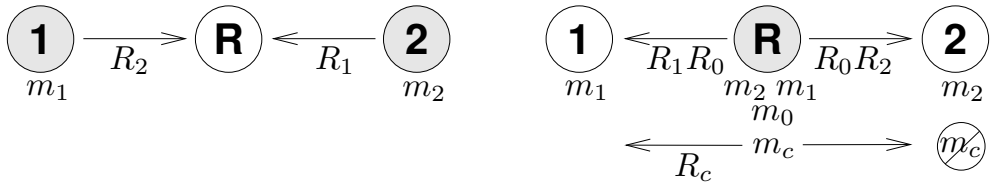
Multicast services can be realized efficiently by common messages; for example the Multimedia Broadcast Multicast Service (MBMS), as specified by the 3GPP organization, benefits from such studies. This substantiates the concern of merging such services efficiently at the physical layer to advantageously exploit the broadcast nature of the wireless medium.

Some work on the SISO Gaussian broadcast channel with common messages and certain side information at the receivers can be found in [Wu07] and [KS07] where the latter assumes degraded message sets. The throughput region of bidirectional multi-hop fading networks with common messages is analyzed in [IS09]. A general model for multi-user settings with correlated sources is given in [GEGP09]. The general broadcast channel with common messages is analyzed in [Tia09] in terms of latent capacity, where the author shows that the achievability of a certain rate vector immediately implies the achievability of a whole non-trivial rate region. However, only the case of symmetric rates for all users is discussed.

All these services are not required to be kept secret from non-legitimate receivers so that they are classified as *public services*. Accordingly, services that have this additional secrecy requirement are classified as *confidential services*. Currently, secrecy techniques usually rely on the assumption of the unproven hardness of certain problems or insufficient computational capabilities of non-legitimate receivers. In contrast, physical layer secrecy techniques do not rely on such assumptions and therefore provide so-called unconditional security, which makes them more and more attractive. This becomes even more important in wireless networks, since, due to the broadcast nature of the wireless medium, a transmitted signal is received by the intended user but can also be overheard by non-intended users. This necessitates the design of systems that enable secure communication to certain legitimate users while keeping non-legitimate users ignorant of the transmission.

In the seminal work [Wyn75] Wyner introduced the *wiretap channel* which characterizes the secure communication problem for a point-to-point link with an additional eavesdropper. Csiszár and Körner generalized this to the *broadcast channel with confidential messages* in [CK78] and characterized the optimal integration of common and confidential services at the physical layer. Recently, there has been growing interest in physical layer secrecy, for current surveys we refer, for example, to [LPS09, LT10, JWG10, BB11] and references therein. Besides the (wireless) point-to-point link [Wyn75, BBRM08, LS09, KW10b, KW10a], there are extensions to multi-user settings as the multiple access channel with confidential messages [LP08], the multiple access wiretap channel [EU08a, TY08], the interference channel with confidential messages [LMSY08], the multi-antenna Gaussian broadcast channel with confidential messages [LP09, LLL10, LLPS10b], the MIMO Gaussian broadcast channel with common and confidential messages [LLPS10a, EU10a], secure communication with relays [EU08b, HY10b], or the two-way wiretap channel [HY10a, EKYE10]. Secrecy for fading channels is discussed for example in [LPS08, KGLP11]. It is shown that secrecy can be improved by cooperation [MYP11] and helping interference [Jor10, TLSP11].

In this chapter we consider physical layer service integration in bidirectional relay networks. Here, the relay node integrates additional common and confidential services in the broadcast phase. More precisely, in addition to the transmission of both individual messages the relay node has the following tasks as shown in Figure 5.1: the transmission of a common message to both nodes and further, the transmission of a confidential message to one node, which should be kept secret from the other, non-legitimate node. This requires the study of the *bidirectional broadcast channel (BBC) with common and confidential messages* which is introduced in Section 5.1. We start with the scenario where the relay integrates only public services and analyze the corresponding BBC with common messages in Section 5.2. The case where the relay integrates confidential messages and no common messages, i.e., the BBC with confidential messages, is addressed in Section 5.3, while Section 5.4 finally considers the most general scenario, i.e., the BBC with common and confidential messages. Then, we analyze the integration of confidential services for MIMO Gaussian channels in Section 5.5 and end a discussion in Section 5.6.



(a) MAC phase

(b) BBC phase

Figure 5.1: Physical layer service integration in decode-and-forward bidirectional relay networks. In the BBC phase, the relay forwards the messages m_1 and m_2 and adds a common message m_0 with rate R_0 to the communication and further a confidential message m_c for node 1 with rate R_c which should be kept secret from node 2.

5.1 Bidirectional Broadcast Channel with Common and Confidential Messages

Let $\mathcal{X}$ and $\mathcal{Y}_i$, $i = 1, 2$, be finite input and output sets. Then for input and output sequences $x^n \in \mathcal{X}^n$ and $y_i^n \in \mathcal{Y}_i^n$, $i = 1, 2$, of length n, the discrete memoryless broadcast channel is given by $W^{\otimes n}(y_1^n, y_2^n|x^n) := \prod_{k=1}^n W(y_{1,k}, y_{2,k}|x_k)$. Since we do not allow any cooperation between the receiving nodes, it is sufficient to consider the marginal transition probabilities $W_i^{\otimes n} := \prod_{k=1}^n W_i(y_{i,k}|x_k)$, $i = 1, 2$ only.

We consider the standard model with a block code of arbitrary but fixed length n. The set of individual messages of node i, $i = 1, 2$, is denoted by $\mathcal{M}_i := \{1, ..., M_i^{(n)}\}$, which is also known at the relay node. Further, the sets of common and confidential messages of the relay node are denoted by $\mathcal{M}_0 := \{1, ..., M_0^{(n)}\}$ and $\mathcal{M}_c := \{1, ..., M_c^{(n)}\}$, respectively. We use the abbreviation $\mathcal{M}_p := \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2$ for the set of all public messages and further $\mathcal{M} := \mathcal{M}_c \times \mathcal{M}_p$.

Definition 5.1. *An* $(n, M_c^{(n)}, M_0^{(n)}, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{m_c,m_0}(W)$ *for the BBC with common and confidential messages consists of one (stochastic) encoder at the relay node*

$$f : \mathcal{M}_c \times \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n$$

and decoders at nodes 1 and 2

$$g_1 : \mathcal{Y}_1^n \times \mathcal{M}_1 \to \mathcal{M}_c \times \mathcal{M}_0 \times \mathcal{M}_2 \cup \{0\}$$
$$g_2 : \mathcal{Y}_2^n \times \mathcal{M}_2 \to \mathcal{M}_0 \times \mathcal{M}_1 \cup \{0\}$$

where the element 0 *in the definition of the decoders plays the role of an erasure symbol and is included for convenience.*

Secure communication may benefit from randomized encoding [CK78, LPS09] so that we allow the encoder f to be stochastic. More precisely, to transmit message $m = (m_c, m_0, m_1, m_2) \in \mathcal{M}$ the corresponding codeword $x^n \in \mathcal{X}^n$ is specified by conditional probabilities $f(x^n|m)$ with $\sum_{x^n \in \mathcal{X}^n} f(x^n|m) = 1$. This means $f(x^n|m)$ is the probability that message $m \in \mathcal{M}$ is encoded as $x^n \in \mathcal{X}^n$.

The quality of such a code is measured by two performance criteria. First, each receiver should successfully decode its intended messages, i.e., the corresponding average probabilities of decoding errors have to be small. In more detail, when the relay has sent the message $m = (m_c, m_0, m_1, m_2)$, and nodes 1 and 2 have received y_1^n and y_2^n, the decoder at node 1 is in error if $g_1(y_1^n, m_1) \neq (m_c, m_0, m_2)$. Accordingly, the decoder at node 2 is in error if $g_2(y_2^n, m_2) \neq (m_0, m_1)$. Then, the average probability of error at node i, $i = 1, 2$ is given by

$$\bar{e}_i \coloneqq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_i(m)$$

where $e_1(m)$ denotes the probability that the decoder at node 1 decodes m_c, m_0, and m_2 incorrectly, and $e_2(m)$ the probability that the decoder at node 2 decodes m_0 and m_1 incorrectly, i.e., $e_1(m) = \mathbb{P}\{g_1(y_1^n, m_1) \neq (m_c, m_0, m_2)|m \text{ has been sent}\}$ and $e_2(m) = \mathbb{P}\{g_2(y_2^n, m_2) \neq (m_0, m_1)|m \text{ has been sent}\}$.

The second criterion is security. Similarly as for example in [Wyn75, CK78] we characterize the secrecy level of the confidential message $m_c \in \mathcal{M}_c$ by the concept of equivocation. Here, the equivocation $H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2)$ describes the uncertainty of node 2 about the confidential message M_c having the received sequence Y_2^n and its own message M_2 as side information available under the assumption that the random variables M_c and M_2 are uniformly distributed over $\mathcal{M}_c$ and $\mathcal{M}_2$. Consequently, the higher the equivocation is, the higher the secrecy level of the confidential message is.

Remark 5.2. *Note that we have to deal with a non-standard encoder/decoder design. Besides the classical task of establishing reliable decoding of the public services at the legitimate receivers, it further has to protect the confidential services from non-legitimate receivers.*

Definition 5.3. *A rate-equivocation tuple* $\boldsymbol{R} = (R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ *is said to be* achievable *for the BBC with common and confidential messages if for any* $\delta > 0$ *there is an* $n(\delta) \in \mathbb{N}$ *and a sequence* $\{\mathcal{C}_{m_c,m_0}^{(n)}(W)\}_{n \in \mathbb{N}}$ *of* $(n, M_c^{(n)}, M_0^{(n)}, M_1^{(n)}, M_2^{(n)})$*-codes such that for all* $n \geq n(\delta)$ *we have*

$$\frac{1}{n} \log M_c^{(n)} \geq R_c - \delta, \quad \frac{1}{n} \log M_0^{(n)} \geq R_0 - \delta,$$
$$\frac{1}{n} \log M_2^{(n)} \geq R_1 - \delta, \quad \frac{1}{n} \log M_1^{(n)} \geq R_2 - \delta,$$

and

$$\tfrac{1}{n}H(\mathrm{M}_c|\mathrm{Y}_2^n,\mathrm{M}_2) \geq R_e - \delta \tag{5.1}$$

while $\bar{e}_1, \bar{e}_2 \to 0$ *as* $n \to \infty$. *The set of all achievable rate tuples is the* capacity-equivocation region *of the BBC with common and confidential messages and is denoted by* $\mathcal{R}_{m_c,m_0}(W)$.

Remark 5.4. *Definition 5.3 includes the case where the confidential rate is higher than the equivocation rate, i.e.,* $R_c > R_e$. *The notion of* perfect secrecy *requires the equivocation rate to be as high as the rate of the confidential message, i.e.,* $R_c = R_e$. *Then, condition (5.1) becomes* $\tfrac{1}{n}H(\mathrm{M}_c|\mathrm{Y}_2^n,\mathrm{M}_2) \geq R_c - \delta$ *which is often equivalently written as*

$$\tfrac{1}{n}I(\mathrm{M}_c;\mathrm{Y}_2^n|\mathrm{M}_2) \leq \delta. \tag{5.2}$$

The definition of the secrecy capacity region $\mathcal{R}^S_{m_c,m_0}(W)$ *of the BBC with common and confidential messages follows immediately.*

5.2 Integration of Common Messages

We start with the scenario where the relay integrates only common messages and no confidential messages. This is the *bidirectional broadcast channel (BBC) with common messages*.

5.2.1 Capacity Region for Discrete Memoryless Channels

The definition of an $(n, M_0^{(n)}, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{m_0}(W)$ can be deduced from Definition 5.1. Similarly, the definitions of an achievable rate triple and the capacity region $\mathcal{R}_{m_0}(W)$ of the BBC with common messages follow immediately from Definition 5.3.

Theorem 5.5. *The capacity region* $\mathcal{R}_{m_0}(W)$ *of the BBC with common messages is the set of all rate triples* $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ *that satisfy*

$$R_0 + R_i \leq I(\mathrm{X};\mathrm{Y}_i|\mathrm{U}), \quad i = 1, 2 \tag{5.3}$$

for random variables $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ *with joint probability distribution* $p_{\mathrm{U}}(u)p_{\mathrm{X}|\mathrm{U}}(x|u)W(y_1, y_2|x)$. *Thereby,* U *is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of* U *can be bounded by* $|\mathcal{U}| \leq 3$.

Remark 5.6. *Clearly, the sum constraints in (5.3) immediately imply that the rate of the common message has to fulfill* $R_0 \leq \min\{I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}), I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})\}$.

Remark 5.7. *Similarly as in [KS07, Theorem 1] it is further possible to get rid of the time-sharing random variable* U *in (5.3) so that the region is given by all rate triples* $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ *that satisfy* $R_0 + R_i \leq I(\mathrm{X}; \mathrm{Y}_i)$*,* $i = 1, 2$*.*

Remark 5.8. *At a first glance, (5.3) suggests a rate splitting approach between the common rate and the individual rates, but one has to be careful since the coding strategy has to be designed in such a way that the common message can be decoded at both receivers. This observation reveals some interesting connections to compound channels [BBT59, Wol60, Wol78], cf. also Chapter 3, where the coding strategy has to ensure that the message to transmit is decodable for a whole set of possible channels.*

Theorem 5.5 is proved in the following.

Proof of Achievability

Here, we present a construction of a coding strategy that achieves all rate triples $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ that satisfy

$$R_0 + R_i \leq I(\mathrm{X}; \mathrm{Y}_i), \quad i = 1, 2 \tag{5.4}$$

for some $p_{\mathrm{X}}(x)W(y_1, y_2|x)$. Then the desired region (5.3) is determined by establishing the convex hull by first introducing an auxiliary random variable U and applying standard arguments as in [OSBB08]. Similarly it follows then from Fenchel-Bunt's extension of Carathéodory's theorem [HUL01] that any rate triple is achievable by time-sharing between three rate triples, i.e., $|\mathcal{U}| \leq 3$ is enough.

The construction is mainly based on the idea of [Cov72] for the classical broadcast channel with common messages, where the whole information sent to each receiver is split into an individual part and a common part. We use this idea to extend the proof idea for the achievability for the BBC without common messages, cf. [OSBB08], to our scenario.

Let us recapitulate the broadcast situation that is considered here. The relay node wants to transmit a common message $m_0 \in \mathcal{M}_0$ with rate R_0 and individual messages $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$ with rates R_2 and R_1, respectively. Node 1 knows its own message m_1 that it transmitted in the previous MAC phase and wants to recover the common message m_0 and the individual message m_2. Similarly, node 2 knows m_2 and wants to recover m_0 and m_1. Having [Cov72] in mind the broadcast situation can also be interpreted in a slightly different way by combining the desired individual messages and the common message. In more detail, node 1 knows its own message m_1 and is interested in the (combined) individual message $m_2' = (m_0, m_2) \in \mathcal{M}_0 \times \mathcal{M}_2 =: \mathcal{M}_2'$ with rate $R_1' = R_0 + R_1$ and, similarly, node 2 knows m_2 and is interested in $m_1' = (m_0, m_1) \in \mathcal{M}_0 \times \mathcal{M}_1 =: \mathcal{M}_1'$ with rate $R_2' = R_0 + R_2$.

Basically, we see that due to this reinterpretation the problem of coding for the BBC with common messages becomes the problem of coding for the BBC without common messages. The only difference is that while in the classical BBC without common messages each receiving node has complementary side information, i.e., it knows exactly the message the other one is interested in, in our scenario each receiving node knows only a part of the information the other one is interested in, i.e., it has only the other individual message and not the common message as side information available. We see that our scenario is not precisely included in [OSBB08], but it is a straightforward extension. Therefore, we go through the proof of achievability and sketch only the differences to [OSBB08, Sec. II-A].

Similarly as in [OSBB08, Sec. II-A] we show by random coding arguments that for given $p_{\mathrm{X}}(x)W(y_1, y_2|x)$ there exists a coding strategy such that all rate pairs $(R'_1, R'_2) \in \mathbb{R}^2_+$ with $R'_i \leq I(\mathrm{X};\mathrm{Y}_i)$, i.e., satisfying $R'_i = R_0 + R_i \leq I(\mathrm{X};\mathrm{Y}_i)$, $i = 1, 2$, cf. also (5.4), are achievable. Therefore, we generate $|\mathcal{M}_p| = |\mathcal{M}_0||\mathcal{M}_1||\mathcal{M}_2|$ independent codewords $x^n_{m_p} \in \mathcal{X}^n$ with $m_p = (m_0, m_1, m_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 = \mathcal{M}_p$ and $M_0^{(n)} := 2^{nR_0}$, $M_1^{(n)} := 2^{nR_2}$, and $M_2^{(n)} := 2^{nR_1}$ according to $p_{\mathrm{X}^n}(x^n) = \prod_{k=1}^{n} p_{\mathrm{X}}(x_k)$. Each receiving node uses typical set decoding in a similar way as in [OSBB08, Sec. II-A]. Now, it is straightforward to show that the probability of a decoding error, averaged over all codewords and all codebooks, at receiving node 1 gets arbitrarily small if the rate of the intended (combined) message $m'_2 = (m_0, m_2) \in \mathcal{M}_0 \times \mathcal{M}_2 = \mathcal{M}'_2$ fulfills $R'_1 = R_0 + R_1 \leq I(\mathrm{X};\mathrm{Y}_1)$. Clearly, the same is also true for receiving node 2 which is able to determine $m'_1 = (m_0, m_1) \in \mathcal{M}_0 \times \mathcal{M}_1 = \mathcal{M}'_1$ if $R'_2 = R_0 + R_2 \leq I(\mathrm{X};\mathrm{Y}_2)$. With the (combined) individual messages $m'_1 \in \mathcal{M}'_1$ and $m'_2 \in \mathcal{M}'_2$ with rates R'_2 and R'_1 the receiving nodes immediately obtain the common message $m_0 \in \mathcal{M}_0$ with rate R_0 and the individual messages $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$ with rate $R'_2 - R_0$ and $R'_1 - R_0$, respectively. Thus, similar to [Cov72, Tia09], all rate triples $(R_0, R'_1 - R_0, R'_2 - R_0) = (R_0, R_1, R_2)$ with $R_0 + R_i \leq I(\mathrm{X};\mathrm{Y}_i)$ are achievable for the BBC with common messages which already proves the achievability. □

Proof of Weak Converse

We have to show that for any given sequence $\{\mathcal{C}^{(n)}_{m_0}(W)\}_{n\in\mathbb{N}}$ of $(n, M_0^{(n)}, M_1^{(n)}, M_2^{(n)})$-codes with $\bar{e}_1, \bar{e}_2 \to 0$ there exist random variables $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ such that

$$\begin{aligned}\tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_2)\big) &\leq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) + o(n^0)\\ \tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_1)\big) &\leq I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) + o(n^0)\end{aligned}$$

are satisfied. For this purpose we need a version of Fano's lemma suitable for the BBC with common messages.

Lemma 5.9. *For the BBC with common messages we have the following versions of Fano's inequality*

$$H(\mathrm{M}_0, \mathrm{M}_2|\mathrm{Y}_1^n, \mathrm{M}_1) \leq \bar{e}_1 \log(M_0^{(n)} M_2^{(n)}) + 1 = n\epsilon_1^{(n)}$$
$$H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) \leq \bar{e}_2 \log(M_0^{(n)} M_1^{(n)}) + 1 = n\epsilon_2^{(n)}$$

with $\epsilon_1^{(n)} = \frac{1}{n}\log(M_0^{(n)} M_2^{(n)})\bar{e}_1 + \frac{1}{n} \to 0$ *and* $\epsilon_2^{(n)} = \frac{1}{n}\log(M_0^{(n)} M_1^{(n)})\bar{e}_2 + \frac{1}{n} \to 0$ *for* $n \to \infty$ *as* $\bar{e}_1, \bar{e}_2 \to 0$.

Proof. In Appendix A.8 we prove Fano's inequality for the BBC with common and confidential messages. The case with only common messages can easily be deduced from this. □

With this, we can bound $H(\mathrm{M}_0) + H(\mathrm{M}_2)$ as follows

$$\begin{aligned} H(\mathrm{M}_0) + H(\mathrm{M}_2) &= H(\mathrm{M}_0|\mathrm{M}_1, \mathrm{M}_2) + H(\mathrm{M}_2|\mathrm{M}_1) \\ &= H(\mathrm{M}_0, \mathrm{M}_2|\mathrm{M}_1) \\ &\leq I(\mathrm{M}_0, \mathrm{M}_2; \mathrm{Y}_1^n|\mathrm{M}_1) + n\epsilon_1^{(n)} \\ &\leq I(\mathrm{M}_0, \mathrm{M}_1, \mathrm{M}_2; \mathrm{Y}_1^n) + n\epsilon_1^{(n)} \\ &\leq I(\mathrm{X}^n; \mathrm{Y}_1^n) + n\epsilon_1^{(n)} \end{aligned} \tag{5.5}$$

where the equalities and inequalities follow from the independence of M_0, M_1, and M_2, the chain rule for entropy, the definition of mutual information, Lemma 5.9, the chain rule for mutual information, the positivity of mutual information, and the data processing inequality. Using the definition of mutual information and dividing by n we get for the rates

$$\begin{aligned} \frac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_2)\big) &\leq \frac{1}{n}\big(H(\mathrm{Y}_1^n) - H(\mathrm{Y}_1^n|\mathrm{X}^n)\big) + \epsilon_1^{(n)} \\ &\leq \frac{1}{n}\sum_{k=1}^{n}\big(H(\mathrm{Y}_{1,k}|\mathrm{Y}_1^{k-1}) - H(\mathrm{Y}_{1,k}|\mathrm{Y}_1^{k-1}, \mathrm{X}_k)\big) + \epsilon_1^{(n)} \\ &\leq \frac{1}{n}\sum_{k=1}^{n}\big(H(\mathrm{Y}_{1,k}) - H(\mathrm{Y}_{1,k}|\mathrm{X}_k)\big) + \epsilon_1^{(n)} \\ &= \frac{1}{n}\sum_{k=1}^{n} I(\mathrm{X}_k; \mathrm{Y}_{1,k}) + \epsilon_1^{(n)} \end{aligned} \tag{5.6}$$

using the chain rule for entropy, the memoryless property of the channel, and again the definition of mutual information. Accordingly, using the same arguments we also obtain

$$\frac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_1)\big) \leq \frac{1}{n}\sum_{k=1}^{n} I(\mathrm{X}_k; \mathrm{Y}_{2,k}) + \epsilon_2^{(n)}. \tag{5.7}$$

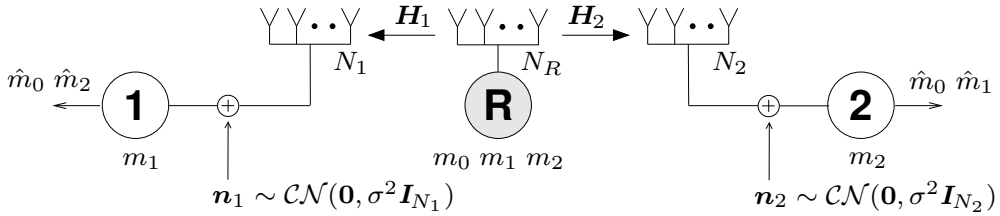


Figure 5.2: MIMO Gaussian BBC with common messages.

Similarly as in [OSBB08, Sec. II-B] for the BBC without common messages we introduce an auxiliary random variable U that is independent of M_0, M_1, M_2, X^n, Y_1^n, and Y_2^n and uniformly distributed over $\{1, ..., n\}$. Further, let $\mathrm{X} \coloneqq \mathrm{X}_{\mathrm{U}}$, $\mathrm{Y}_1 \coloneqq \mathrm{Y}_{1,\mathrm{U}}$, and $\mathrm{Y}_2 \coloneqq \mathrm{Y}_{2,\mathrm{U}}$. With this, for (5.6) and (5.7) we get

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} I(\mathrm{X}_k; \mathrm{Y}_{i,k}) &= \sum_{k=1}^{n} \mathbb{P}\{\mathrm{U}=k\} I(\mathrm{X}_k; \mathrm{Y}_{i,k}|\mathrm{U}=k) \\ &= I(\mathrm{X}_{\mathrm{U}}; \mathrm{Y}_{i,\mathrm{U}}|\mathrm{U}) \\ &= I(\mathrm{X}; \mathrm{Y}_i|\mathrm{U})\end{aligned}$$

which establishes the desired rates as states in (5.3). This completes the proof of the weak converse. □

5.2.2 Capacity Region for MIMO Gaussian Channels

We assume N_R antennas at the relay node and N_i antennas at node i, $i = 1, 2$, as shown in Figure 5.2. In the bidirectional broadcast phase, the discrete-time complex-valued input-output relation between the relay node and node i, $i = 1, 2$, is given by

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x} + \boldsymbol{n}_i, \tag{5.8}$$

where $\boldsymbol{y}_i \in \mathbb{C}^{N_i \times 1}$ denotes the output at node i, $\boldsymbol{H}_i \in \mathbb{C}^{N_i \times N_R}$ the multiplicative channel matrix, $\boldsymbol{x} \in \mathbb{C}^{N_R \times 1}$ the input of the relay node, and $\boldsymbol{n}_i \in \mathbb{C}^{N_i \times 1}$ the independent additive noise according to a circular symmetric complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I}_{N_i})$. We assume perfect channel state information at all nodes and an average transmit power constraint $\mathrm{tr}(\boldsymbol{Q}) \leq P$ with $\boldsymbol{Q} = \mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^H\}$.

Theorem 5.10. *The capacity region $\mathcal{R}_{m_0}(\boldsymbol{H}_1, \boldsymbol{H}_2|P)$ of the MIMO Gaussian BBC with common messages and average power constraint P is the set of all rate triples $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ that satisfy*

$$\begin{aligned}R_0 + R_1 &\leq \log\det(\boldsymbol{I}_{N_1} + \tfrac{1}{\sigma^2}\boldsymbol{H}_1 \boldsymbol{Q} \boldsymbol{H}_1^H) \\ R_0 + R_2 &\leq \log\det(\boldsymbol{I}_{N_2} + \tfrac{1}{\sigma^2}\boldsymbol{H}_2 \boldsymbol{Q} \boldsymbol{H}_2^H)\end{aligned} \tag{5.9}$$

for some $\boldsymbol{Q} \succeq \mathbf{0}$ *with* $tr(\boldsymbol{Q}) \leq P$.

Since the $\log\det$ function is concave in $\boldsymbol{Q}$ [HJ99, Theorem 7.6.7], the region in (5.9) is already convex. Hence, an auxiliary random variable that realizes an additional time-sharing operation is not necessary since such an operation will not enlarge the region.

Proof of Achievability

To show the achievability of all rate triples $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ satisfying (5.9) for a given covariance matrix $\boldsymbol{Q}$, we follow the proof of achievability of its discrete counterpart, cf. Section 5.2.1. The only difference to the discrete case is that we generate each entry of all codewords independently according to $\mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$. With this, the achievability of all rate triples satisfying (5.9) is straightforward to show. Then the desired region given in Theorem 5.10 is immediately obtained by taking the union over all covariance matrices that satisfy the input power constraint which finishes the proof of achievability. □

Proof of Weak Converse

We have to show that for any given sequence of $(n, M_0^{(n)}, M_1^{(n)}, M_2^{(n)})$-codes with $\bar{e}_1, \bar{e}_2 \to 0$ there exists a covariance matrix $\boldsymbol{Q}$ satisfying the average power constraint $\mathrm{tr}(\boldsymbol{Q}) \leq P$ such that

$$\begin{aligned}\tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_2)\big) &\leq \log\det(\boldsymbol{I}_{N_1} + \tfrac{1}{\sigma^2}\boldsymbol{H}_1\boldsymbol{Q}\boldsymbol{H}_1^H) + o(n^0)\\ \tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_1)\big) &\leq \log\det(\boldsymbol{I}_{N_2} + \tfrac{1}{\sigma^2}\boldsymbol{H}_2\boldsymbol{Q}\boldsymbol{H}_2^H) + o(n^0)\end{aligned}$$

are satisfied.

Following the proof of its discrete counterpart, cf. Section 5.2.1, it is straightforward to show that we can bound the entropies by

$$H(\mathrm{M}_0) + H(\mathrm{M}_2) \leq I(\mathbf{X}^n; \mathbf{Y}_1^n) + n\epsilon_1^{(n)} \tag{5.10a}$$

$$H(\mathrm{M}_0) + H(\mathrm{M}_1) \leq I(\mathbf{X}^n; \mathbf{Y}_2^n) + n\epsilon_2^{(n)} \tag{5.10b}$$

cf also (5.5). Note that this immediately implies that $H(\mathrm{M}_0) \leq \min\{I(\mathbf{X}^n; \mathbf{Y}_1^n) + n\epsilon_1^{(n)}, I(\mathbf{X}^n; \mathbf{Y}_2^n) + n\epsilon_2^{(n)}\}$, cf. also Remark 5.6.

For the rest of the proof it remains to bound the term $I(\mathbf{X}^n; \mathbf{Y}_i^n)$, $i = 1, 2$, in such a way that we obtain the expected $\log\det$ expression. With the definition of the mutual information and

the memoryless property of the channel, we have

$$\begin{aligned} I(\mathbf{X}^n;\mathbf{Y}_i^n) &= h(\mathbf{Y}_i^n) - h(\mathbf{Y}_i^n|\mathbf{X}^n) \\ &\leq \sum_{k=1}^{n} \big(h(\mathbf{Y}_{i,k}) - h(\mathbf{Y}_{i,k}|\mathbf{X}_k)\big) \\ &= \sum_{k=1}^{n} I(\mathbf{Y}_{i,k};\mathbf{X}_k) \\ &= \sum_{k=1}^{n} \big(h(\mathbf{Y}_{i,k}) - h(\mathbf{N}_{i,k})\big) \end{aligned}$$

with $\mathbf{Y}_{i,k} = \boldsymbol{H}_i\mathbf{X}_k + \mathbf{N}_{i,k}$, $i = 1,2$. Note that the random variables $\mathbf{X}_k$ and $\mathbf{N}_{i,k}$ with $h(\mathbf{N}_{i,k}) = \log\det(\pi\,\mathrm{e}\,\sigma^2\boldsymbol{I}_{N_i})$ are independent. Then, from the entropy maximization theorem follows that $h(\mathbf{Y}_{i,k}) \leq \log\det(\pi\,\mathrm{e}(\sigma^2\boldsymbol{I}_{N_i} + \boldsymbol{H}_i\boldsymbol{Q}_k\boldsymbol{H}_i^H))$ with equality if the input is Gaussian, i.e., $\mathbf{X}_k \sim \mathcal{CN}(\mathbf{0},\boldsymbol{Q}_k)$. With this, we have $h(\mathbf{Y}_{i,k}) - h(\mathbf{N}_{i,k}) \leq \log\det(\boldsymbol{I}_{N_i} + \frac{1}{\sigma^2}\boldsymbol{H}_i\boldsymbol{Q}_k\boldsymbol{H}_i^H)$ which immediately implies

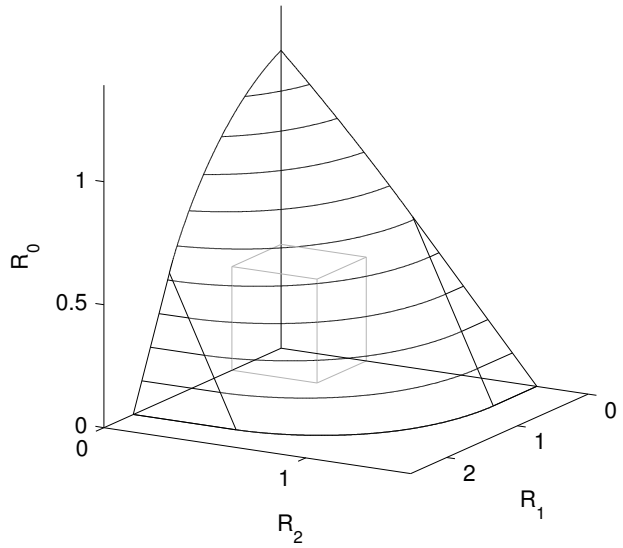
$$\begin{aligned} \frac{1}{n}I(\mathbf{X}^n;\mathbf{Y}_i^n) &\leq \frac{1}{n}\sum_{k=1}^{n} \log\det\big(\boldsymbol{I}_{N_i} + \tfrac{1}{\sigma^2}\boldsymbol{H}_i\boldsymbol{Q}_k\boldsymbol{H}_i^H\big) \\ &\leq \log\det\big(\boldsymbol{I}_{N_i} + \tfrac{1}{\sigma^2}\boldsymbol{H}_i\big(\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{Q}_k\big)\boldsymbol{H}_i^H\big) \end{aligned} \tag{5.11}$$

where the last inequality follows from the concavity of the $\log\det$ function [HJ99, Th. 7.6.7].

Obviously, $R_0 + R_1 = \liminf_{n\to\infty}\frac{1}{n}(\log M_0^{(n)} + \log M_2^{(n)}) \leq \limsup_{n\to\infty}\frac{1}{n}(\log M_0^{(n)} + \log M_2^{(n)})$ and $R_0 + R_2 = \liminf_{n\to\infty}\frac{1}{n}(\log M_0^{(n)} + \log M_1^{(n)}) \leq \limsup_{n\to\infty}\frac{1}{n}(\log M_0^{(n)} + \log M_1^{(n)})$ are always satisfied. We assume M_i to be uniformly distributed so that we have $\frac{1}{n}\log M_i^{(n)} = \frac{1}{n}H(\mathrm{M}_i)$, $i = 0,1,2$, which yields together with (5.10a), (5.10b), and (5.11)

$$R_0 + R_i \leq \limsup_{n\to\infty}\Big[\log\det\big(\boldsymbol{I}_{N_i} + \tfrac{1}{\sigma^2}\boldsymbol{H}_i\big(\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{Q}_k\big)\boldsymbol{H}_i^H\big) + \epsilon_i^{(n)}\Big], \tag{5.12}$$

$i = 1,2$. Next, let us define the compact set $\mathcal{Q} \coloneqq \{\boldsymbol{Q} \in \mathbb{C}^{N_R\times N_R} : \mathrm{tr}(\boldsymbol{Q}) \leq P, \boldsymbol{Q} \succeq \mathbf{0}\}$ and observe that $\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{Q}_k \in \mathcal{Q}$ since $\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{Q}_k \succeq \mathbf{0}$ and $\frac{1}{n}\sum_{k=1}^{n}\mathrm{tr}(\boldsymbol{Q}_k) = \mathrm{tr}(\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{Q}_k) \leq P$ hold. This implies that there exists a subsequence $(n_l)_{l\in\mathbb{N}}$ such that $\frac{1}{n_l}\sum_{k=1}^{n_l}\boldsymbol{Q}_k \to \boldsymbol{Q}$ as $n_l \to \infty$ with $\boldsymbol{Q} \in \mathcal{Q}$. Therewith and with the continuity of the



$$\boldsymbol{H}_1 = \begin{bmatrix} 0.2 - 0.5i & 0.5i \\ -1 & 1.3 + 0.3i \end{bmatrix}$$

$$\boldsymbol{H}_2 = \begin{bmatrix} 0.7 - 0.2i & 0.7 \\ 0 & -0.8i \end{bmatrix}$$

$$P = \sigma^2 = 1$$

Figure 5.3: Capacity region of the MIMO Gaussian BBC with common messages (black) and a comparable TDMA approach (gray) with $N_R = N_1 = N_2 = 2$.

$\log\det$ function we get

$$\begin{aligned} \limsup_{n_l \to \infty} \Big[\log\det\big(\boldsymbol{I}_{N_i} + \tfrac{1}{\sigma^2}\boldsymbol{H}_i\big(\tfrac{1}{n_l}\sum_{k=1}^{n_l}\boldsymbol{Q}_k\big)\boldsymbol{H}_i^H\big) + \epsilon_i^{(n_l)}\Big] \\ = \log\det\big(\boldsymbol{I}_{N_i} + \tfrac{1}{\sigma^2}\boldsymbol{H}_i\boldsymbol{Q}\boldsymbol{H}_i^H\big). \end{aligned} \tag{5.13}$$

Combining (5.12) and (5.13) we obtain $R_0 + R_1 \leq \log\det\big(\boldsymbol{I}_{N_1} + \frac{1}{\sigma^2}\boldsymbol{H}_1\boldsymbol{Q}\boldsymbol{H}_1^H\big)$. Using the same subsequence $(n_l)_{l\in\mathbb{N}}$ and arguments we also obtain $R_0 + R_2 \leq \log\det\big(\boldsymbol{I}_{N_2} + \frac{1}{\sigma^2}\boldsymbol{H}_2\boldsymbol{Q}\boldsymbol{H}_2^H\big)$ which finishes the proof. □

Example

As an example Figure 5.3 depicts the capacity region of a MIMO Gaussian BBC with common messages and illustrates how the optimal strategy outperforms a simple TDMA approach that realizes the same routing task with three orthogonal time slots.

5.2.3 Covariance Optimization Problem

Since the capacity region $\mathcal{R}_{m_0}(\boldsymbol{H}_1, \boldsymbol{H}_2|P)$ is convex, the rate triples on the dominant surface characterize the capacity region completely. Therefore, we are is interested in finding the optimal transmit covariance matrices that achieve the rate triples on the dominant surface since they constitute the basis for further cross-layer designs such as stability-optimal scheduling policies.

A rate triple on the dominant surface of the capacity region is a solution of a weighted rate sum problem so that we consider the corresponding convex optimization problem

$$
\begin{aligned}
R_\Sigma(\boldsymbol{w}) = \max_{R_0,R_1,R_2} \quad & \sum_{i=0}^{2} w_i R_i && \text{(5.14a)}\\
\text{s.t.} \quad & R_0 + R_i \leq C_i(\boldsymbol{Q}), \quad i = 1,2 && \text{(5.14b)}\\
& R_i \geq 0, \quad i = 0,1,2 && \text{(5.14c)}\\
& \text{tr}(\boldsymbol{Q}) \leq P,\ \boldsymbol{Q} \succeq 0 && \text{(5.14d)}
\end{aligned}
$$

with $\boldsymbol{w} = (w_0, w_1, w_2) \in \mathbb{R}_+^3$ the weight vector and $C_i(\boldsymbol{Q}) = \log\det(\boldsymbol{I}_{N_i} + \frac{1}{\sigma^2}\boldsymbol{H}_i\boldsymbol{Q}\boldsymbol{H}_i^H)$, $i = 1, 2$, in the following.

For the optimal rate triples the constraints in (5.14b) will be satisfied with equality for positive weights. Since otherwise, if $R_0 + R_i < C_i(\boldsymbol{Q})$, we can increase the rate R_i up to the point where we have equality, i.e., $R_0 + R_i = C_i(\boldsymbol{Q})$, without affecting the other rates and therewith increasing the weighted rate sum $R_\Sigma(\boldsymbol{w})$. On the other hand if some weights are zero, there exists also an optimal solution where (5.14b) will be satisfied with equality. Therefore, we concentrate on those rate triples that satisfy (5.14b) with equality and rewrite the optimization problem as follows

$$
\begin{aligned}
\max_{\boldsymbol{Q},R_0} \quad & (w_0 - w_1 - w_2)R_0 + w_1 C_1(\boldsymbol{Q}) + w_2 C_2(\boldsymbol{Q}) && \text{(5.15)}\\
\text{s.t.} \quad & 0 \leq R_0 \leq C_i(\boldsymbol{Q}), \quad i = 1,2\\
& \text{tr}(\boldsymbol{Q}) \leq P,\ \boldsymbol{Q} \succeq 0.
\end{aligned}
$$

Then the Lagrangian for the corresponding minimization problem is given by

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{Q}, R_0, \boldsymbol{\nu}, \xi, \mu, \boldsymbol{\Psi}) = -(w_0 - w_1 - w_2)R_0 - \sum_{i=1}^{2} w_i C_i(\boldsymbol{Q})\\
+ \nu_1\big(R_0 - C_1(\boldsymbol{Q})\big) + \nu_2\big(R_0 - C_2(\boldsymbol{Q})\big)\\
- \xi R_0 + \mu\big(\text{tr}(\boldsymbol{Q}) - P\big) - \text{tr}(\boldsymbol{Q}\boldsymbol{\Psi})
\end{aligned}
$$

with Lagrange multipliers $\xi, \mu \in \mathbb{R}$, $\boldsymbol{\nu} = (\nu_1, \nu_2) \in \mathbb{R}^2$, and $\boldsymbol{\Psi} \in \mathbb{C}^{N_R \times N_R}$, from which we get the Karush-Kuhn-Tucker (KKT) conditions with $C_i'(\boldsymbol{Q}) = \boldsymbol{H}_i^H(\sigma^2 \boldsymbol{I}_{N_i} + \boldsymbol{H}_i \boldsymbol{Q} \boldsymbol{H}_i^H)^{-1} \boldsymbol{H}_i$, $i = 1, 2$, as

$$\mu \boldsymbol{I}_{N_R} - \boldsymbol{\Psi} = (w_1 + \nu_1) C_1'(\boldsymbol{Q}) + (w_2 + \nu_2) C_2'(\boldsymbol{Q}) \tag{5.16a}$$

$$w_0 = w_1 + w_2 + \nu_1 + \nu_2 - \xi \tag{5.16b}$$

$$0 \leq R_0 \leq C_i(\boldsymbol{Q}),\ i = 1, 2 \tag{5.16c}$$

$$\boldsymbol{Q} \succeq 0, \qquad \operatorname{tr}(\boldsymbol{Q}) \leq P \tag{5.16d}$$

$$\boldsymbol{\Psi} \succeq 0, \qquad \nu_1, \nu_2, \xi, \mu \geq 0 \tag{5.16e}$$

$$\operatorname{tr}(\boldsymbol{Q}\boldsymbol{\Psi}) = 0, \qquad \mu\big(\operatorname{tr}(\boldsymbol{Q}) - P\big) = 0 \tag{5.16f}$$

$$\xi R_0 = 0, \qquad \nu_i\big(R_0 - C_i(\boldsymbol{Q})\big) = 0,\ i = 1, 2 \tag{5.16g}$$

with primal, dual, and complementary slackness conditions (5.16c)-(5.16d), (5.16e), and (5.16f)-(5.16g) respectively. Since the constraint functions satisfy a generalized version of Slater's condition [BV04, Sec. 5.9], the KKT conditions (5.16a)-(5.16g) are necessary and sufficient and therefore characterize the optimal transmit covariance matrix for a certain weight vector $\boldsymbol{w} = (w_0, w_1, w_2)$.

Although the optimization problem (5.14) is a convex optimization problem and can therefore be efficiently solved using interior point method, further insights can be obtained by studying its structure in more detail as done in the following.

5.2.4 Capacity Achieving Transmit Strategies

Already the proof of achievability, cf. Section 5.2.1, indicates that the BBC with common messages is closely related to the BBC without common messages. Motivated by this observation, we analyze the optimization problem from Section 5.2.3 in detail and establish a strong connection between these two cases in the following so that the results such as transmit strategies obtained for one case will also be applicable for the other one.

It is reasonable to distinguish three different kinds of weight vectors based on the relation between the weight of the common message and the weights of the individual messages. For notational convenience we collect the corresponding weight vectors in three sets

$$\mathcal{W}^{(<)} := \{(w_0, w_1, w_2) \in \mathbb{R}_+^3 : w_0 < w_1 + w_2\}$$
$$\mathcal{W}^{(>)} := \{(w_0, w_1, w_2) \in \mathbb{R}_+^3 : w_0 > w_1 + w_2\}$$
$$\mathcal{W}^{(=)} := \{(w_0, w_1, w_2) \in \mathbb{R}_+^3 : w_0 = w_1 + w_2\}.$$

In the following three subsections we analyze the optimization problem for each set of weight vectors separately. This will be a reasonable division since they characterize the cases with no common message rate, full common message rate, and the case with a trade-off between the common rate and the individual rates.

Zero Common Message Rate

If $w_0 < w_1 + w_2$, the formulation (5.15) of the optimization problem already shows that the weighted rate sum is maximized by setting $R_0 = 0$. Since otherwise, an increasing common rate R_0 would result in a decreasing weighted rate sum. Our intuition is confirmed by the following results.

Proposition 5.11. *Let $\boldsymbol{w} \in \mathcal{W}^{(<)}$ be a weight vector for the BBC with common messages. Then for the weighted rate sum optimal rate triple we have $R_0 = 0$.*

Proof. Since $\nu_1, \nu_2 \geq 0$, cf. (5.16e), condition (5.16b) shows that for $w_0 < w_1 + w_2$ we must have $\xi > 0$ which indeed implies $R_0 = 0$ by (5.16g). □

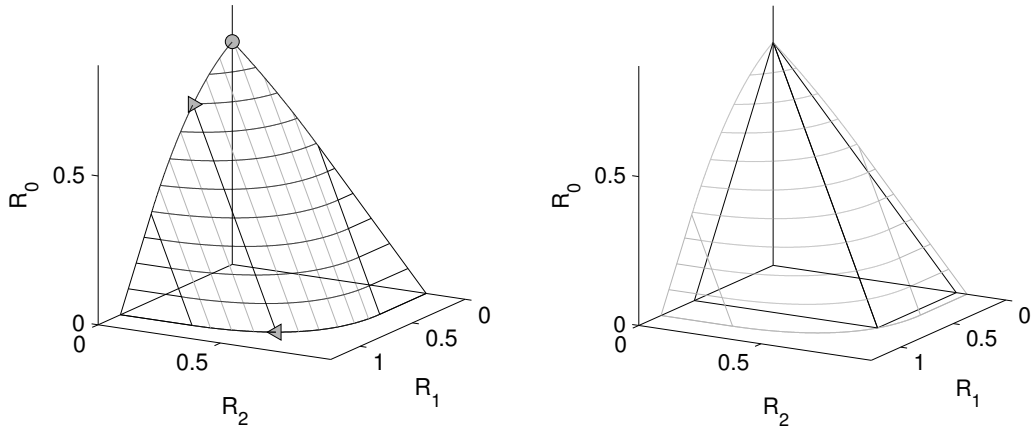
Clearly, if the rate of the common message is zero, the BBC with common messages reduces to the BBC without common messages. Therefore, for weight vector $\boldsymbol{w}' = (0, w_1', w_2') \in \mathbb{R}_+^3$ let $\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')$ be the optimal transmit covariance matrix for the BBC without common messages. Thereby, we know from [OWB09a, OJWB09] that it suffices to consider normalized weight vectors only, i.e., $w_1' + w_2' = 1$, since the optimal transmit covariance matrix only depends on the relation between the two individual weights and not on the exact values. Then the optimum for the BBC with common messages is achieved by the same transmit covariance matrix, i.e., $\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}) = \boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')$, for all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(<)}$ with $w_i = w_i'$, $i = 1, 2$, as long as $w_0 < w_1' + w_2'$ is satisfied. This means the BBC without common messages immediately determines the capacity achieving transmit strategies for the BBC with common messages.

Another important issue is to characterize the optimal rate triples $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w}) = (R_0(\boldsymbol{w}), R_1(\boldsymbol{w}), R_2(\boldsymbol{w}))$ for given weight vector $\boldsymbol{w} = (w_0, w_1, w_2)$.

Proposition 5.12. *For weight vector $\boldsymbol{w} \in \mathcal{W}^{(<)}$ let $\boldsymbol{Q}_{opt}(\boldsymbol{w})$ be the optimal transmit covariance matrix for the BBC with common messages. Then the weighted rate sum optimal rate triple $\boldsymbol{R}_{opt}(\boldsymbol{w})$ is*

$$R_0(\boldsymbol{w}) = 0, \tag{5.17a}$$

$$R_i(\boldsymbol{w}) = C_i\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big),\ i = 1, 2. \tag{5.17b}$$



(a) Capacity region of the MISO Gaussian BBC with common messages

(b) Achievable rate region for the XOR coding approach (black)

Figure 5.4: MISO Gaussian BBC with common messages with $N_R = 2$, $N_1 = N_2 = 1$ for $\boldsymbol{h}_1 = [1.3 \quad 1.3i]^T$, $\boldsymbol{h}_2 = [1 \quad -ie^{i\frac{\pi}{3}}]^T$, $P = 1$, and $\sigma^2 = 1$. For fixed individual weights w_1 and w_2, in Fig. 5.4(a) the point $\triangleleft$ characterizes $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ for all $\boldsymbol{w} \in \mathcal{W}^{(<)}$ with the fixed individual weights. The solid dashed line between $\triangleleft$ and $\triangleright$ corresponds to all $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ for $\boldsymbol{w} \in \mathcal{W}^{(=)}$. For all $\boldsymbol{w} \in \mathcal{W}^{(>)}$ the optimal rate triples moves along the curved section and tends to the XOR solution in $\circ$.

Proof. The weighted rate sum optimal rate triple $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ follows immediately from Proposition 5.11. □

Remark 5.13. *It shows that the weights for the common and individual messages have a strong impact on further cross-layer designs. We see that for all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(<)}$ it is optimal to allocate no resources to the common message and to transmit solely the individual messages which indeed influences the scheduling policy at the relay node.*

Figure 5.4(a) depicts the capacity region of a MISO Gaussian BBC with common messages, where the relay node has multiple antennas, while each receiving node is only equipped with a single antenna. For all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(<)}$ we see from (5.17) that $R_0(\boldsymbol{w}) = 0$ and the weighted rate sum optimal rate triples $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ describe the boundary of the capacity region on the R_1/R_2-plane as further studied in [OWB09a, OJWB09].

Full Common Message Rate

Next, we turn to the more interesting case where the weight of the common message exceeds the weights of the individual messages, i.e., $w_0 > w_1 + w_2$.

Proposition 5.14. *Let $\boldsymbol{w} \in \mathcal{W}^{(>)}$ be a weight vector for the BBC with common messages. Then for the weighted rate sum optimal rate triple we have $R_0 > 0$.*

Proof. We prove the proposition by contradiction. Let us assume $R_0 = 0$ so that $\nu_1 = 0$ and $\nu_2 = 0$ by (5.16g). Since $w_0 > w_1 + w_2$, condition (5.16b) can only be satisfied if $\xi < 0$ which is a contradiction to (5.16e). Therefore, if $w_0 > w_1 + w_2$, we must have $R_0 > 0$ which proves the proposition. □

Intuitively we would expect that it is optimal to allocate all available resources to the common message. But the rate of the common message is limited by a $\min$ operation, cf. Remark 5.6, so that this might not maximize the weighted rate sum in general. Therefore, similar to the previous case we want to know when a given transmit covariance matrix that is optimal for the BBC without common messages is also optimal for the BBC with common messages.

Theorem 5.15. *For weight vector $\boldsymbol{w}' = (0, w_1', w_2') \in \mathbb{R}_+^3$ let $\boldsymbol{Q}'_{opt}(\boldsymbol{w}')$ be the optimal transmit covariance matrix for the BBC without common messages. For all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(>)}$ that further satisfy, if $C_1(\boldsymbol{Q}'_{opt}(\boldsymbol{w}')) < C_2(\boldsymbol{Q}'_{opt}(\boldsymbol{w}'))$, the condition*

$$w_0 = w_1' + w_2',\ w_1 < w_1',\ w_2 = w_2', \tag{5.18a}$$

or, if $C_1(\boldsymbol{Q}'_{opt}(\boldsymbol{w}')) > C_2(\boldsymbol{Q}'_{opt}(\boldsymbol{w}'))$, the condition

$$w_0 = w_1' + w_2',\ w_1 = w_1',\ w_2 < w_2', \tag{5.18b}$$

or, if $C_1(\boldsymbol{Q}'_{opt}(\boldsymbol{w}')) = C_2(\boldsymbol{Q}'_{opt}(\boldsymbol{w}'))$, the condition

$$w_0 = w_1' + w_2',\ w_1 \leq w_1',\ w_2 \leq w_2', \tag{5.18c}$$

the optimum for the BBC with common messages is achieved by the same transmit covariance matrix, i.e., $\boldsymbol{Q}_{opt}(\boldsymbol{w}) = \boldsymbol{Q}'_{opt}(\boldsymbol{w}')$.

Proof. We start with case (5.18a) and note that we have $\xi = 0$ by (5.16g) since $R_0 > 0$ which follows from Proposition 5.14. If $C_1(\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')) < C_2(\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}'))$, then from (5.16c) follows that $R_0 < C_2(\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}'))$ which immediately implies together with (5.16g) that $\nu_2 = 0$. With this, (5.16a) reads as

$$\mu \boldsymbol{I}_{N_R} - \boldsymbol{\Psi} = (w_1 + \nu_1) C_1'\big(\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')\big) + w_2 C_2'\big(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}')\big)$$

which is exactly the same structure as the MIMO Gaussian BBC without common messages has, cf. for example [OJWB09, Eq. (2a)]. Consequently, the optimization problem of the BBC with common messages becomes the optimization problem of the BBC without common messages but with modified individual weights $w_2' = w_2$ and $w_1' = w_1 + \nu_1 = w_0 - w_2$ where the last equality follows from (5.16b). Hence, the optimal transmit covariance matrix $\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')$ for the BBC without common messages and weight vector $\boldsymbol{w}' = (0, w_1', w_2')$ is also a solution of the corresponding problem of the BBC with common messages for all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(>)}$ that further satisfy $w_0 = w_1' + w_2'$, $w_1 < w_1'$, and $w_2 = w_2'$ which proves the first assertion (5.18a).

Now, the case (5.18b) follows accordingly using the same arguments. Furthermore, the third assertion (5.18c) follows immediately from (5.16a) and (5.16b) and $\nu_i \geq 0$, $i = 1, 2$. □

Remark 5.16. *A given weight vector $\boldsymbol{w}'$ uniquely characterizes the optimal transmit covariance matrix $\boldsymbol{Q}'_{opt}(\boldsymbol{w}')$ for the BBC without common messages, cf. [OWB09a, OJWB09] and immediately determines the maximal unidirectional rates $C_1(\boldsymbol{Q}'_{opt}(\boldsymbol{w}'))$ and $C_2(\boldsymbol{Q}'_{opt}(\boldsymbol{w}'))$ for given $\boldsymbol{Q}'_{opt}(\boldsymbol{w}')$. But more important, it directly affects the common rate, since it is restricted by the minimum of the two maximal unidirectional rates, cf. also Remark 5.6. This substantiates the result that an optimal transmit covariance matrix $\boldsymbol{Q}'_{opt}(\boldsymbol{w}')$ for the BBC without common messages is also optimal for the BBC with common messages for three different sets of weight vectors based on the relation between the maximal unidirectional rates $C_1(\boldsymbol{Q}'_{opt}(\boldsymbol{w}'))$ and $C_2(\boldsymbol{Q}'_{opt}(\boldsymbol{w}'))$, respectively.*

Furthermore, the results so far allow to characterize the weighted rate sum optimal rate triples $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ for weight vectors $\boldsymbol{w} \in \mathcal{W}^{(>)}$ in detail. Similarly as in Theorem 5.15 we have to distinguish between three cases.

Proposition 5.17. *For weight vector $\boldsymbol{w} \in \mathcal{W}^{(>)}$ let $\boldsymbol{Q}_{opt}(\boldsymbol{w})$ be the optimal transmit covariance matrix for the BBC with common messages. If $C_1(\boldsymbol{Q}_{opt}(\boldsymbol{w})) < C_2(\boldsymbol{Q}_{opt}(\boldsymbol{w}))$, then the weighted rate sum optimal rate triple $\boldsymbol{R}_{opt}(\boldsymbol{w})$ is*

$$R_0(\boldsymbol{w}) = C_1\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big) \tag{5.19a}$$

$$R_1(\boldsymbol{w}) = 0 \tag{5.19b}$$

$$R_2(\boldsymbol{w}) = C_2\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big) - C_1\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big). \tag{5.19c}$$

If $C_1(\boldsymbol{Q}_{opt}(\boldsymbol{w})) > C_2(\boldsymbol{Q}_{opt}(\boldsymbol{w}))$, then the weighted rate sum optimal rate triple is $\boldsymbol{R}_{opt}(\boldsymbol{w}) = (C_2(\boldsymbol{Q}_{opt}(\boldsymbol{w})), C_1(\boldsymbol{Q}_{opt}(\boldsymbol{w})) - C_2(\boldsymbol{Q}_{opt}(\boldsymbol{w})), 0)$.

If $C_1(\boldsymbol{Q}_{opt}(\boldsymbol{w})) = C_2(\boldsymbol{Q}_{opt}(\boldsymbol{w}))$, then $\boldsymbol{R}_{opt}(\boldsymbol{w})$ is

$$R_0(\boldsymbol{w}) = C_1\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big) \tag{5.20a}$$

$$R_i(\boldsymbol{w}) = 0, \quad i = 1, 2. \tag{5.20b}$$

Proof. We start with the proof of the first case, then the second one follows accordingly using the same arguments. If $C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) < C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$, then $R_0(\boldsymbol{w}) < C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$ so that $\nu_2 = 0$ by (5.16g). Further, from Proposition 5.14 we know that in the optimal rate triple we have $R_0(\boldsymbol{w}) > 0$ so that $\xi = 0$ by (5.16g). With this, (5.16b) reads as $w_0 = w_1 + w_2 + \nu_1$ which implies that $\nu_1 > 0$ since $w_0 > w_1 + w_2$ by assumption. From this follows $R_0(\boldsymbol{w}) = C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$ by (5.16g) so that the weighted rate sum optimal rate triple $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ is given by (5.19).

It remains to prove the third case $C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) = C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$. Since $R_0(\boldsymbol{w}) > 0$, we have $\xi = 0$ so that (5.16b) becomes $w_0 = w_1 + w_2 + \nu_1 + \nu_2$. Since $w_0 > w_1 + w_2$ by assumption, this immediately implies that $\nu_1 > 0$ or $\nu_2 > 0$. If $\nu_1 > 0$ then $R_0(\boldsymbol{w}) = C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) = C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$ by (5.16g). Similarly, $\nu_2 > 0$ leads to $R_0(\boldsymbol{w}) = C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) = C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$ so that the weighted rate sum optimal rate triple $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ is given by (5.20). □

Going back to our example in Figure 5.4(a) we see that for all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(>)}$ the weighted rate sum optimal rate triples $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ describe the boundaries of the capacity region on the R_0/R_1- and R_0/R_2-plane respectively. In more detail, all rate triples $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ with $C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) < C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$ lie on the R_0/R_2-plane and with $C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) > C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$ on the R_0/R_1-plane. For equality, i.e., $C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) = C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$, the rate triple $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ characterizes the XOR solution on the R_0-axis (denoted by point $\circ$ in Figure 5.4(a)).

This substantiates the fact that an optimal transmit strategy for the BBC without common messages for one specific weight vector is optimal for the BBC with common messages for a whole set of weight vectors as specified in (5.18). For example consider the following. For all $\boldsymbol{w} \in \mathcal{W}^{(>)}$ it is optimal to allocate as much rate to the common message as possible. If $C_1(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})) < C_2(\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}))$, then $R_1(\boldsymbol{w}) = 0$ which implies that $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ is the same for all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(>)}$ with fixed w_0 and w_2 as long as $w_1 < w_0 - w_2$. Moreover, it follows that the boundary of the capacity region on the R_0/R_2-plane is solely characterized by the relation between the weights w_0 and w_2.

Dominant Surface

Already in (5.15) we see that if $w_0 = w_1 + w_2$, the weighted rate sum is independent of the common rate. This indicates that we can interchange the rate of the common message and the rates of the individual messages.

Theorem 5.18. *For weight vector $\boldsymbol{w}' = (0, w_1', w_2') \in \mathbb{R}_+^3$ let $\boldsymbol{Q}'_{opt}(\boldsymbol{w}')$ be the optimal transmit covariance matrix for the BBC without common messages. Then for all weight vectors $\boldsymbol{w} \in \mathcal{W}^{(=)}$ with $w_0 = w_1' + w_2'$ and $w_i = w_i'$, $i = 1, 2$, the optimum for the BBC*

with common messages is achieved by the same transmit covariance matrix, i.e., $\boldsymbol{Q}_{opt}(\boldsymbol{w}) = \boldsymbol{Q}'_{opt}(\boldsymbol{w}')$.

Proof. If $R_0 = 0$, then $\nu_1 = \nu_2 = 0$ by (5.16g) which implies that (5.16a) becomes $\mu \boldsymbol{I}_{N_R} - \boldsymbol{\Psi} = w_1 C'_1(\boldsymbol{Q}) + w_2 C'_2(\boldsymbol{Q})$. Again, this is the BBC without common messages and individual weights $w'_i = w_i$, $i = 1, 2$, so that the optimal transmit covariance matrix for the BBC without common messages immediately determines the one for the BBC with common messages.

If $R_0 > 0$, then $\xi = 0$ by (5.16g) so that (5.16b) reads as $w_0 = w_1 + w_2 + \nu_1 + \nu_2$. Since $\nu_i \geq 0$, $i = 1, 2$, by (5.16e), Equation (5.16b) is only valid if $\nu_1 = \nu_2 = 0$. Consequently, (5.16a) becomes $\mu \boldsymbol{I}_{N_R} - \boldsymbol{\Psi} = w_1 C'_1(\boldsymbol{Q}) + w_2 C'_2(\boldsymbol{Q})$. The same arguments as in the first case finish the proof. □

Proposition 5.19. *For weight vector* $\boldsymbol{w} \in \mathcal{W}^{(=)}$ *let* $\boldsymbol{Q}_{opt}(\boldsymbol{w})$ *be the optimal transmit covariance matrix for the BBC with common messages. Then the weighted rate sum optimal rate triples* $\boldsymbol{R}_{opt}(\boldsymbol{w})$ *are*

$$\begin{aligned} R_0(\boldsymbol{w}) &\leq \min\left\{C_1\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big), C_2\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big)\right\}, \\ R_i(\boldsymbol{w}) &= C_i\big(\boldsymbol{Q}_{opt}(\boldsymbol{w})\big) - R_0(\boldsymbol{w}), \quad i = 1, 2. \end{aligned}$$

Proof. $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ follows immediately from Theorem 5.18. □

We see that for the weighted rate sum optimal rate triples there is a trade-off between the common rate and the individual rates as illustrated in Figure 5.4(a). For a given weight vector $\boldsymbol{w} \in \mathcal{W}^{(=)}$ the optimal rate triples $\boldsymbol{R}_{\text{opt}}(\boldsymbol{w})$ correspond to a line that begins on the boundary on the R_1/R_2-plane and ends on the R_0/R_1- or R_0/R_2-plane (visualized by gray dashed-dotted lines). Consequently, the weight vectors $\boldsymbol{w} \in \mathcal{W}^{(=)}$ characterizes the dominant surface of the capacity region completely.

Interpretation

There is another interpretation. If we fix some individual weights w_1 and w_2, the optimal rate triple is always on the R_1/R_2-plane (for example given by the point $\triangleleft$ in Figure 5.4(a)), as long as the common weight fulfills $w_0 < w_1 + w_2$. This immediately implies that the sum rate performance is the same as for the corresponding BBC without common messages.

In the case of equality, i.e., $w_0 = w_1 + w_2$, all rate triples on the connecting line between the points $\triangleleft$ and $\triangleright$ are optimal. If $w_0 > w_1 + w_2$ the optimal rate triple is on the R_0/R_1- or R_0/R_2-plane and with increasing common weight w_0 the optimal rate triple moves along

the corresponding boundary and tends to the XOR solution in point $\circ$ as $w_0 \to \infty$. Figure 5.4(b) depicts the achievable rate region that is characterized by the XOR solution. Since the common message is transmitted to both receiving nodes, a positive common rate reduces the maximal achievable rates for both individual messages, cf. also (5.9). Moreover, for the XOR solution in point $\circ$ the common rate uses all available resources so that both individual rates are zero.

5.2.5 Applications

In the previous section we established a strong connection between the transmit covariance matrix optimization problems for the BBC with and without common messages. This was done by showing that an optimal transmit covariance matrix for the BBC without common messages is also a solution for certain optimization problems for the BBC with common messages.

We indicate that this connection can be exploited to easily transfer results from one case to the other one which shows that the results obtained in the previous section are not only relevant in itself. In the following we briefly review results from [OWB09a, OJWB09] where we assume that the reader is familiar with these references. But we want to accent that these are only examples and that there are much more results which can be transfered in a similar way. The aim of this section is to demonstrate the usefulness of the established connection between the BBC with and without common messages.

First, we consider a MISO scenario, where the relay node is equipped with multiple antennas, while the two other nodes each have a single antenna. Then we know from [OWB09a] that beamforming into the subspace spanned by the channels is always optimal for the BBC without common messages. This means that for all weight vectors $\boldsymbol{w}'$ the optimal transmit covariance matrix $\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')$ is of rank one. From the previous section we know that for certain weights the transmit strategy $\boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')$ is also optimal for the BBC with common messages. More precisely, this means that for weight vectors $\boldsymbol{w}$ as specified by the results from the previous section, the optimal transmit covariance matrix $\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})$ for the BBC with common messages is immediately determined by $\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w}) = \boldsymbol{Q}'_{\text{opt}}(\boldsymbol{w}')$. Consequently, $\boldsymbol{Q}_{\text{opt}}(\boldsymbol{w})$ is also of rank one and beamforming into the subspace spanned by the channels is optimal for the BBC with common messages.

Furthermore, it is shown that the normalized capacity-achieving beamforming vector for the BBC without common messages can be expressed as a linear combination of the two channel directions with a fixed phase difference between the coefficients. This transfers to the BBC with common messages in a similar fashion. Interestingly, it is shown that the correlation between the channels is advantageous.

Moreover, this allows to characterize the transmit strategy that realizes equal sum rates. Again, the previous section determines when the corresponding transmit strategy for the BBC without common messages transfers to the one with common messages. In particular, this is an interesting transmit strategy since it characterizes the rate region that is achievable using (suboptimal) network coding strategies such as the XOR coding approach [LJS05, HKE$^+$07] as depicted in Figure 5.4(b).

For the MIMO scenario, the situation is much more complicated since in general there exist different equivalent transmit strategies with different ranks. But for the special case where the ranks of the channels is equal to the number of antennas at the relay node and a full-rank transmission is optimal, the optimal transmit covariance matrix for the BBC without common messages can be obtained in closed-form from [OJWB09]. Accordingly, once we obtained the optimal covariance matrix, it immediately transfers to the BBC with common messages under certain weight conditions. The same is true for the case of parallel channels. In particular this is a relevant scenario since it immediately provides also solutions for the power allocation problem of a single-antenna OFDM system.

5.3 Integration of Confidential Messages

In this section we consider the scenario where the relay integrates only confidential messages for one receiving node which should be kept secret from the other, non-legitimate node. This is the *bidirectional broadcast channel (BBC) with confidential messages*.

5.3.1 Bidirectional Broadcast Channel with Confidential Messages

The definition of an $(n, M_c^{(n)}, M_1^{(n)}, M_2^{(n)})$-code $\mathcal{C}_{m_c}(W)$ deduces from Definition 5.1. Similarly, the definitions of an achievable rate-equivocation tuple, capacity-equivocation region $\mathcal{R}_{m_c}(W)$, and secrecy capacity region $\mathcal{R}_{m_c}^S(W)$ of the BBC with confidential messages follow immediately from Definition 5.3 and Remark 5.4.

Theorem 5.20. *The capacity-equivocation region $\mathcal{R}_{m_c}(W)$ of the BBC with confidential messages is a closed convex set of those rate-equivocation tuples $(R_c, R_e, R_1, R_2) \in \mathbb{R}_+^4$ that satisfy*

$$\begin{aligned} 0 \leq R_e &\leq R_c \\ R_e &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathrm{Y}_2|\mathrm{U}) \\ R_c + R_i &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \\ R_i &\leq I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \end{aligned}$$

for random variables $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ *with joint probability distribution* $p_{\mathrm{U}}(u)p_{\mathrm{V}|\mathrm{U}}(v|u)p_{\mathrm{X}|\mathrm{V}}(x|v)W(y_1, y_2|x)$*. Moreover, the cardinalities of the ranges of* U *and* V *can be bounded by*

$$|\mathcal{U}| \leq |\mathcal{X}| + 3, \qquad |\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3.$$

Remark 5.21. *While for the BBC without confidential messages the auxiliary random variable* U *only enables a time-sharing operation and carries no information, cf. Theorem 2.5 and 5.5, for the BBC with confidential messages we will see that* U *carries the public information and* V *realizes an additional randomization.*

From Theorem 5.20 follows immediately the *secrecy capacity region* $\mathcal{R}_{m_c}^S(W)$ of the BBC with confidential messages which is the set of rate triples $(R_c, R_1, R_2) \in \mathbb{R}_+^3$ such that $(R_c, R_c, R_1, R_2) \in \mathcal{R}_{m_c}(W)$.

Corollary 5.22. *The secrecy capacity region* $\mathcal{R}_{m_c}^S(W)$ *of the BBC with confidential messages is the set of all rate triples* $(R_c, R_1, R_2) \in \mathbb{R}_+^3$ *that satisfy*

$$\begin{aligned} R_c &\leq I(\mathrm{V}; \mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V}; \mathrm{Y}_2|\mathrm{U}) \\ R_i &\leq I(\mathrm{U}; \mathrm{Y}_i), \quad i = 1, 2 \end{aligned}$$

with perfect secrecy, i.e., (5.2) is satisfied, for random variables $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ *with joint probability distribution* $p_{\mathrm{U}}(u)p_{\mathrm{V}|\mathrm{U}}(v|u)p_{\mathrm{X}|\mathrm{V}}(x|v)W(y_1, y_2|x)$.

The capacity-equivocation region in Theorem 5.20 describes the scenario where the confidential message is transmitted with rate R_c at a certain secrecy level R_e. Thereby, the equivocation rate R_e can be interpreted as the amount of information of the confidential message that can be kept secret from the non-legitimate node. Therefore, Theorem 5.20 includes the case where the non-legitimate node has some partial knowledge about the confidential information, namely if $R_c > R_e$. The secrecy capacity region in Corollary 5.22 characterizes the scenario with perfect secrecy which is, from today's point of view, the practically more relevant case. Since $R_c = R_e$, the confidential message can be kept completely hidden from the non-legitimate node.

In the following we prove Theorem 5.20 and therewith establish the capacity-equivocation region $\mathcal{R}_{m_c}(W)$ of the BBC with confidential messages.

5.3.2 Secrecy-Achieving Coding Strategy

In this subsection we present a coding strategy that achieves the desired rates with the required secrecy level and therewith we prove the achievability part of Theorem 5.20.

Codebook Design

A crucial part is the construction of a suitable codebook with a specific structure consisting of two layers: one for the public and one for the confidential communication. This is done in the following Lemma 5.23.

The first layer corresponds to a codebook that is suitable for the relay to transmit public (bidirectional) messages $m_2' \in \mathcal{M}_2'$ and $m_1' \in \mathcal{M}_1'$ to nodes 1 and 2 as well as a public common message $m_0' \in \mathcal{M}_0'$ to both nodes. This corresponds to the coding problem for the BBC with common messages, cf. Section 5.2.

Then, for each codeword there is a sub-codebook with a product structure similarly as in [CK78] for the classical broadcast channel with confidential messages. The legitimate receiver for the confidential message, i.e., node 1, can decode each codeword regardless to which column and row index it corresponds. But the main idea behind such a codebook design is that the non-legitimate receiver, i.e., node 2, has to decode the column index of the transmitted codeword with the maximum rate its channel provides, and therefore is not able to decode the remaining row index [LPS09].

Lemma 5.23. *For any $\delta > 0$ let $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ be a Markov chain of random variables and $I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) > I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})$.*

i) *There exists a set of (public) codewords $u^n_{m'} \in \mathcal{U}^n$, $m' = (m_0', m_1', m_2') \in \mathcal{M}_0' \times \mathcal{M}_1' \times \mathcal{M}_2' =: \mathcal{M}'$, with*

$$\tfrac{1}{n}\big(\log|\mathcal{M}_0'| + \log|\mathcal{M}_2'|\big) \geq I(\mathrm{U};\mathrm{Y}_1) - \delta \tag{5.21a}$$
$$\tfrac{1}{n}\big(\log|\mathcal{M}_0'| + \log|\mathcal{M}_1'|\big) \geq I(\mathrm{U};\mathrm{Y}_2) - \delta \tag{5.21b}$$

such that

$$\frac{1}{|\mathcal{M}'|}\sum_{m'\in\mathcal{M}'} e_1(m_0', m_2'|m_1') \leq \epsilon^{(n)} \tag{5.22a}$$
$$\frac{1}{|\mathcal{M}'|}\sum_{m'\in\mathcal{M}'} e_2(m_0', m_1'|m_2') \leq \epsilon^{(n)} \tag{5.22b}$$

and $\epsilon^{(n)} \to 0$ as $n \to \infty$. Thereby, $e_1(m_0', m_2'|m_1')$ denotes the probability that node 1 decodes $(m_0', m_2') \in \mathcal{M}_0' \times \mathcal{M}_2'$ incorrectly if $m_1' \in \mathcal{M}_1'$ is given. The probability of error $e_2(m_0', m_1'|m_2')$ for node 2 is defined accordingly.

ii) *For each $u^n_{m'} \in \mathcal{U}^n$ there exists (confidential) codewords $x^n_{jlm'} \in \mathcal{X}^n$, $j \in \mathcal{J}$, $l \in \mathcal{L}$, $m' \in \mathcal{M}'$, with*

$$\tfrac{1}{n}\log|\mathcal{J}| \geq I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) - \delta, \tag{5.23a}$$
$$\tfrac{1}{n}\log|\mathcal{L}| \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) - \delta, \tag{5.23b}$$

such that

$$\frac{1}{|\mathcal{J}||\mathcal{L}||\mathcal{M}'|}\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}}\sum_{m'\in\mathcal{M}'} e_1(j,l|m') \leq \epsilon^{(n)} \tag{5.24a}$$

$$\frac{1}{|\mathcal{J}||\mathcal{L}||\mathcal{M}'|}\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}}\sum_{m'\in\mathcal{M}'} e_2(j|l,m') \leq \epsilon^{(n)} \tag{5.24b}$$

and $\epsilon^{(n)} \to 0$ as $n \to \infty$. Here, $e_1(j,l|m')$ is the probability that node 1 decodes $j \in \mathcal{J}$ or $l \in \mathcal{L}$ incorrectly if $m' \in \mathcal{M}'$ is known. Similarly, $e_2(j|l,m')$ is the probability that node 2 decodes $j \in \mathcal{J}$ incorrectly if $l \in \mathcal{L}$ and $m' \in \mathcal{M}'$ are given.

Proof. The proof exploits ideas from the BBC with common messages for the first part, cf. also Section 5.2, and from the classical broadcast channel with confidential messages [CK78] for the second part. For completeness, the details can be found in Appendix A.9. □

Of course, the communication of confidential information and especially the codebook design above is only meaningful, if the channel from the relay node to the intended receiver provides higher rates than the one to the non-legitimate node. From Lemma 5.23 we see that $I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) > I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})$ is the limiting criterion that decides if confidential communication is possible or not.

Achievable Equivocation-Rate Region

Next, we use the codebook from Lemma 5.23 to construct suitable encoder and decoders for the BBC with confidential messages.

Lemma 5.24. *Let $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ be a Markov chain of random variables and $I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) > I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})$. Using the codebook from Lemma 5.23 all rate-equivocation tuples $(R_c, R_e, R_1, R_2) \in \mathbb{R}_+^4$ that satisfy*

$$0 \leq R_e = I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) \leq R_c \tag{5.25a}$$

$$R_c + R_i \leq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \tag{5.25b}$$

$$R_i \leq I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \tag{5.25c}$$

are achievable for the BBC with confidential messages.

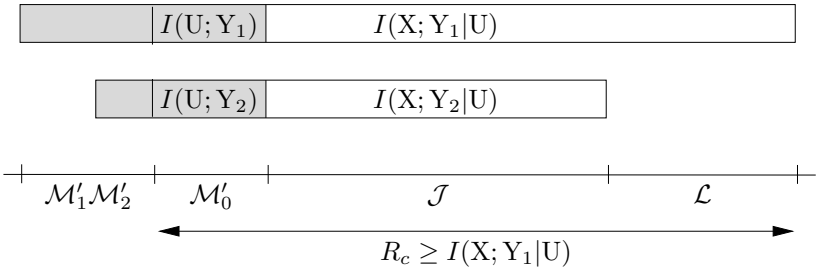


Figure 5.5: The two bars visualize the available resources of both links. Each one is split up into two parts: one designated for the public communication (gray) and one for the confidential communication (white). Since $R_c \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, some resources of the bidirectional communication have to be spent for the confidential message as well (realized by a common message).

Proof. For any $\delta > 0$ and given rate-equivocation tuple $(R_c, R_e, R_1, R_2) \in \mathbb{R}_+^4$ satisfying (5.25a)-(5.25c) we have to construct message sets, encoder, and decoders with

$$\tfrac{1}{n}\log|\mathcal{M}_c| \geq R_c - \delta \tag{5.26a}$$
$$\tfrac{1}{n}\log|\mathcal{M}_2| \geq R_1 - \delta \tag{5.26b}$$
$$\tfrac{1}{n}\log|\mathcal{M}_1| \geq R_2 - \delta \tag{5.26c}$$

and further, cf. also (5.1),

$$\tfrac{1}{n}H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) - \delta. \tag{5.27}$$

The following construction is mainly based on the one for the classical broadcast channel with confidential messages [CK78]. Thereby, we have to distinguish between two cases as visualized in Figures 5.5 and 5.6.

If $R_c \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, cf. Figure 5.5, we construct the set of confidential messages as

$$\mathcal{M}_c \coloneqq \mathcal{J} \times \mathcal{L} \times \mathcal{M}_0'$$

where the sets $\mathcal{J}$ and $\mathcal{L}$ are chosen as in Lemma 5.23 and $\mathcal{M}_0'$ is an arbitrary set of common messages such that (5.26a) is satisfied. The sets $\mathcal{M}_1 = \mathcal{M}_1'$ and $\mathcal{M}_2 = \mathcal{M}_2'$ are arbitrary such that (5.26b)-(5.26c) hold. Finally, we define the deterministic encoder f that maps the confidential message $(j, l, m_0') \in \mathcal{M}_c$ and the individual messages $m_i \in \mathcal{M}_i$, $i = 1, 2$, into the codeword $x_{jlm'}^n \in \mathcal{X}^n$ with $m' = (m_0', m_1', m_2')$ and $m_i' = m_i$, $i = 1, 2$.

Remark 5.25. *Since $R_c \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, a part of the confidential message must be transmitted as a common message using resources designated for the public communication, cf. Figure 5.5. It is not possible to simply "add" the remaining part to the individual message for node 1, since this would require that this part of the confidential message is already available a priori as side information at node 2.*

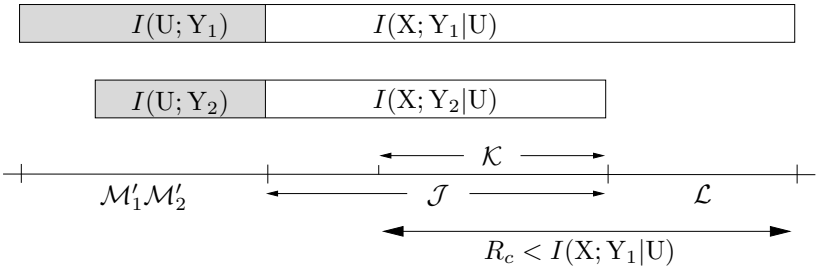


Figure 5.6: Since $R_c < I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, there are more resources for the confidential communication available than needed. This allows the relay to enable a stochastic coding strategy that exploits all the available resources by introducing a mapping from $\mathcal{J}$ to $\mathcal{K}$.

If $R_c < I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, cf. Figure 5.6, we set $\mathcal{M}_c \coloneqq \mathcal{K} \times \mathcal{L}$ where $\mathcal{K}$ is an arbitrary set such that (5.26a) holds. Further, we define a mapping $h : \mathcal{J} \to \mathcal{K}$ that partitions $\mathcal{J}$ into subsets of "nearly equal size" [CK78], which means

$$|h^{-1}(k)| \leq 2|h^{-1}(k')|, \quad \text{for all } k, k' \in \mathcal{K}.$$

Moreover, since $R_c < I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, there is no need for a set of common messages so that $\mathcal{M}_0' = \emptyset$. The sets $\mathcal{M}_1 = \mathcal{M}_1'$ and $\mathcal{M}_2 = \mathcal{M}_2'$ are arbitrary such that (5.26b)-(5.26c) hold. Finally, we define the stochastic encoder f that maps the confidential message $(k, l) \in \mathcal{M}_c$ and the individual messages $m_i \in \mathcal{M}_i$, $i = 1, 2$, into the codeword $x^n_{jlm'} \in \mathcal{X}^n$ with $m' = (0, m_1', m_2')$, where j is uniformly drawn from the set $h^{-1}(k) \subset \mathcal{J}$ and $m_i' = m_i$, $i = 1, 2$.

Remark 5.26. *This time, set $\mathcal{J}$ is not needed in total for the confidential communication. However, to force the non-legitimate receiver, i.e., node 2, to decode at its maximum rate, we define a stochastic encoder that "spreads" the confidential messages over the whole set $\mathcal{J}$. Moreover, if $R_c \leq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})$, the whole set $\mathcal{J}$ is used for additional randomization.*

Up to now we defined message sets and the encoder. In both cases the decoders are immediately determined by the decoding sets given in Lemma 5.23, cf. Appendix A.9 for more details. Hence, the achievability of the rates as specified in (5.25a)-(5.25c) follows immediately from Lemma 5.23.

To complete the proof it remains to show that this coding strategy achieves the required secrecy level (5.27) at node 2. Proceeding as in [CK78] let X^n be the input random variable of the channel, whose realizations are the codewords $x^n_{jlm'} \in \mathcal{X}^n$ (as specified by the encoder above). Further, let $\mathrm{M}' = (\mathrm{M}_0', \mathrm{M}_1', \mathrm{M}_2')$ be the random variable that corresponds to the third

index of the realization of X^n. With $\mathrm{M}_i = \mathrm{M}'_i$, $i = 1, 2$, from the definition of the encoder above, we get for the equivocation

$$\begin{aligned} H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) &\geq H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}') \\ &= H(\mathrm{M}_c, \mathrm{Y}_2^n|\mathrm{M}') - H(\mathrm{Y}_2^n|\mathrm{M}') \\ &= H(\mathrm{M}_c, \mathrm{Y}_2^n, \mathrm{X}^n|\mathrm{M}') - H(\mathrm{X}^n|\mathrm{M}_c, \mathrm{M}', \mathrm{Y}_2^n) - H(\mathrm{Y}_2^n|\mathrm{M}') \\ &= H(\mathrm{M}_c, \mathrm{X}^n|\mathrm{M}') + H(\mathrm{Y}_2^n|\mathrm{M}_c, \mathrm{M}', \mathrm{X}^n) - H(\mathrm{X}^n|\mathrm{M}_c, \mathrm{M}', \mathrm{Y}_2^n) - H(\mathrm{Y}_2^n|\mathrm{M}') \\ &\geq H(\mathrm{X}^n|\mathrm{M}') + H(\mathrm{Y}_2^n|\mathrm{X}^n) - H(\mathrm{X}^n|\mathrm{M}_c, \mathrm{M}', \mathrm{Y}_2^n) - H(\mathrm{Y}_2^n|\mathrm{M}'). \end{aligned} \tag{5.28}$$

In the following we bound all terms in (5.28) separately. We start with the first term and observe that for given $\mathrm{M}' = m'$ the random variable X^n has $|\mathcal{J}||\mathcal{L}|$ possible values. Since we assume X^n to be independently and uniformly distributed, we have $H(\mathrm{X}^n|\mathrm{M}') = \log|\mathcal{J}| + \log|\mathcal{L}|$. With the definition of the sets $\mathcal{J}$ and $\mathcal{L}$, cf. (5.23) of Lemma 5.23, we obtain

$$\tfrac{1}{n} H(\mathrm{X}^n|\mathrm{M}') \underset{n\to\infty}{\longrightarrow} I(\mathrm{X}; \mathrm{Y}_1|\mathrm{U}). \tag{5.29}$$

For the second term in (5.28) we get from the weak law of large numbers

$$\tfrac{1}{n} H(\mathrm{Y}_2^n|\mathrm{X}^n) \underset{n\to\infty}{\longrightarrow} H(\mathrm{Y}_2|\mathrm{X}). \tag{5.30}$$

If $R_c \geq I(\mathrm{X}; \mathrm{Y}_1|\mathrm{U})$, the third term in (5.28) vanishes, since given M_c and M' the deterministic encoder already determines X^n. If $R_c < I(\mathrm{X}; \mathrm{Y}_1|\mathrm{U})$, we have a stochastic encoder and define

$$\varphi(k, l, m', y_2^n) = \begin{cases} x_{klm'}^n & \text{if } (u_{m'}^n, x_{jlm'}^n, y_2^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_2),\ h(j) = k \\ \text{arbitrary} & \text{otherwise.} \end{cases}$$

Then we have $\mathbb{P}\{\mathrm{X}^n \neq \varphi(\mathrm{M}_c, \mathrm{M}', \mathrm{Y}_2^n)\} \leq \epsilon^{(n)}$ with $\epsilon^{(n)} \to 0$ as $n \to \infty$ and therefore, by Fano's lemma, cf. also [CK78, LPS09],

$$\tfrac{1}{n} H(\mathrm{X}^n|\mathrm{M}_c, \mathrm{M}', \mathrm{Y}_2^n) \underset{n\to\infty}{\longrightarrow} 0 \tag{5.31}$$

so that the third term vanishes also in this case. For the last term in (5.28) we define

$$\hat{y}_2^n = \begin{cases} y_2^n & \text{if } (u_{m'}^n, y_2^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_2) \\ \text{arbitrary} & \text{otherwise} \end{cases}$$

so that

$$H(\mathrm{Y}_2^n|\mathrm{M}') \leq H(\mathrm{Y}_2^n|\hat{\mathrm{Y}}_2^n) + H(\hat{\mathrm{Y}}_2^n|\mathrm{M}').$$

For the first term we have $\mathbb{P}\{\mathrm{Y}_2^n \neq \hat{\mathrm{Y}}_2^n\} \leq \epsilon^{(n)}$ with $\epsilon^{(n)} \to 0$ as $n \to \infty$ by Fano's lemma, cf. [CK78, LPS09], so that it is negligible. Moreover, following [CK78, LPS09] for given $\mathrm{M}' = m'$ we get for the conditional entropy

$$\begin{aligned} H(\hat{\mathrm{Y}}_2^n|\mathrm{M}' = m') &\leq \log|\mathcal{A}_\epsilon^{(n)}(\mathrm{Y}_2|u_{m'}^n)| \\ &\leq \log(2^{n(H(\mathrm{Y}_2|\mathrm{U})+2\epsilon)}) = n(H(\mathrm{Y}_2|\mathrm{U}) + 2\epsilon) \end{aligned}$$

where the second inequality follows from the definition of the decoding sets, cf. also [CT06, Theorem 15.2.2]. With this we obtain

$$\tfrac{1}{n} H(\hat{\mathrm{Y}}_2^n|\mathrm{M}') \underset{n\to\infty}{\longrightarrow} H(\mathrm{Y}_2|\mathrm{U}). \tag{5.32}$$

Finally, by substituting (5.29)-(5.32) into (5.28) we obtain (5.27) which establishes the desired secrecy level at node 2 and therewith proves the lemma. □

Randomization and Convexity

Here, we complete the proof of achievability of Theorem 5.20 where the argumentation goes along with the one for the classical broadcast channel with confidential messages [CK78].

To obtain the whole region as given in Theorem 5.20, we follow [CK78] and introduce an auxiliary channel that enables an additional randomization.

Lemma 5.27. *Let* $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ *be a Markov chain of random variables and* $I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) > I(\mathrm{V};\mathrm{Y}_2|\mathrm{U})$. *Then all rate-equivocation tuples* $(R_c, R_e, R_1, R_2) \in \mathbb{R}_+^4$ *that satisfy*

$$0 \leq R_e \leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathrm{Y}_2|\mathrm{U}) \leq R_c \tag{5.33a}$$

$$R_c + R_i \leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \tag{5.33b}$$

$$R_i \leq I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \tag{5.33c}$$

are achievable for the BBC with confidential messages. The corresponding rate region is denoted by $\widetilde{\mathcal{R}}_{m_c}(W)$.

Proof. The prefixing realized by the random variable V is exactly the same as in [CK78, Lemma 4]. Moreover, it is obvious that if the rate-equivocation tuple (R_c, R_e, R_1, R_2) is achievable, than each rate-equivocation tuple (R_c, R_e', R_1, R_2) with $0 \leq R_e' \leq R_e$ is also achievable. Consequently, we can further replace the equality in (5.25a) by an inequality in (5.33a). Then the desired region follows immediately from Lemma 5.24. □

Lemma 5.28. *The rate region* $\widetilde{\mathcal{R}}_{m_c}(W)$ *is convex.*

Proof. Exactly as in [CK78, Lemma 5] it is easy to show that any linear combination of two rate-equivocation tuples in $\widetilde{\mathcal{R}}_{m_c}(W)$ is contained in $\widetilde{\mathcal{R}}_{m_c}(W)$.

In more detail, we follow [CK78, Lemma 5] and define rate tuples (R_c', R_e', R_1', R_2') and $(R_c'', R_e'', R_1'', R_2'')$ that satisfy (5.33) with corresponding random variables $\mathrm{U}_1 - \mathrm{V}_1 - \mathrm{X}_1 - (\mathrm{Y}_{1,1}, \mathrm{Y}_{2,1})$ and $\mathrm{U}_2 - \mathrm{V}_2 - \mathrm{X}_2 - (\mathrm{Y}_{1,2}, \mathrm{Y}_{2,2})$. Further, let J be a time-sharing random variable that is independent of all other random variables and distributed over $\{1, 2\}$ with probabilities α and $1 - \alpha$. Now, we define

$$\mathrm{U} \coloneqq (\mathrm{U}_\mathrm{J}, \mathrm{J}), \quad \mathrm{V} \coloneqq \mathrm{V}_\mathrm{J}, \quad \mathrm{X} \coloneqq \mathrm{X}_\mathrm{J}, \quad \mathrm{Y}_1 \coloneqq \mathrm{Y}_{1,\mathrm{J}}, \quad \mathrm{Y}_2 \coloneqq \mathrm{Y}_{2,\mathrm{J}}.$$

Then, we have $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ and

$$\begin{aligned} I(\mathrm{V}; \mathrm{Y}_i|\mathrm{U}) &= \alpha I(\mathrm{V}_1; \mathrm{Y}_{i,1}|\mathrm{U}_1) + (1-\alpha) I(\mathrm{V}_2; \mathrm{Y}_{i,2}|\mathrm{U}_2) \\ I(\mathrm{U}; \mathrm{Y}_i) \geq I(\mathrm{U}; \mathrm{Y}_i|\mathrm{J}) &= \alpha I(\mathrm{U}_1; \mathrm{Y}_{i,1}) + (1-\alpha) I(\mathrm{U}_2; \mathrm{Y}_{i,2}), \end{aligned}$$

$i = 1, 2$, which implies that

$$\alpha(R_c', R_e', R_1', R_2') + (1-\alpha)(R_c'', R_e'', R_1'', R_2'') \in \widetilde{\mathcal{R}}_{m_c}(W)$$

proving the lemma. □

It remains to show that $\widetilde{\mathcal{R}}_{m_c}(W)$ describes the same rate region as the one specified by Theorem 5.20.

Lemma 5.29. *The rate region $\widetilde{\mathcal{R}}_{m_c}(W)$ equals the capacity region $\mathcal{R}_{m_c}(W)$ given in Theorem 5.20.*

Proof. It is obvious that $\widetilde{\mathcal{R}}_{m_c}(W) \subseteq \mathcal{R}_{m_c}(W)$ holds. To show the reversed inclusion, i.e., $\mathcal{R}_{m_c}(W) \subseteq \widetilde{\mathcal{R}}_{m_c}(W)$, let $(R_c, R_e, R_1, R_2) \in \mathcal{R}_{m_c}(W)$ be any rate-equivocation tuple. For this, we construct as in [CK78] the maximal achievable confidential and equivocation rates that are possible for given individual rates R_1 and R_2 as

$$\begin{aligned} R_c^* &\coloneqq I(\mathrm{V}; \mathrm{Y}_1|\mathrm{U}) + \min\left\{I(\mathrm{U}; \mathrm{Y}_1) - R_1, I(\mathrm{U}; \mathrm{Y}_2) - R_2\right\} \\ R_e^* &\coloneqq I(\mathrm{V}; \mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V}; \mathrm{Y}_2|\mathrm{U}). \end{aligned}$$

Then we have $R_e \leq R_e^*$, $R_e^* \leq R_c \leq R_c^*$, and therewith also $(R_c^*, R_e^*, R_1, R_2) \in \widetilde{\mathcal{R}}_{m_c}(W)$. Now, from the definition of $\widetilde{\mathcal{R}}_{m_c}(W)$ follows that the rate-equivocation tuples (R_c^*, R_e^*, R_1, R_2), $(R_c^*, 0, R_1, R_2)$, and $(0, 0, R_1, R_2)$ belong to $\widetilde{\mathcal{R}}_{m_c}(W)$ as well. Finally, from the convexity of $\widetilde{\mathcal{R}}_{m_c}(W)$, cf. Lemma 5.28, follows that $(R_c, R_e, R_1, R_2) \in \widetilde{\mathcal{R}}_{m_c}(W)$ which proves the lemma. □

To complete the proof of achievability it remains to bound the cardinalities of the ranges of U and V. Since the bounds of the cardinalities depend only on the structure of the random variables, the result follows immediately from [CK78, Appendix] where the same bounds are established for the classical broadcast channel with confidential messages. □

5.3.3 Converse

Already the presented coding strategy indicates that, basically, the BBC with confidential messages exploits ideas of the BBC (with common messages), cf. Section 5.2 and [OSBB08], and also of the classical broadcast channel with confidential messages [CK78]. Based on this observation it is easy to establish the weak converse by extending the converse of the classical broadcast channel with confidential messages [CK78] using standard arguments for the BBC, cf. Section 5.2 and [OSBB08].

We have to show that for any given sequence $\{\mathcal{C}_{m_c}^{(n)}(W)\}_{n\in\mathbb{N}}$ of $(n, M_c^{(n)}, M_1^{(n)}, M_2^{(n)})$-codes with $\bar{e}_1, \bar{e}_2 \to 0$ there exist random variables $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ such that

$$\begin{aligned}
\tfrac{1}{n}H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathrm{Y}_2|\mathrm{U}) + o(n^0)\\
\tfrac{1}{n}H(\mathrm{M}_2) &\leq I(\mathrm{U};\mathrm{Y}_1) + o(n^0)\\
\tfrac{1}{n}H(\mathrm{M}_1) &\leq I(\mathrm{U};\mathrm{Y}_2) + o(n^0)\\
\tfrac{1}{n}\big(H(\mathrm{M}_c) + H(\mathrm{M}_2)\big) &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_1) + o(n^0)\\
\tfrac{1}{n}\big(H(\mathrm{M}_c) + H(\mathrm{M}_1)\big) &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_2) + o(n^0)
\end{aligned}$$

are satisfied. For this purpose we need a version of Fano's lemma suitable for the BBC with confidential messages.

Lemma 5.30 (Fano's inequality)**.** *For the BBC with confidential messages we have the following versions of Fano's inequality*

$$\begin{aligned}
H(\mathrm{M}_c, \mathrm{M}_2|\mathrm{Y}_1^n, \mathrm{M}_1) &\leq \bar{e}_1 \log(M_c^{(n)} M_2^{(n)}) + 1 = n\epsilon_1^{(n)},\\
H(\mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) &\leq \bar{e}_2 \log M_1^{(n)} + 1 = n\epsilon_2^{(n)},
\end{aligned}$$

with $\epsilon_1^{(n)} = \frac{1}{n}\log(M_c^{(n)} M_2^{(n)})\bar{e}_1 + \frac{1}{n} \to 0$ *and* $\epsilon_2^{(n)} = \frac{1}{n}\log(M_1^{(n)})\bar{e}_2 + \frac{1}{n} \to 0$ *for* $n \to \infty$ *as* $\bar{e}_1, \bar{e}_2 \to 0$.

Proof. In Appendix A.8 we prove Fano's inequality for the BBC with common and confidential messages. The case with only confidential messages can easily be deduced from this. □

We start with some upper bounds on the entropy terms. Using the fact that M_c, M_1, M_2 are independent, the definition of mutual information, the chain rule for entropy, and Fano's inequality, cf. Lemma 5.30, we obtain similarly as in [CK78, Eqs. (35)-(37)]

$$H(\mathrm{M}_c) \leq I(\mathrm{M}_c; \mathrm{Y}_1^n|\mathrm{M}_1, \mathrm{M}_2) + n\epsilon_1^{(n)} \tag{5.34a}$$

$$H(\mathrm{M}_2) \leq I(\mathrm{M}_1, \mathrm{M}_2; \mathrm{Y}_1^n) + n\epsilon_1^{(n)} \tag{5.34b}$$

$$H(\mathrm{M}_1) \leq I(\mathrm{M}_1, \mathrm{M}_2; \mathrm{Y}_2^n) + n\epsilon_2^{(n)} \tag{5.34c}$$

and further for the equivocation we get

$$\begin{aligned} H(\mathrm{M}_c|\mathrm{Y}_2^n,\mathrm{M}_2) &= H(\mathrm{M}_c|\mathrm{Y}_2^n,\mathrm{M}_1,\mathrm{M}_2) + I(\mathrm{M}_c;\mathrm{M}_1|\mathrm{Y}_2^n,\mathrm{M}_2) \\ &= H(\mathrm{M}_c|\mathrm{M}_1,\mathrm{M}_2) - I(\mathrm{M}_c;\mathrm{Y}_2^n|\mathrm{M}_1,\mathrm{M}_2) + I(\mathrm{M}_c;\mathrm{M}_1|\mathrm{Y}_2^n,\mathrm{M}_2) \\ &= I(\mathrm{M}_c;\mathrm{Y}_1^n|\mathrm{M}_1,\mathrm{M}_2) - I(\mathrm{M}_c;\mathrm{Y}_2^n|\mathrm{M}_1,\mathrm{M}_2) \\ &\qquad + H(\mathrm{M}_c|\mathrm{Y}_1^n,\mathrm{M}_1,\mathrm{M}_2) + I(\mathrm{M}_c;\mathrm{M}_1|\mathrm{Y}_2^n,\mathrm{M}_2) \\ &\leq I(\mathrm{M}_c;\mathrm{Y}_1^n|\mathrm{M}_1,\mathrm{M}_2) - I(\mathrm{M}_c;\mathrm{Y}_2^n|\mathrm{M}_1,\mathrm{M}_2) + n\epsilon_1^{(n)} + n\epsilon_2^{(n)} \end{aligned} \tag{5.35}$$

where the last inequality follows from Fano's inequality, cf. Lemma 5.30, and $H(\mathrm{M}_c|\mathrm{Y}_1^n,\mathrm{M}_1,\mathrm{M}_2) \leq H(\mathrm{M}_c,\mathrm{M}_2|\mathrm{Y}_1^n,\mathrm{M}_1) \leq n\epsilon_1^{(n)}$ and $I(\mathrm{M}_c;\mathrm{M}_1|\mathrm{Y}_2^n,\mathrm{M}_2) = H(\mathrm{M}_1|\mathrm{Y}_2^n,\mathrm{M}_2) - H(\mathrm{M}_1|\mathrm{Y}_2^n,\mathrm{M}_c,\mathrm{M}_2) \leq H(\mathrm{M}_1|\mathrm{Y}_2^n,\mathrm{M}_2) \leq n\epsilon_2^{(n)}$.

The next step is to expand the mutual information terms in (5.34)-(5.35) by making extensive use of the chain rule for mutual information. By replacing the common message in [CK78, Sec. V] with our (bidirectional) individual messages, it is straightforward to show that, similarly as in [CK78, Eqs. (38)-(41)], the mutual information terms in (5.34)-(5.35) can be expressed as

$$I(\mathrm{M}_c;\mathrm{Y}_1^n|\mathrm{M}_1,\mathrm{M}_2) = \sum_{k=1}^{n} I(\mathrm{M}_c;\mathrm{Y}_{1,k}|\mathrm{Y}_1^{k-1},\mathrm{Y}_{2,k+1}^n,\mathrm{M}_1,\mathrm{M}_2) + \Sigma_1 - \Sigma_2 \tag{5.36a}$$

$$I(\mathrm{M}_c;\mathrm{Y}_2^n|\mathrm{M}_1,\mathrm{M}_2) = \sum_{k=1}^{n} I(\mathrm{M}_c;\mathrm{Y}_{2,k}|\mathrm{Y}_1^{k-1},\mathrm{Y}_{2,k+1}^n,\mathrm{M}_1,\mathrm{M}_2) + \Sigma_1^* - \Sigma_2^* \tag{5.36b}$$

and

$$I(\mathrm{M}_1,\mathrm{M}_2;\mathrm{Y}_1^n) \leq \sum_{k=1}^{n} I(\mathrm{Y}_1^{k-1},\mathrm{Y}_{2,k+1}^n,\mathrm{M}_1,\mathrm{M}_2;\mathrm{Y}_{1,k}) - \Sigma_1 \tag{5.37a}$$

$$I(\mathrm{M}_1,\mathrm{M}_2;\mathrm{Y}_2^n) \leq \sum_{k=1}^{n} I(\mathrm{Y}_1^{k-1},\mathrm{Y}_{2,k+1}^n,\mathrm{M}_1,\mathrm{M}_2;\mathrm{Y}_{2,k}) - \Sigma_1^* \tag{5.37b}$$

where

$$\Sigma_1 = \sum_{k=1}^{n} I(\mathrm{Y}_{2,k+1}^n;\mathrm{Y}_{1,k}|\mathrm{Y}_1^{k-1},\mathrm{M}_1,\mathrm{M}_2)$$

$$\Sigma_1^* = \sum_{k=1}^{n} I(\mathrm{Y}_1^{k-1};\mathrm{Y}_{2,k}|\mathrm{Y}_{2,k+1}^n,\mathrm{M}_1,\mathrm{M}_2)$$

and the analogous terms Σ_2 and Σ_2^* with $\mathrm{M}_1,\mathrm{M}_2$ replaced by $\mathrm{M}_c,\mathrm{M}_1,\mathrm{M}_2$.

Lemma 5.31. *We have the following identities:* $\Sigma_1 = \Sigma_1^*$ *and* $\Sigma_2 = \Sigma_2^*$.

Proof. In [CK78, Lemma 7] a similar result for the classical broadcast channel with confidential messages is given. Our result follows immediately by simply replacing the common message in [CK78, Lemma 7] by our two (bidirectional) individual messages M_1 and M_2. □

As in [CK78, Sec. V] we introduce an auxiliary random variable J that is independent of M_c, M_1, M_2, X^n, Y_1^n, and Y_2^n and uniformly distributed over $\{1, ..., n\}$. Further, let

$$\mathrm{U} \coloneqq (\mathrm{Y}_1^{\mathrm{J}-1}, \mathrm{Y}_{2,\mathrm{J}+1}^n, \mathrm{M}_1, \mathrm{M}_2, \mathrm{J}),\ \mathrm{V} \coloneqq (\mathrm{U}, \mathrm{M}_c),\ \mathrm{X} \coloneqq \mathrm{X}_{\mathrm{J}},\ \mathrm{Y}_1 \coloneqq \mathrm{Y}_{1,\mathrm{J}},\ \mathrm{Y}_2 \coloneqq \mathrm{Y}_{2,\mathrm{J}}$$

so that

$$\frac{1}{n}\sum_{k=1}^{n} I(\mathrm{M}_c; \mathrm{Y}_{1,k}|\mathrm{Y}_1^{k-1}, \mathrm{Y}_{2,k+1}^n, \mathrm{M}_1, \mathrm{M}_2) = I(\mathrm{M}_c; \mathrm{Y}_1|\mathrm{U}) = I(\mathrm{V}; \mathrm{Y}_1|\mathrm{U})$$
$$\frac{1}{n}\sum_{k=1}^{n} I(\mathrm{M}_c; \mathrm{Y}_{2,k}|\mathrm{Y}_1^{k-1}, \mathrm{Y}_{2,k+1}^n, \mathrm{M}_1, \mathrm{M}_2) = I(\mathrm{M}_c; \mathrm{Y}_2|\mathrm{U}) = I(\mathrm{V}; \mathrm{Y}_2|\mathrm{U})$$

and

$$\frac{1}{n}\sum_{k=1}^{n} I(\mathrm{Y}_1^{k-1}, \mathrm{Y}_{2,k+1}^n, \mathrm{M}_1, \mathrm{M}_2; \mathrm{Y}_{1,k}) = I(\mathrm{U}; \mathrm{Y}_1|\mathrm{J}) \leq I(\mathrm{U}; \mathrm{Y}_1)$$
$$\frac{1}{n}\sum_{k=1}^{n} I(\mathrm{Y}_1^{k-1}, \mathrm{Y}_{2,k+1}^n, \mathrm{M}_1, \mathrm{M}_2; \mathrm{Y}_{2,k}) = I(\mathrm{U}; \mathrm{Y}_2|\mathrm{J}) \leq I(\mathrm{U}; \mathrm{Y}_2).$$

Now, to complete the proof it remains to put all ingredients together. Therefore, we substitute this into (5.36)-(5.37), apply Lemma 5.31, so that with (5.34)-(5.35) the weak converse is established. □

5.4 Integration of Common and Confidential Messages

Finally we are able to address the most general scenario as depicted in Figure 5.1 where the relay integrates common and confidential messages. This is the *bidirectional broadcast channel (BBC) with common and confidential messages*.

Theorem 5.32. *The capacity-equivocation region $\mathcal{R}_{m_c,m_0}(W)$ of the discrete memoryless BBC with common and confidential messages is the set of all rate-equivocation tuples $(R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ that satisfy*

$$0 \leq R_e \leq R_c \tag{5.38a}$$
$$R_e \leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathrm{Y}_2|\mathrm{U}) \tag{5.38b}$$
$$R_c + R_0 + R_i \leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \tag{5.38c}$$
$$R_0 + R_i \leq I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \tag{5.38d}$$

for random variables $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ with joint probability distribution $p_{\mathrm{U}}(u)p_{\mathrm{V}|\mathrm{U}}(v|u)p_{\mathrm{X}|\mathrm{V}}(x|v)W(y_1, y_2|x)$. Moreover, the cardinalities of the ranges of U and V can be bounded by

$$|\mathcal{U}| \leq |\mathcal{X}| + 3, \qquad |\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3.$$

From Theorem 5.32 follows immediately the *secrecy capacity region* $\mathcal{R}_{m_c,m_0}^S(W)$ of the BBC with common and confidential messages which is the set of rate tuples $(R_c, R_0, R_1, R_2) \in \mathbb{R}_+^4$ such that $(R_c, R_c, R_0, R_1, R_2) \in \mathcal{R}_{m_c,m_0}(W)$.

Corollary 5.33. *The secrecy capacity region $\mathcal{R}_{m_c,m_0}^S(W)$ of the discrete memoryless BBC with common and confidential messages is the set of all rate tuples $(R_c, R_0, R_1, R_2) \in \mathbb{R}_+^4$ that satisfy*

$$R_c \leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathrm{Y}_2|\mathrm{U}) \tag{5.39a}$$
$$R_0 + R_i \leq I(\mathrm{U};\mathrm{Y}_i), \quad i = 1,2 \tag{5.39b}$$

with perfect secrecy, i.e., (5.2) is satisfied, for random variables $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ with joint probability distribution $p_{\mathrm{U}}(u)p_{\mathrm{V}|\mathrm{U}}(v|u)p_{\mathrm{X}|\mathrm{V}}(x|v)W(y_1, y_2|x)$.

Remark 5.34. *This result unifies the previous results obtained so far. It is obvious that the BBC with confidential messages (and no common messages), cf. Corollary 5.22, is included in (5.39). To obtain the region without confidential messages, i.e., the BBC with common messages, cf. Theorem 5.5, we observe that there is no need for the auxiliary random variables anymore, since there are no confidential messages to transmit. Therefore, we set $\mathrm{U} = \mathrm{V} = \mathrm{X}$ in (5.39b) and obtain the region given by (5.3), cf. especially Remark 5.7. A more detailed discussion for the MIMO Gaussian case is given in Section 5.5.3.*

With the knowledge that we obtained for the previous (partial) scenarios where the relay either integrates only a common message or a confidential message, we are able to address the general scenario with common and confidential messages and therewith are able to prove Theorem 5.32.

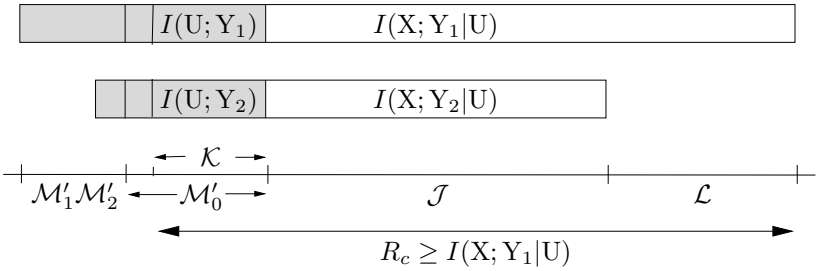


Figure 5.7: Available resources of each link are split up into two parts: one designated for the public common and bidirectional communication (gray) and one for the confidential communication (white). Since $R_c \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, the confidential message needs some resources designated for the common communication.

5.4.1 Achievability

To prove the achievability part of Theorem 5.32 we especially benefit from the codebook design given in Lemma 5.23. Although this codebook was originally designed for the BBC with only confidential messages, it already has the properties that are needed for an additional common message. With Lemma 5.23 we can show the following.

Lemma 5.35. *Let* $\mathrm{U} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ *and* $I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) > I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})$. *Using the codebook from Lemma 5.23 all rate-equivocation tuples* $(R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ *that satisfy*

$$0 \leq R_e = I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) \leq R_c \tag{5.40a}$$
$$R_c + R_0 + R_i \leq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_i), \quad i = 1, 2 \tag{5.40b}$$
$$R_0 + R_i \leq I(\mathrm{U};\mathrm{Y}_i), \quad i = 1, 2 \tag{5.40c}$$

are achievable for the BBC with common and confidential messages.

Proof. For given rate-equivocation tuple $(R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ that satisfies (5.40a)-(5.40c) we have to construct message sets, encoder, and decoders with

$$\tfrac{1}{n} \log |\mathcal{M}_c| \geq R_c - \delta \tag{5.41a}$$
$$\tfrac{1}{n} \log |\mathcal{M}_0| \geq R_0 - \delta \tag{5.41b}$$
$$\tfrac{1}{n} \log |\mathcal{M}_1| \geq R_2 - \delta \tag{5.41c}$$
$$\tfrac{1}{n} \log |\mathcal{M}_2| \geq R_1 - \delta \tag{5.41d}$$

and further, cf. also (5.1),

$$\tfrac{1}{n} H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{X};\mathrm{Y}_2|\mathrm{U}) - \delta. \tag{5.42}$$

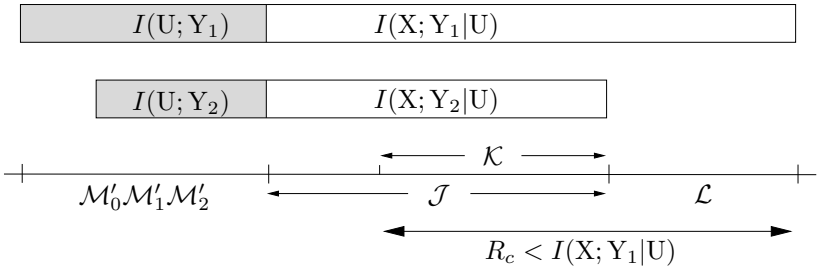


Figure 5.8: Since $R_c < I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, there are more resources for the confidential message available than needed. This allows the relay to enable a stochastic encoding strategy which exploits all the available resources by introducing a mapping from $\mathcal{J}$ to $\mathcal{K}$.

The construction is an extension of the one given in Lemma 5.24 by further integrating the common message. Similarly, we have to distinguish between two cases as shown in Figures 5.7 and 5.8.

If $R_c \geq I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, cf. Figure 5.7, the set of confidential messages is given by

$$\mathcal{M}_c \coloneqq \mathcal{J} \times \mathcal{L} \times \mathcal{K}$$

where $\mathcal{J}$ and $\mathcal{L}$ are chosen according to Lemma 5.23 and $\mathcal{K}$ is an arbitrary set such that (5.41a) holds. The sets $\mathcal{M}_1' = \mathcal{M}_1$, $\mathcal{M}_2' = \mathcal{M}_2$, and $\mathcal{M}_0' = \mathcal{M}_0 \times \mathcal{K}$ are chosen such that (5.41b)-(5.41d) are satisfied. The deterministic encoder f maps the confidential message $(j,l,k) \in \mathcal{M}_c$, and the common and individual messages $m_i \in \mathcal{M}_i$, $i = 0,1,2$, into the codeword $x^n_{jlm'} \in \mathcal{X}^n$ with $m' = (m_0', m_1', m_2')$ with $m_0' = (m_0, k)$ and $m_i' = m_i$, $i = 1,2$.

If $R_c < I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})$, cf. Figure 5.8, the set of confidential messages is given by $\mathcal{M}_c \coloneqq \mathcal{K}\times\mathcal{L}$ where $\mathcal{K}$ is arbitrary chosen such that (5.41a) is satisfied. In addition, we define a mapping $h : \mathcal{J} \to \mathcal{K}$ which partitions the set $\mathcal{J}$ into subsets of "nearly equal size" [CK78], i.e.,

$$|h^{-1}(k)| \leq 2|h^{-1}(k')|, \qquad \text{for all } k,k' \in \mathcal{K}.$$

The sets $\mathcal{M}_i'$ $=$ $\mathcal{M}_i$, i $=$ $0,1,2$, are arbitrary such that (5.41b)-(5.41d) are satisfied. The stochastic encoder f maps the confidential message $(k,l) \in \mathcal{M}_c$ and the common and individual messages $m_i \in \mathcal{M}_i$, $i = 0,1,2$, into the codeword $x^n_{jlm'} \in \mathcal{X}^n$ with $m' = (m_0', m_1', m_2')$ and $m_i' = m_i$, $i = 0,1,2$. The index j is uniformly drawn from the set $h^{-1}(k) \subset \mathcal{J}$.

In both cases the decoders are immediately determined by the codebook design given in Lemma 5.23. It remains to show that the equivocation rate fulfills (5.42). Since the confidential message is encoded in the same way as in Lemma 5.24 for the BBC with confidential messages (and no common message), we omit the details for brevity. □

Once we have established the achievable rate-equivocation region in Lemma 5.35, it is straightforward to show that this region equals the capacity-equivocation region stated in Theorem 5.32. Since the argumentation follows exactly the one presented in Section 5.3.2, we omit the details for brevity.

5.4.2 Converse

We have to show that for any given sequence $\{\mathcal{C}_{m_c,m_0}^{(n)}(W)\}_{n\in\mathbb{N}}$ of $(n, M_c^{(n)}, M_0^{(n)}, M_1^{(n)}, M_2^{(n)})$-codes with $\bar{e}_1, \bar{e}_2 \to 0$ there exist random variables $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ such that

$$\begin{aligned}
\tfrac{1}{n} H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathrm{Y}_2|\mathrm{U}) + o(n^0) \\
\tfrac{1}{n}\big(H(\mathrm{M}_c) + H(\mathrm{M}_0) + H(\mathrm{M}_2)\big) &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_1) + o(n^0) \\
\tfrac{1}{n}\big(H(\mathrm{M}_c) + H(\mathrm{M}_0) + H(\mathrm{M}_1)\big) &\leq I(\mathrm{V};\mathrm{Y}_1|\mathrm{U}) + I(\mathrm{U};\mathrm{Y}_2) + o(n^0) \\
\tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_2)\big) &\leq I(\mathrm{U};\mathrm{Y}_1) + o(n^0) \\
\tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_1)\big) &\leq I(\mathrm{U};\mathrm{Y}_2) + o(n^0)
\end{aligned}$$

are satisfied. For this purpose we need a version of Fano's lemma that is suitable for the BBC with common and confidential messages.

Lemma 5.36 (Fano's inequality)**.** *For the BBC with common and confidential messages we have the following versions of Fano's inequality*

$$\begin{aligned}
H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2|\mathrm{Y}_1^n, \mathrm{M}_1) &\leq \bar{e}_1 \log(M_c^{(n)} M_0^{(n)} M_2^{(n)}) + 1 = n\epsilon_1^{(n)} \\
H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) &\leq \bar{e}_2 \log(M_0^{(n)} M_1^{(n)}) + 1 = n\epsilon_2^{(n)}
\end{aligned}$$

with $\epsilon_1^{(n)} = \frac{1}{n}\log(M_c^{(n)} M_0^{(n)} M_2^{(n)})\bar{e}_1 + \frac{1}{n} \to 0$ *and* $\epsilon_2^{(n)} = \frac{1}{n}\log(M_0^{(n)} M_1^{(n)})\bar{e}_2 + \frac{1}{n} \to 0$ *for* $n \to \infty$ *as* $\bar{e}_1, \bar{e}_2 \to 0$.

Proof. The proof can be found in Appendix A.8. □

For notational convenience we introduce the abbreviation $\mathrm{M}_p = (\mathrm{M}_0, \mathrm{M}_1, \mathrm{M}_2)$ for the public communication. From the independence of M_c, M_0, M_1, M_2, the chain rule for entropy, the definition of mutual information, Fano's inequality, cf. Lemma 5.36, and the chain rule for mutual information we get for the entropies of the public messages

$$\begin{aligned}
H(\mathrm{M}_0) + H(\mathrm{M}_2) &= H(\mathrm{M}_0, \mathrm{M}_2|\mathrm{M}_1) \\
&= I(\mathrm{M}_0, \mathrm{M}_2; \mathrm{Y}_1^n|\mathrm{M}_1) + H(\mathrm{M}_0, \mathrm{M}_2|\mathrm{Y}_1^n, \mathrm{M}_1) \\
&\leq I(\mathrm{M}_0, \mathrm{M}_2; \mathrm{Y}_1^n|\mathrm{M}_1) + n\epsilon_1^{(n)} \\
&\leq I(\mathrm{M}_p; \mathrm{Y}_1^n) + n\epsilon_1^{(n)} \qquad (5.43)
\end{aligned}$$

and similarly

$$H(\mathrm{M}_0) + H(\mathrm{M}_1) \leq I(\mathrm{M}_p; \mathrm{Y}_2^n) + n\epsilon_2^{(n)}. \tag{5.44}$$

For the entropy of the confidential message we obtain

$$\begin{aligned} H(\mathrm{M}_c) &= H(\mathrm{M}_c|\mathrm{M}_p) \\ &= I(\mathrm{M}_c; \mathrm{Y}_1^n|\mathrm{M}_p) + H(\mathrm{M}_c|\mathrm{Y}_1^n, \mathrm{M}_p) \\ &\leq I(\mathrm{M}_c; \mathrm{Y}_1^n|\mathrm{M}_p) + H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2|\mathrm{Y}_1^n, \mathrm{M}_1) \\ &\leq I(\mathrm{M}_c; \mathrm{Y}_1^n|\mathrm{M}_p) + n\epsilon_1^{(n)} \end{aligned} \tag{5.45}$$

and further for the equivocation at the non-legitimate node

$$\begin{aligned} H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) &= H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_p) + I(\mathrm{M}_c; \mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) \\ &= H(\mathrm{M}_c|\mathrm{M}_p) - I(\mathrm{M}_c; \mathrm{Y}_2^n|\mathrm{M}_p) + I(\mathrm{M}_c; \mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) \\ &= I(\mathrm{M}_c; \mathrm{Y}_1^n|\mathrm{M}_p) - I(\mathrm{M}_c; \mathrm{Y}_2|\mathrm{M}_p) + H(\mathrm{M}_c|\mathrm{Y}_1^n, \mathrm{M}_p) + I(\mathrm{M}_c; \mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) \\ &\leq I(\mathrm{M}_c; \mathrm{Y}_1^n|\mathrm{M}_p) - I(\mathrm{M}_c; \mathrm{Y}_2^n|\mathrm{M}_p) + n\epsilon_1^{(n)} + n\epsilon_2^{(n)} \end{aligned} \tag{5.46}$$

where the last inequality follows from $H(\mathrm{M}_c|\mathrm{Y}_1^n, \mathrm{M}_p) \leq H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2|\mathrm{Y}_1^n, \mathrm{M}_1) \leq n\epsilon_1^{(n)}$, $I(\mathrm{M}_c; \mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) = H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) - H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_c, \mathrm{M}_2) \leq H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) \leq n\epsilon_2^{(n)}$, and Fano's inequality, cf. Lemma 5.36.

Once we have established the bounds (5.43)-(5.46), the rest of the proof goes along with Section 5.3.3. Starting from Equation (5.36) and introducing auxiliary random variables U and V that satisfy the Markov chain relation $\mathrm{U} - \mathrm{V} - \mathrm{X} - (\mathrm{Y}_1, \mathrm{Y}_2)$ it is straightforward to show that the rates are bounded by the desired conditions given in (5.38). □

5.5 Confidential Messages in MIMO Gaussian Bidirectional Relay Networks

In this section we prove the corresponding result for MIMO Gaussian channels. Therefore we assume N_R antennas at the relay node and N_i antennas at node i, $i = 1, 2$, as shown in Figure 5.9. The discrete-time real-valued input-output relation between the relay and node i, $i = 1, 2$, can now be modeled as

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x} + \boldsymbol{n}_i, \tag{5.47}$$

where $\boldsymbol{y}_i \in \mathbb{R}^{N_i \times 1}$ denotes the output at node i, $\boldsymbol{H}_i \in \mathbb{R}^{N_i \times N_R}$ the multiplicative channel matrix, $\boldsymbol{x} \in \mathbb{R}^{N_R \times 1}$ the input of the relay, and $\boldsymbol{n}_i \in \mathbb{R}^{N_i \times 1}$ the independent additive noise according to a Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{N_i})$ with zero mean and identity covariance matrix. We assume perfect channel state information at all nodes.

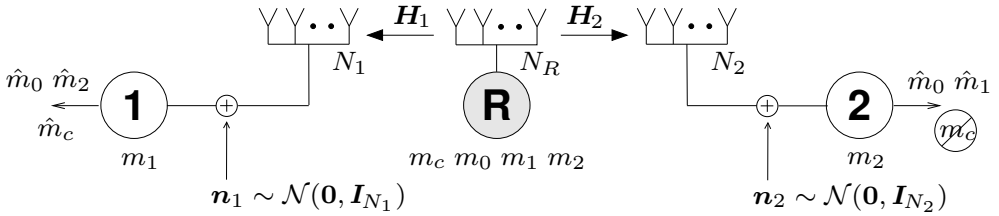


Figure 5.9: General MIMO Gaussian BBC with common and confidential messages.

As in [WSS06b, LS09, LLL10] we consider two different kinds of power constraints: an average power constraint and a more general matrix power constraint. An input sequence $\boldsymbol{x}^n = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n)$ of length n satisfies an average power constraint P if

$$\frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}_k^T \boldsymbol{x}_k \leq P \tag{5.48}$$

holds. Similarly, a sequence $\boldsymbol{x}^n$ satisfies a matrix power constraint $\boldsymbol{S}$ if

$$\frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}_k \boldsymbol{x}_k^T \preceq \boldsymbol{S} \tag{5.49}$$

where $\boldsymbol{S} \succeq \boldsymbol{0}$ is a positive semidefinite matrix.[1]

In principle, the secrecy capacity region of the MIMO Gaussian BBC with common and confidential messages is computable by evaluating the corresponding region of the discrete case, cf. Corollary 5.33, for MIMO Gaussian channels. But a direct evaluation is almost intractable due to the presence of the auxiliary random variables U and V so that we establish a precise matrix characterization in the following.

Theorem 5.37. *The secrecy capacity region $\mathcal{R}^S_{m_c,m_0}(\boldsymbol{H}_1, \boldsymbol{H}_2|\boldsymbol{S})$ of the MIMO Gaussian BBC with common and confidential messages under the matrix power constraint $\boldsymbol{S}$ is the set of all rate tuples $\boldsymbol{R} = (R_c, R_0, R_1, R_2) \in \mathbb{R}^4_+$ that satisfy*

$$R_c \leq \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_1} + \boldsymbol{H}_1\boldsymbol{Q}^{(c)}\boldsymbol{H}_1^T\right) - \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_2} + \boldsymbol{H}_2\boldsymbol{Q}^{(c)}\boldsymbol{H}_2^T\right) \tag{5.50a}$$

$$R_0 + R_i \leq \frac{1}{2}\log\det\left(\frac{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i\boldsymbol{S}\boldsymbol{H}_i^T}{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i\boldsymbol{Q}^{(c)}\boldsymbol{H}_i^T}\right), \quad i = 1, 2 \tag{5.50b}$$

for some $\boldsymbol{0} \preceq \boldsymbol{Q}^{(c)} \preceq \boldsymbol{S}$.

[1] The notation $\boldsymbol{A} \succeq \boldsymbol{B}$ means the matrix $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

Having [WSS06b, Lemma 1] in mind, we immediately obtain from the secrecy capacity region under the matrix power constraint (5.49) the corresponding region under the average power constraint (5.48) which usually characterizes the practically more relevant case.

Corollary 5.38. *The secrecy capacity region $\mathcal{R}^S_{m_c,m_0}(\boldsymbol{H}_1, \boldsymbol{H}_2|P)$ of the MIMO Gaussian BBC with common and confidential messages under the average power constraint P is the set of all rate tuples $\boldsymbol{R} \in \mathbb{R}^4_+$ that satisfy*

$$\begin{aligned} R_c &\leq \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_1} + \boldsymbol{H}_1\boldsymbol{Q}^{(c)}\boldsymbol{H}_1^T\right) - \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_2} + \boldsymbol{H}_2\boldsymbol{Q}^{(c)}\boldsymbol{H}_2^T\right) \\ R_0 + R_i &\leq \frac{1}{2}\log\det\left(\frac{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i(\boldsymbol{Q}^{(c)} + \boldsymbol{Q}^{(p)})\boldsymbol{H}_i^T}{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i\boldsymbol{Q}^{(c)}\boldsymbol{H}_i^T}\right), \quad i = 1, 2 \end{aligned}$$

for some $\boldsymbol{Q}^{(c)} \succeq \boldsymbol{0}$, $\boldsymbol{Q}^{(p)} \succeq \boldsymbol{0}$ with tr$(\boldsymbol{Q}^{(c)} + \boldsymbol{Q}^{(p)}) \leq P$.

Theorem 5.37 is proved in the following subsections. First, we consider the special case of square and invertible channel matrices and establish the secrecy capacity region for this case using channel-enhancement arguments. Then, we extend this to arbitrary (possibly non-square and non-invertible) channel matrices using standard approximation arguments as in [WSS06b, LS09, LLL10] to finally end up with the desired result.

5.5.1 Aligned MIMO Bidirectional Broadcast Channel

In this section we consider the case where the channel matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are square and invertible. Then, multiplying both sides (5.47) by $\boldsymbol{H}_i^{-1}$, an equivalent channel model is given by

$$\boldsymbol{y}_i = \boldsymbol{x} + \boldsymbol{n}_i \tag{5.51}$$

where $\boldsymbol{y}_i, \boldsymbol{x}, \boldsymbol{n}_i \in \mathbb{R}^{N_R \times 1}$ but the additive noise $\boldsymbol{n}_i$ is now Gaussian distributed with zero mean and covariance matrix

$$\boldsymbol{\Sigma}_i = \boldsymbol{H}_i^{-1}\boldsymbol{H}_i^{-T} \in \mathbb{R}^{N_R \times N_R}, \tag{5.52}$$

i.e., $\boldsymbol{n}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, $i = 1, 2$. We adopt the notation used in [WSS06b, LLL10] and call the channel model (5.51) the *aligned* MIMO Gaussian BBC and (5.47) the *general* MIMO Gaussian BBC. The main result for the aligned case is summarized in the following theorem.

Theorem 5.39. *The secrecy capacity region $\mathcal{R}^S_{m_c,m_0}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2|\boldsymbol{S})$ of the aligned MIMO Gaussian BBC with common and confidential messages under the matrix power constraint $\boldsymbol{S}$ is*

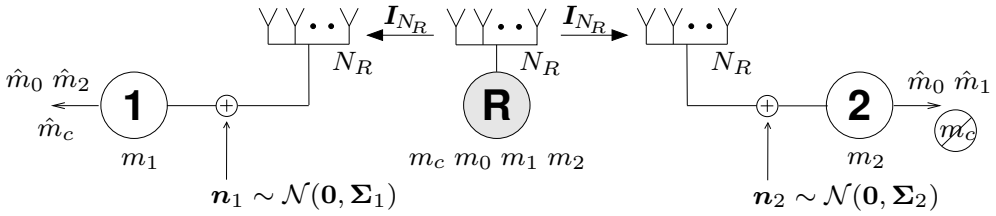


Figure 5.10: Aligned MIMO Gaussian BBC with common and confidential messages.

the set of all rate tuples $\boldsymbol{R} \in \mathbb{R}_+^4$ *that satisfy*

$$\begin{aligned} R_c &\leq \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}^{(c)}+\boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1}\right) - \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}^{(c)}+\boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2}\right) \\ R_0 + R_i &\leq \frac{1}{2}\log\det\left(\frac{\boldsymbol{S}+\boldsymbol{\Sigma}_i}{\boldsymbol{Q}^{(c)}+\boldsymbol{\Sigma}_i}\right), \quad i = 1,2 \end{aligned} \tag{5.53}$$

for some $\boldsymbol{0} \preceq \boldsymbol{Q}^{(c)} \preceq \boldsymbol{S}$.

The theorem is proved in the following subsections.

Proof of Achievability

Similarly as for the classical aligned MIMO Gaussian broadcast channel with common and confidential messages [LLL10] the proof of achievability is a straightforward extension of its discrete counterpart. To obtain the desired region (5.53) we follow the proof of the discrete case with a proper choice of auxiliary and input random variables. More precisely, with $\mathrm{G} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{(c)})$ and $\mathrm{U} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S} - \boldsymbol{Q}^{(c)})$ with G and U are independent, and further $\mathrm{V} = \mathbf{X} = \mathrm{U} + \mathrm{G}$, the region (5.53) follows immediately from Corollary 5.33. Therefore we omit the details for brevity.

Remark 5.40. *Interestingly, a simple superposition strategy that superimposes two signals, one for the public messages and one for the confidential message, suffices to achieve the secrecy capacity. Moreover, an additional randomization as in the discrete case, realized by the auxiliary random variable* V *in Theorem 5.32, is no longer needed for MIMO Gaussian channels.*

Proof of Converse

To establish the converse it remains to show that no other rate tuples than characterized by (5.53) are achievable for some $\boldsymbol{0} \preceq \boldsymbol{Q}^{(c)} \preceq \boldsymbol{S}$. Without loss of generality it suffices at this

point to consider only matrix power constraints that satisfy $\boldsymbol{S} \succ \mathbf{0}$.[2]

We prove the optimality by contradiction. Therefore, we construct a rate tuple $\boldsymbol{R}^o = (R_c^o, R_0^o, R_1^o, R_2^o) \notin \mathcal{R}^S_{m_c,m_0}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2|\boldsymbol{S})$ that lies outside the desired region (5.53) and assume that this rate tuple is achievable for the aligned MIMO Gaussian BBC with common and confidential messages.

First, we observe that achievable public rates R_0^o, R_1^o, and R_2^o are always bounded from above by

$$R_0^o + R_i^o \leq \frac{1}{2} \log \det \left(\frac{\boldsymbol{S} + \boldsymbol{\Sigma}_i}{\boldsymbol{\Sigma}_i} \right), \quad i = 1, 2.$$

We note that for $R_c^o = 0$ and $\boldsymbol{Q}^{(c)} = \mathbf{0}$ in (5.53) there are public rates that actually achieve this upper bound. Further, for given achievable public rates R_0^o, R_1^o, and R_2^o the maximal achievable confidential rate $R_{c,\text{opt}}$ is characterized by the following optimization problem:

$$\begin{aligned} \max_{\boldsymbol{Q}^{(c)}} \quad & \frac{1}{2} \log \det \left(\frac{\boldsymbol{Q}^{(c)} + \boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1} \right) - \frac{1}{2} \log \det \left(\frac{\boldsymbol{Q}^{(c)} + \boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2} \right) \\ \text{s.t.} \quad & \frac{1}{2} \log \det \left(\frac{\boldsymbol{S} + \boldsymbol{\Sigma}_i}{\boldsymbol{Q}^{(c)} + \boldsymbol{\Sigma}_i} \right) \geq R_0^o + R_i^o, \quad i = 1, 2 \\ & \mathbf{0} \preceq \boldsymbol{Q}^{(c)} \preceq \boldsymbol{S}. \end{aligned} \tag{5.54}$$

Finally, we set $R_c^o = R_{c,\text{opt}} + \delta$ for some $\delta > 0$ to ensure that this rate tuple lies outside the region (5.53) as required, i.e., $\boldsymbol{R}^o \notin \mathcal{R}^S_{m_c,m_0}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2|\boldsymbol{S})$.

The optimization problem (5.54) can be written as a minimization problem in standard form as

$$\begin{aligned} \min_{\boldsymbol{Q}^{(c)}} \quad & \frac{1}{2} \log \det \left(\frac{\boldsymbol{Q}^{(c)} + \boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2} \right) - \frac{1}{2} \log \det \left(\frac{\boldsymbol{Q}^{(c)} + \boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1} \right) \\ \text{s.t.} \quad & R_0^o + R_i^o - \frac{1}{2} \log \det \left(\frac{\boldsymbol{S} + \boldsymbol{\Sigma}_i}{\boldsymbol{Q}^{(c)} + \boldsymbol{\Sigma}_i} \right) \leq 0, \quad i = 1, 2 \\ & -\boldsymbol{Q}^{(c)} \preceq \mathbf{0} \\ & \boldsymbol{Q}^{(c)} - \boldsymbol{S} \preceq \mathbf{0}. \end{aligned}$$

[2] For the validity and a detailed proof of this restriction we refer to [WSS06b, Lemma 2]. The argumentation is as follows: if we are confronted with a matrix power constraint $\boldsymbol{S} \succeq \mathbf{0}$ that is positive semidefinite with $\det(\boldsymbol{S}) = 0$, we can define an equivalent aligned MIMO Gaussian BBC with $N_R' = \text{rank}(\boldsymbol{S}) < N_R$ transmit and receive antennas at all nodes and a modified matrix power constraint $\boldsymbol{S}' \succ \mathbf{0}$ that is strictly positive definite. Thus, the case $\boldsymbol{S} \succeq \mathbf{0}$ with $\det(\boldsymbol{S}) = 0$ can be converted to the case $\boldsymbol{S}' \succ \mathbf{0}$ without changing the secrecy capacity region.

Then the Lagrangian for this minimization problem is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{Q}^{(c)},\boldsymbol{\mu},\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_2) = \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}^{(c)}+\boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2}\right) - \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}^{(c)}+\boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1}\right)\\ + \sum_{i=1}^{2}\mu_i\left[R_0^o + R_i^o - \frac{1}{2}\log\det\left(\frac{\boldsymbol{S}+\boldsymbol{\Sigma}_i}{\boldsymbol{Q}^{(c)}+\boldsymbol{\Sigma}_i}\right)\right]\\ - \operatorname{tr}(\boldsymbol{Q}^{(c)}\boldsymbol{\Psi}_1) + \operatorname{tr}\big((\boldsymbol{Q}^{(c)}-\boldsymbol{S})\boldsymbol{\Psi}_2\big)\end{aligned}$$

with Lagrange multipliers $\boldsymbol{\mu} = (\mu_1,\mu_2)$, $\mu_i \geq 0$, and $\boldsymbol{\Psi}_i \succeq \mathbf{0}$, $i = 1,2$. Then we know from the Karush-Kuhn-Tucker (KKT) conditions, cf. for example [BV04], that the derivative of the Lagrangian must vanish at an optimal $\boldsymbol{Q}_{\text{opt}}^{(c)}$ which yields[3]

$$\frac{1}{2}\big(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1\big)^{-1} + \boldsymbol{\Psi}_1 = \frac{\mu_1}{2}\big(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1\big)^{-1} + \frac{\mu_2+1}{2}\big(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2\big)^{-1} + \boldsymbol{\Psi}_2 \tag{5.55}$$

while the optimal $\boldsymbol{Q}_{\text{opt}}^{(c)}$ further has to satisfy the complementary slackness conditions

$$\mu_i\left[R_0^o + R_i^o - \frac{1}{2}\log\det\left(\frac{\boldsymbol{S}+\boldsymbol{\Sigma}_i}{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_i}\right)\right] = 0, \quad i = 1,2 \tag{5.56}$$

$$\boldsymbol{Q}_{\text{opt}}^{(c)}\boldsymbol{\Psi}_1 = \mathbf{0}, \qquad (\boldsymbol{S}-\boldsymbol{Q}_{\text{opt}}^{(c)})\boldsymbol{\Psi}_2 = \mathbf{0}. \tag{5.57}$$

By combining (5.54) and (5.56) we get for the weighted secrecy sum-capacity of the constructed rate tuple $\boldsymbol{R}^o$ the following

$$\begin{aligned}&R_c^o + \mu_1(R_0^o+R_1^o) + \mu_2(R_0^o+R_2^o)\\ &\quad= R_{c,\text{opt}} + \delta + \mu_1(R_0^o+R_1^o) + \mu_2(R_0^o+R_2^o)\\ &\quad= \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1}\right) - \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2}\right) + \sum_{i=1}^{2}\frac{\mu_i}{2}\log\det\left(\frac{\boldsymbol{S}+\boldsymbol{\Sigma}_i}{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_i}\right) + \delta.\end{aligned} \tag{5.58}$$

But we show that for any achievable rate tuple $\boldsymbol{R} \in \mathbb{R}_+^4$ the weighted secrecy sum-capacity is bounded from above by

$$\begin{aligned}&R_c + \mu_1(R_0+R_1) + \mu_2(R_0+R_2)\\ &\quad\leq \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1}\right) - \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2}\right) + \sum_{i=1}^{2}\frac{\mu_i}{2}\log\det\left(\frac{\boldsymbol{S}+\boldsymbol{\Sigma}_i}{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_i}\right)\end{aligned}$$

which establishes the desired contradiction to (5.58).

[3] As in [WSS06b, Appendix IV] or [LLL10] we can easily show that a set of constraint qualifications hold for the optimization problem (5.54). This implies that the KKT conditions hold and are necessary for characterizing the optimal transmit covariance matrix.

Reinterpretation of Legitimate Receiver

For the following analysis it will be beneficial to reinterpret this scenario by splitting the legitimate node 1 into two virtual receivers: one designated for the public and one for the confidential communication. Then, an equivalent aligned MIMO Gaussian BBC can be represented by

$$\boldsymbol{y}_{1\mathrm{a}} = \boldsymbol{x} + \boldsymbol{n}_{1\mathrm{a}} \tag{5.59a}$$

$$\boldsymbol{y}_{1\mathrm{b}} = \boldsymbol{x} + \boldsymbol{n}_{1\mathrm{b}} \tag{5.59b}$$

$$\boldsymbol{y}_{2} = \boldsymbol{x} + \boldsymbol{n}_{2} \tag{5.59c}$$

with $\boldsymbol{n}_{1\mathrm{a}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$, $\boldsymbol{n}_{1\mathrm{b}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$, and $\boldsymbol{n}_2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_2)$. Each (virtual) receiver is only interested in either the public or the confidential messages. Receiver 1a wants to know the confidential message m_c, receiver 1b the public messages m_0 and m_2, and receiver 2 the public messages m_0 and m_1. Here, the confidential message has to be kept secret only from receiver 2, but, of course, need not be kept secret from (virtual) receiver 1b.

Note that (virtual) receivers 1a and 1b in (5.59a) are affected by noise that has the same covariance matrix $\boldsymbol{\Sigma}_1$, cf. (5.52), which is the same as of the noise at the legitimate receiver 1 in the original aligned BBC (5.51). Similarly, the noise at receiver 2 in (5.59c) is according to the same covariance matrix $\boldsymbol{\Sigma}_2$, cf. (5.52), corresponding to the noise at the non-legitimate receiver 2 in (5.51). Therefore, any strategy that achieves a certain rate tuple for (5.51) will do likewise for (5.59), and vice versa, so that both scenarios share the same secrecy capacity region.

Channel Enhancement

Next, with the reinterpretation (5.59) of the communication scenario as a starting point, we enhance the channel designated for the confidential message, i.e., (virtual) receiver 1a. For this purpose let $\widetilde{\boldsymbol{\Sigma}}_1$ be a real symmetric matrix that satisfies

$$\frac{1}{2}(\boldsymbol{Q}_{\mathrm{opt}}^{(c)} + \widetilde{\boldsymbol{\Sigma}}_1)^{-1} = \frac{1}{2}(\boldsymbol{Q}_{\mathrm{opt}}^{(c)} + \boldsymbol{\Sigma}_1)^{-1} + \boldsymbol{\Psi}_1. \tag{5.60}$$

Then we know from [WSS06b, Lemma 11] that

$$\mathbf{0} \prec \widetilde{\boldsymbol{\Sigma}}_1 \preceq \boldsymbol{\Sigma}_1 \tag{5.61}$$

and

$$\det\left(\frac{\boldsymbol{Q}_{\mathrm{opt}}^{(c)} + \widetilde{\boldsymbol{\Sigma}}_1}{\widetilde{\boldsymbol{\Sigma}}_1}\right) = \det\left(\frac{\boldsymbol{Q}_{\mathrm{opt}}^{(c)} + \boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1}\right) \tag{5.62}$$

hold. With (5.60), Equation (5.55) becomes

$$\frac{1}{2}(\boldsymbol{Q}_{\text{opt}}^{(c)}+\widetilde{\boldsymbol{\Sigma}}_1)^{-1}=\frac{\mu_1}{2}(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1)^{-1}+\frac{\mu_2+1}{2}(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2)^{-1}+\boldsymbol{\Psi}_2. \tag{5.63}$$

Since the matrices $(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1)^{-1}$, $(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2)^{-1}$, and $\boldsymbol{\Psi}_2$ on the right hand side of (5.63) are all positive semidefinite, it follows immediately that $\frac{1}{2}(\boldsymbol{Q}_{\text{opt}}^{(c)}+\widetilde{\boldsymbol{\Sigma}}_1)^{-1}\succeq\frac{1}{2}(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2)^{-1}$ and consequently

$$\widetilde{\boldsymbol{\Sigma}}_1\preceq\boldsymbol{\Sigma}_2. \tag{5.64}$$

This allows us to construct an enhanced MIMO Gaussian BBC by replacing the noise covariance matrix $\boldsymbol{\Sigma}_1$ at the (virtual) receiver 1a with its enhanced version $\widetilde{\boldsymbol{\Sigma}}_1$, cf. (5.61). Then, (5.59a) becomes

$$\tilde{\boldsymbol{y}}_{1\text{a}}=\boldsymbol{x}+\tilde{\boldsymbol{n}}_{1\text{a}} \tag{5.65}$$

with $\tilde{\boldsymbol{n}}_{1\text{a}}\sim\mathcal{N}(\boldsymbol{0},\widetilde{\boldsymbol{\Sigma}}_1)$, while the channels for receiver 1b and 2 remain the same. Figure 5.11 shows the communication scenario of the enhanced MIMO Gaussian BBC. Since $\widetilde{\boldsymbol{\Sigma}}_1\preceq\boldsymbol{\Sigma}_1$, cf. also (5.61), the covariance matrix of the noise for receiving the confidential message for the enhanced BBC (5.65) is "*smaller*" than for the original BBC (5.59). Hence, its secrecy capacity region is at least as large as for the aligned MIMO Gaussian BBC. Moreover, from (5.61) and (5.64) we get

$$\boldsymbol{0}\preceq\widetilde{\boldsymbol{\Sigma}}_1\preceq\boldsymbol{\Sigma}_i,\quad i=1,2 \tag{5.66}$$

which means that both received signals $\boldsymbol{y}_{1\text{b}}$ and $\boldsymbol{y}_2$ at the public receivers are (stochastically) degraded with respect to the received signal $\tilde{\boldsymbol{y}}_{1\text{a}}$ at the confidential receiver. For the discrete memoryless counterpart of the enhanced BBC the following proposition characterizes the corresponding secrecy capacity region.

Proposition 5.41. *For a discrete memoryless BBC with common and confidential messages and transition probability $\widetilde{W}(\tilde{y}_{1a},y_{1b},y_2|x)$ that satisfies the Markov chain conditions $\mathrm{X}-\widetilde{\mathrm{Y}}_{1a}-\mathrm{Y}_{1b}$ and $\mathrm{X}-\widetilde{\mathrm{Y}}_{1a}-\mathrm{Y}_2$, the secrecy capacity region is given by the set of all rate tuples $\boldsymbol{R}\in\mathbb{R}_+^4$ that satisfy*

$$\begin{aligned}R_c&\le I(\mathrm{X};\widetilde{\mathrm{Y}}_{1a}|\mathrm{U})-I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})\\ R_0+R_1&\le I(\mathrm{U};\mathrm{Y}_{1b})\\ R_0+R_2&\le I(\mathrm{U};\mathrm{Y}_2)\end{aligned}$$

for random variables $\mathrm{U}-\mathrm{X}-\widetilde{\mathrm{Y}}_{1a}-(\mathrm{Y}_{1b},\mathrm{Y}_2)$ with joint probability distribution $p_{\mathrm{U}}(u)p_{\mathrm{X}|\mathrm{U}}(x|u)\widetilde{W}(\tilde{y}_{1a},y_{1b},y_2|x)$.

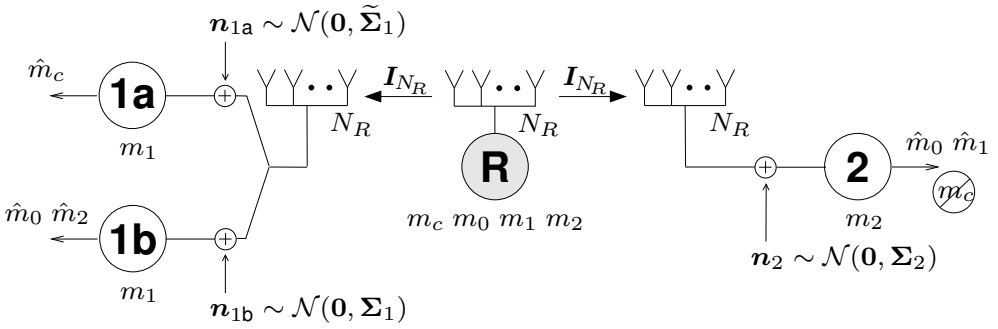


Figure 5.11: Enhanced MIMO Gaussian BBC with common and confidential messages. Node 1 is split up into two virtual receivers, one enhanced for the confidential message and one for the public messages. For receiver 1a the noise covariance matrix $\boldsymbol{\Sigma}_1$ is replaced by $\widetilde{\boldsymbol{\Sigma}}_1$ to enhance the channel for the confidential message.

Proof. The achievability follows immediately from the non-degraded case using the same ideas and techniques, cf. Section 5.4.1. Similarly, the converse follows the one in Section 5.4.2 while exploiting the degradedness as in [LLL10, Proposition 1]. Since the converse part is the more relevant one in the following, the details of the proof are given in Appendix A.10 for completeness. □

Remark 5.42. *In contrast to the non-degraded case, cf. Theorem 5.32, we only need one auxiliary random variable* U *in the degraded scenario instead of both* U *and* V*. This makes the evaluation of the secrecy capacity region for MIMO Gaussian channels tractable as done in the following.*

Equivalence of Weighted Secrecy Sum-Capacity

To establish the desired contradiction it remains to bound the weighted secrecy sum-capacity of the enhanced MIMO Gaussian BBC. As in [LLL10] for the classical MIMO Gaussian broadcast channel with common and confidential messages we use an extremal entropy inequality that is a special case of [WLS$^+$09, Corollary 4].

Proposition 5.43 ([WLS$^+$09, Corollary 4])**.** *Let $\tilde{\boldsymbol{n}}_{1a} \sim \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_1)$, $\boldsymbol{n}_{1b} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$, and $\boldsymbol{n}_2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_2)$ be given which satisfy $\mathbf{0} \succeq \widetilde{\boldsymbol{\Sigma}}_1 \succeq \boldsymbol{\Sigma}_i$, $i = 1, 2$, cf. (5.66). Further, let $\boldsymbol{S} \succ \mathbf{0}$ be given. If there exists a $N_R \times N_R$ real symmetric matrix $\boldsymbol{Q}_{opt}^{(c)}$ such that $\mathbf{0} \preceq \boldsymbol{Q}_{opt}^{(c)} \preceq \boldsymbol{S}$*

and satisfying

$$\frac{1}{2}(\boldsymbol{Q}_{opt}^{(c)}+\widetilde{\boldsymbol{\Sigma}}_1)^{-1}=\frac{\mu\lambda}{2}(\boldsymbol{Q}_{opt}^{(c)}+\boldsymbol{\Sigma}_1)^{-1}+\frac{\mu(1-\lambda)}{2}(\boldsymbol{Q}_{opt}^{(c)}+\boldsymbol{\Sigma}_2)^{-1}+\boldsymbol{\Psi}_2,$$
$$(\boldsymbol{S}-\boldsymbol{Q}_{opt}^{(c)})\boldsymbol{\Psi}_2=\mathbf{0}$$

for some $\boldsymbol{\Psi}_2\succeq\mathbf{0}$ *and real scalars* $\mu\geq 0$ *and* $0\leq\lambda\leq 1$*, then*

$$\begin{aligned}h(\mathbf{X}+\widetilde{\mathbf{N}}_{1a}|\mathrm{U})&-\mu\lambda h(\mathbf{X}+\mathbf{N}_{1b}|\mathrm{U})-\mu(1-\lambda)h(\mathbf{X}+\mathbf{N}_2|\mathrm{U})\\&\leq\frac{1}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{opt}^{(c)}+\widetilde{\mathbf{N}}_1)\right)-\frac{\mu\lambda}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{opt}^{(c)}+\mathbf{N}_1)\right)\\&\quad-\frac{\mu(1-\lambda)}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{opt}^{(c)}+\mathbf{N}_2)\right)\end{aligned}$$

for any $(\mathrm{U},\mathbf{X})$ *independent of* $(\widetilde{\mathbf{N}}_{1a},\mathbf{N}_{1b},\mathbf{N}_2)$ *such that* $\mathbb{E}\{\mathbf{X}\mathbf{X}^T\}\preceq\boldsymbol{S}$*.*

By Proposition 5.41 we get for the weighted secrecy sum-capacity of any rate tuple $\boldsymbol{R}\in\mathbb{R}_+^4$ for the enhanced BBC (5.65)

$$\begin{aligned}&R_c+\mu_1(R_0+R_1)+\mu_2(R_0+R_2)\\&\quad\leq I(\mathbf{X};\widetilde{\mathbf{Y}}_{1\mathrm{a}}|\mathrm{U})-I(\mathbf{X};\mathbf{Y}_2|\mathrm{U})+\mu_1 I(\mathrm{U};\mathbf{Y}_{1\mathrm{b}})+\mu_2 I(\mathrm{U};\mathbf{Y}_2)\\&\quad=h(\mathbf{N}_2)-h(\widetilde{\mathbf{N}}_{1\mathrm{a}})+\mu_1 h(\mathbf{X}+\mathbf{N}_{1\mathrm{b}})+\mu_2 h(\mathbf{X}+\mathbf{N}_2)\\&\qquad+\Big[h(\mathbf{X}+\widetilde{\mathbf{N}}_{1\mathrm{a}}|\mathrm{U})-\mu_1 h(\mathbf{X}+\mathbf{N}_{1\mathrm{b}}|\mathrm{U})-(\mu_2+1)h(\mathbf{X}+\mathbf{N}_2|\mathrm{U})\Big]\\&\quad\leq\frac{1}{2}\log\det\left(2\pi e\boldsymbol{\Sigma}_2\right)-\frac{1}{2}\log\det\left(2\pi e\widetilde{\boldsymbol{\Sigma}}_1\right)+\sum_{i=1}^{2}\frac{\mu_i}{2}\log\det\left(2\pi e(\boldsymbol{S}+\boldsymbol{\Sigma}_i)\right)\\&\qquad+\Big[h(\mathbf{X}+\widetilde{\mathbf{N}}_{1\mathrm{a}}|\mathrm{U})-\mu_1 h(\mathbf{X}+\mathbf{N}_{1\mathrm{b}}|\mathrm{U})-(\mu_2+1)h(\mathbf{X}+\mathbf{N}_2|\mathrm{U})\Big]\end{aligned}\tag{5.67}$$

where the last inequality follows from $h(\widetilde{\mathbf{N}}_{1\mathrm{a}})=\frac{1}{2}\log\det(2\pi e\widetilde{\boldsymbol{\Sigma}}_1)$, $h(\mathbf{N}_2)=\frac{1}{2}\log\det(2\pi e\boldsymbol{\Sigma}_2)$ and $h(\mathbf{X}+\mathbf{N}_{1\mathrm{b}})\leq\frac{1}{2}\log\det(2\pi e(\boldsymbol{S}+\boldsymbol{\Sigma}_1))$, $h(\mathbf{X}+\mathbf{N}_2)\leq\frac{1}{2}\log\det(2\pi e(\boldsymbol{S}+\boldsymbol{\Sigma}_2))$.

Now with $\mu=\mu_1+\mu_2+1$ and $\lambda=\frac{\mu_1}{\mu_1+\mu_2+1}$ we get from (5.63) together with Proposition 5.43

$$\begin{aligned}h(\mathbf{X}+\widetilde{\mathbf{N}}_{1\mathrm{a}}|\mathrm{U})&-\mu_1 h(\mathbf{X}+\mathbf{N}_{1\mathrm{b}}|\mathrm{U})-(\mu_2+1)h(\mathbf{X}+\mathbf{N}_2|\mathrm{U})\\&\leq\frac{1}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{\mathrm{opt}}^{(c)}+\widetilde{\boldsymbol{\Sigma}}_1)\right)-\frac{\mu_1}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{\mathrm{opt}}^{(c)}+\boldsymbol{\Sigma}_1)\right)\\&\quad-\frac{\mu_2+1}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{\mathrm{opt}}^{(c)}+\boldsymbol{\Sigma}_2)\right).\end{aligned}$$

Substituting this into (5.67) we end up with

$$\begin{aligned}
&R_c + \mu_1(R_0 + R_1) + \mu_2(R_0 + R_2) \\
&\quad \leq \frac{1}{2}\log\det\left(2\pi e\boldsymbol{\Sigma}_2\right) - \frac{1}{2}\log\det\left(2\pi e\widetilde{\boldsymbol{\Sigma}}_1\right) \\
&\qquad + \sum_{i=1}^{2}\frac{\mu_i}{2}\log\det\left(2\pi e(\boldsymbol{S}+\boldsymbol{\Sigma}_i)\right) + \frac{1}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{\text{opt}}^{(c)}+\widetilde{\boldsymbol{\Sigma}}_1)\right) \\
&\qquad - \frac{\mu_1}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1)\right) - \frac{\mu_2+1}{2}\log\det\left(2\pi e(\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2)\right) \\
&\quad = \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\widetilde{\boldsymbol{\Sigma}}_1}{\widetilde{\boldsymbol{\Sigma}}_1}\right) - \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2}\right) \\
&\qquad + \sum_{i=1}^{2}\frac{\mu_i}{2}\log\det\left(\frac{\boldsymbol{S}+\boldsymbol{\Sigma}_i}{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_i}\right) \\
&\quad = \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_1}{\boldsymbol{\Sigma}_1}\right) - \frac{1}{2}\log\det\left(\frac{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_2}{\boldsymbol{\Sigma}_2}\right) \\
&\qquad + \sum_{i=1}^{2}\frac{\mu_i}{2}\log\det\left(\frac{\boldsymbol{S}+\boldsymbol{\Sigma}_i}{\boldsymbol{Q}_{\text{opt}}^{(c)}+\boldsymbol{\Sigma}_i}\right)
\end{aligned} \tag{5.68}$$

where the last equality follows from (5.62), cf. [WSS06b, Lemma 11].

Since the secrecy capacity region of the aligned MIMO Gaussian BBC (5.47) is contained in the corresponding region of the enhanced MIMO Gaussian BBC (5.65), cf. also Section 5.5.1, it is clear that for any rate tuple $\boldsymbol{R} \in \mathbb{R}_+^4$ the upper bound on the weighted secrecy sum-capacity (5.68) – established above for the enhanced MIMO Gaussian BBC – must hold, of course, also for the non-enhanced aligned MIMO Gaussian BBC. But since $\delta > 0$, this contradicts (5.58). This completes the proof of converse and therewith establishes the secrecy capacity region $\mathcal{R}_{m_c,m_0}^{S}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2|\boldsymbol{S})$. □

5.5.2 General MIMO Bidirectional Broadcast Channel

To prove the secrecy capacity region of the general MIMO Gaussian BBC with common and confidential messages, cf. Theorem 5.37, we extend the secrecy capacity region of the corresponding aligned MIMO Gaussian BBC (5.51), cf. Theorem 5.39, to the general case (5.47) where the channel matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ need not be necessarily square and invertible. The proof follows the one of the classical MIMO Gaussian broadcast channel with common and confidential messages [LLL10] which is based on [WSS06b, LS09]. Although our proof is almost identical to [LLL10, Section IV] we present it for completeness in the following.

Proof of Achievability

Similarly as for the aligned case, cf. Theorem 5.39, the achievability follows from the discrete result in Corollary 5.33 with the same choice of auxiliary and input random variables, i.e., $\mathrm{G} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{(c)})$ and $\mathrm{U} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{S} - \boldsymbol{Q}^{(c)})$ with G and U are independent, and further $\mathrm{V} = \mathbf{X} = \mathrm{U} + \mathrm{G}$. We omit the details for brevity.

Proof of Converse

As argued in the previous subsection the case of square and invertible channel matrices can easily be transformed into an aligned MIMO Gaussian BBC whose secrecy capacity region is known from Theorem 5.39. Thus, the goal is to approximate any general MIMO Gaussian BBC (with possibly non-square and non-invertible channel matrices) by an appropriate aligned MIMO Gaussian BBC. We follow [LLL10, Section IV] where similar approximation arguments are presented for the classical MIMO Gaussian broadcast channel with common and confidential messages.

First, we consider the case where $\boldsymbol{H}_1 \in \mathbb{R}^{N_1 \times N_R}$ and $\boldsymbol{H}_2 \in \mathbb{R}^{N_2 \times N_R}$ are not square. Using the singular value decomposition (SVD) it can be shown that there exists equivalent square channel matrices $\widetilde{\boldsymbol{H}}_i \in \mathbb{R}^{N_R \times N_R}$, $i = 1, 2$, that yield the same secrecy capacity region. For details we refer to [WSS06b, Section V-B].

Consequently, we can assume without loss of generality that the channel matrices are square. It remains to check when these matrices are not invertible. We can apply the SVD to write

$$\boldsymbol{H}_i = \boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{V}_i^T$$

with $\boldsymbol{U}_i$, $\boldsymbol{V}_i$ unitary matrices and $\boldsymbol{\Lambda}_i$ diagonal, $i = 1, 2$. Next, we define channel matrices that are definitely invertible as

$$\overline{\boldsymbol{H}}_i = \boldsymbol{U}_i (\boldsymbol{\Lambda}_i + \alpha \boldsymbol{I}_{N_R}) \boldsymbol{V}_i^T \tag{5.69}$$

for (small) $\alpha > 0$ and the corresponding MIMO Gaussian BBC

$$\overline{\boldsymbol{y}}_i = \overline{\boldsymbol{H}}_i \boldsymbol{x} + \boldsymbol{n}_i, \quad i = 1, 2. \tag{5.70}$$

Since $\overline{\boldsymbol{H}}_1$ and $\overline{\boldsymbol{H}}_2$ are square and invertible, we know from Theorem 5.39 that the secrecy capacity region $\mathcal{R}^S_{m_c, m_0}(\overline{\boldsymbol{H}}_1, \overline{\boldsymbol{H}}_2 | \boldsymbol{S})$ is given by the set of all rate tuples $\boldsymbol{R} \in \mathbb{R}^4_+$ that satisfy

$$\begin{aligned} R_c &\leq \frac{1}{2} \log\det\left(\boldsymbol{I}_{N_1} + \overline{\boldsymbol{H}}_1 \boldsymbol{Q}^{(c)} \overline{\boldsymbol{H}}_1^T\right) - \frac{1}{2} \log\det\left(\boldsymbol{I}_{N_2} + \overline{\boldsymbol{H}}_2 \boldsymbol{Q}^{(c)} \overline{\boldsymbol{H}}_2^T\right) \\ R_0 + R_i &\leq \frac{1}{2} \log\left(\frac{\boldsymbol{I}_{N_i} + \overline{\boldsymbol{H}}_i \boldsymbol{S} \overline{\boldsymbol{H}}_i^T}{\boldsymbol{I}_{N_i} + \overline{\boldsymbol{H}}_i \boldsymbol{Q}^{(c)} \overline{\boldsymbol{H}}_i^T}\right), \quad i = 1, 2 \end{aligned}$$

for some $\mathbf{0} \preceq \boldsymbol{Q}^{(c)} \preceq \boldsymbol{S}$.

Following [LS09, LLL10] we define

$$\boldsymbol{D}_i = \boldsymbol{U}_i \boldsymbol{\Lambda}_i (\boldsymbol{\Lambda}_i + \alpha \boldsymbol{I}_{N_R})^{-1} \boldsymbol{U}_i^T,$$

so that $\boldsymbol{H}_i = \boldsymbol{D}_i \overline{\boldsymbol{H}}_i$, $i = 1, 2$. Since $\boldsymbol{D}_i \boldsymbol{D}_i^T \prec \boldsymbol{I}_{N_R}$, we know from [WLS+09, Definition 1] that $\boldsymbol{D}_i$ defines a degradedness order. More precisely, we have the following Markov chains

$$\mathbf{X} - \overline{\mathbf{Y}}_i - \mathbf{Y}_i, \quad i = 1, 2. \tag{5.71}$$

From (5.71) we see that both receivers – the legitimate node 1 and the non-legitimate node 2 – receive a stronger signal in the new BBC (5.70) than in the original BBC (5.47). Since the channel to the non-legitimate node 2 is enhanced as well, it is by no means self-evident that this leads to an increased secrecy capacity region. This differs from the classical MIMO Gaussian broadcast channel [WSS06b]. However, the secrecy capacity region can be bounded as follows.

Lemma 5.44. *For the secrecy capacity regions we have the following relation*

$$\mathcal{R}^S_{m_c,m_0}(\boldsymbol{H}_1, \boldsymbol{H}_2 | \boldsymbol{S}) \subseteq \mathcal{R}^S_{m_c,m_0}(\overline{\boldsymbol{H}}_1, \overline{\boldsymbol{H}}_2 | \boldsymbol{S}) + \Delta(\boldsymbol{H}_2, \overline{\boldsymbol{H}}_2 | \boldsymbol{S})$$

with $\Delta(\boldsymbol{H}_2, \overline{\boldsymbol{H}}_2 | \boldsymbol{S})$ *the set of all rate tuples* $(R_c, 0, 0, 0) \in \mathbb{R}_+^4$ *that satisfy*

$$R_c \leq \frac{1}{2} \log \det \left(\boldsymbol{I}_{N_R} + \overline{\boldsymbol{H}}_2 \boldsymbol{S} \overline{\boldsymbol{H}}_2^T \right) - \frac{1}{2} \log \det \left(\boldsymbol{I}_{N_R} + \boldsymbol{H}_2 \boldsymbol{S} \boldsymbol{H}_2^T \right).$$

Proof. The proof is almost identical to the corresponding result for classical MIMO Gaussian broadcast channel with common and confidential message, cf. [LLL10, Section IV]. For completeness the details can be found in Appendix A.11. □

The following observation concludes the proof of the secrecy capacity region of the general MIMO Gaussian BBC with common and confidential messages. For $\alpha \searrow 0$ we have $\overline{\boldsymbol{H}}_i \to \boldsymbol{H}_i$, $i = 1, 2$, cf. (5.69), so that

$$\mathcal{R}^S_{m_c,m_0}(\overline{\boldsymbol{H}}_1, \overline{\boldsymbol{H}}_2 | \boldsymbol{S}) \to \mathcal{R}^S_{m_c,m_0}(\boldsymbol{H}_1, \boldsymbol{H}_2 | \boldsymbol{S})$$

and

$$\Delta(\boldsymbol{H}_2, \overline{\boldsymbol{H}}_2 | \boldsymbol{S}) \to \{(0, 0, 0, 0)\}.$$

This together with Lemma 5.44 finishes the proof of Theorem 5.37. □

5.5.3 Numerical Example and Discussion

In the previous sections we established the secrecy capacity region of the BBC with common and confidential messages. This unifies previous results such as the BBC with confidential messages, cf. Section 5.3, the BBC with common messages, cf. Section 5.2, or the classical broadcast channel with common and confidential messages [CK78, LLL10]. For the case of no common messages we get the following.

Corollary 5.45. *The secrecy capacity region of the MIMO Gaussian BBC with confidential messages under the average power constraint P is the set of all rate triples $(R_c, R_1, R_2) \in \mathbb{R}_+^3$ that satisfy*

$$R_c \leq \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_1} + \boldsymbol{H}_1\boldsymbol{Q}^{(c)}\boldsymbol{H}_1^T\right) - \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_2} + \boldsymbol{H}_2\boldsymbol{Q}^{(c)}\boldsymbol{H}_2^T\right)$$
$$R_i \leq \frac{1}{2}\log\det\left(\frac{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i(\boldsymbol{Q}^{(c)} + \boldsymbol{Q}^{(p)})\boldsymbol{H}_i^T}{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i\boldsymbol{Q}^{(c)}\boldsymbol{H}_i^T}\right), \quad i = 1, 2$$

for some $\boldsymbol{Q}^{(c)} \succeq \boldsymbol{0}$, $\boldsymbol{Q}^{(p)} \succeq \boldsymbol{0}$ with tr$(\boldsymbol{Q}^{(c)} + \boldsymbol{Q}^{(p)}) \leq P$.

If there are no confidential services for the relay to integrate, it solely transmits public messages and the scenario reduces to the BBC with common messages.

Corollary 5.46. *The capacity region of the MIMO Gaussian BBC with common messages under the average power constraint P is the set of all rate triples $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ that satisfy*

$$R_0 + R_i \leq \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_i} + \boldsymbol{H}_i\boldsymbol{Q}^{(p)}\boldsymbol{H}_i^T\right), \quad i = 1, 2$$

for some $\boldsymbol{Q}^{(p)} \succeq \boldsymbol{0}$ with tr$(\boldsymbol{Q}^{(p)}) \leq P$.

For the case of no bidirectional messages we end up with the classical broadcast channel with common and confidential messages.

Corollary 5.47 ([LLL10])**.** *The secrecy capacity region of the MIMO Gaussian broadcast channel with common and confidential messages under the average power constraint P is the set of all rate pairs $(R_c, R_0) \in \mathbb{R}_+^2$ that satisfy*

$$R_c \leq \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_1} + \boldsymbol{H}_1\boldsymbol{Q}^{(c)}\boldsymbol{H}_1^T\right) - \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_2} + \boldsymbol{H}_2\boldsymbol{Q}^{(c)}\boldsymbol{H}_2^T\right)$$
$$R_0 \leq \min_{i\in\{1,2\}}\left\{\frac{1}{2}\log\det\left(\frac{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i(\boldsymbol{Q}^{(c)} + \boldsymbol{Q}^{(p)})\boldsymbol{H}_i^T}{\boldsymbol{I}_{N_i} + \boldsymbol{H}_i\boldsymbol{Q}^{(c)}\boldsymbol{H}_i^T}\right)\right\}$$

for some $\boldsymbol{Q}^{(c)} \succeq \boldsymbol{0}$, $\boldsymbol{Q}^{(p)} \succeq \boldsymbol{0}$ with tr$(\boldsymbol{Q}^{(c)} + \boldsymbol{Q}^{(p)}) \leq P$.

Remark 5.48. *Clearly, the whole discussion and Corollaries 5.45-5.47 also hold for the general matrix power constraint (5.49).*

The optimal transmit covariance matrices are determined by non-convex optimization problems and so the weighted rate sum optimal rate tuples as well. Hence, obtaining the boundary of the secrecy capacity region is in general non-trivial.

For the MISO scenario we can reformulate the optimization problem in such a way that it becomes convex and therewith tractable. Since the relay has multiple transmit antennas but nodes 1 and 2 have only single receive antennas, the channel matrices $\boldsymbol{H}_i$ become vectors $\boldsymbol{h}_i$, $i = 1, 2$, and the region of Theorem 5.37 can be written as

$$R_c \leq \frac{1}{2} \log \left(1 + \frac{\boldsymbol{h}_1 \boldsymbol{Q}^{(c)} \boldsymbol{h}_1^T - \boldsymbol{h}_2 \boldsymbol{Q}^{(c)} \boldsymbol{h}_2^T}{1 + \boldsymbol{h}_2 \boldsymbol{Q}^{(c)} \boldsymbol{h}_2^T} \right) \tag{5.72a}$$

$$R_0 + R_i \leq \frac{1}{2} \log \left(1 + \frac{\boldsymbol{h}_i (\boldsymbol{S} - \boldsymbol{Q}^{(c)}) \boldsymbol{h}_i^T}{1 + \boldsymbol{h}_i \boldsymbol{Q}^{(c)} \boldsymbol{h}_i^T} \right), \quad i = 1, 2. \tag{5.72b}$$

Next, we follow [WSS06a] or [LLL10, Section V] and consider a re-parametrization of the rates as

$$R_c = \log(1 + \alpha \gamma_c) \tag{5.73a}$$

$$R_0 + R_i = \log(1 + \alpha \gamma_i), \quad i = 1, 2 \tag{5.73b}$$

where α is an auxiliary parameter and $\gamma_c, \gamma_1, \gamma_2$ can be interpreted as received SNR "weights". Combining (5.72) and (5.73) we end up with

$$\boldsymbol{h}_1 \boldsymbol{Q}^{(c)} \boldsymbol{h}_1^T - \boldsymbol{h}_2 \boldsymbol{Q}^{(c)} \boldsymbol{h}_2^T \geq \alpha \gamma_c (1 + \boldsymbol{h}_2 \boldsymbol{Q}^{(c)} \boldsymbol{h}_2^T) \tag{5.74a}$$

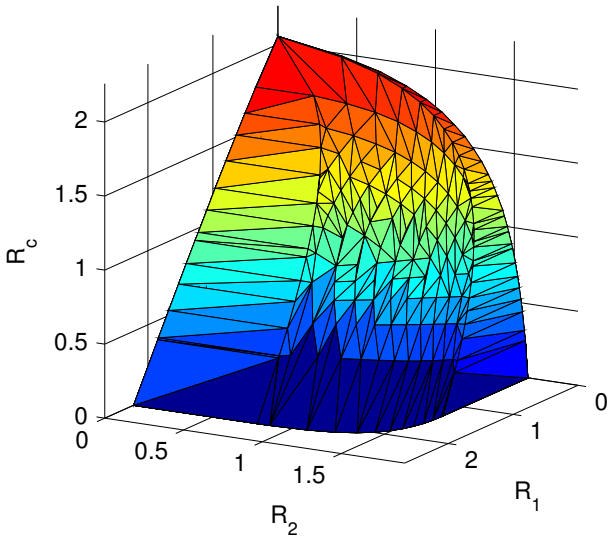
$$\boldsymbol{h}_1 (\boldsymbol{S} - \boldsymbol{Q}^{(c)}) \boldsymbol{h}_1^T \geq \alpha \gamma_1 (1 + \boldsymbol{h}_1 \boldsymbol{Q}^{(c)} \boldsymbol{h}_1^T) \tag{5.74b}$$

$$\boldsymbol{h}_2 (\boldsymbol{S} - \boldsymbol{Q}^{(c)}) \boldsymbol{h}_2^T \geq \alpha \gamma_2 (1 + \boldsymbol{h}_2 \boldsymbol{Q}^{(c)} \boldsymbol{h}_2^T) \tag{5.74c}$$

$$\boldsymbol{S} \succeq \boldsymbol{Q}^{(c)} \succeq \boldsymbol{0}. \tag{5.74d}$$

Instead of using (5.50) to check if a rate tuple is in the capacity region, i.e., $\boldsymbol{R} \in \mathcal{R}^S_{m_c, m_0}(\boldsymbol{h}_1, \boldsymbol{h}_2 | \boldsymbol{S})$, we can alternatively look for a positive semidefinite matrix $\boldsymbol{Q}^{(c)}$ that satisfies the conditions (5.74a)-(5.74d). Since all these conditions are linear in $\boldsymbol{Q}^{(c)}$, this belongs to the class of convex optimization problems which can be solved efficiently.

Obviously, all rates increase as the auxiliary parameter α increases. Thus we obtain the weighted rate sum optimal rate triple on the boundary of the secrecy capacity region $\mathcal{R}^S_{m_c, m_0}(\boldsymbol{h}_1, \boldsymbol{h}_2 | \boldsymbol{S})$ for fixed weights $\gamma_c, \gamma_1, \gamma_2$ by finding the maximum α such that (5.74a)-(5.74d) provide at least one feasible solution, cf. also [WSS06a, LLL10]. Finally, running



$$\boldsymbol{h}_1 = \begin{bmatrix} 2 \\ 0.4 \end{bmatrix}$$

$$\boldsymbol{h}_2 = \begin{bmatrix} 0.4 \\ 1.2 \end{bmatrix}$$

$$P = 5$$

Figure 5.12: Secrecy capacity region of the MISO Gaussian BBC with confidential messages with $N_R = 2$ and $N_1 = N_2 = 1$.

through all weight vectors with $\gamma_c + \gamma_1 + \gamma_2 = 1$ yields all weighted rate sum optimal rate tuples and characterizes the boundary of $\mathcal{R}^S_{m_c,m_0}(\boldsymbol{h}_1, \boldsymbol{h}_2|\boldsymbol{S})$.

Similarly, in the case of an average power constraint P we obtain the conditions

$$\begin{aligned}
&\boldsymbol{h}_1\boldsymbol{Q}^{(c)}\boldsymbol{h}_1^T - \boldsymbol{h}_2\boldsymbol{Q}^{(c)}\boldsymbol{h}_2^T \geq \alpha\gamma_c(1 + \boldsymbol{h}_2\boldsymbol{Q}^{(c)}\boldsymbol{h}_2^T) \\
&\boldsymbol{h}_1\boldsymbol{Q}^{(p)}\boldsymbol{h}_1^T \geq \alpha\gamma_1(1 + \boldsymbol{h}_1\boldsymbol{Q}^{(c)}\boldsymbol{h}_1^T) \\
&\boldsymbol{h}_2\boldsymbol{Q}^{(p)}\boldsymbol{h}_2^T \geq \alpha\gamma_2(1 + \boldsymbol{h}_2\boldsymbol{Q}^{(c)}\boldsymbol{h}_2^T) \\
&\mathrm{tr}(\boldsymbol{Q}^{(c)} + \boldsymbol{Q}^{(p)}) \leq P \\
&\boldsymbol{Q}^{(c)} \succeq \boldsymbol{0},\ \boldsymbol{Q}^{(p)} \succeq \boldsymbol{0}
\end{aligned}$$

which again allows to compute the boundary of the secrecy capacity region $\mathcal{R}^S_{m_c,m_0}(\boldsymbol{h}_1, \boldsymbol{h}_2|P)$.

For visual feasibility we consider the case with no common messages and depict in Figure 5.12 the secrecy capacity region of the MISO Gaussian BBC with confidential messages. For plots of the BBC with common messages and of the classical broadcast channel with common and confidential messages we refer to Section 5.2 and [LLL10], respectively.

5.6 Discussion

In this chapter we studied the efficient integration of public and confidential services in bidirectional relay networks at the physical layer. This necessitated the analysis of the BBC with common and confidential messages which completely characterizes the integration of bidirectional, common, and confidential services at the physical layer. This solves not only the optimal processing for bidirectional relay networks, but also gives us first insights for larger and more complex networks so that our results are not only relevant for itself. It gives valuable insights how services should be merged from an information-theoretic point of view. This is beneficial since it enables a joint resource allocation policy and it is expected that this will result in a significantly reduced complexity and an improved energy efficiency.

The optimal integration of common services in bidirectional relay networks shows that there are strong connections between the BBC with and without common messages. The strong connection begins with the optimal coding strategy for the BBC with common messages, since basically, it becomes coding without common messages. The common message is treated as a part of both individual messages and as a result, the coding idea of the case without common messages is applicable. All messages are combined into a single data stream based on the network coding idea which allows the receiving nodes to decode the intended individual and common messages using their own message as side information.

In retrospect it is not surprising that the connection carries over to the transmit covariance optimization problem. In contrast to suboptimal strategies such as superposition coding approaches [OB08b], where the messages are associated with several transmit covariance matrices, in the optimal strategy there is only one transmit covariance matrix that has to be optimized. This is similar to the BBC without common messages [OWB09a, OJWB09], and as expected, the optimal transmit strategies transfer as well.

Then we considered the optimal integration of confidential services in bidirectional relay networks. In this scenario the relay has some public information for both receivers as well as confidential information for one receiver that should be kept secret from the other one. The task is to enable additional secure communication *within* such a network. We want to stress the fact that this differs from the wiretap scenario where the (bidirectional) communication itself should be secure from possible eavesdroppers *outside* the network. Some work on the corresponding bidirectional broadcast wiretap channel can be found for example in [ASS10, MS10] or [WSB11].

This scenario addresses the problem of realizing additional confidential communication within a network that exploits principles from network coding for the public communication; hence, the optimal processing is by no means self-evident. Interestingly, it is shown that superimposing two signals – one for the public services and one for the confidential service – is optimal for MIMO Gaussian networks.

6 Conclusion

The use of relays is currently becoming more and more attractive since they have the potential to improve the performance and coverage of wireless networks significantly by exploiting the broadcast nature of the wireless medium. Relay communication suffers from the fact that it needs orthogonal resources for transmission and reception. This half-duplex constraint usually results in a loss in spectral efficiency. In this thesis we study the concept of bidirectional relaying which has the ability to reduce the inherent loss in spectral efficiency by exploiting the bidirectional property of the communication.

Bidirectional relaying applies to three-node networks where a half-duplex relay node establishes a bidirectional communication between two other nodes using a decode-and-forward protocol. Due to the separation of the communication into two phases, the concept of bidirectional relaying is particularly suitable to be integrated in conventional wireless networks such as ad-hoc, sensor, and even cellular networks. For example, in [ODS10] it is discussed how bidirectional relaying can be efficiently embedded in a cellular downlink, while [GGY11] addresses the problem of interference mitigation in femto-macro coexistence with bidirectional relaying. Accordingly, the study of bidirectional relaying is of great interest.

Previous studies considered the case of perfect channel state information at all nodes. But in practical wireless communication systems channel uncertainty is a ubiquitous phenomenon due to changing channel conditions or imperfect channel estimation. The question must be asked if it is worth to improve the available CSI at the nodes or if the nodes should be left with the uncertainty. In Chapter 3 we answer this question by analyzing bidirectional relaying for compound channels. This models the scenario where the exact channel realization is not known. Rather, it is only known to the nodes that this realization remains constant during the whole time of transmission and that it belongs to a pre-specified set of channels. Interestingly, it is shown that an improvement in CSI at the receivers does not lead to an increased capacity region. On the other hand, improved CSI at the transmitter can increase the corresponding capacity region.

Reliable communication in bidirectional relay networks is still possible for compound channels, but, of course, at reduced rates compared to the case of perfect CSI. In Chapter 4 it is shown that for arbitrarily varying channels the impact is much more dramatic. If the channel varies from symbol to symbol in an unknown and arbitrary manner, it might happen that no communication is possible if conventional deterministic coding strategies are applied.

This necessitates the use of more sophisticated strategies based on a common randomness. This additional resource allows the transmitter and the receivers to coordinate their choice of encoder and decoders and to establish a reliable communication even in scenarios where deterministic strategies fail. This discussion becomes especially relevant if bidirectional relaying is applied in uncoordinated wireless networks where the communication is disturbed by unknown varying interference from outside the bidirectional relay network.

Another important issue is the efficient integration of multiple services at the physical layer. Already in current cellular systems operators offer not only (bidirectional) voice communication, but also further multicast or confidential services which are subject to certain secrecy constraints. Since bidirectional relaying is a promising candidate to increase the spectral efficiency of next generation cellular systems, it is important to study the efficient integration of additional services in bidirectional relay networks as done in Chapter 5.

Interestingly, it is shown that optimal coding and transmit strategies for bidirectional relaying with and without additional multicast services are strongly connected. Accordingly, existing policies and algorithms for bidirectional relaying without multicast must only slightly be adapted and extended by including an additional discussion for the weight of the common message to apply also to networks with additional multicast services.

After that the efficient integration of additional confidential services in bidirectional relay networks is discussed. Here, the relay integrates an additional confidential message for one node which has to be kept secret from the other, non-legitimate node. Interestingly, the optimal processing for enabling confidential services within a MIMO Gaussian bidirectional relay network is to superimpose two signals – one for the public services and one for the confidential service.

The results of physical layer service integration in bidirectional relay networks are not only relevant for itself, since they solve the optimal processing for such networks, but also, since they give us valuable insights for larger and more complex networks.

More detailed and explicit discussions of the results are given in Section 3.6 for bidirectional relaying for compound channels, in Section 4.8 for bidirectional relaying in uncoordinated wireless networks, and in Section 5.6 for physical layer service integration in bidirectional relay networks.

Future Work and Open Problems

In this thesis we addressed two different important research directions. Firstly, we analyzed the impact of channel uncertainty on bidirectional relaying and secondly, we studied the efficient integration of different services at the physical layer. Consequently, the next natural

step is to bring both research directions together and to study robust physical layer service integration in bidirectional relay networks.

With the results obtained in this thesis the integration of common messages in bidirectional relay networks for compound and arbitrarily varying channels should be straightforward. For perfect channel state information we have seen that the optimal processing of bidirectional relaying with and without additional common messages are strongly connected. We expect that this connection carries over to the case of channel uncertainty. Nevertheless the integration of common messages for compound and arbitrarily varying channels should be explicitly characterized.

We expect the integration of confidential messages under channel uncertainty to be much more involved. Even for the compound wiretap channel, which can be regarded as the simplest scenario for physical layer security under channel uncertainty, the capacity is not known in general. While the scenario with CSI at the transmitter is solved [BBS11a], the case with channel uncertainty remains open. To the best of our knowledge only bounds on the capacity are known [LKPS09] or a multi-letter expression is established [BBS11a].

So far we considered the integration of one confidential message for one node which has to be kept secret from the other one. Similarly as in [LLPS10a, LLPS10b, EU10a] this scenario can be extended by further integrating a second confidential message for the other node. Then, each node receives a bidirectional and a confidential message where each confidential message has to be kept secret from its non-legitimate node.

In this thesis we considered secrecy within the bidirectional relay network. But there might also be eavesdroppers outside the bidirectional relay network, which has to be kept ignorant of the communication. Therefore, another research direction would be to protect the bidirectional communication itself from eavesdroppers outside the network. As a first step it is reasonable to analyze both phases separately and protect them against possible eavesdroppers. This necessitates the study of the multiple access wiretap channel [EU08a, TY08] and the bidirectional broadcast wiretap channel [ASS10, MS10, WSB11] for which first results are available. Even if an eavesdropper is not be able to intercept the communication in one phase, he might be able to get enough information in each phase so that the composition of them suffices to conclude on the transmitted messages. Therefore, the main goal must be to protect both phases together.

The ignorance of the non-legitimate node about the confidential message was measured using the criterion of weak secrecy, i.e., we require $\frac{1}{n} I(\mathrm{M}_c; \mathrm{Y}_2^n | \mathrm{M}_2)$ to be small, cf. (5.2). There exists a stronger notion of secrecy where the division by n is dropped, i.e., $I(\mathrm{M}_c; \mathrm{Y}_2^n | \mathrm{M}_2)$ has to be small. This notion is not only stronger by dropping the division by n, but also has the advantage that it has the following operational meaning. The output at the non-legitimate node will be almost independent of the confidential message. Moreover, if this requirement is satisfied the decoding error at the non-legitimate node will tend to 1. For further details we

refer for example to [MW00, BBRM08, BB08, BBS11b]. Thus, it is reasonable to extend the results for the integration of confidential messages to this stronger notion of secrecy.

Throughout the thesis we considered a restricted decode-and-forward bidirectional relay network. A possible extension would be to drop some of these assumptions. For example, we assume the relay to decode both messages in the first phase, but in the end the relay is only interested in establishing a bidirectional communication between the two other nodes. Therefore, the question must be asked if it is necessary or even suboptimal for the relay to decode both messages. There are other schemes which weaken the strict assumption of decoding at the relay node. For example there are compress-and-forward strategies [SOS07, GTN08] or compute-and-forward schemes [WNPS10, NCL10, BC07, NG11, OKJ10] based on structured codes where the relay decodes a function of both messages. It would be interesting if we can establish similar results for such schemes. Another extension would be to drop the assumption of a restricted bidirectional relay network and, accordingly, allow both nodes to cooperate and to use feedback.

A Additional Proofs

A.1 Proof of Theorem 4.26

We follow [Hug97, Theorem 1] where a similar result for the single-user AVC is proved. We start with the trivial case where $R_{i,\max} = 0$, $i = 1, 2$, which implies that $\inf_{q \in \mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{i,q}) = 0$ for every $P_{\mathrm{X}} \in \mathcal{P}_0(n, \mathcal{X})$. If $P_{\mathrm{X}}(x) > 0$ for all $x \in \mathcal{X}$, this implies the existence of a distribution $P_{\mathrm{XSY}_i} = P_{\mathrm{X}} \otimes P_{\mathrm{S}} \otimes W_i$ such that the input X and the output Y_i are independent, which means

$$\sum_s W_i(y_i|x, s) P_{\mathrm{S}}(s) = P_{\mathrm{Y}_i}(y_i).$$

Now, for any $t_i \geq 1$ we set

$$U_i(s|x_1, ..., x_{t_i}) \coloneqq P_{\mathrm{S}}(s)$$

for all $s, x_1, ..., x_{t_i}$ to obtain a channel which is symmetric in $x, x_1, ..., x_{t_i}$, cf. also (4.3). This implies immediately that the AVBBC $\mathfrak{W}^n$ is $(\mathcal{Y}_i, t_i)$-symmetrizable for all $t_i \geq 1$, $i = 1, 2$.

Next, we assume that $R_{i,\max} > 0$ and that the AVBBC $\mathfrak{W}^n$ is $(\mathcal{Y}_i, t_i)$-symmetrizable and further $\inf_{q \in \mathcal{P}(\mathcal{S})} I(P_{\mathrm{X}}, \overline{W}_{i,q}) > 0$, $i = 1, 2$. Consequently, there is a channel $U_i : \mathcal{X}^{t_i} \to \mathcal{P}(\mathcal{S})$ such that

$$\widetilde{W}_i(y_i|x_1, ..., x_{t_i+1}) \coloneqq \sum_s W_i(y_i|x_1, s) U_i(s|x_2, ..., x_{t_i+1}) \tag{A.1}$$

is symmetric in $x_1, ..., x_{t_i+1}$. Let $\mathrm{X}^{t_i+1} = (\mathrm{X}_1, ..., \mathrm{X}_{t_i+1})$ be a sequence of independent random variables each with distribution P_{X}. Further, denote the output of the auxiliary channel U_i by S' corresponding to the input $\mathrm{X}_2, ..., \mathrm{X}_{t_i+1}$, and the output of the channel W_i by Y_i' for the inputs X_1 and S'. As in [Hug97] for the single-user AVC we observe that $\mathrm{X}^{t_i+1} - (\mathrm{X}_1, \mathrm{S}') - \mathrm{Y}_i'$ forms a Markov chain, so that the Data Processing Inequality [CK81, p. 55] gives

$$\begin{aligned} I(\mathrm{X}_1, \mathrm{S}'; \mathrm{Y}_i') &\geq I(\mathrm{X}^{t_i+1}; \mathrm{Y}_i') \\ &\geq \sum_{k=1}^{t_i+1} I(\mathrm{X}_k; \mathrm{Y}_i') \\ &= (t_i + 1) I(\mathrm{X}_1; \mathrm{Y}_i') \end{aligned}$$

where the second inequality follows from the independence of $\mathrm{X}_1, ..., \mathrm{X}_{t_i+1}$ and the non-negativity of the (conditional) mutual information and the last equality from $P_{\mathrm{X}_k\mathrm{Y}'_i} = P_{\mathrm{X}_1\mathrm{Y}'_i}$, which is a consequence of the symmetry of $\widetilde{W}_i$, cf. (A.1). If we subtract $I(\mathrm{X}_1;\mathrm{Y}'_i)$ from both sides, having $I(\mathrm{X}_1;\mathrm{Y}'_i) \geq \inf_{q\in\mathcal{P}(\mathcal{S})} I(P_\mathrm{X}, \overline{W}_{i,q}) > 0$ in mind, we get

$$\begin{aligned} t_i &\leq \frac{I(\mathrm{S}';\mathrm{Y}'_i|\mathrm{X}_1)}{I(\mathrm{X}_1;\mathrm{Y}'_i)} \\ &\leq \max_{\substack{P_{\mathrm{X}\mathrm{S}\mathrm{Y}_i}:P_{\mathrm{X}\mathrm{S}\mathrm{Y}_i}=P_\mathrm{X}\otimes P_\mathrm{S}\otimes W_i \\ \text{for some S}}} \frac{I(\mathrm{S};\mathrm{Y}_i|\mathrm{X})}{I(\mathrm{X};\mathrm{Y}_i)}. \end{aligned}$$

Clearly, this holds for all P_X so that we finally obtain

$$\begin{aligned} t_i &\leq \min_{P_\mathrm{X}} \max_{\substack{P_{\mathrm{X}\mathrm{S}\mathrm{Y}_i}:P_{\mathrm{X}\mathrm{S}\mathrm{Y}_i}=P_\mathrm{X}\otimes P_\mathrm{S}\otimes W_i \\ \text{for some S}}} \frac{I(\mathrm{S};\mathrm{Y}_i|\mathrm{X})}{I(\mathrm{X};\mathrm{Y}_i)} \\ &\leq \frac{\log(\min\{|\mathcal{S}||\mathcal{Y}_i|\})}{R_{i,\max}} \end{aligned}$$

which proves (4.19). □

A.2 Proof of Lemma 4.27

The lemma follows immediately from [Hug97, Lemma 4], where a similar result for the single-user AVC is proved. But for completeness we present the proof in the following. We carry out the analysis for receiving node 1. The analysis for node 2 follows accordingly.

First, we observe that it suffices to consider $K_1 \geq 1$, since otherwise there is nothing to prove. We consider any list code $\mathcal{C}_{\text{list}}(\mathfrak{W}^n)$ with codewords $x^n_{m_1,m_2} = (x_{m_1,m_2,1}, ..., x_{m_1,m_2,n}) \in \mathcal{X}^n$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, and a list decoder at node 1 with list size $L_1 \leq T_1$. Since $K_1 \leq T_1$, the AVBBC $\mathfrak{W}^n$ is $(\mathcal{Y}_1, K_1)$-symmetrizable so that there exists a channel $U_1 : \mathcal{X}^{K_1} \to \mathcal{P}(\mathcal{S})$ such that

$$\sum_{s\in\mathcal{S}} W_1(y_1|x_1, s)U_1(s|x_2, ..., x_{K_1+1})$$

is symmetric in $x_1, ..., x_{K_1+1}$, cf. also (4.3).

Then for each $m_1 \in \mathcal{M}_1$ the following holds. For any set $\mathcal{J} = \{j_1, ..., j_{K_1}\} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)$ of K_1 messages, let $\mathrm{S}^n_{m_1,\mathcal{J}} = (\mathrm{S}_{m_1,\mathcal{J},1}, ..., \mathrm{S}_{m_1,\mathcal{J},n}) \in \mathcal{S}^n$ be a random state sequence with

$$\mathbb{P}\{\mathrm{S}_{m_1,\mathcal{J},k} = s\} = U_1(s|x_{m_1,j_1,k}, ..., x_{m_1 j,K_1,k}).$$

For any $i \in \mathcal{M}_2$ and any $\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)$ as defined above it follows that

$$\begin{aligned}
\mathbb{E}[W_1^n(y_1^n|x_{m_1,i}^n, \mathrm{S}_{m_1,\mathcal{J}}^n)] &= \prod_{k=1}^{n} \mathbb{E}[W_1(y_{1,k}|x_{m_1,i,k}, \mathrm{S}_{m_1,\mathcal{J},k})] \\
&= \prod_{k=1}^{n} \sum_{s\in\mathcal{S}} W_1(y_{1,k}|x_{m_1,i,k}, s)\mathbb{P}\{\mathrm{S}_{m_1,\mathcal{J},k} = s\} \\
&= \prod_{k=1}^{n} \sum_{s\in\mathcal{S}} W_1(y_{1,k}|x_{m_1,i,k}, s)U_1(s|x_{m_1,j_1,k}, ..., x_{m_1,j_{K_1},k}) \\
&= \sum_{s^n\in\mathcal{S}^n} W_1^n(y_1^n|x_{m_1,i}^n, s^n)U_1^n(s^n|x_{m_1,j_1}^n, ..., x_{m_1,j_{K_1}}^n)
\end{aligned}$$

is symmetric in $i, j_1, ..., j_{K_1}$. Consequently, for any $\mathcal{J}' \in \mathfrak{P}_{K_1+1}(\mathcal{M}_2)$ and any fixed $i_0 \in \mathcal{J}'$, we have

$$\mathbb{E}[W_1^n(y_1^n|x_{m_1,i}^n, \mathrm{S}_{m_1,\mathcal{J}'\setminus\{i\}}^n)] = \mathbb{E}[W_1^n(y_1^n|x_{m_1,i_0}^n, \mathrm{S}_{m_1,\mathcal{J}'\setminus\{i_0\}}^n)]$$

for all $i \in \mathcal{J}'$. Since the list size of the list decoder at node 1 is L_1, the received y_1^n can be decoded in at most L_1 ways so that it follows for the probability of error that

$$\begin{aligned}
&\sum_{i\in\mathcal{J}'} \mathbb{E}[e_1((m_1,i), \mathrm{S}_{m_1,\mathcal{J}'\setminus\{i\}}^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))] \\
&\qquad= \sum_{i\in\mathcal{J}'} \left(1 - \sum_{y_1^n: i\in\mathcal{L}^{(1)}(y_1^n,m_1)} \mathbb{E}[W_1^n(y_1^n|x_{m_1,i}^n, \mathrm{S}_{m_1,\mathcal{J}'\setminus\{i\}}^n)]\right) \\
&\qquad= K_1 + 1 - \sum_{i\in\mathcal{J}'} \sum_{y_1^n: i\in\mathcal{L}^{(1)}(y_1^n,m_1)} \mathbb{E}[W_1^n(y_1^n|x_{m_1,i_0}^n, \mathrm{S}_{m_1,\mathcal{J}'\setminus\{i_0\}}^n)] \\
&\qquad\geq K_1 + 1 - L_1 \sum_{y_1^n\in\mathcal{Y}_1^n} \mathbb{E}[W_1^n(y_1^n|x_{m_1,i_0}^n, \mathrm{S}_{m_1,\mathcal{J}'\setminus\{i_0\}}^n)] \\
&\qquad= K_1 + 1 - L_1.
\end{aligned}$$

For a fixed $m_1 \in \mathcal{M}_1$ this leads to

$$\begin{aligned}
&\frac{1}{|\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \sum_{\mathcal{J}\in\mathfrak{P}_{K_1}(\mathcal{M}_2)} \mathbb{E}[\bar{e}_1(\mathrm{S}_{m_1,\mathcal{J}}^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))] \\
&\qquad= \frac{1}{M_1^{(n)} M_2^{(n)} |\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \times \\
&\qquad\qquad \sum_{\mathcal{J}\in\mathfrak{P}_{K_1}(\mathcal{M}_2)} \sum_{m_1'=1}^{M_1^{(n)}} \sum_{m_2'=1}^{M_2^{(n)}} \mathbb{E}[e_1((m_1', m_2'), \mathrm{S}_{m_1,\mathcal{J}}^n|\mathcal{C}_{\text{list}}(\mathfrak{W}^n))]
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{1}{M_1^{(n)} M_2^{(n)} \binom{M_2^{(n)}}{K_1}} \times \\
&\qquad \sum_{m_1'=1}^{M_1^{(n)}} \sum_{\mathcal{J}' \in \mathfrak{P}_{K_1+1}(\mathcal{M}_2)} \sum_{m_2' \in \mathcal{J}'} \mathbb{E}[e_1((m_1', m_2'), \mathrm{S}^n_{m_1, \mathcal{J}' \setminus \{m_2'\}} | \mathcal{C}_{\text{list}}(\mathfrak{W}^n))] \\
&\geq \frac{1}{M_1^{(n)}} \sum_{m_1'=1}^{M_1^{(n)}} \frac{\binom{M_2^{(n)}}{K_1+1}(K_1+1-L_1)}{M_2^{(n)} \binom{M_2^{(n)}}{K_1}} \\
&= \left(1 - \frac{L_1}{K_1+1}\right) \left(\frac{M_2^{(n)} - K_1}{M_2^{(n)}}\right).
\end{aligned}
$$

Thus, we obtain for the average probability of error

$$
\begin{aligned}
&\frac{1}{M_1^{(n)} |\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \sum_{m_1=1}^{M_1^{(n)}} \sum_{\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)} \mathbb{E}[\bar{e}_1(\mathrm{S}^n_{m_1, \mathcal{J}} | \mathcal{C}_{\text{list}}(\mathfrak{W}^n))] \\
&\qquad = \frac{1}{M_1^{(n)}} \sum_{m_1=1}^{M_1^{(n)}} \left(\frac{1}{|\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \sum_{\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)} \mathbb{E}[\bar{e}_1(\mathrm{S}^n_{m_1, \mathcal{J}} | \mathcal{C}_{\text{list}}(\mathfrak{W}^n))] \right) \\
&\qquad \geq \frac{1}{M_1^{(n)}} \sum_{m_1=1}^{M_1^{(n)}} \left(1 - \frac{L_1}{K_1+1}\right) \left(\frac{M_2^{(n)} - K_1}{M_2^{(n)}}\right) \\
&\qquad = \left(1 - \frac{L_1}{K_1+1}\right) \left(\frac{M_2^{(n)} - K_1}{M_2^{(n)}}\right)
\end{aligned}
$$

which implies the existence of at least one $m_1 \in \mathcal{M}_2$ and $\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)$ with

$$
\mathbb{E}[\bar{e}_1(\mathrm{S}^n_{m_1, \mathcal{J}} | \mathcal{C}_{\text{list}}(\mathfrak{W}^n))] \geq \left(1 - \frac{L_1}{K_1+1}\right) \left(\frac{M_2^{(n)} - K_1}{M_2^{(n)}}\right).
$$

Consequently, there is a realization s^n of $\mathrm{S}^n_{m_1, \mathcal{J}}$ with $\bar{e}_1(s^n | \mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \geq \left(1 - \frac{L_1}{K_1+1}\right) \left(\frac{M_2^{(n)} - K_1}{M_2^{(n)}}\right)$ which finally implies

$$
\max_{s^n \in \mathcal{S}^n} \bar{e}_1(s^n | \mathcal{C}_{\text{list}}(\mathfrak{W}^n)) \geq \left(1 - \frac{L_1}{K_1+1}\right) \left(\frac{M_2^{(n)} - K_1}{M_2^{(n)}}\right)
$$

so that the first part of the lemma is proved. Clearly, the analysis for node 2 follows accordingly using the same argumentation. □

A.3 Proof of Lemma 4.28

In the following we show that $M_1^{(n)} M_2^{(n)}$ with[1] $M_1^{(n)} = \exp(nR_2)$ and $M_2^{(n)} = \exp(nR_1)$ randomly selected codewords will possess, with probability arbitrarily close to one, the properties (4.20a)-(4.20j) as stated in Lemma 4.28. We follow [Hug97] and extend the proof idea of [CN88b] for list size one to arbitrary list sizes. But first, we restate two lemmas which are essential to proof the desired properties of the codewords.

Lemma A.1. *Let $\mathrm{Z}_1^n, ..., \mathrm{Z}_N^n$ be arbitrary random variables and let $f_i(\mathrm{Z}_1^n, ..., \mathrm{Z}_i^n)$ be arbitrary with $0 \leq f_i \leq 1$, $i = 1, ..., N$. Then the condition*

$$\mathbb{E}[f_i(\mathrm{Z}_1^n, ..., \mathrm{Z}_i^n) | \mathrm{Z}_1^n, ..., \mathrm{Z}_{i-1}^n] \leq a \quad \textit{almost surely}, \qquad i = 1, ..., N, \tag{A.2}$$

implies

$$\mathbb{P}\Big\{ \frac{1}{N} \sum_{i=1}^{N} f_i(\mathrm{Z}_1^n, ..., \mathrm{Z}_i^n) > t \Big\} \leq \exp\big(-N(t - a \log e) \big). \tag{A.3}$$

Proof. The proof can be found in [Ahl80a] or [CN88b]. □

Further, we will need a covering lemma for $\mathfrak{P}_{L_1}(\mathcal{M}_2)$ and $\mathfrak{P}_{L_2}(\mathcal{M}_1)$. Therefore, let

$$\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2 \backslash \{i\}) \coloneqq \{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{i\}) : j < i \text{ for all } j \in \mathcal{J}\} \tag{A.4a}$$

$$\overline{\mathfrak{P}}_{L_2}(\mathcal{M}_1 \backslash \{i\}) \coloneqq \{\mathcal{J} \in \mathfrak{P}_{L_2}(\mathcal{M}_1 \backslash \{i\}) : j < i \text{ for all } j \in \mathcal{J}\}. \tag{A.4b}$$

Moreover, let $\Pi_{M_k^{(n)}}$ be the set of all permutations acting on $(1, ..., M_k^{(n)})$, $k = 1, 2$. For any permutation $\pi \in \Pi_{M_k^{(n)}}$ let

$$\pi\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2 \backslash \{i\}) \coloneqq \{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{i\}) : \pi(j) < \pi(i) \text{ for all } j \in \mathcal{J}\} \tag{A.5a}$$

$$\pi\overline{\mathfrak{P}}_{L_2}(\mathcal{M}_1 \backslash \{i\}) \coloneqq \{\mathcal{J} \in \mathfrak{P}_{L_2}(\mathcal{M}_1 \backslash \{i\}) : \pi(j) < \pi(i) \text{ for all } j \in \mathcal{J}\}. \tag{A.5b}$$

Then, the following lemma shows that $\mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{i\})$ and $\mathfrak{P}_{L_2}(\mathcal{M}_1 \backslash \{i\})$ can be covered by a small number of "ordered" sets $\pi\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2 \backslash \{i\})$ and $\pi\overline{\mathfrak{P}}_{L_2}(\mathcal{M}_1 \backslash \{i\})$ respectively.

Lemma A.2. *For all $n \geq \log(2L_1)$, there exist $p_1 \leq n(L_1 + 1)^2(R_1 + 1)$ permutations $\pi_1, ..., \pi_{p_1} \in \Pi_{M_2^{(n)}}$ such that for all $1 \leq i \leq M_2^{(n)}$*

$$\mathfrak{P}_{L_1}(\mathcal{M}_2 \backslash \{i\}) = \bigcup_{k=1}^{p_1} \pi_k \overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2 \backslash \{i\}).$$

[1] In this proof we have to deal with terms which decrease double exponentially fast. For notational convenience we will use the notation $\exp(\cdot)$ instead of $2^{(\cdot)}$. But recall that all exponentials are still to the basis 2.

Proof. The proof can be found in [Hug97, Lemma A2]. □

Clearly, a similar relation for receiving node 2 follows accordingly. Now, we turn to the proof of Lemma 4.28. Therefore, we fix an $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n$, $P_{\mathrm{X}} \in \mathcal{P}_0(n, \mathcal{X})$, and any joint types $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}}$ and $P_{\mathrm{X}\mathrm{X}^{L_2}\mathrm{S}}$. We assume that $P_{\mathrm{XS}} = P_{x^n,s^n}$ and $P_{\mathrm{X}_k} = P_{\mathrm{X}}$, $k = 1, ..., \max\{L_1, L_2\}$. Further, let $\mathrm{Z}^n_{m_1,m_2}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$ be independent random variables, each uniformly distributed on $\mathcal{T}_{\mathrm{X}}^{(n)}$.

In the following, we show that for each $m_1 \in \mathcal{M}_1$ the properties (4.20a)-(4.20e) are satisfied. Then, the second part of the properties, i.e., (4.20f)-(4.20j) for each $m_2 \in \mathcal{M}_2$, follows accordingly.

To show (4.20a)-(4.20e), we fix an arbitrary $m_1 \in \mathcal{M}_1$ for the following analysis. First, we estimate the size of the sets

$$\{i : (x^n, \mathrm{Z}^n_{m_1,i}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}\}, \qquad 1 \le k \le L_1.$$

We define

$$f_{m_1,i}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i}) = \begin{cases} 1, & \text{if } \mathrm{Z}^n_{m_1,i} \in \mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{XS}}(x^n, s^n) \\ 0, & \text{otherwise} \end{cases} \tag{A.6}$$

and apply Lemma A.1. The condition (A.2) of Lemma A.1 is now fulfilled with

$$\begin{aligned} a &= \mathbb{P}\{\mathrm{Z}^n_{m_1,i} \in \mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{XS}}(x^n, s^n)\} \\ &= \frac{|\mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{XS}}(x^n, s^n)|}{|\mathcal{T}^{(n)}_{\mathrm{X}}|} \\ &\le \frac{\exp\big(nH(\mathrm{X}_k|\mathrm{X},\mathrm{S})\big)}{(n+1)^{-|\mathcal{X}|}\exp\big(nH(\mathrm{X})\big)} \\ &= (n+1)^{|\mathcal{X}|}\exp\big(-nI(\mathrm{X}_k;\mathrm{X},\mathrm{S})\big) \end{aligned}$$

where the inequality follows from Lemma B.6 and the last equality from $H(\mathrm{X}_k) = H(\mathrm{X})$, $1 \le k \le L_1$. For $R_1 = \frac{1}{n}\log\frac{M_2^{(n)}}{L_1}$ and $L_1 \ge 1$ we set

$$t = \frac{1}{M_2^{(n)}}\exp\big(n(|R_1 - I(\mathrm{X}_k;\mathrm{X},\mathrm{S})|^+ + \epsilon)\big)$$

so that $M_2^{(n)}(t - a\log e) \ge \frac{1}{2}\exp(n\epsilon)$ if $n \ge n_1(\epsilon, L_1)$, with

$$n_1(\epsilon, L_1) \coloneqq \min\Big\{n : L_1(n+1)^{|\mathcal{X}|}\log(\mathrm{e}) \le \frac{1}{2}\exp(n\epsilon)\Big\}.$$

Then (A.3) results in

$$\mathbb{P}\Big\{\big|\{i:\mathrm{Z}^n_{m_1,i}\in\mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{XS}}(x^n,s^n)\}\big|>\exp\big(n(|R_1-I(\mathrm{X}_k;\mathrm{X},\mathrm{S})|^+\epsilon)\big)\Big\} < \exp\Big(-\frac{1}{2}\exp(n\epsilon)\Big). \tag{A.7}$$

If we replace $\mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{XS}}(x^n,s^n)$ by $\mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{S}}(s^n)$ in (A.6), the same reasoning leads similarly to

$$\mathbb{P}\Big\{\big|\{i:\mathrm{Z}^n_{m_1,i}\in\mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{S}}(s^n)\}\big|>\exp\big(n(|R_1-I(\mathrm{X}_k;\mathrm{S})|^+\epsilon)\big)\Big\} < \exp\Big(-\frac{1}{2}\exp(n\epsilon)\Big) \tag{A.8}$$

for $n\geq n_1(\epsilon,L_1)$. Further, if we replace $\mathcal{T}^{(n)}_{\mathrm{X}_k|\mathrm{S}}(s^n)$ by $\mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{S}}(s^n)$ and ϵ by $(\frac{\epsilon}{2}+\frac{\log(L_1)}{n})$ in (A.8) we get

$$\mathbb{P}\Big\{\big|\{i:\mathrm{Z}^n_{m_1,i}\in\mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{S}}(s^n)\}\big|>L_1\exp\big(n(|R_1-I(\mathrm{X};\mathrm{S})|^+\tfrac{\epsilon}{2})\big)\Big\} < \exp\Big(-\frac{L_1}{2}\exp\Big(n\frac{\epsilon}{2}\Big)\Big)$$

for all $n\geq n_1(\frac{\epsilon}{2},1)$. In particular, if $I(\mathrm{X};\mathrm{S})\geq\epsilon$ (and recall that $R_1\geq\epsilon$ since $R_1=\frac{1}{n}\log\frac{M_2^{(n)}}{L_1}$ and $M_2^{(n)}\geq L_1\exp(n\epsilon)$ as assumed) then

$$|R_1-I(\mathrm{X};\mathrm{S})|^+=R_1-\min\{R_1,I(\mathrm{X};\mathrm{S})\}\leq R_1-\epsilon$$

so that

$$\mathbb{P}\Big\{\frac{1}{M_2^{(n)}}\big|\{i:\mathrm{Z}^n_{m_1,i}\in\mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{S}}(s^n)\}\big|>\exp\Big(-n\frac{\epsilon}{2}\Big)\Big\}<\exp\Big(-\frac{L_1}{2}\exp\Big(n\frac{\epsilon}{2}\Big)\Big). \tag{A.9}$$

The doubly exponential bounds in (A.7) and (A.9) will suffice to establish the desired properties (4.20a) and (4.20b). Now we turn to property (4.20d). To prove this we need an elementary probability bound similarly as in [Hug97]. Therefore, let $\mathrm{V}_1,...,\mathrm{V}_{L_1}$ be nonnegative random variables and v a nonnegative constant. Then

$$\prod_{k=1}^{L_1}\mathrm{V}_k\geq v^{L_1}$$

is only satisfied if $\mathrm{V}_k\geq v$ for some $1\leq k\leq L_1$. Hence

$$\mathbb{P}\Big\{\Big(\prod_{k=1}^{L_1}\mathrm{V}_k\Big)^{\frac{1}{L_1}}\geq v\Big\}\leq\mathbb{P}\Big\{\bigcup_{k=1}^{L_1}\{\mathrm{V}_k\geq v\}\Big\}\leq\sum_{k=1}^{L_1}\mathbb{P}\{\mathrm{V}_k\geq v\}. \tag{A.10}$$

Then for $R_1 < \min_k I(\mathrm{X}_k;\mathrm{S})$ we get

$$\big|\{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2) : (x^n, \mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}}\}\big| \leq \prod_{k=1}^{L_1} \big|\{i : (x^n, \mathrm{Z}^n_{m_1,i}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}\}\big|.$$

If we apply (A.10) with $\mathrm{V}_k \coloneqq |\{i : (x^n, \mathrm{Z}^n_{m_1,i}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}\}|$ and $v \coloneqq \exp(n\frac{\epsilon}{L_1})$, we get

$$\begin{aligned}
&\mathbb{P}\big\{|\{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2) : (x^n, \mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}}\}| > \exp(n\epsilon)\big\} \\
&\qquad \leq \sum_{k=1}^{L_1} \mathbb{P}\Big\{|\{i : (x^n, \mathrm{Z}^n_{m_1,i}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}_k\mathrm{S}}\}| > \exp\Big(n\frac{\epsilon}{L_1}\Big)\Big\} \\
&\qquad < L_1 \exp\Big(-\frac{1}{2}\exp\Big(n\frac{\epsilon}{L_1}\Big)\Big) \qquad \text{(A.11)}
\end{aligned}$$

for all $n \geq n_1(\frac{\epsilon}{L_1}, L_1)$, where the last inequality follows from (A.7) by observing that

$$R_1 < \min_k I(\mathrm{X}_k;\mathrm{S}) \leq \min_k I(\mathrm{X}_k;\mathrm{X},\mathrm{S}).$$

The doubly exponential bound in (A.11) will suffice to establish the desired property (4.20d). To establish the remaining (4.20c) and (4.20e) we proceed as follows. Let $\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})$ be as introduced in (A.4a) and define

$$\mathcal{B}_{m_1,i} = \mathcal{B}_{m_1,i}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i-1}) \coloneqq \{\mathcal{J} \in \overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\}) : \mathrm{Z}^n_{m_1,\mathcal{J}} \in \mathcal{T}^{(n)}_{\mathrm{X}^{L_1}|\mathrm{S}}(s^n)\}.$$

Further, we set $\mathcal{A}_{m_1,i} \coloneqq \mathcal{B}_{m_1,i}$ if $|\mathcal{B}_{m_1,i}| \leq \exp(n\epsilon_1)$ and $\mathcal{A}_{m_1,i} \coloneqq \emptyset$ otherwise, with ϵ_1 specified later. Let

$$f_{m_1,i}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i}) \coloneqq \begin{cases} 1, & \text{if } \mathrm{Z}^n_{m_1,i} \in \bigcup_{\mathcal{J}\in\mathcal{A}_{m_1,i}} \mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{X}^{L_1}\mathrm{S}}(\mathrm{Z}^n_{m_1\mathcal{J}}, s^n) \\ 0, & \text{otherwise.} \end{cases} \qquad \text{(A.12)}$$

If

$$\big|\{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2) : \mathrm{Z}^n_{m_1,\mathcal{J}} \in \mathcal{T}^{(n)}_{\mathrm{X}^{L_1}|\mathrm{S}}(s^n)\}\big| \leq \exp(n\epsilon_1)$$

then $\mathcal{A}_{m_1,i} = \mathcal{B}_{m_1,i}$ for each i and therefore

$$\begin{aligned}
&\big|\{i : (\mathrm{Z}^n_{m_1,i}, \mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}} \text{ for some } \mathcal{J} \in \overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})\}\big| \\
&\qquad\qquad = \sum_{i=1}^{M_2^{(n)}} f_{m_1,i}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i}).
\end{aligned}$$

Now, from (A.11) it follows that

$$\begin{aligned}
&\mathbb{P}\Big\{\big|\{i : (\mathrm{Z}^n_{m_1,i}, \mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}} \\
&\qquad\qquad \text{for some } \mathcal{J} \in \overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})\}\big| \neq \sum_{i=1}^{M_2^{(n)}} f_{m_1,i}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i})\Big\} \\
&\leq \mathbb{P}\big\{\big|\{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2) : (\mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \in \mathcal{T}^{(n)}_{\mathrm{X}^{L_1}\mathrm{S}}\}\big| > \exp(n\epsilon_1)\big\} \\
&\leq L_1 \exp\Big(-\frac{1}{2}\exp\Big(n\frac{\epsilon_1}{L_1}\Big)\Big) \qquad \text{(A.13)}
\end{aligned}$$

for $n \geq n_1(\frac{\epsilon_1}{L_1}, L_1)$. Further, we have

$$\begin{aligned}
&\mathbb{E}[f_{m_1,i}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i})|\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i-1}] \\
&\quad = \mathbb{P}\Big\{\mathrm{Z}^n_{m_1,i} \in \bigcup_{\mathcal{J}\in\mathcal{A}_{m_1,i}} \mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{X}^{L_1}\mathrm{S}}(\mathrm{Z}^n_{m_1,\mathcal{J}}, s^n)\Big|\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i-1}\Big\} \\
&\quad \leq |\mathcal{A}_{m_1,i}| \frac{\exp\big(nH(\mathrm{X}|\mathrm{X}^{L_1}, \mathrm{S})\big)}{(n+1)^{-|\mathcal{X}|}\exp\big(nH(\mathrm{X})\big)} \\
&\quad \leq (n+1)^{|\mathcal{X}|}\exp\big(-n(I(\mathrm{X};\mathrm{X}^{L_1}, \mathrm{S}) - \epsilon_1)\big). \qquad \text{(A.14)}
\end{aligned}$$

If we assume that $\epsilon_1 = \frac{\epsilon}{4}$ and $I(\mathrm{X};\mathrm{X}^{L_1}, \mathrm{S}) \geq \epsilon$ then $f_{m_1,i}$ in (A.12) satisfies (A.2) with

$$a \coloneqq (n+1)^{|\mathcal{X}|}\exp\Big(-n\frac{3\epsilon}{4}\Big). \qquad \text{(A.15)}$$

Then, (A.3) with $t \coloneqq \exp(-n\frac{2\epsilon}{3})$ becomes

$$\begin{aligned}
\mathbb{P}\Big\{\frac{1}{M_2^{(n)}}\sum_{i=1}^{M_2^{(n)}} f_{m_1,i}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,i}) > \exp\Big(-n\frac{2\epsilon}{3}\Big)\Big\} &\leq \exp\Big(-\frac{M_2^{(n)}}{2}\exp\Big(-n\frac{2\epsilon}{3}\Big)\Big) \\
&\leq \exp\Big(-\frac{L_1}{2}\exp\Big(n\frac{\epsilon}{3}\Big)\Big)
\end{aligned}$$

for $n \geq n_1(\frac{\epsilon}{12}, 1)$, where the last inequality follows from $R_1 \geq \epsilon$. Together with (A.13), we get for all $n \geq n_1(\frac{\epsilon}{12L_1}, 1)$

$$\begin{aligned}
&\mathbb{P}\Big\{\frac{1}{M_2^{(n)}}\big|\{i : \mathrm{Z}^n_{m_1,i} \in \mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{X}^{L_1}\mathrm{S}}(\mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \text{ for some } \mathcal{J} \in \overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})\}\big| > \exp\Big(-n\frac{2\epsilon}{3}\Big)\Big\} \\
&\quad < L_1\exp\Big(-\frac{1}{2}\exp\Big(n\frac{\epsilon}{4L_1}\Big)\Big) + \exp\Big(-\frac{L_1}{2}\exp\Big(n\frac{\epsilon}{3}\Big)\Big) \\
&\quad \leq (L_1+1)\exp\Big(-\frac{1}{2}\exp\Big(n\frac{\epsilon}{4L_1}\Big)\Big). \qquad \text{(A.16)}
\end{aligned}$$

For the following analysis let $\pi \in \Pi_{M_2^{(n)}}$ be any permutation. We note that (A.16) remains valid when we replace $\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})$ with $\pi\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})$, cf. (A.5a), which can be easily shown by replacing $\mathrm{Z}^n_{m_1,i}$ everywhere with $\mathrm{Z}^n_{m_1,\pi(i)}$. Further, let $\pi_1, ..., \pi_{p_1}$ be the $p_1 \leq n(L_1+1)^2(\log|\mathcal{X}|+1)$ permutations from Lemma A.2, which satisfy

$$\mathfrak{P}_{L_1}(\mathcal{M}_2\backslash\{i\}) = \bigcup_{k=1}^{p_1} \pi_k \overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\}).$$

From the union bound we get for $R_1 < \min_k I(\mathrm{X}_k;\mathrm{S})$ and $I(\mathrm{X};\mathrm{X}^{L_1},\mathrm{S}) \geq \epsilon$

$$\begin{aligned}
&\mathbb{P}\Big\{ \frac{1}{M_2^{(n)}} \big| \{i : \mathrm{Z}^n_{m_1,i} \in \mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{X}^{L_1}\mathrm{S}}(\mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \\
&\qquad\qquad \text{for some } \mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2\backslash\{i\})\} \big| > \exp\Big(-n\frac{\epsilon}{2}\Big)\Big\} \\
&\leq \mathbb{P}\Big\{ \frac{1}{M_2^{(n)}} \sum_{k=1}^{p_1} \big| \{i : \mathrm{Z}^n_{m_1,i} \in \mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{X}^{L_1}\mathrm{S}}(\mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \\
&\qquad\qquad \text{for some } \mathcal{J} \in \pi^k\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})\} \big| > \exp\Big(-n\frac{\epsilon}{2}\Big)\Big\} \\
&\leq \sum_{k=1}^{p_1} \mathbb{P}\Big\{ \frac{1}{M_2^{(n)}} \big| \{i : \mathrm{Z}^n_{m_1,i} \in \mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{X}^{L_1}\mathrm{S}}(\mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \\
&\qquad\qquad \text{for some } \mathcal{J} \in \pi^k\overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})\} \big| > \frac{1}{p_1}\exp\Big(-n\frac{\epsilon}{2}\Big)\Big\} \\
&\leq \exp\Big(n\frac{\epsilon}{6}\Big) \cdot \mathbb{P}\Big\{ \frac{1}{M_2^{(n)}} \big| \{i : \mathrm{Z}^n_{m_1,i} \in \mathcal{T}^{(n)}_{\mathrm{X}|\mathrm{X}^{L_1}\mathrm{S}}(\mathrm{Z}^n_{m_1,\mathcal{J}}, s^n) \\
&\qquad\qquad \text{for some } \mathcal{J} \in \overline{\mathfrak{P}}_{L_1}(\mathcal{M}_2\backslash\{i\})\} \big| > \exp\Big(-n\frac{2\epsilon}{3}\Big)\Big\} \\
&\leq (L_1+1)\exp\Big(n\frac{\epsilon}{6} - \frac{1}{2}\exp\Big(n\frac{\epsilon}{4L_1}\Big)\Big)
\end{aligned} \tag{A.17}$$

for all $n \geq n_1(\frac{\epsilon}{12L_1}, 1), \log(2L_1)$, such that

$$p_1 \leq n(L_1+1)^2(\log|\mathcal{X}|+1) \leq \exp\Big(n\frac{\epsilon}{6}\Big).$$

The second inequality follows from an analogous version of (A.10) for the arithmetic mean, and the last inequality follows from (A.16).

It remains to establish property (4.20c). Therefore, we observe that if we set $L_1 = 1$ and

$$\epsilon_1 = |R_1 - I(\mathrm{X}_k;\mathrm{S})|^+ + \frac{\epsilon}{4}$$

and apply (A.8) with $\frac{\epsilon}{4}$ replacing ϵ, the probability in (A.11) is bounded by $\exp\big(-\frac{1}{2}\exp\big(n\frac{\epsilon}{4}\big)\big)$ for all $n > n_1(\frac{\epsilon}{4}, 1)$. Moreover, if we assume that

$$I(\mathrm{X};\mathrm{X}_k,\mathrm{S}) - |R_1 - I(\mathrm{X}_k;\mathrm{S})|^+ \geq \epsilon$$

the $f_{m_1,i}$ in (A.12) again satisfy (A.2) with the a as given in (A.15). If we proceed as in (A.17), we finally obtain for $n > n_1(\frac{\epsilon}{12}, 1)$

$$\mathbb{P}\Big\{\frac{1}{M_2^{(n)}}\big|\{i : \mathrm{Z}_{m_1,i}^n \in \mathcal{T}_{\mathrm{X}|\mathrm{X}_1\mathrm{S}}^{(n)}(\mathrm{Z}_{m_1,j}^n, s^n) \text{ for some } j \neq i\}\big| > \exp\Big(-n\frac{\epsilon}{2}\Big)\Big\} < 2\exp\Big(n\frac{\epsilon}{6} - \frac{1}{2}\exp\Big(n\frac{\epsilon}{4}\Big)\Big). \tag{A.18}$$

Now we are in the position to complete the proof of the properties (4.20a)-(4.20e). As the total number of all possible combinations of sequences $x^n \in \mathcal{T}_{\mathrm{X}}^{(n)}$, states $s^n \in \mathcal{S}^n$, and joint types $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}}$ grow only exponentially in n, the doubly exponentially probability bounds (A.7), (A.9), (A.11), (A.17), and (A.18) ensure that with a probability close to 1 all the inequalities (4.20a)-(4.20e) hold simultaneously if $n \geq n_0(\epsilon, L_1)$ is sufficiently large.

The second part of the lemma, more precisely that for each $m_2 \in \mathcal{M}_2$ the properties (4.20f)-(4.20j) hold for $n \geq n_0(\epsilon, L_2)$, can be shown analogously using the same argumentation. Hence, if $n \geq \max\{n_0(\epsilon, L_1), n_0(\epsilon, L_2)\}$, there exist codewords $x_{m_1,m_2}^n \in \mathcal{T}_{\mathrm{X}}^{(n)}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, that simultaneously satisfy all the properties (4.20a)-(4.20j) for all choices of $x^n \in \mathcal{T}_{\mathrm{X}}^{(n)}$, $s^n \in \mathcal{S}^n$, and joint types $P_{\mathrm{X}\mathrm{X}^{L_1}\mathrm{S}}$ and $P_{\mathrm{X}\mathrm{X}^{L_2}\mathrm{S}}$. □

A.4 Proof of Lemma 4.30

The lemma is proved by contradiction. For receiving node i, $i = 1, 2$, suppose that the ensemble $(\mathrm{X}^{T_i+2}, \mathrm{S}^{T_i+2}, \mathrm{Y}_i)$ satisfies the conditions given in (4.21). Now, consider the divergences

$$D_{i,k} \coloneqq D(P_{\mathrm{X}^{T_i+2}\mathrm{S}_k\mathrm{Y}_i} \| P_{\mathrm{X}_k} \otimes P_{\mathrm{X}_k^{T_i+2}\mathrm{S}_k} \otimes W_{i,k}), \qquad k = 1, ..., T_i + 2$$

where $P_{\mathrm{X}_k} \otimes P_{\mathrm{X}_k^{T_i+2}\mathrm{S}_k} \otimes W_{i,k}$ is the distribution on $\mathcal{X}^{T_i+2} \times \mathcal{S} \times \mathcal{Y}_i$ with probability mass function

$$P(x_k) P_{\mathrm{X}_k^{T_i+2}\mathrm{S}_k}(x_k^{T_i+2}, s) W_i(y_i|x_k, s).$$

Next, we apply (B.1) to $D_{i,k}$ with $\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i$ playing the role of X and obtain

$$\begin{aligned} D_{i,k} &= D(P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i} \| P_{\mathrm{X}_k} \otimes P_{\mathrm{S}_k} \otimes W_{i,k}) + D(P_{\mathrm{X}_k^{T_i+2}|\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i} \| P_{\mathrm{X}_k^{T_i+2}|\mathrm{S}_k} | P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i}) \\ &= D(P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i} \| P_{\mathrm{X}_k} \otimes P_{\mathrm{S}_k} \otimes W_{i,k}) + I(\mathrm{X}_k, \mathrm{Y}_i; \mathrm{X}_k^{T_i+2} | \mathrm{S}_k) \end{aligned}$$

where $P_{\mathrm{X}_k} \otimes P_{\mathrm{S}_k} \otimes W_{i,k}$ is the distribution on $\mathcal{X} \times \mathcal{S} \times \mathcal{Y}_i$ with probability mass function $P(x_k)P_{\mathrm{S}_k}(s)W_i(y_i|x_k,s)$. Thereby, the last equality follows from (B.2). By assumption (4.21) holds so that we have $P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i} \in \mathcal{D}_{\eta_i}$ and the first term on the right side is bounded by η_i. Further, from (4.21) follows that the second term is also bounded by η_i. Thus, we have

$$\begin{aligned} 2\eta_i &\geq D_{i,k} \\ &= D(P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i} \| P_{\mathrm{X}_k} \otimes P_{\mathrm{S}_k} \otimes W_{i,k}) + I(\mathrm{X}_k, \mathrm{Y}_i; \mathrm{X}_k^{T_i+2} | \mathrm{S}_k) \\ &= \sum_{x_k,s,y_i} P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i}(x_k,s,y_i) \log \frac{P_{\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i}(x_k,s,y_i)}{P(x_k)P_{\mathrm{S}_k}(s)W_i(y_i|x_k,s)} \\ &\qquad + \sum_{x^{T_i+2},s,y_i} P_{\mathrm{X}^{T_i+2}\mathrm{S}_k\mathrm{Y}_i}(x^{T_i+2},s,y_i) \log \frac{P_{\mathrm{X}_k^{T_i+2}|\mathrm{X}_k\mathrm{S}_k\mathrm{Y}_i}(x_k^{T_i+2}|x_k,s,y_i)}{P_{\mathrm{X}_k^{T_i+2}|\mathrm{S}_k}(x_k^{T_i+2}|s)} \\ &= \sum_{x^{T_i+2},s,y_i} P_{\mathrm{X}^{T_i+2}\mathrm{S}_k\mathrm{Y}_i}(x^{T_i+2},s,y_i) \log \frac{P_{\mathrm{X}^{T_i+2}\mathrm{S}_k\mathrm{Y}_i}(x^{T_i+2},s,y_i)}{P(x_k)P_{\mathrm{X}_k^{T_i+2}\mathrm{S}_k}(x_k^{T_i+2},s)W_i(y_i|x_k,s)} \\ &= \sum_{x^{T_i+2},y_i} \sum_s P_{\mathrm{X}^{T_i+2}\mathrm{S}_k\mathrm{Y}_i}(x^{T_i+2},s,y_i) \times \\ &\qquad \log \frac{P_{\mathrm{X}^{T_i+2}\mathrm{S}_k\mathrm{Y}_i}(x^{T_i+2},s,y_i)}{P(x_k)P_{\mathrm{X}_k^{T_i+2}}(x_k^{T_i+2})P_{\mathrm{S}_k|\mathrm{X}_k^{T_i+2}}(s|x_k^{T_i+2})W_i(y_i|x_k,s)} \\ &\geq \sum_{x^{T_i+2},y_i} P_{\mathrm{X}^{T_i+2}\mathrm{Y}_i}(x^{T_i+2},y_i) \times \\ &\qquad \log \frac{P_{\mathrm{X}^{T_i+2}\mathrm{Y}_i}(x^{T_i+2},y_i)}{P(x_k)P_{\mathrm{X}_k^{T_i+2}}(x_k^{T_i+2}) \sum_s P_{\mathrm{S}_k|\mathrm{X}_k^{T_i+2}}(s|x_k^{T_i+2})W_i(y_i|x_k,s)} \\ &= D(P_{\mathrm{X}^{T_i+2}\mathrm{Y}_i} \| P_{\mathrm{X}_k} \otimes P_{\mathrm{X}_k^{T_i+2}} \otimes V_{i,k}) \end{aligned} \tag{A.19}$$

with $V_{i,k}(y_i|x^{T_i+2}) = \sum_s P_{\mathrm{S}_k|\mathrm{X}_k^{T_i+2}}(s|x_k^{T_i+2})W_i(y_i|x_k,s)$. The last inequality follows from the log-sum inequality.

From [CK81, p. 58] we know that we can bound the variational distance between two probability distributions from above by the square root of their divergence times an absolute

constant.[2] Therefore it follows from (A.19) that

$$\sum_{x^{T_i+2},y_i} \left|P_{\mathrm{X}^{T_i+2}\mathrm{Y}_i}(x^{T_i+2},y_i) - P(x_k)P_{\mathrm{X}_k^{T_i+2}}(x_k^{T_i+2})V_{i,k}(y_i|x^{T_i+2})\right| \leq c\sqrt{D(P_{\mathrm{X}^{T_i+2}\mathrm{Y}_i}\|P_{\mathrm{X}_k}\otimes P_{\mathrm{X}_k^{T_i+2}}\otimes V_{i,k})} \leq c\sqrt{2\eta_i} \tag{A.20}$$

with $c=\sqrt{2\ln 2}$ for all $i=1,...,T_i+2$. It follows from the triangle-inequality that

$$\max_{j\neq k}\sum_{x^{T_i+2},y_i} \left|P(x_k)P_{\mathrm{X}_k^{T_i+2}}(x_k^{T_i+2})V_{i,k}(y_i|x_k,x_k^{T_i+2}) - P(x_j)P_{\mathrm{X}_j^{T_i+2}}(x_j^{T_i+2})V_{i,j}(y_i|x_j,x_j^{T_i+2})\right| \leq 2c\sqrt{2\eta_i}. \tag{A.21}$$

Lemma A.3. *Let $\beta>0$ and the symmetrizability T_i, $i=1,2$, of the AVBBC $\mathfrak{W}^n$ be finite. Then there exists a $\xi>0$ such that every $P\in\mathcal{P}(\mathcal{X})$ with $\min_x P(x)\geq\beta$ and every collection of T_i+2 probability distributions $U_i\in\mathcal{P}(\mathcal{X}^{T_i+1}\times\mathcal{S})$, $1\leq k\leq T_i+2$, satisfy*

$$\max_{j\neq k}\sum_{x^{T_i+2},y_i}\left|\sum_s W_i(y_i|x_k,s)U_k(x_k^{T_i+2},s)P(x_k) - \sum_s W_i(y_i|x_j,s)U_j(x_j^{T_i+2},s)P(x_j)\right| \geq \xi. \tag{A.22}$$

Proof. The proof is identical to [Hug97, Lemma A4] and is therefore omitted. □

In particular, for the choice of $U_k=P_{\mathrm{X}_k^{T_i+2}\mathrm{S}_k}$, $1\leq k\leq T_i+2$, we obtain from (A.21) and (A.22)

$$\eta_i \geq \frac{\xi^2}{8c^2}$$

which contradicts the assumption that η_i can be chosen arbitrarily small. □

A.5 Proof of Theorem 4.36

The proof is a modification of the corresponding one in [Ahl86], where a similar result is given without constraints on the sequences of states. First, we observe that (4.50) is equivalent to

$$\sum_{s^n\in\mathcal{S}^n}\Big(1-f(s^n)\Big)q^{\otimes n}(s^n)\leq\alpha \qquad \text{for all } q\in\mathcal{P}_0(n,\mathcal{S},\Lambda). \tag{A.23}$$

[2] This bound with a worse constant was first given by Pinsker [Pin64] and is therefore also known as Pinsker's inequality.

Since each $\pi \in \Pi_n$ is bijective and because $q^{\otimes n}(\pi(s^n)) = q^{\otimes n}(s^n)$ for all $s^n \in \mathcal{S}^n$, we obtain from (A.23)

$$\begin{aligned}\alpha &\geq \sum_{s^n \in \mathcal{S}^n} \Big(1 - f\big(\pi(s^n)\big)\Big) q^{\otimes n}\big(\pi(s^n)\big) \\ &= \sum_{s^n \in \mathcal{S}^n} \Big(1 - f\big(\pi(s^n)\big)\Big) q^{\otimes n}(s^n) \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda) \text{ and all } \pi \in \Pi_n. \quad \text{(A.24)}\end{aligned}$$

Therefore, averaging (A.24) over Π_n yields

$$\sum_{s^n \in \mathcal{S}^n} \frac{1}{n!} \sum_{\pi \in \Pi_n} \Big(1 - f\big(\pi(s^n)\big)\Big) q^{\otimes n}(s^n) \leq \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda). \qquad \text{(A.25)}$$

Since $1 - \frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi(s^n)) \geq 0$, restricting the state sequences to $\mathcal{T}_q^{(n)}$ we get from (A.25)

$$\sum_{s^n \in \mathcal{T}_q^{(n)}} \frac{1}{n!} \sum_{\pi \in \Pi_n} \Big(1 - f\big(\pi(s^n)\big)\Big) q^{\otimes n}(s^n) \leq \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda)$$

which is equivalent to

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} \Big(1 - f\big(\pi(s^n)\big)\Big) q^{\otimes n}(\mathcal{T}_q^{(n)}) \leq \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda),\ s^n \in \mathcal{T}_q^{(n)} \quad \text{(A.26)}$$

because for $s^n \in \mathcal{T}_q^{(n)}$, the term $\frac{1}{n!} \sum_{\pi \in \Pi_n} (1 - f(\pi(s^n)))$ does not depend on s^n. Since $\mathcal{T}_q^{(n)} \geq (n+1)^{-|\mathcal{S}|}$, cf. [CK81, p. 30], (A.26) implies

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} \Big(1 - f\big(\pi(s^n)\big)\Big) \leq (n+1)^{|\mathcal{S}|} \alpha \quad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda),\ s^n \in \mathcal{T}_q^{(n)}. \quad \text{(A.27)}$$

Obviously, we have $\mathcal{S}_\Lambda^n = \bigcup_{q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda)} \mathcal{T}_q^{(n)}$ so that (A.27) shows that

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} f\big(\pi(s^n)\big) > 1 - (n+1)^{|\mathcal{S}|} \alpha \qquad \text{for all } s^n \in \mathcal{S}_\Lambda^n$$

which completes the proof of the theorem. □

A.6 Proof of Lemma 4.39

The lemma follows from Lemma 4.27, where the AVBBC with list decoding but without constraints on input and states is treated, and from [CN88b, Lemma 1], where the single-user

AVC with constraints is analyzed. Using the same ideas we are able to prove the corresponding result for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ. Thereby, we carry out the analysis for the case where $\Lambda_1(P_{\mathrm{X}}) < \Lambda$ for given type P_{X}, then the case $\Lambda_2(P_{\mathrm{X}}) < \Lambda$ follows accordingly.

We consider any deterministic code $\mathcal{C}_{\text{det}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ with codewords $x^n_{m_1,m_2} = (x_{m_1,m_2,1}, ..., x_{m_1,m_2,n}) \in \mathcal{X}^n$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, and the corresponding decoding sets $\mathcal{D}^{(1)}_{m_2|m_1} \subset \mathcal{Y}_1^n$ at node 1. Next, for any channel $U_1 \in \mathcal{U}_1$ which symmetrizes the AVBBC $\mathfrak{W}^n$, we define random variables $\mathrm{S}^n_{m_1,m_2} = (\mathrm{S}_{m_1,m_2,1}, ..., \mathrm{S}_{m_1,m_2,n}) \in \mathcal{S}^n$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$ with statistically independent elements and

$$\mathbb{P}\{\mathrm{S}_{m_1,m_2,k} = s\} = U_1(s|x_{m_1,m_2,k}).$$

In Lemma 4.27 the AVBBC with list decoding is analyzed. If we set the list sizes at the decoders and the symmetrizability of the channel to one, i.e., $L_i = T_i = 1$, $i = 1, 2$, we immediately obtain from Lemma 4.27 that there exists at least one $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$ such that

$$\mathbb{E}\big[\bar{e}_1(\mathrm{S}^n_{m_1,m_2}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))\big] \geq \frac{M_2^{(n)} - 1}{2M_2^{(n)}} \geq \frac{1}{4}. \tag{A.28}$$

Next, we restrict to codewords of type P_{X}, i.e., $x^n_{m_1,m_2} \in \mathcal{T}_{\mathrm{X}}^{(n)}$, $m_1 = 1, ..., M_1^{(n)}$, $m_2 = 1, ..., M_2^{(n)}$, with $\Lambda_1(P_{\mathrm{X}}) < \Lambda$. Further, we choose $U_1 \in \mathcal{U}_1$ such that it attains the minimum in (4.47). Then, with (4.46b) we get for the expectation

$$\begin{aligned}
\mathbb{E}[l(\mathrm{S}^n_{m_1,m_2})] &= \frac{1}{n}\sum_{k=1}^{n}\sum_{s\in\mathcal{S}} l(s)U_1(s|x_{m_1,m_2,k}) \\
&= \sum_{x\in\mathcal{X}}\sum_{s\in\mathcal{S}} P_{\mathrm{X}}(x)U_1(s|x)l(s) \\
&= \Lambda_1(P_{\mathrm{X}})
\end{aligned}$$

and the variance

$$\mathrm{var}[l(\mathrm{S}^n_{m_1,m_2})] \leq \frac{l^2_{\max}}{n}.$$

From Chebyshev's inequality we obtain

$$\begin{aligned}
\mathbb{P}\big\{l(\mathrm{S}^n_{m_1,m_2}) > \Lambda\big\} &= \mathbb{P}\big\{l(\mathrm{S}^n_{m_1,m_2}) - \mathbb{E}[l(\mathrm{S}^n_{m_1,m_2})] > \Lambda - \Lambda_1(P_{\mathrm{X}})\big\} \\
&\leq \frac{1}{n}\frac{l^2_{\max}}{(\Lambda - \Lambda_1(P_{\mathrm{X}}))^2}.
\end{aligned} \tag{A.29}$$

Finally, since

$$\mathbb{E}[\bar{e}_1(\mathrm{S}^n_{m_1,m_2}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))] \leq \max_{s^n:l(s^n)\leq\Lambda} \bar{e}_1(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) + \mathbb{P}\{l(\mathrm{S}^n_{m_1,m_2}) > \Lambda\},$$

we get from (A.28) and (A.29)

$$\begin{aligned}\max_{s^n:l(s^n)\leq\Lambda} \bar{e}_1(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) &\geq \mathbb{E}[\bar{e}_1(\mathrm{S}^n_{m_1,m_2}|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))] - \mathbb{P}\{l(\mathrm{S}^n_{m_1,m_2}) > \Lambda\} \\ &\geq \frac{M_2^{(n)}-1}{2M_2^{(n)}} - \frac{1}{n}\frac{l_{\text{max}}^2}{(\Lambda-\Lambda_1(P_{\mathrm{X}}))^2}\end{aligned}$$

which proves the first part of the lemma. Clearly, the second part where $\Lambda_2(P_{\mathrm{X}}) < \Lambda$ for given type P_{X} follows accordingly using the same argumentation. □

A.7 Proof of Lemma 4.43

The lemma is proved by contradiction as done in [CN88b, Lemma 4] for the single-user AVC. For receiving node i, $i = 1, 2$, we suppose that the quintuple $(\mathrm{X}, \mathrm{X}', \mathrm{S}, \mathrm{S}', \mathrm{Y}_i)$ satisfies the conditions given in (4.54). Since $P_{\mathrm{XSY}_i} \in \mathcal{D}_{\eta_i}(\Lambda)$ and $I(\mathrm{X}, \mathrm{Y}_i; \mathrm{X}'|\mathrm{S}) \leq \eta_i$, we have

$$\begin{aligned}2\eta_i &\geq D(P_{\mathrm{XSY}_i}\|P_{\mathrm{X}} \otimes P_{\mathrm{S}} \otimes W_i) + I(\mathrm{X}, \mathrm{Y}_i; \mathrm{X}'|\mathrm{S}) \\ &= \sum_{x,s,y_i} P_{\mathrm{XSY}_i}(x,s,y_i) \log \frac{P_{\mathrm{XSY}_i}(x,s,y_i)}{P_{\mathrm{X}}(x)P_{\mathrm{S}}(s)W_i(y_i|x,s)} \\ &\qquad + \sum_{x,x',s,y_i} P_{\mathrm{XX'SY}_i}(x,x',s,y_i) \log \frac{P_{\mathrm{X'|XSY}_i}(x'|x,s,y_i)}{P_{\mathrm{X'|S}}(x'|s)} \\ &= \sum_{x,x',s,y_i} P_{\mathrm{XX'SY}_i}(x,x',s,y_i) \log \frac{P_{\mathrm{XX'SY}_i}(x,x',s,y_i)}{P_{\mathrm{X}}(x)P_{\mathrm{X'S}}(x',s)W_i(y_i|x,s)} \\ &= \sum_{x,x',y_i}\sum_{s} P_{\mathrm{XX'SY}_i}(x,x',s,y_i) \log \frac{P_{\mathrm{XX'SY}_i}(x,x',s,y_i)}{P_{\mathrm{X}}(x)P_{\mathrm{X'}}(x')P_{\mathrm{S|X'}}(s|x')W_i(y_i|x,s)} \\ &\geq \sum_{x,x',y_i} P_{\mathrm{XX'Y}_i}(x,x',y_i) \log \frac{P_{\mathrm{XX'Y}_i}(x,x',y_i)}{P_{\mathrm{X}}(x)P_{\mathrm{X'}}(x')\sum_s P_{\mathrm{S|X'}}(s|x')W_i(y_i|x,s)} \\ &= D(P_{\mathrm{XX'Y}_i}\|P_{\mathrm{X}} \otimes P_{\mathrm{X'}} \otimes V_i') \qquad \text{(A.30)}\end{aligned}$$

with $V_i'(y_i|x,x') = \sum_s P_{\mathrm{S|X'}}(s|x')W_i(y_i|x,s)$ and the last inequality follows from the log-sum inequality.

From [CK81, p. 58] we know that we can bound the variational distance between two probability distributions from above by the square root of their divergence times an absolute constant. With this and (A.30) we get

$$\sum_{x,x',y_i} \left|P_{\mathrm{X}\mathrm{X}'\mathrm{Y}_i}(x,x',y_i) - P_{\mathrm{X}}(x)P_{\mathrm{X}'}(x')V_i'(y_i|x,x')\right| \\ \leq c\sqrt{D(P_{\mathrm{X}\mathrm{X}'\mathrm{Y}_i}\|P_{\mathrm{X}}\otimes P_{\mathrm{X}'}\otimes V_i')} \leq c\sqrt{2\eta_i} \qquad \text{(A.31)}$$

with $c = \sqrt{2\ln 2}$. Similarly, since $P_{\mathrm{X}'\mathrm{S}'\mathrm{Y}_i} \in \mathcal{D}_{\eta_i}(\Lambda)$ and $I(\mathrm{X}',\mathrm{Y}_i;\mathrm{X}|\mathrm{S}') \leq \eta_i$, cf. (4.54), we obtain

$$\sum_{x,x',y_i} \left|P_{\mathrm{X}\mathrm{X}'\mathrm{Y}_i}(x,x',y_i) - P_{\mathrm{X}'}(x')P_{\mathrm{X}}(x)V_i(y_i|x',x)\right| \leq c\sqrt{2\eta_i} \qquad \text{(A.32)}$$

with $c = \sqrt{2\ln 2}$ and $V_i(y_i|x',x) = \sum_s P_{\mathrm{S}'|\mathrm{X}}(s|x)W_i(y_i|x',s)$. Next, (A.31) and (A.32) together imply

$$\sum_{x,x',y_i} P_{\mathrm{X}}(x)P_{\mathrm{X}'}(x')\left|V_i(y_i|x',x) - V_i'(y_i|x,x')\right| \leq 2c\sqrt{2\eta_i}.$$

Since $\min_x P_{\mathrm{X}}(x) \geq \beta$, it immediately follows that

$$\max_{x,x',y_i} \left|V_i(y_i|x',x) - V_i'(y_i|x,x')\right| \leq \frac{2c\sqrt{2\eta_i}}{\beta^2}. \qquad \text{(A.33)}$$

Lemma A.4. *For any AVBBC $\mathfrak{W}^n$ with state constraint Λ and any input P_{X} with $\Lambda_i(P_{\mathrm{X}}) \geq \Lambda + \alpha$, $\alpha > 0$, $i = 1, 2$, for which each pair $U_1 : \mathcal{X} \to \mathcal{P}(\mathcal{S})$ and $U_2 : \mathcal{X} \to \mathcal{P}(\mathcal{S})$ satisfies*

$$\sum_{x,s} P_{\mathrm{X}}(x)U_1(s|x)l(s) \leq \Lambda \qquad \text{(A.34a)}$$

$$\sum_{x,s} P_{\mathrm{X}}(x)U_2(s|x)l(s) \leq \Lambda \qquad \text{(A.34b)}$$

there exists some $\xi > 0$ such that

$$\max_{x,x',y_i} \left|\sum_s W_i(y_i|x,s)U_1(s|x') - \sum_s W_i(y_i|x',s)U_2(s|x)\right| \geq \xi \qquad i = 1, 2. \qquad \text{(A.35)}$$

Proof. As in [CN88b, Lemma A2] we can interchange the two sums and then x and x' without changing the maximum in (A.35). Thus we can write for all $U_1, U_2 : \mathcal{X} \to \mathcal{P}(\mathcal{S})$

$$\max_{x,x',y_i} \left|\sum_s W_i(y_i|x,s)U_1(s|x') - \sum_s W_i(y_i|x',s)U_2(s|x)\right|$$

as

$$\max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U_2(s|x') - \sum_s W_i(y_i|x',s)U_1(s|x) \right|$$

so that we get

$$\begin{aligned}
&\max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U_1(s|x') - \sum_s W_i(y_i|x',s)U_2(s|x) \right| \\
&= \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)\frac{U_1(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_2(s|x)}{2} \right| \\
&\qquad + \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)\frac{U_2(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_1(s|x)}{2} \right| \\
&\geq \max_{x,x',y_i} \left\{ \left| \sum_s W_i(y_i|x,s)\frac{U_1(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_2(s|x)}{2} \right| \right. \\
&\qquad \left. + \left| \sum_s W_i(y_i|x,s)\frac{U_2(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_1(s|x)}{2} \right| \right\} \\
&\geq \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)\frac{U_1(s|x') + U_2(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_1(s|x) + U_2(s|x)}{2} \right| \\
&= \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U(s|x') - \sum_s W_i(y_i|x',s)U(s|x) \right| \qquad \text{(A.36)}
\end{aligned}$$

with $U = \frac{1}{2}(U_1 + U_2)$. Further, since U_1 and U_2 satisfy (A.34) for some P_{X}, then also U satisfies

$$\sum_{x,s} P_{\mathrm{X}}(x)U(s|x)l(s) \leq \Lambda. \qquad \text{(A.37)}$$

Since (A.36) can be considered as a continuous function of the pair (P_{X}, U) on the compact set of all channels $U : \mathcal{X} \to \mathcal{P}(\mathcal{S})$, it attains its minimum for some (P_{X}^*, U^*), where the minimization is taken over all channels U that satisfy (A.37). Additionally, since (P_{X}^*, U^*) satisfies (A.37), U^* cannot satisfy (4.4) which in turn implies that $\max_{x,x',y_i} |\sum_s W_i(y_i|x,s)U(s|x') - \sum_s W_i(y_i|x',s)U(s|x)| > 0$ completing the proof. □

If we choose $U_1 = P_{\mathrm{S}|\mathrm{X}'}$ and $U_2 = P_{\mathrm{S}'|\mathrm{X}}$, we obtain from (A.35)

$$\max_{x,x',y_i} \left| V_i(y_i|x',x) - V_i'(y_i|x,x') \right| \geq \xi. \qquad \text{(A.38)}$$

Finally, (A.33) and (A.38) yield

$$\eta_i \geq \frac{\xi^2 \beta^4}{8c^2}, \qquad i = 1, 2,$$

which contradicts the assumption that η_i can be chosen arbitrarily small. □

A.8 Fano's Inequality

For the BBC with common and confidential messages we have the following versions of Fano's inequality

$$H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2 | \mathrm{Y}_1^n, \mathrm{M}_1) \leq \bar{e}_1 \log(M_c^{(n)}, M_0^{(n)}, M_2^{(n)}) + 1 = n\epsilon_1^{(n)}$$
$$H(\mathrm{M}_0, \mathrm{M}_1 | \mathrm{Y}_2^n, \mathrm{M}_2) \leq \bar{e}_2 \log(M_0^{(n)}, M_1^{(n)}) + 1 = n\epsilon_2^{(n)}$$

with $\epsilon_1^{(n)} = \frac{1}{n}\log(M_c^{(n)} M_0^{(n)} M_2^{(n)})\bar{e}_1 + \frac{1}{n} \to 0$ and $\epsilon_2^{(n)} = \frac{1}{n}\log(M_0^{(n)} M_1^{(n)})\bar{e}_2 + \frac{1}{n} \to 0$ for $n \to \infty$ as $\bar{e}_1, \bar{e}_2 \to 0$.

Proof. We present the analysis for receiving node 1. Then, the other case follows accordingly using the same arguments. From the received sequence Y_1^n and its own message M_1 node 1 estimates the indices M_c, M_0, and M_2 from the sent codeword $\mathrm{X}^n(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_1, \mathrm{M}_2)$. We define the event of an error at node 1 as

$$\mathrm{E}_1 \coloneqq \begin{cases} 1, & \text{if } g_1(\mathrm{Y}_1^n, \mathrm{M}_1) \neq (\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2) \\ 0, & \text{if } g_1(\mathrm{Y}_1^n, \mathrm{M}_1) = (\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2) \end{cases}$$

so that we have for the average probability of error $\bar{e}_1 = \mathbb{P}\{\mathrm{E}_1 = 1\}$. From the chain rule for entropies we have

$$\begin{aligned} &H(\mathrm{E}_1, \mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2 | \mathrm{Y}_1^n, \mathrm{M}_1) \\ &\quad = H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2 | \mathrm{Y}_1^n, \mathrm{M}_1) + H(\mathrm{E}_1 | \mathrm{Y}_1^n, \mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_1, \mathrm{M}_2) \\ &\quad = H(\mathrm{E}_1 | \mathrm{Y}_1^n, \mathrm{M}_1) + H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2 | \mathrm{Y}_1^n, \mathrm{M}_1, \mathrm{E}_1). \end{aligned}$$

Since E_1 is a function of M_c, M_0, M_1, M_2, and Y_1^n, we have $H(\mathrm{E}_1 | \mathrm{Y}_1^n, \mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_1, \mathrm{M}_2) = 0$. Further, since E_1 is a binary-valued random variable, we get $H(\mathrm{E}_1 | \mathrm{Y}_1^n, \mathrm{M}_1) \leq H(\mathrm{E}_1) \leq 1$. So that finally with the next inequality

$$\begin{aligned} &H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2 | \mathrm{Y}_1^n, \mathrm{M}_1, \mathrm{E}_1) \\ &\quad = \mathbb{P}\{\mathrm{E}_1 = 0\} H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2 | \mathrm{Y}_1^n, \mathrm{M}_1, \mathrm{E}_1 = 0) + \\ &\qquad\qquad \mathbb{P}\{\mathrm{E}_1 = 1\} H(\mathrm{M}_c, \mathrm{M}_0, \mathrm{M}_2 | \mathrm{Y}_1^n, \mathrm{M}_1, \mathrm{E}_1 = 1) \\ &\quad \leq (1 - \bar{e}_1) 0 + \bar{e}_1 \log((M_0^{(n)} - 1)(M_2^{(n)} - 1)) \\ &\quad \leq \bar{e}_1 \log(M_0^{(n)} M_2^{(n)}) \end{aligned}$$

we get the desired version of Fano's inequality for the BBC with common and confidential messages. □

A.9 Proof of Lemma 5.23

Here we present the proof of Lemma 5.23. As in [LPS09] we prove the existence of a codebook with the desired properties by random coding arguments.

Random codebook generation and encoding

For the public communication we define (bidirectional) message sets $\mathcal{M}'_i$, $i = 0, 1, 2$, such that $|\mathcal{M}'_0||\mathcal{M}'_2| = \lfloor 2^{n(I(\mathrm{U};\mathrm{Y}_1)-\delta/2)} \rfloor$ and $|\mathcal{M}'_0||\mathcal{M}'_1| = \lfloor 2^{n(I(\mathrm{U};\mathrm{Y}_2)-\delta/2)} \rfloor$ are fulfilled. We generate $|\mathcal{M}'| = |\mathcal{M}'_0||\mathcal{M}'_1||\mathcal{M}'_2|$ independent codewords $u^n_{m'} \in \mathcal{U}^n$ with $m' = (m'_0, m'_1, m'_2)$ according to $p_{\mathrm{U}^n}(u^n) = \prod_{k=1}^n p_{\mathrm{U}}(u_k)$.

Further, for the confidential communication we choose (confidential) message sets $\mathcal{J}$ and $\mathcal{L}$ with $|\mathcal{J}| = \lfloor 2^{n(I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})-\delta/2)} \rfloor$ and $|\mathcal{L}| = \lfloor 2^{n(I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})-I(\mathrm{X};\mathrm{Y}_2|\mathrm{U})-\delta/2)} \rfloor$. Obviously, these sets satisfy the conditions (5.21) and (5.23). In the following, we consider only the case where these sets are non-empty[3] and set $\epsilon := \delta/8$. Then, for each $u^n_{m'} \in \mathcal{U}^n$ we generate $|\mathcal{J}||\mathcal{L}|$ independent codewords $x^n_{jlm'} \in \mathcal{X}^n$ according to $p_{\mathrm{X}^n|\mathrm{U}^n}(x^n|u^n) = \prod_{k=1}^n p_{\mathrm{X}|\mathrm{U}}(x_k|u_k)$.

Decoding

The receiving nodes use typical set decoding where each node uses its received sequence and its side information to create the decoding sets. In more detail, if $x^n_{jlm'} \in \mathcal{X}^n$ has been sent, node 1 uses the received sequence $y^n_1 \in \mathcal{Y}^n_1$ and its own message $m'_1 \in \mathcal{M}'_1$ to create

$$\mathcal{D}_{11}(m'_1, y^n_1) := \big\{(m'_0, m'_2) \in \mathcal{M}'_0 \times \mathcal{M}'_2 : (u^n_{m'}, y^n_1) \in \mathcal{A}^{(n)}_\epsilon(\mathrm{U}, \mathrm{Y}_1)\big\}.$$

If $\mathcal{D}_{11}(m'_1, y^n_1)$ is empty or contains more than one element, node 1 maps to the symbol 0, cf. also Definition 5.1, and declares an error. Otherwise, in a second step it uses the unique $(m'_0, m'_2) \in \mathcal{D}_{11}(m'_1, y^n_1)$ and its own $m'_1 \in \mathcal{M}'_1$ to create

$$\mathcal{D}_{12}(m', y^n_1) := \big\{(j, l) \in \mathcal{J} \times \mathcal{L} : (u^n_{m'}, x^n_{jlm'}, y^n_1) \in \mathcal{A}^{(n)}_\epsilon(\mathrm{U}, \mathrm{X}, \mathrm{Y}_1)\big\}.$$

[3] We need not consider the trivial cases of zero rates since they are always achievable.

Again, if $\mathcal{D}_{12}(m', y_1^n)$ is empty or contains more than one element, node 1 maps to 0 and declares an error. Otherwise, if there is a unique $(j, l) \in \mathcal{D}_{12}(m', y_1^n)$, it declares that $(j, l, m') \in \mathcal{J} \times \mathcal{L} \times \mathcal{M}'$ has been sent.

Similarly, node 2 uses $y_2^n \in \mathcal{Y}_2^n$ and $m_2' \in \mathcal{M}_2'$ to define

$$\mathcal{D}_{21}(m_2', y_2^n) \coloneqq \big\{(m_0', m_1') \in \mathcal{M}_0' \times \mathcal{M}_1' : (u_{m'}^n, y_2^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_2)\big\}.$$

If there is a unique $(m_0', m_1') \in \mathcal{D}_{21}(m_2', y_2^n)$, with its own $m_2' \in \mathcal{M}_2'$ and given $l \in \mathcal{L}$ it creates

$$\mathcal{D}_{22}(l, m', y_2^n) \coloneqq \big\{j \in \mathcal{J} : (u_{m'}^n, x_{jlm'}^n, y_2^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_2)\big\}.$$

It declares that $(j, l, m') \in \mathcal{J} \times \mathcal{L} \times \mathcal{M}'$ has been sent if there is a unique $j \in \mathcal{D}_{22}(l, m', y_2^n)$. The events of an error are defined accordingly as for node 1.

Analysis of probability of error

For the following analysis we introduce for any $(j, l, m') \in \mathcal{J} \times \mathcal{L} \times \mathcal{M}'$ the random error events for node 1:

$$\begin{aligned}
E_{11}(m_0', m_2'|m_1') &\coloneqq \big\{(u_{m'}^n, y_1^n) \notin \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_1)\big\} \\
E_{12}(m_0', m_2'|m_1') &\coloneqq \big\{\exists\, (\hat{m}_0, \hat{m}_2) \neq (m_0', m_2') : (u_{\hat{m}_0 m_1' \hat{m}_2}^n, y_1^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_1)\big\} \\
E_{13}(j, l|m') &\coloneqq \big\{(u_{m'}^n, x_{jlm'}^n, y_1^n) \notin \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_1)\big\} \\
E_{14}(j, l|m') &\coloneqq \big\{\exists\, (\hat{j}, \hat{l}) \neq (j, l) : (u_{m'}^n, x_{\hat{j}\hat{l}m'}^n, y_1^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_1)\big\}.
\end{aligned}$$

From the union bound we get for the probabilities of error

$$e_1(m_0', m_2'|m_1') \leq \mathbb{P}\big\{E_{11}(m_0', m_2'|m_1')\big\} + \mathbb{P}\big\{E_{12}(m_0', m_2'|m_1')\big\} \tag{A.39a}$$

$$e_1(j, l|m') \leq \mathbb{P}\big\{E_{13}(j, l|m')\big\} + \mathbb{P}\big\{E_{14}(j, l|m')\big\} \tag{A.39b}$$

where each one is bounded separately using standard arguments, cf. Appendix B.2.2 or standard literature such as [CT06].

For $\mathbb{P}\{E_{11}(m_0', m_2'|m_1')\}$ we know from the definition of the decoding sets, cf. Lemma B.15 in Appendix B.2.2, that for increasing n we have

$$\mathbb{P}\big\{(u_{m'}^n, y_1^n) \notin \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_1)\big\} \underset{n\to\infty}{\longrightarrow} 0. \tag{A.40}$$

With $\hat{m} = (\hat{m}_0, m_1', \hat{m}_2)$ we get for the second event

$$\begin{aligned}\mathbb{P}\big\{E_{12}(m_0', m_2'|m_1')\big\} &\leq |\mathcal{M}_0'||\mathcal{M}_2'|\,\mathbb{P}\big\{(u^n_{\hat{m}_0 m_1' \hat{m}_2}, y_1^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_1)\big\} \\ &= |\mathcal{M}_0'||\mathcal{M}_2'| \sum_{(u^n_{\hat{m}}, y_1^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U},\mathrm{Y}_1)} p_{\mathrm{Y}_1^n}(y_1^n) p_{\mathrm{U}^n}(u^n_{\hat{m}}) \\ &\leq 2^{n(I(\mathrm{U};\mathrm{Y}_1)-\delta/2)} 2^{n(H(\mathrm{U},\mathrm{Y}_1)+\epsilon)} 2^{-n(H(\mathrm{Y}_1)-\epsilon)} 2^{-n(H(\mathrm{U})-\epsilon)} \\ &= 2^{-n\epsilon} \underset{n\to\infty}{\longrightarrow} 0 \end{aligned} \tag{A.41}$$

where the first inequality follows from the union bound, the second one from the definition of the sets $\mathcal{M}_0'$, $\mathcal{M}_2'$ and $|\mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_1)| \leq 2^{n(H(\mathrm{U},\mathrm{Y}_1)+\epsilon)}$, cf. Lemma B.15, and the last equality from $\delta = 8\epsilon$. Substituting (A.40)-(A.41) into (A.39a) we conclude that $e_1(m_0', m_2'|m_1') \to 0$ as $n \to \infty$.

For $\mathbb{P}\{E_{13}(j, l|m')\}$ follows, similarly as in the first event, from the definition of the decoding sets that for increasing n

$$\mathbb{P}\big\{(u^n_{m'}, x^n_{jlm'}, y_1^n) \notin \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_1)\big\} \underset{n\to\infty}{\longrightarrow} 0. \tag{A.42}$$

It remains to bound $\mathbb{P}\{E_{14}(j, l|m')\}$. Therefore, we proceed as in the second event and obtain

$$\begin{aligned}&\mathbb{P}\{E_{13}(j, l|m')\} \\ &\quad\leq |\mathcal{J}||\mathcal{L}| \sum_{(u^n_{m'}, x^n_{\hat{j}\hat{l}m'}, y_1^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U},\mathrm{X},\mathrm{Y}_1)} p_{\mathrm{Y}_1^n|\mathrm{U}^n}(y_1^n|u^n_{m'}) p_{\mathrm{X}^n|\mathrm{U}^n}(x^n_{\hat{j}\hat{l}m'}|u^n_{m'}) p_{\mathrm{U}^n}(u^n_{m'}) \\ &\quad\leq |\mathcal{J}||\mathcal{L}| 2^{n(H(\mathrm{U},\mathrm{X},\mathrm{Y}_1)+\epsilon)} 2^{-n(H(\mathrm{Y}_1|\mathrm{U})-\epsilon)} 2^{-n(H(\mathrm{X}|\mathrm{U})-\epsilon)} 2^{-n(H(\mathrm{U})-\epsilon)} \\ &\quad\leq 2^{-n4\epsilon} \underset{n\to\infty}{\longrightarrow} 0 \end{aligned} \tag{A.43}$$

where the second inequality follows from $|\mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_1)| \leq 2^{n(H(\mathrm{U},\mathrm{X},\mathrm{Y}_1)+\epsilon)}$ and the third from $|\mathcal{J}||\mathcal{L}| \leq 2^{n(I(\mathrm{X};\mathrm{Y}_1|\mathrm{U})-\delta)}$ and $\delta = 8\epsilon$. Substituting (A.42)-(A.43) into (A.39b) we end up with $e_1(j, l|m') \to 0$ as $n \to \infty$.

The analysis for the probability of error at node 2 follows accordingly with the random error events

$$\begin{aligned} E_{21}(m_0', m_1'|m_2') &\coloneqq \big\{(u^n_{m'}, y_2^n) \notin \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_2)\big\} \\ E_{22}(m_0', m_1'|m_2') &\coloneqq \big\{\exists(\hat{m}_0, \hat{m}_1) \neq (m_0', m_1') : (u^n_{\hat{m}_0 \hat{m}_1 m_2'}, y_2^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{Y}_2)\big\} \\ E_{23}(j|l, m') &\coloneqq \big\{(u^n_{m'}, x^n_{jlm'}, y_2^n) \notin \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_2)\big\} \\ E_{24}(j|l, m') &\coloneqq \big\{\exists\hat{j} \neq j : (u^n_{m'}, x^n_{\hat{j}lm'}, y_2^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{U}, \mathrm{X}, \mathrm{Y}_2)\big\}. \end{aligned}$$

Using the same arguments, it is straightforward to show that the probabilities of error fulfill

$$e_2(m_0', m_1'|m_2') \leq \mathbb{P}\big\{E_{21}(m_0', m_1'|m_2')\big\} + \mathbb{P}\big\{E_{22}(m_0', m_1'|m_2')\big\} \underset{n\to\infty}{\longrightarrow} 0 \tag{A.44}$$

$$e_2(j|l, m') \leq \mathbb{P}\big\{E_{23}(j|l, m')\big\} + \mathbb{P}\big\{E_{24}(j|l, m')\big\} \underset{n\to\infty}{\longrightarrow} 0. \tag{A.45}$$

From (A.40)-(A.45) we conclude that the probabilities of error, averaged over all codewords and codebooks, get arbitrarily small. Finally, from random coding arguments follows that for n large enough there exists a codebook with the desired rates (5.21) and (5.23) that satisfies the conditions on the probabilities of error (5.22) and (5.24) proving the lemma. □

A.10 Proof of Converse of Proposition 5.41

We have to show that any given sequence of $(n, M_c^{(n)}, M_0^{(n)}, M_1^{(n)}, M_2^{(n)})$-codes with $\bar{e}_{1\mathrm{a}}, \bar{e}_{1\mathrm{b}}, \bar{e}_2 \to 0$ there exist random variables $\mathrm{U} - \mathrm{X} - \widetilde{\mathrm{Y}}_{1\mathrm{a}} - (\mathrm{Y}_{1\mathrm{b}}, \mathrm{Y}_2)$ such that

$$\begin{aligned}
\tfrac{1}{n} H(\mathrm{M}_c) &\leq I(\mathrm{X}; \widetilde{\mathrm{Y}}_{1\mathrm{a}}|\mathrm{U}) - I(\mathrm{X}; \mathrm{Y}_2|\mathrm{U}) + o(n^0) \\
\tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_2)\big) &\leq I(\mathrm{U}; \mathrm{Y}_{1\mathrm{b}}) + o(n^0) \\
\tfrac{1}{n}\big(H(\mathrm{M}_0) + H(\mathrm{M}_1)\big) &\leq I(\mathrm{U}; \mathrm{Y}_2) + o(n^0)
\end{aligned}$$

are satisfied. For this purpose we need an appropriate version of Fano's lemma.

Lemma A.5 (Fano's inequality)**.** *For the BBC with common and confidential messages that satisfies* $\mathrm{X} - \widetilde{\mathrm{Y}}_{1a} - \mathrm{Y}_{1b}$ *and* $\mathrm{X} - \widetilde{\mathrm{Y}}_{1a} - \mathrm{Y}_2$ *we have the following versions of Fano's inequality*

$$\begin{aligned}
H(\mathrm{M}_c|\widetilde{\mathrm{Y}}_{1a}^n, \mathrm{M}_1) &\leq \bar{e}_{1a} \log M_c^{(n)} + 1 = n\epsilon_{1a}^{(n)} \\
H(\mathrm{M}_0, \mathrm{M}_2|\mathrm{Y}_{1b}^n, \mathrm{M}_1) &\leq \bar{e}_{1b} \log(M_0^{(n)} M_2^{(n)}) + 1 = n\epsilon_{1b}^{(n)} \\
H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) &\leq \bar{e}_2 \log(M_0^{(n)} M_1^{(n)}) + 1 = n\epsilon_2^{(n)}
\end{aligned}$$

with $\epsilon_{1a}^{(n)} = \frac{1}{n}\log(M_c^{(n)})\bar{e}_{1a} + \frac{1}{n} \to 0$, $\epsilon_{1b}^{(n)} = \frac{1}{n}\log(M_0^{(n)} M_2^{(n)})\bar{e}_{1b} + \frac{1}{n} \to 0$, *and* $\epsilon_2^{(n)} = \frac{1}{n}\log(M_0^{(n)} M_1^{(n)})\bar{e}_2 + \frac{1}{n} \to 0$ *for* $n \to \infty$ *as* $\bar{e}_{1a}, \bar{e}_{1b}, \bar{e}_2 \to 0$.

Proof. The proof is quite similar to the one given in Appendix A.8 and is therefore omitted for brevity. □

For notational convenience we introduce the abbreviation $\mathrm{M}_p = (\mathrm{M}_0, \mathrm{M}_1, \mathrm{M}_2)$ for the public messages and define the auxiliary random variable $\mathrm{U}_k := (\mathrm{M}_p, \widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1})$ that satisfies $\mathrm{U}_k - \mathrm{X}_k - (\widetilde{\mathrm{Y}}_{1\mathrm{a}k}, \mathrm{Y}_{1\mathrm{b}k}, \mathrm{Y}_{2k})$.

We follow [LLL10, Proposition 1] and bound the entropies of the public messages using the independence of M_0, M_1, M_2, the definition of mutual information, Fano's inequality, cf. Lemma A.5, and the chain rule for mutual information. We get

$$\begin{aligned}
H(\mathrm{M}_0) + H(\mathrm{M}_2) &= H(\mathrm{M}_0, \mathrm{M}_2|\mathrm{M}_1) \\
&\leq I(\mathrm{M}_0, \mathrm{M}_2; \mathrm{Y}_{1\mathrm{b}}^n|\mathrm{M}_1) + n\epsilon_{1\mathrm{b}}^{(n)} \\
&\leq I(\mathrm{M}_p; \mathrm{Y}_{1\mathrm{b}}^n) + n\epsilon_{1\mathrm{b}}^{(n)} \\
&= \sum_{k=1}^{n} I(\mathrm{M}_p; \mathrm{Y}_{1\mathrm{b}k}|\mathrm{Y}_{1\mathrm{b}}^{k-1}) + n\epsilon_{1\mathrm{b}}^{(n)} \\
&\leq \sum_{k=1}^{n} I(\mathrm{M}_p, \widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}; \mathrm{Y}_{1\mathrm{b}k}|\mathrm{Y}_{1\mathrm{b}}^{k-1}) + n\epsilon_{1\mathrm{b}}^{(n)} \\
&\leq \sum_{k=1}^{n} I(\mathrm{M}_p, \widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{Y}_{1\mathrm{b}}^{k-1}; \mathrm{Y}_{1\mathrm{b}k}) + n\epsilon_{1\mathrm{b}}^{(n)} \\
&\leq \sum_{k=1}^{n} I(\mathrm{M}_p, \widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}; \mathrm{Y}_{1\mathrm{b}k}) + n\epsilon_{1\mathrm{b}}^{(n)} \\
&\leq \sum_{k=1}^{n} I(\mathrm{U}_k; \mathrm{Y}_{1\mathrm{b}k}) + n\epsilon_{1\mathrm{b}}^{(n)}
\end{aligned} \tag{A.46}$$

where the second last inequality follows from the Markov chain $(\mathrm{M}_p, \mathrm{Y}_{1\mathrm{b}k}) - \widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1} - \mathrm{Y}_{1\mathrm{b}}^{k-1}$, i.e., the degradedness of the channel. Using the same arguments we similarly obtain

$$H(\mathrm{M}_0) + H(\mathrm{M}_1) \leq \sum_{k=1}^{n} I(\mathrm{U}_k; \mathrm{Y}_{2k}) + n\epsilon_2^{(n)}. \tag{A.47}$$

Next, we consider the confidential message and proceed exactly as in [LLL10, Proposition 1]. Since the confidential message has to be kept secret from receiver 2 but needs not be kept secret from (virtual) receiver 1b, we have to consider the term $H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2)$ due to the perfect secrecy condition (5.2). We obtain

$$\begin{aligned}
H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) &= H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_p) + I(\mathrm{M}_c; \mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) \\
&\leq H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_p) + n\epsilon_2^{(n)} \\
&= I(\mathrm{M}_c; \widetilde{\mathrm{Y}}_{1\mathrm{a}}^n|\mathrm{Y}_2^n, \mathrm{M}_p) + H(\mathrm{M}_c|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^n, \mathrm{Y}_2^n, \mathrm{M}_p) + n\epsilon_2^{(n)} \\
&\leq I(\mathrm{M}_c; \widetilde{\mathrm{Y}}_{1\mathrm{a}}^n|\mathrm{Y}_2^n, \mathrm{M}_p) + n\epsilon_{1\mathrm{a}}^{(n)} + n\epsilon_2^{(n)} \\
&\leq I(\mathrm{M}_c, \mathrm{X}^n; \widetilde{\mathrm{Y}}_{1\mathrm{a}}^n|\mathrm{Y}_2^n, \mathrm{M}_p) + n\epsilon_{1\mathrm{a}}^{(n)} + n\epsilon_2^{(n)} \\
&= I(\mathrm{X}^n; \widetilde{\mathrm{Y}}_{1\mathrm{a}}^n|\mathrm{Y}_2^n, \mathrm{M}_p) + n\epsilon_{1\mathrm{a}}^{(n)} + n\epsilon_2^{(n)}
\end{aligned}$$

$$
\begin{aligned}
&= H(\mathrm{X}^n|\mathrm{Y}_2^n, \mathrm{M}_p) - H(\mathrm{X}^n|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^n, \mathrm{Y}_2^n, \mathrm{M}_p) + n\epsilon_{1\mathrm{a}}^{(n)} + n\epsilon_2^{(n)} \\
&= H(\mathrm{X}^n|\mathrm{Y}_2^n, \mathrm{M}_p) - H(\mathrm{X}^n|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^n, \mathrm{M}_p) + n\epsilon_{1\mathrm{a}}^{(n)} + n\epsilon_2^{(n)} \\
&= I(\mathrm{X}^n; \widetilde{\mathrm{Y}}_{1\mathrm{a}}^n|\mathrm{M}_p) - I(\mathrm{X}^n; \mathrm{Y}_2^n|\mathrm{M}_p) + n\epsilon_{1\mathrm{a}}^{(n)} + n\epsilon_2^{(n)}
\end{aligned}
$$

where we made extensively use of the definition of mutual information and Fano's inequality, cf. Lemma A.5. In more detail, the first inequality follows from $I(\mathrm{M}_c; \mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) = H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) - H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_c, \mathrm{M}_2) \leq H(\mathrm{M}_0, \mathrm{M}_1|\mathrm{Y}_2^n, \mathrm{M}_2) \leq n\epsilon_2^{(n)}$, the second inequality from $H(\mathrm{M}_c|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^n, \mathrm{Y}_2^n, \mathrm{M}_p) \leq H(\mathrm{M}_c|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^n, \mathrm{M}_1) \leq n\epsilon_{1\mathrm{a}}^{(n)}$, the third equality from the Markov chain $(\mathrm{M}_c, \mathrm{M}_p) - (\mathrm{X}^n, \mathrm{Y}_2^n) - \widetilde{\mathrm{Y}}_{1\mathrm{a}}^n$, the second last equality from the degradedness of the channel, i.e., $(\mathrm{X}^n, \mathrm{M}_p) - \widetilde{\mathrm{Y}}_{1\mathrm{a}}^n - \mathrm{Y}_2^n$, and the last equality from the addition of the "zero" term $H(\mathrm{X}^n|\mathrm{M}_p) - H(\mathrm{X}^n|\mathrm{M}_p)$. Next, with $\epsilon^{(n)} = \epsilon_{1\mathrm{a}}^{(n)} + \epsilon_2^{(n)}$ and using the chain rule for mutual information we get similar to [LLL10, Proposition 1]

$$
\begin{aligned}
H(\mathrm{M}_c|\mathrm{Y}_2^n, \mathrm{M}_2) &\leq \sum_{k=1}^{n} \Big[I(\mathrm{X}^n; \widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p) - I(\mathrm{X}^n; \mathrm{Y}_{2k}|\mathrm{Y}_2^{k-1}, \mathrm{M}_p) \Big] + n\epsilon^{(n)} \\
&= \sum_{k=1}^{n} \Big[H(\widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p) - H(\widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p, \mathrm{X}^n) \\
&\qquad - H(\mathrm{Y}_{2k}|\mathrm{Y}_2^{k-1}, \mathrm{M}_p) + H(\mathrm{Y}_{2k}|\mathrm{Y}_2^{k-1}, \mathrm{M}_p, \mathrm{X}^n) \Big] + n\epsilon^{(n)} \\
&\leq \sum_{k=1}^{n} \Big[H(\widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p) - H(\widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p, \mathrm{X}_k) \\
&\qquad - H(\mathrm{Y}_{2k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{Y}_2^{k-1}, \mathrm{M}_p) + H(\mathrm{Y}_{2k}|\mathrm{Y}_2^{k-1}, \mathrm{M}_p, \mathrm{X}_k) \Big] + n\epsilon^{(n)} \\
&\leq \sum_{k=1}^{n} \Big[H(\widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p) - H(\widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p, \mathrm{X}_k) \\
&\qquad - H(\mathrm{Y}_{2k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p) + H(\mathrm{Y}_{2k}|\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p, \mathrm{X}_k) \Big] + n\epsilon^{(n)} \\
&= \sum_{k=1}^{n} \Big[I(\mathrm{X}_k; \widetilde{\mathrm{Y}}_{1\mathrm{a}k}|\mathrm{U}_k) - I(\mathrm{X}_k; \mathrm{Y}_{2k}|\mathrm{U}_k) \Big] + n\epsilon^{(n)}
\end{aligned} \tag{A.48}
$$

where the second inequality follows from the Markov chain $(\widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p, \mathrm{X}^n) - \mathrm{X}_k - \widetilde{\mathrm{Y}}_{1\mathrm{a}k}$ and the third inequality from the Markov chains $(\mathrm{Y}_{2k}, \mathrm{M}_p) - \widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1} - \mathrm{Y}_2^{k-1}$ and $(\mathrm{Y}_2^{k-1}, \widetilde{\mathrm{Y}}_{1\mathrm{a}}^{k-1}, \mathrm{M}_p) - \mathrm{X}_k - \mathrm{Y}_{2k}$.

To complete the proof, we introduce an auxiliary random variable J that is independent of M_c, M_p, X^n, $\widetilde{\mathrm{Y}}_{1\mathrm{a}}^n$, $\mathrm{Y}_{1\mathrm{b}}^n$, and Y_2^n and uniformly distributed over $\{1, ..., n\}$. Further, let

$$
\mathrm{U} \coloneqq (\mathrm{U}_\mathrm{J}, \mathrm{J}), \quad \mathrm{X} \coloneqq \mathrm{X}_\mathrm{J}, \quad \widetilde{\mathrm{Y}}_{1\mathrm{a}} \coloneqq \widetilde{\mathrm{Y}}_{1\mathrm{a}\mathrm{J}}, \quad \mathrm{Y}_{1\mathrm{b}} \coloneqq \mathrm{Y}_{1\mathrm{b}\mathrm{J}}, \quad \mathrm{Y}_2 \coloneqq \mathrm{Y}_{2\mathrm{J}}.
$$

Substituting this into (A.46)-(A.48) and dividing by n yields the desired bounds. □

A.11 Proof of Lemma 5.44

Similarly as in [LLL10] for the classical MIMO Gaussian broadcast channel with common and confidential messages let $(R_c, R_0, R_1, R_2) \in \mathbb{R}_+^4$ be an achievable rate tuple for the MIMO Gaussian BBC with common and confidential messages that satisfies

$$\begin{aligned} R_c &\leq I(\mathrm{V};\mathbf{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathbf{Y}_2|\mathrm{U}) \\ R_0 + R_i &\leq I(\mathrm{U};\mathbf{Y}_i), \quad i = 1,2, \end{aligned}$$

with $\mathrm{U} - \mathrm{V} - \mathbf{X} - (\mathbf{Y}_1, \mathbf{Y}_2)$, cf. Corollary 5.33. Further, let $(\overline{R}_c, \overline{R}_0, \overline{R}_1, \overline{R}_2) \in \mathbb{R}_+^4$ be an achievable rate tuple that satisfies

$$\begin{aligned} \overline{R}_c &\leq I(\mathrm{V};\overline{\mathbf{Y}}_1|\mathrm{U}) - I(\mathrm{V};\overline{\mathbf{Y}}_2|\mathrm{U}) \\ \overline{R}_0 + \overline{R}_i &\leq I(\mathrm{U};\overline{\mathbf{Y}}_i), \quad i = 1,2, \end{aligned}$$

with $\mathrm{U}-\mathrm{V}-\mathbf{X}-(\mathbf{Y}_1, \mathbf{Y}_2)$ for the new MIMO Gaussian BBC with common and confidential messages, cf. (5.70). Next, we bound the differences between the rates for both channels as in [LLL10]. Due to the Markov chains (5.71) we have $I(\mathrm{U};\mathbf{Y}_i) \leq I(\mathrm{U};\overline{\mathbf{Y}}_i)$, $i = 1,2$, so that we get for the difference of the (individual) bidirectional rates

$$R_i - \overline{R}_i \leq I(\mathrm{U};\mathbf{Y}_i) - I(\mathrm{U};\overline{\mathbf{Y}}_i) \leq 0, \quad i = 1,2. \tag{A.49}$$

Similarly, because of (5.71) we have $I(\mathrm{V};\mathbf{Y}_i|\mathrm{U}) \leq I(\mathrm{V};\overline{\mathbf{Y}}_i|\mathrm{U})$ and $I(\mathbf{X};\mathbf{Y}_i|\mathrm{U},\mathrm{V}) \leq I(\mathbf{X};\overline{\mathbf{Y}}_i|\mathrm{U},\mathrm{V})$, $i = 1,2$, so that the difference of the confidential rates is given by, cf. also [LLL10, Section IV],

$$\begin{aligned} R_c - \overline{R}_c &\leq I(\mathrm{V};\mathbf{Y}_1|\mathrm{U}) - I(\mathrm{V};\mathbf{Y}_2|\mathrm{U}) - \big[I(\mathrm{V};\overline{\mathbf{Y}}_1|\mathrm{U}) - I(\mathrm{V};\overline{\boldsymbol{Y}}_2|\mathrm{U})\big] \\ &= I(\mathrm{V};\overline{\mathbf{Y}}_2|\mathrm{U}) - I(\mathrm{V};\mathbf{Y}_2|\mathrm{U}) - \big[I(\mathrm{V};\overline{\mathbf{Y}}_1|\mathrm{U}) - I(\mathrm{V};\mathbf{Y}_1|\mathrm{U})\big] \\ &\leq I(\mathrm{V};\overline{\mathbf{Y}}_2|\mathrm{U}) - I(\mathrm{V};\mathbf{Y}_2|\mathrm{U}) \\ &= I(\mathrm{U},\mathrm{V};\overline{\mathbf{Y}}_2) - I(\mathrm{U},\mathrm{V};\mathbf{Y}_2) - \big[I(\mathrm{U};\overline{\mathbf{Y}}_2) - I(\mathrm{U};\mathbf{Y}_2)\big] \\ &\leq I(\mathrm{U},\mathrm{V};\overline{\mathbf{Y}}_2) - I(\mathrm{U},\mathrm{V};\mathbf{Y}_2) \\ &= I(\mathbf{X};\overline{\mathbf{Y}}_2) - I(\mathbf{X};\mathbf{Y}_2) - \big[I(\mathbf{X};\overline{\mathbf{Y}}_2|\mathrm{U},\mathrm{V}) - I(\mathbf{X};\mathbf{Y}_2|\mathrm{U},\mathrm{V})\big] \\ &\leq I(\mathbf{X};\overline{\mathbf{Y}}_2) - I(\mathbf{X};\mathbf{Y}_2) \\ &= I(\mathbf{X};\overline{\mathbf{Y}}_2|\mathbf{Y}_2) \\ &\leq \max_{\mathbf{0}\preceq\boldsymbol{Q}^{(c)}\preceq\boldsymbol{S}} \left[\frac{1}{2}\log\det\left(\boldsymbol{I}_{N_R} + \overline{\boldsymbol{H}}_2\boldsymbol{Q}^{(c)}\overline{\boldsymbol{H}}_2^T\right) - \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_R} + \boldsymbol{H}_2\boldsymbol{Q}^{(c)}\boldsymbol{H}_2^T\right)\right] \\ &= \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_R} + \overline{\boldsymbol{H}}_2\boldsymbol{S}\overline{\boldsymbol{H}}_2^T\right) - \frac{1}{2}\log\det\left(\boldsymbol{I}_{N_R} + \boldsymbol{H}_2\boldsymbol{S}\boldsymbol{H}_2^T\right) \end{aligned} \tag{A.50}$$

where the second last equality follows from the Markov condition (5.71), the last inequality from [Tho87, Lemma 1], and the last equality from the fact that $\boldsymbol{H}_2^T\boldsymbol{H}_2 \prec \overline{\boldsymbol{H}}_2^T\overline{\boldsymbol{H}}_2$. Finally, (A.49) and (A.50) establish the desired result. □

B Types and Typical Sequences

In this work we make extensively use of the concept of *types* and *typical sequences* which are briefly reviewed in the following. For a more comprehensive and detailed treatment we refer to standard books such as [Wol78, CK81, CT06, Kra08].

B.1 Types

In the following we briefly review the concept of *types* which is based on combinatorial properties of sequences.

Definition B.1. *The* type *(or* empirical distribution*) of a sequence* $x^n = (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ *of length* n *is the distribution* $P_{x^n} \in \mathcal{P}(\mathcal{X})$ *defined by*

$$P_{x^n}(a) \coloneqq \frac{N(a|x^n)}{n} \qquad \text{for every } \ a \in \mathcal{X}$$

where $N(a|x^n)$ *denotes the number of indices* i *such that* $x_i = a$*,* $i = 1, ..., n$*. The subset* $\mathcal{P}_0(n, \mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ *consists all possible types of sequences in* $\mathcal{X}^n$ *and is given by*

$$\mathcal{P}_0(n, \mathcal{X}) = \{P \in \mathcal{P}(\mathcal{X}) : P \text{ is type of sequences in } \mathcal{X}^n\}.$$

The notation of types extends to joint types in a natural way.

Definition B.2. *The* joint type *of sequences* $x^n = (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ *and* $y^n = (y_1, y_2, ..., y_n) \in \mathcal{Y}^n$ *of length* n *is the distribution* $P_{x^n,y^n} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ *defined by*

$$P_{x^n,y^n}(a, b) \coloneqq \frac{N(a, b|x^n, y^n)}{n} \qquad \text{for every } \ a \in \mathcal{X}, b \in \mathcal{Y}$$

where $N(a, b|x^n, y^n)$ *denotes the number of indices* i *such that* $(x_i, y_i) = (a, b)$*,* $i = 1, ..., n$*.*

Alternatively, joint types of sequences $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$ are often described by the type P_{x^n} of the sequence $x^n \in \mathcal{X}^n$ and a stochastic matrix $V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ such that

$$P_{x^n,y^n}(a, b) = P_{x^n}(a)V(b|a) \qquad \text{for every } \ a \in \mathcal{X}, b \in \mathcal{Y}.$$

Note that the stochastic matrix $V(b|a)$ is uniquely determined by the joint type P_{x^n,y^n} for all $a \in \mathcal{X}$ with $P_{x^n}(a) > 0$, i.e., for all $a \in \mathcal{X}$ which do occur in the sequence $x^n \in \mathcal{X}^n$, cf. [CK81, Sec. 1.2]. This leads to the following definition of conditional types.

Definition B.3. *The sequence $y^n \in \mathcal{Y}^n$ has* conditional type $V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *for given $x^n \in \mathcal{X}^n$ if*

$$N(a, b|x^n, y^n) = N(a|x^n)V(b|a) \qquad \text{for every } a \in \mathcal{X}, b \in \mathcal{Y}.$$

We follow [CK81] and represent types of sequences of length n by distributions of dummy random variables. For instance, the random variables X and Y represent a joint type, i.e., $P_{\mathrm{XY}} = P_{x^n,y^n}$ for some $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$.

The set of all sequences of type P_{x^n} is denoted by

$$\mathcal{T}^{(n)}_{\mathrm{X}} = \{x^n : x^n \in \mathcal{X}^n, P_{x^n} = P_{\mathrm{X}}\}.$$

Of course, this notation extends to joint types and sections in a self-explanatory way, e.g.,

$$\begin{aligned}\mathcal{T}^{(n)}_{\mathrm{XY}} &= \{(x^n, y^n) : x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n, P_{x^n,y^n} = P_{\mathrm{XY}}\}\\ \mathcal{T}^{(n)}_{\mathrm{Y|X}}(x^n) &= \{y^n : (x^n, y^n) \in \mathcal{T}^{(n)}_{\mathrm{XY}}\}.\end{aligned}$$

Remark B.4. *For clarity of presentation it is often useful to write $\mathcal{T}^{(n)}_{\mathrm{X}} = \mathcal{T}^{(n)}_{P_{\mathrm{X}}}$, $\mathcal{T}^{(n)}_{\mathrm{XY}} = \mathcal{T}^{(n)}_{P_{\mathrm{XY}}}$, and $\mathcal{T}^{(n)}_{\mathrm{Y|X}}(x^n) = \mathcal{T}^{(n)}_{V}(x^n)$ interchangeably to emphasize the dependency on the corresponding types. The last term created the name V-*shell*, cf. [CK81, Sec. 1.2].*

Next, we state some elementary properties of types; for more details we refer for example to [CK81, Sec. 1.2] or [Csi98].

Lemma B.5. *The number of different types of sequences in $\mathcal{X}^n$ is bounded by*

$$|\mathcal{P}_0(n, \mathcal{X})| \leq (n+1)^{|\mathcal{X}|},$$

i.e., it is a polynomial in n.

Proof. See for example [CK81, Lemma 2.2]. □

Lemma B.6. *For any sequence $x^n \in \mathcal{X}^n$, type $P_{\mathrm{X}} \in \mathcal{P}_0(n, \mathcal{X})$, and stochastic matrix $V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ such that $\mathcal{T}^{(n)}_{\mathrm{Y|X}}(x^n) \neq \emptyset$, we have*

$$\begin{aligned}(n+1)^{-|\mathcal{X}|}2^{nH(\mathrm{X})} &\leq |\mathcal{T}^{(n)}_{\mathrm{X}}| \leq 2^{nH(\mathrm{X})}\\ (n+1)^{-|\mathcal{X}||\mathcal{Y}|}2^{nH(\mathrm{Y|X})} &\leq |\mathcal{T}^{(n)}_{\mathrm{Y|X}}(x^n)| \leq 2^{nH(\mathrm{Y|X})}.\end{aligned}$$

Proof. See for example [CK81, Lemma 2.3 and 2.5]. □

For the next lemma we need the following properties of the mutual information and information divergence. Therefore, let X, Y and X', Y' be two pairs of random variables on $\mathcal{X} \times \mathcal{Y}$ with probability distributions $p_{\mathrm{XY}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $p_{\mathrm{X'Y'}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, respectively. Then we have the following relations

$$D(p_{\mathrm{XY}} \| p_{\mathrm{X'Y'}}) = D(p_{\mathrm{X}} \| p_{\mathrm{X'}}) + D(p_{\mathrm{Y|X}} \| p_{\mathrm{Y'|X'}} | p_{\mathrm{X}}) \tag{B.1}$$

$$I(\mathrm{X};\mathrm{Y}) = D(p_{\mathrm{XY}} \| p_{\mathrm{X}} \otimes p_{\mathrm{Y}}) = D(p_{\mathrm{Y|X}} \| p_{\mathrm{Y}} | p_{\mathrm{X}}). \tag{B.2}$$

Lemma B.7. *For any sequence $x^n \in \mathcal{X}^n$, type $P_{\mathrm{X}} \in \mathcal{P}_0(n, \mathcal{X})$, and stochastic matrix $V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ such that $\mathcal{T}^{(n)}_{\mathrm{Y|X}}(x^n) \neq \emptyset$, we have*

$$V^{\otimes n}\big(\mathcal{T}^{(n)}_{\mathrm{Y|X}}(x^n) | x^n\big) = \sum_{y^n \in \mathcal{T}^{(n)}_{\mathrm{Y|X}}(x^n)} V^{\otimes n}(y^n|x^n) \leq 2^{-nD(P_{\mathrm{XY}} \| P_{\mathrm{X}} \otimes V)} \tag{B.3}$$

where $V^{\otimes n}(y^n|x^n) \coloneqq \prod_{k=1}^n V(y_k|x_k)$ and $P_{\mathrm{X}} \otimes V$ denotes the distribution on $\mathcal{X} \times \mathcal{Y}$ with probability mass function $P_{\mathrm{X}}(x)V(y|x)$.

Further for some given $s^n \in \mathcal{S}^n$,

$$V^{\otimes n}(\mathcal{T}^{(n)}_{\mathrm{Y|XS}}(x^n, s^n) | x^n) = \sum_{y^n \in \mathcal{T}^{(n)}_{\mathrm{Y|XS}}(x^n, s^n)} V^{\otimes n}(y^n|x^n) \leq 2^{-nI(\mathrm{Y};\mathrm{S|X})}. \tag{B.4}$$

Proof. For (B.3) confer for example [CK81, Lemma 2.6]. Relation (B.4) follows immediately from (B.1) and (B.2) by

$$\begin{aligned} D(P_{\mathrm{XSY}} \| P_{\mathrm{XS}} \otimes V) &= D(P_{\mathrm{Y|XS}} \| V | P_{\mathrm{XS}}) \\ &= I(\mathrm{Y};\mathrm{S|X}) + D(P_{\mathrm{Y|X}} \| V | P_{\mathrm{X}}) \\ &\geq I(\mathrm{Y};\mathrm{S|X}), \end{aligned}$$

cf. also [Hug97]. □

B.2 Typical Sequences

The notion of *typical sequences* was originally introduced by Shannon in his seminal work "A Mathematical Theory of Communication" [Sha48]. But he did it in a more intuitive rather than in a precise and technical sense.

There are different approaches to define typicality. One definition is based on the entropy of a random variable. Such sequences are known as *weakly typical* or *entropy-typical* sequences. A more natural definition of typicality is based on the empirical distribution of the sequence. Such sequences are known as *strongly typical* or *letter-typical* sequences. The latter concept can give stronger results but has the drawback that it can only be applied to discrete random variables, while the former also works for continuous random variables.

These two concepts are briefly reviewed in the following.

B.2.1 Strong Typicality

Definition B.8. *For any distribution $p \in \mathcal{P}(\mathcal{X})$ a sequence $x^n \in \mathcal{X}^n$ is said to be* typical *(or* strongly typical *or* p-typical*) with constant ϵ if*

$$\left|\frac{1}{n}N(a|x^n) - p(a)\right| \leq \epsilon \qquad \text{for every } \; a \in \mathcal{X}$$

and, in addition, $N(a|x^n) = 0$ if $p(a) = 0$. The set of all such typical sequences is denoted by $\mathcal{T}_{p,\epsilon}^{(n)}$.

Definition B.9. *For a stochastic matrix $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ a sequence $y^n \in \mathcal{Y}^n$ is called* W-typical under the condition $x^n \in \mathcal{X}^n$ *(or* W-generated by the sequence $x^n \in \mathcal{X}^n$*) if*

$$\left|\frac{1}{n}N(a,b|x^n,y^n) - \frac{1}{n}N(a|x^n)W(b|a)\right| \leq \epsilon \qquad \text{for every } \; b \in \mathcal{Y}$$

and, in addition, $N(a,b|x^n,y^n) = 0$ if $W(b|a) = 0$. The set of all such sequences is denoted by $\mathcal{T}_{W,\epsilon}^{(n)}(x^n)$.

Moreover, we need the following lemmas which give us exponential rates of convergence for typical sequences.

Lemma B.10. *For every $\epsilon > 0$ and every $p \in \mathcal{P}(\mathcal{X})$ the following*

$$p^{\otimes n}(\mathcal{T}_{p,\epsilon}^{(n)}) \geq 1 - (n+1)^{|\mathcal{X}|}2^{-nc\epsilon^2} \tag{B.5}$$

with $c = \frac{1}{2\ln 2}$ always holds for all $n \in \mathbb{N}$.

Proof. The proof can be found in [Shi96, Lemma III.1.3] but it is given in the following for completeness. From [CK81, Lemma 2.6] we have $p^{\otimes n}(\mathcal{T}_p^{(n)}) \leq 2^{-nD(p_{x^n}\|p)}$. The bad set $(\mathcal{T}_{p,\epsilon}^{(n)})^c \coloneqq \{x^n \in \mathcal{X}^n : \exists a \in \mathcal{X} : |\frac{N(a|x^n)}{n} - p(a)| > \epsilon\}$ can be partitioned into disjoint sets

of the form $(\mathcal{T}_{p,\epsilon}^{(n)})^c \cap \mathcal{T}_p^{(n)}$. Since there are at most $(n+1)^{|\mathcal{X}|}$ type classes, cf. Lemma B.5, we get

$$p^{\otimes n}((\mathcal{T}_{p,\epsilon}^{(n)})^c) \leq (n+1)^{|\mathcal{X}|} 2^{-nD_*}$$

where $D_* = \min\{D(p_{x^n} \| p) : \exists a \in \mathcal{X} : |\frac{N(a|x^n)}{n} - p(a)| > \epsilon\}$. From Pinsker's inequality [Pin64] we have $D(p_{x^n} \| p) \geq \frac{1}{2\ln 2}(\sum_a |p_{x^n}(a) - p(a)|)^2$ so that $D_* \geq \frac{1}{2\ln 2}|p_{x^n}(a) - p(a)|^2 = \frac{1}{2\ln 2}\epsilon^2$ for all $a \in \mathcal{X}$. From this we have

$$p^{\otimes n}((\mathcal{T}_{p,\epsilon}^{(n)})^c) \leq (n+1)^{|\mathcal{X}|} 2^{-nc\epsilon^2}$$

with $c = \frac{1}{2\ln 2}$ which proves the lemma. □

Lemma B.11. *For every $\epsilon > 0$ and every $x^n \in \mathcal{X}^n$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ the following*

$$W^{\otimes n}(\mathcal{T}_{W,\epsilon}^{(n)}(x^n)|x^n) \geq 1 - (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-nc\epsilon^2} \tag{B.6}$$

with $c = \frac{1}{2\ln 2}$ always holds for all $n \in \mathbb{N}$.

Proof. The proof follows [Shi96, Lemma III.1.3]. We define $V_{y^n|x^n}(b|a) = \frac{p_{x^n,y^n}(a,b)}{p_{x^n}(a)}$ for empirical distributions $p_{x^n}(a) = \frac{1}{n}N(a|x^n)$ and $p_{x^n,y^n}(a,b) = \frac{1}{n}N(a,b|x^n,y^n)$ for all $a \in \mathcal{X}$, $b \in \mathcal{Y}$. Then from [CK81, Lemma 2.6] we have $W^{\otimes n}(\mathcal{T}_W^{(n)}(x^n)|x^n) \leq 2^{-nD(V_{y^n|x^n}\|W|p_{x^n})}$ with

$$\begin{aligned} D(V_{y^n|x^n}\|W|p_{x^n}) &= \sum_{(a,b)\in\mathcal{X}\times\mathcal{Y}} p_{x^n}(a) V_{y^n|x^n}(b|a) \log \frac{V_{y^n|x^n}(b|a)}{W(b|a)} \\ &= \sum_{(a,b)\in\mathcal{X}\times\mathcal{Y}} p_{x^n,y^n}(a,b) \log \frac{p_{x^n,y^n}(a,b)}{W(b|a)p_{x^n}(a)} \\ &= D(p_{x^n,y^n}\|p_{x^n}W). \end{aligned} \tag{B.7}$$

Observe that $\mathcal{T}_W^{(n)}(x^n)$ has at most $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$ V-shells, cf. also [CK81, Sec. 1.2], so that

$$W^{\otimes n}((\mathcal{T}_{W,\epsilon}^{(n)})^c(x^n)|x^n) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-nD_*} \tag{B.8}$$

with $D_* = \min\{D(p_{x^n,y^n}\|p_{x^n}W) : \exists (a,b) \in \mathcal{X}\times\mathcal{Y} : |p_{x^n,y^n}(a,b) - p_{x^n}(a)W(b|a)| > \epsilon\}$. From [CK81] we have[1] $D(p_{x^n,y^n}\|p_{x^n}W) \geq \frac{1}{2\ln 2}(\sum_{(a,b)\in\mathcal{X}\times\mathcal{Y}} |p_{x^n,y^n}(a,b) - p_{x^n}(a)W(b|a)|)^2$ so that $D_* \geq \frac{1}{2\ln 2}|p_{x^n,y^n}(a,b) - p_{x^n}(a)W(b|a)|^2 = \frac{\epsilon^2}{2\ln 2}$ for all $a \in \mathcal{X}$, $b \in \mathcal{Y}$. From this we have

$$W^{\otimes n}((\mathcal{T}_{W,\epsilon}^{(n)})^c(x^n)|x^n) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-nc\delta^2} \tag{B.9}$$

with $c = \frac{1}{2\ln 2}$ which proves the lemma. □

[1] This bound with a worse constant was first given by Pinsker [Pin64] and is therefore also known as Pinsker's inequality.

The next lemma relates typical sequences generated by different (input) distributions to the same output distribution. It plays a crucial role especially in the proof of the optimal coding strategy for compound channels with CSIT as done in Section 3.4.2.

Lemma B.12. *Let $p, \tilde{p} \in \mathcal{P}(\mathcal{X})$ be input distributions, $W, \widetilde{W} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $q, \tilde{q} \in \mathcal{P}(\mathcal{Y})$ the corresponding output distributions. Further, let $\epsilon \in (0, \frac{1}{4|\mathcal{X}||\mathcal{Y}|})$. Then for every $n \in \mathbb{N}$ and all $\tilde{x}^n \in \mathcal{T}^{(n)}_{\tilde{p},\epsilon}$ and $x^n \in \mathcal{T}^{(n)}_{p,\epsilon}$ it holds*

$$q^{\otimes n}(\mathcal{T}^{(n)}_{\widetilde{W},\epsilon}(\tilde{x}^n)) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-n(I(\tilde{p},\widetilde{W})-\varphi(\epsilon)-\psi(\epsilon))} \tag{B.10a}$$

$$q^{\otimes n}(\mathcal{T}^{(n)}_{W,\epsilon}(x^n)) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-n(I(p,W)-\varphi(\epsilon)-\psi(\epsilon))} \tag{B.10b}$$

with universal $\varphi(\epsilon), \psi(\epsilon) > 0$ and $\lim_{\epsilon \searrow 0} \varphi(\epsilon) = 0 = \lim_{\epsilon \searrow 0} \psi(\epsilon)$.

Proof. Let $y^n \in \mathcal{T}^{(n)}_{\widetilde{W},\epsilon}(\tilde{x}^n)$ where $\tilde{x}^n \in \mathcal{T}^{(n)}_{\tilde{p},\epsilon}$. Then we know from [CK81, Lemma 2.6] that $q^{\otimes n}(y^n) = 2^{-n(D(p_{y^n}\|q)+H(p_{y^n}))}$ with $p_{y^n}(b) = \frac{N(b|y^n)}{n}$. Since $D(p_{y^n}\|q) \geq 0$ we have

$$q^{\otimes n}(y^n) \leq 2^{-nH(p_{y^n})}. \tag{B.11}$$

Since $y^n \in \mathcal{T}^{(n)}_{\widetilde{W},\epsilon}(\tilde{x}^n)$ and $\tilde{x}^n \in \mathcal{T}^{(n)}_{\tilde{p},\epsilon}$, it follows from [CK81, Lemma 2.10] that $y^n \in \mathcal{T}^{(n)}_{\tilde{q},2|\mathcal{X}|\epsilon}$ and therewith $\sum_{b\in\mathcal{Y}} |p_{y^n}(b) - \tilde{q}(b)| \leq 2|\mathcal{X}||\mathcal{Y}|\epsilon < 1/2$ so that [CK81, Lemma 2.7] implies

$$|H(p_{y^n}) - H(\tilde{q})| \leq -2|\mathcal{X}||\mathcal{Y}|\epsilon \log \frac{2|\mathcal{X}||\mathcal{Y}|\epsilon}{|\mathcal{Y}|} =: \varphi(\epsilon) \tag{B.12}$$

with $\lim_{\epsilon \searrow 0} \varphi(\epsilon) = 0$. If we combine (B.11) and (B.12), we obtain $q^{\otimes n}(y^n) \leq 2^{-n(H(\tilde{q})-\varphi(\epsilon))}$ and therewith

$$q^{\otimes n}(\mathcal{T}^{(n)}_{\widetilde{W},\epsilon}(\tilde{x}^n)) \leq |\mathcal{T}^{(n)}_{\widetilde{W},\epsilon}(\tilde{x}^n)| 2^{-n(H(\tilde{q})-\varphi(\epsilon))}. \tag{B.13}$$

Since $\tilde{x}^n \in \mathcal{T}^{(n)}_{\tilde{p},\epsilon}$, it follows from [CK81, Lemma 2.13] that $|\mathcal{T}^{(n)}_{\widetilde{W},\epsilon}(\tilde{x}^n)| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{n(H(\widetilde{W}|\tilde{p})+\psi(\epsilon))}$ holds with universal $\psi(\epsilon) > 0$ and $\lim_{\epsilon \searrow 0} \psi(\epsilon) = 0$. Inserting this in (B.13) we obtain

$$q^{\otimes n}(\mathcal{T}^{(n)}_{\widetilde{W},\epsilon}(\tilde{x}^n)) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-n(I(\tilde{p},\widetilde{W})-\varphi(\epsilon)-\psi(\epsilon))}. \tag{B.14}$$

The relation $q^{\otimes n}(\mathcal{T}^{(n)}_{W,\epsilon}(x^n)) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-n(I(p,W)-\varphi(\epsilon)-\psi(\epsilon))}$ follows immediately for $W = \widetilde{W}$ and $p = \tilde{p}$ which proves the lemma. □

B.2.2 Weak Typicality

Definition B.13. *For any distribution $p_{\mathrm{X}} \in \mathcal{P}(\mathcal{X})$ a sequence $x^n \in \mathcal{X}^n$ is said to be* weakly typical *(or* entropy-typical*) with constant ϵ if*

$$\left| -\frac{1}{n} \log p_{\mathrm{X}}^{\otimes n}(x^n) - H(\mathrm{X}) \right| < \epsilon.$$

The set of all such weakly typical sequences is denoted by $\mathcal{A}_\epsilon^{(n)}(\mathrm{X})$.

This concept extends to jointly weakly typical sequences as follows. As in [CT06, Section 15.2] let $\mathrm{S} \subseteq \{\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_k\}$ be an ordered subset of the random variables.

Definition B.14. *For any distribution $p_{\mathrm{X}_1^n \mathrm{X}_2^n ... \mathrm{X}_k^n}^{\otimes n}(x_1^n, x_2^n, ..., x_k^n)$ sequences $(x_1^n, x_2^n, ..., x_k^n)$ are said to be* jointly weakly typical *with constant ϵ if*

$$\left| -\frac{1}{n} \log p_{\mathrm{S}^n}^{\otimes n}(s^n) - H(\mathrm{S}) \right| < \epsilon, \quad \forall\, \mathrm{S} \subseteq \{\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_k\}.$$

The set of all such weakly typical sequences is denoted by $\mathcal{A}_\epsilon^{(n)}(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_k)$.

For example the set $\mathcal{A}_\epsilon^{(n)}(\mathrm{X}_1, \mathrm{X}_2)$ is given by all sequences (x_1^n, x_2^n) that satisfy

$$\left| -\frac{1}{n} \log p_{\mathrm{X}_1^n \mathrm{X}_2^n}^{\otimes n}(x_1^n, x_2^n) - H(\mathrm{X}_1, \mathrm{X}_2) \right| < \epsilon$$
$$\left| -\frac{1}{n} \log p_{\mathrm{X}_1^n}^{\otimes n}(x_1^n) - H(\mathrm{X}_1) \right| < \epsilon$$
$$\left| -\frac{1}{n} \log p_{\mathrm{X}_2^n}^{\otimes n}(x_2^n) - H(\mathrm{X}_2) \right| < \epsilon.$$

Let $\mathrm{S}, \mathrm{S}_1, \mathrm{S}_2 \subseteq \{\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_k\}$. Then we have the following properties for weakly typical sequences.

Lemma B.15. *For any $\epsilon > 0$ and sufficiently large n we have*

(a) $\mathbb{P}\{s^n \notin \mathcal{A}_\epsilon^{(n)}(\mathrm{S})\} \to 0$ *as* $n \to \infty$

(b) *If* $s^n \in \mathcal{A}_\epsilon^{(n)}(\mathrm{S})$ *then* $p_{\mathrm{S}^n}^{\otimes n}(s^n) \leq 2^{-n(H(\mathrm{S})-\epsilon)}$

(c) $|\mathcal{A}_\epsilon^{(n)}(\mathrm{S})| \leq 2^{n(H(\mathrm{S})+\epsilon)}$

(d) *If* $(s_1^n, s_2^n) \in \mathcal{A}_\epsilon^{(n)}(\mathrm{S}_1, \mathrm{S}_2)$ *then* $p_{\mathrm{S}_1^n|\mathrm{S}_2^n}^{\otimes n}(s_1^n|s_2^n) \leq 2^{-n(H(\mathrm{S}_1|\mathrm{S}_2)-\epsilon)}$.

Proof. See for example [CT06, Theorem 15.2.1]. □

Publication List

[BW10] Holger Boche and Rafael F. Wyrembelski. Degrees of Coordination in Cognitive Networks. In *Proc. Conf. on Cognitive Radio Oriented Wireless Networks and Commun.*, pages 1–5, Cannes, France, June 2010.

[OJWB09] Tobias J. Oechtering, Eduard A. Jorswieck, Rafael F. Wyrembelski, and Holger Boche. On the Optimal Transmit Strategy for the MIMO Bidirectional Broadcast Channel. *IEEE Trans. Commun.*, 57(12):3817–3826, December 2009.

[OWB08a] Tobias J. Oechtering, Rafael F. Wyrembelski, and Holger Boche. Optimal Time-Division of Two-Phase Decode-and-Forward Bidirectional Relaying. In *Proc. Int. Symp. Inf. Theory Applications*, pages 829–834, Auckland, New Zealand, December 2008.

[OWB08b] Tobias J. Oechtering, Rafael F. Wyrembelski, and Holger Boche. Optimal Transmit Strategy for the 2x1 MISO Bidirectional Broadcast Channel. In *Proc. IEEE Signal Process. Adv. Wireless Commun.*, pages 316–320, Recife, Brazil, July 2008.

[OWB09a] Tobias J. Oechtering, Rafael F. Wyrembelski, and Holger Boche. Multiantenna Bidirectional Broadcast Channels - Optimal Transmit Strategies. *IEEE Trans. Signal Process.*, 57(5):1948–1958, May 2009.

[OWB09b] Tobias J. Oechtering, Rafael F. Wyrembelski, and Holger Boche. On the Optimal Transmission for the MIMO Bidirectional Broadcast Channel. In *Proc. IEEE Int. Conf. Commun.*, pages 1–5, Dresden, Germany, June 2009.

[WB11a] Rafael F. Wyrembelski and Holger Boche. Bidirectional Broadcast Channels with Common and Confidential Messages. In *Proc. IEEE Inf. Theory Workshop*, pages 713–717, Paraty, Brazil, October 2011.

[WB11b] Rafael F. Wyrembelski and Holger Boche. How to Achieve Privacy in Bidirectional Relay Networks. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1891–1895, Saint Petersburg, Russia, July 2011.

[WB11c] Rafael F. Wyrembelski and Holger Boche. Physical Layer Service Integration in Bidirectional Relay Networks. *IEEE Trans. Wireless Commun.*, 2011. submitted.

[WB11d] Rafael F. Wyrembelski and Holger Boche. Secrecy in MIMO Gaussian Bidirectional Broadcast Channels. In *Proc. IEEE Signal Process. Adv. Wireless Commun.*, pages 361–365, San Francisco, CA, USA, June 2011.

[WB11e] Rafael F. Wyrembelski and Holger Boche. Service Integration in Multiantenna Bidirectional Relay Networks: Public and Confidential Services. In *Proc. IEEE Global Commun. Conf. Workshops*, pages 884–888, Houston, TX, USA, December 2011.

[WB12a] Rafael F. Wyrembelski and Holger Boche. Privacy in Bidirectional Relay Networks. *IEEE Trans. Commun.*, 2012. accepted.

[WB12b] Rafael F. Wyrembelski and Holger Boche. Strong Secrecy in Compound Broadcast Channels with Confidential Messages. In *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, July 2012. accepted.

[WBB09a] Rafael F. Wyrembelski, Igor Bjelaković, and Holger Boche. Coding Strategies for Bidirectional Relaying for Arbitrarily Varying Channels. In *Proc. IEEE Global Commun. Conf.*, pages 1–6, Honolulu, HI, USA, December 2009.

[WBB09b] Rafael F. Wyrembelski, Igor Bjelaković, and Holger Boche. On the Capacity of Bidirectional Relaying with Unknown Varying Channels. In *Proc. IEEE Workshop Comp. Adv. Multi-Sensor Adaptive Processing*, pages 269–272, Aruba, Dutch Antilles, December 2009.

[WBB10a] Rafael F. Wyrembelski, Igor Bjelaković, and Holger Boche. Bidirectional Relaying in Wireless Networks - Impact of Degree of Coordination. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, pages 3234–3237, Dallas, TX, USA, March 2010.

[WBB10b] Rafael F. Wyrembelski, Igor Bjelaković, and Holger Boche. List Decoding for Bidirectional Broadcast Channels with Unknown Varying Channels. In *Proc. IEEE Int. Conf. Commun.*, pages 1–6, Cape Town, South Africa, May 2010.

[WBB10c] Rafael F. Wyrembelski, Igor Bjelaković, and Holger Boche. On Arbitrarily Varying Bidirectional Broadcast Channels with Constraints on Input and States. In *Proc. Int. Symp. Inf. Theory Applications*, pages 410–415, Taichung, Taiwan, October 2010.

[WBB11] Rafael F. Wyrembelski, Igor Bjelaković, and Holger Boche. Arbitrarily Varying Bidirectional Broadcast Channels under List Decoding. *IEEE Trans. Inf. Theory*, 2011. submitted.

[WBB12] Rafael F. Wyrembelski, Igor Bjelaković, and Holger Boche. Arbitrarily Varying Channels - A Model for Uncoordinated Wireless Networks with Applications to Bidirectional Relaying. *Entropy*, 2012. submitted.

[WBOB09] Rafael F. Wyrembelski, Igor Bjelaković, Tobias J. Oechtering, and Holger Boche. On the Capacity of Bidirectional Broadcast Channels under Channel Uncertainty. In *Proc. IEEE Int. Conf. Commun.*, pages 1–5, Dresden, Germany, June 2009.

[WBOB10] Rafael F. Wyrembelski, Igor Bjelaković, Tobias J. Oechtering, and Holger Boche. Optimal Coding Strategies for Bidirectional Broadcast Channels under Channel Uncertainty. *IEEE Trans. Commun.*, 58(10):2984–2994, October 2010.

[WOB+08a] Rafael F. Wyrembelski, Tobias J. Oechtering, Igor Bjelaković, Clemens Schnurr, and Holger Boche. Capacity of Gaussian MIMO Bidirectional Broadcast Channels. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 584–588, Toronto, Canada, July 2008.

[WOB08b] Rafael F. Wyrembelski, Tobias J. Oechtering, and Holger Boche. Decode-and-Forward Strategies for Bidirectional Relaying. In *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun.*, pages 1–6, Cannes, France, September 2008.

[WOB08c] Rafael F. Wyrembelski, Tobias J. Oechtering, and Holger Boche. The Asymptotic Performance of Three-Hop Relay Communication using Amplify-and-Forward. In *Proc. Int. ITG Conf. Source and Channel Coding*, Ulm, Germany, January 2008.

[WOB10] Rafael F. Wyrembelski, Tobias J. Oechtering, and Holger Boche. MIMO Bidirectional Broadcast Channels with Common Message. In *Proc. IEEE Global Commun. Conf.*, pages 1–5, Miami, FL, USA, December 2010.

[WOB11] Rafael F. Wyrembelski, Tobias J. Oechtering, and Holger Boche. MIMO Gaussian Bidirectional Broadcast Channels with Common Messages. *IEEE Trans. Wireless Commun.*, 10(9):2950–2959, September 2011.

[WOBS12] Rafael F. Wyrembelski, Tobias J. Oechtering, Holger Boche, and Mikael Skoglund. Robust Transmit Strategies for Multiantenna Bidirectional Broadcast Channels. In *Proc. ITG Workshop Smart Antennas*, pages 46–53, Dresden, Germany, March 2012.

[WSB11] Rafael F. Wyrembelski, Aydin Sezgin, and Holger Boche. Secrecy in Broadcast Channels with Receiver Side Information. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 290–294, Pacific Grove, CA, USA, November 2011.

[WWB11] Rafael F. Wyrembelski, Moritz Wiese, and Holger Boche. Strong Secrecy in Bidirectional Relay Networks. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 217–221, Pacific Grove, CA, USA, November 2011. invited.

References

[AC99] Rudolf Ahlswede and Ning Cai. Arbitrarily Varying Multiple-Access Channels Part I – Ericson's Symmetrizability Is Adequate, Gubner's Conjecture Is True. *IEEE Trans. Inf. Theory*, 45(2):742–749, March 1999.

[ACLY00] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network Information Flow. *IEEE Trans. Inf. Theory*, 46(4):1204–1216, July 2000.

[AGK76] Rudolf Ahlswede, Péter Gács, and János Körner. Bounds on Conditional Probabilities with Applications in Multi-User Communication. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 34:157–177, 1976.

[AH94] Robert J. Aumann and Sergiu Hart. *Handbook of Game Theory with Economic Applications Volume 2*. Elsevier Science & Technology, 1994.

[Ahl71] Rudolf Ahlswede. Multi-Way Communication Channels. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 23–52, Tsahkadsor, Armenian, September 1971.

[Ahl74] Rudolf Ahlswede. The Capacity Region of a Channel with Two Senders and Two Receivers. *Annals Probability*, 2(5):805–814, October 1974.

[Ahl78] Rudolf Ahlswede. Elimination of Correlation in Random Codes for Arbitrarily Varying Channels. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 44:159–175, 1978.

[Ahl80a] Rudolf Ahlswede. A Method of Coding and an Application to Arbitrarily Varying Channels. *J. Comb. Inform. Syst. Sci.*, 5(5):10–35, 1980.

[Ahl80b] Rudolf Ahlswede. Coloring Hypergraphs: A New Approach to Multi-User Source Coding–II. *J. Comb. Inform. Syst. Sci.*, 5(3):220–268, 1980.

[Ahl86] Rudolf Ahlswede. Arbitrarily Varying Channels with States Sequence Known to the Sender. *IEEE Trans. Inf. Theory*, 32(5):621–629, September 1986.

[ASS10] Sara Al-Sayed and Aydin Sezgin. Secrecy in Gaussian MIMO Bidirectional Broadcast Wiretap Channels: Transmit Strategies. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 285–289, Pacific Grove, CA, USA, November 2010.

[AST10] A. Salman Avestimehr, Aydin Sezgin, and David N. C. Tse. Capacity of the Two-Way Relay Channel Within a Constant Gap. *European Trans. Telecommun.*, 21(4):363–374, June 2010.

[AW69] Rudolf Ahlswede and Jacob Wolfowitz. The Structure of Capacity Functions for Compound Channels. In *Proc. Int. Symp. on Prob. and Inf. Theory*, pages 12–54, McMaster University, Canada, April 1969.

[BB08] João Barros and Matthieu Bloch. Strong Secrecy for Wireless Channels. In *Int. Conf. on Information-Theoretic Security*, pages 40–53, Calgary, Canada, August 2008. invited.

[BB11] Matthieu Bloch and João Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[BBRM08] Matthieu Bloch, João Barros, Miguel R. D. Rodrigues, and Steven W. McLaughlin. Wireless Information-Theoretic Security. *IEEE Trans. Inf. Theory*, 54(6):2515–2534, June 2008.

[BBS11a] Igor Bjelaković, Holger Boche, and Jochen Sommerfeld. Capacity Results for Compound Wiretap Channels. In *Proc. IEEE Inf. Theory Workshop*, pages 60–64, Paraty, Brazil, October 2011.

[BBS11b] Igor Bjelaković, Holger Boche, and Jochen Sommerfeld. Secrecy Results for Compound Wiretap Channels. submitted 2011. available at http://arxiv.org/abs/1106.2013.

[BBT59] David Blackwell, Leo Breiman, and A. J. Thomasian. The Capacity of a Class of Channels. *Ann. Math. Stat.*, 30(4):1229–1241, December 1959.

[BBT60] David Blackwell, Leo Breiman, and A. J. Thomasian. The Capacities of Certain Channel Classes under Random Coding. *Ann. Math. Stat.*, 31(3):558–567, 1960.

[BC07] Ihn-Jung Baik and Sae-Young Chung. Network Coding for Two-Way Relay Channels using Lattices. *Telecommunications review*, 17(17):1009–1021, 2007.

[BCC+07] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, and H. Vincent Poor. *MIMO Wireless Communications*. Cambridge University Press, 2007.

[BNP95] Vladimir Blinovsky, Prakash Narayan, and M. Pinsker. Capacity of the Arbitrarily Varying Channel Under List Decoding. *Probl. Pered. Inform.*, 31(2):99–113, 1995.

[BO98] Tamer Basar and Geert J. Olsder. *Dynamic Noncooperative Game Theory*. Classics In Applied Mathematics. SIAM, 2 edition, 1998.

[BV04] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[CE79] Thomas M. Cover and Abbas El Gamal. Capacity Theorems for the Relay Channel. *IEEE Trans. Inf. Theory*, 25(5):572–584, September 1979.

[CHK09] Tao Cui, Tracey Ho, and Jörg Kliewer. Memoryless Relay Strategies for Two-Way Relay Channels. *IEEE Trans. Commun.*, 57(10):3132–3143, October 2009.

[CK78] Imre Csiszár and János Körner. Broadcast Channels with Confidential Messages. *IEEE Trans. Inf. Theory*, 24(3):339–348, May 1978.

[CK81] Imre Csiszár and János Körner. *Information Theory - Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1 edition, 1981.

[CN88a] Imre Csiszár and Prakash Narayan. Arbitrarily Varying Channels with Constrained Inputs and States. *IEEE Trans. Inf. Theory*, 34(1):27–34, January 1988.

[CN88b] Imre Csiszár and Prakash Narayan. The Capacity of the Arbitrarily Varying Channel Revisited: Positivity, Constraints. *IEEE Trans. Inf. Theory*, 34(2):181–193, March 1988.

[CN91] Imre Csiszár and Prakash Narayan. Capacity of the Gaussian Arbitrarily Varying Channel. *IEEE Trans. Inf. Theory*, 37(1):18–26, January 1991.

[Cov72] Thomas M. Cover. Broadcast Channels. *IEEE Trans. Inf. Theory*, 18(1):2–14, January 1972.

[Csi98] Imre Csiszár. The Method of Types. *IEEE Trans. Inf. Theory*, 44(6):2505–2523, October 1998.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley & Sons, 2 edition, 2006.

[CY09] Min Chen and Aylin Yener. Multiuser Two-Way Relaying: Detection and Interference Management Strategies. *IEEE Trans. Wireless Commun.*, 8(8):4296–4305, August 2009.

[DS75] Roland L. Dobrushin and S. Z. Stambler. Coding Theorems for Classes of Arbitrarily Varying Discrete Memoryless Channels. *Probl. Pered. Inform.*, 11(2):3–22, 1975. (Englisch translation).

[EK11] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2011.

[EKYE10] Aly El Gamal, O. Ozan Koyluoglu, Moustafa Youssef, and Hesham El Gamal. New Achievable Secrecy Rate Regions for the Two Way Wiretap Channel. In *Proc. IEEE Inf. Theory Workshop*, pages 1–5, Cairo, Egypt, January 2010.

[Eli57] Peter Elias. List Decoding for Noisy Channels. *IRE WESCON Conv. Rec.*, 2:94–104, 1957.

[Eri85] Thomas Ericson. Exponential Error Bounds for Random Codes in the Arbitrarily Varying Channel. *IEEE Trans. Inf. Theory*, 31(1):42–48, January 1985.

[EU08a] Ersen Ekrem and Sennur Ulukus. On the Secrecy of Multiple Access Wiretap Channel. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 1014–1021, Urbana-Champaign, IL, USA, September 2008.

[EU08b] Ersen Ekrem and Sennur Ulukus. Secrecy in Cooperative Relay Broadcast Channels. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2217–2221, Toronto, Canada, July 2008.

[EU10a] Ersen Ekrem and Sennur Ulukus. Gaussian MIMO Broadcast Channels with Common and Confidential Messages. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2583–2587, Austin, TX, USA, June 2010.

[EU10b] Ersen Ekrem and Sennur Ulukus. On Gaussian MIMO Compound Wiretap Channels. In *Proc. Conf. Inf. Sciences and Systems*, pages 1–6, Baltimore, MD, USA, March 2010.

[FS07] Christina Fragouli and Emina Soljanin. Network Coding Fundamentals / Network Coding Applications. *Foundations and Trends in Networking*, 2(1–2):1–269, 2007.

[Gal68] Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley & Sons, 1968.

[GEGP09] Deniz Gündüz, Elza Erkip, Andrea Goldsmith, and H. Vincent Poor. Source and Channel Coding for Correlated Sources Over Multiuser Channels. *IEEE Trans. Inf. Theory*, 55(9):3927–3944, September 2009.

[GGY11] Kiran T. Gowda, David Gesbert, and Erhan Yilmaz. Interference Mitigation in Femto-Macro Coexistence with Two-Way Relay Channel. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1599–1603, Saint Petersburg, Russia, August 2011.

[GH95] John A. Gubner and Brian L. Hughes. Nonconvexity of the Capacity Region of the Multiple-Access Arbitrarily Varying Channel Subject to Constraints. *IEEE Trans. Inf. Theory*, 41(1):3–13, January 1995.

[GJJV03] Andrea Goldsmith, Syed A. Jafar, Nihar Jindal, and Sriram Vishwanath. Capacity Limits of MIMO Channels. *IEEE J. Sel. Areas Commun.*, 21(5):684–702, June 2003.

[GTN08] Deniz Gündüz, Ertem Tuncel, and Jayanth Nayak. Rate Regions for the Separated Two-Way Relay Channel. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 1333–1340, September 2008.

[Gub90] John A. Gubner. On the Deterministic-Code Capacity of the Multiple-Access Arbitrarily Varying Channel. *IEEE Trans. Inf. Theory*, 36(2):262–275, March 1990.

[Gub91] John A. Gubner. State Constraints for the Multiple-Access Arbitrarily Varying Channel. *IEEE Trans. Inf. Theory*, 37(1):27–35, January 1991.

[Gub92] John A. Gubner. On the Capacity Region of the Discrete Additive Multiple-Access Arbitrarily Varying Channel. *IEEE Trans. Inf. Theory*, 38(4):1344–1347, July 1992.

[Han98] Te Sun Han. An Information-Spectrum Approach to Capacity Theorems for the General Multiple-Access Channel. *IEEE Trans. Inf. Theory*, 44(7):2773–2795, November 1998.

[Han03] Te Sun Han. *Information-Spectrum Methods in Information Theory*. Stochastic Modelling and Applied Probability. Springer, 2003.

[HB06] Eran Hof and Shraga I. Bross. On the Deterministic-Code Capacity of the Two-User Discrete Memoryless Arbitrarily Varying General Broadcast Channel With Degraded Message Sets. *IEEE Trans. Inf. Theory*, 52(11):5023–5044, November 2006.

[HGS09] Chin Keong Ho, Kiran T. Gowda, and Sumei Sun. A Generalized Two-way Relay Channel with Private Information for the Relay. In *Proc. IEEE Int. Conf. Commun.*, pages 1–6, Dresden, Germany, June 2009.

[HJ99] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1999.

[HKE+07] Ingmar Hammerström, Marc Kuhn, Celal Eşli, Jian Zhao, Armin Wittneben, and Gerhard Bauch. MIMO Two-Way Relaying with Transmit CSI at the Relay. In *Proc. IEEE Signal Process. Adv. Wireless Commun.*, pages 1–5, Helsinki, Finland, June 2007.

[HN87] Brian Hughes and Prakash Narayan. Gaussian Arbitrarily Varying Channels. *IEEE Trans. Inf. Theory*, 33(2):267–284, March 1987.

[Hug97] Brian L. Hughes. The Smalles List for the Arbitrarily Varying Channel. *IEEE Trans. Inf. Theory*, 43(3):803–815, May 1997.

[HUL01] Jean-Baptiste Hiriart-Urruty and Claude Lemarèchal. *Fundamentals of Convex Analysis*. Springer-Verlag, 2001.

[HY10a] Xiang He and Aylin Yener. A New Outer Bound for the Secrecy Capacity Region of the Gaussian Two-Way Wiretap Channel. In *Proc. IEEE Int. Conf. Commun.*, pages 1–5, Cape Town, South Africa, May 2010.

[HY10b] Xiang He and Aylin Yener. Cooperation with an Untrusted Relay: A Secrecy Perspective. *IEEE Trans. Inf. Theory*, 56(8):3807–3827, August 2010.

[ILH10] Onurcan Iscan, Imran Latif, and Christoph Hausl. Network Coded Multi-way Relaying with Iterative Decoding. In *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun.*, pages 482–487, Istanbul, Turkey, September 2010.

[IS09] Fabio Iannello and Osvaldo Simeone. On the Throughput Region of Single and Two-Way Multi-Hop Fading Networks with Relay Piggybacking. In *Proc. IEEE Signal Process. Adv. Wireless Commun.*, pages 484–488, Perugia, Italy, June 2009.

[Jah81] Johann-Heinrich Jahn. Coding of Arbitrarily Varying Multiuser Channels. *IEEE Trans. Inf. Theory*, 27(2):212–226, March 1981.

[Jor10] Eduard A. Jorswieck. Secrecy Capacity of Single- and Multi-Antenna Channels with Simple Helpers. In *Proc. Int. ITG Conf. Source and Channel Coding*, pages 1–6, Siegen, Germany, January 2010.

[JWG10] Eduard A. Jorswieck, Anne Wolf, and Sabrina Gerbracht. Secrecy on the Physical Layer in Wireless Networks. *Trends in Telecommunications Technologies*, pages 413–435, March 2010.

[KGG05] Gerhard Kramer, Michael Gastpar, and Piyush Gupta. Cooperative Strategies and Capacity Theorems for Relay Networks. *IEEE Trans. Inf. Theory*, 51(9):3037–3063, September 2005.

[KGLP11] O. Ozan Koyluoglu, Hesham El Gamal, Lifeng Lai, and H. Vincent Poor. Interference Alignment for Secrecy. *IEEE Trans. Inf. Theory*, 57(6):3323–3332, June 2011.

[Khi11] Ashish Khisti. Interference Alignment for the Multiantenna Compound Wiretap Channel. *IEEE Trans. Inf. Theory*, 57(5):2976–2993, May 2011.

[KMT08] Sang Joon Kim, Patrick Mitran, and Vahid Tarokh. Performance Bounds for Bidirectional Coded Cooperation Protocols. *IEEE Trans. Inf. Theory*, 54(11):5235–5241, November 2008.

[KMY06] Gerhard Kramer, Ivana Marić, and Roy D. Yates. Cooperative Communications. *Foundations and Trends in Communications and Information Theory*, 1(3-4):271–425, 2006.

[Kno06] Raymond Knopp. Two-Way Radio Networks With a Star Topology . In *Proc. Int. Zurich Seminar on Commun.*, pages 154–157, Zurich, Switzerland, February 2006.

[KP11] Tung T. Kim and H. Vincent Poor. Diversity-Multiplexing Trade-off in Adaptive Two-Way Relaying. *IEEE Trans. Inf. Theory*, 57(7):4235–4254, July 2011.

[Kra08] Gerhard Kramer. Topics in Multi-User Information Theory. *Foundations and Trends in Communications and Information Theory*, 4(4–5):265–444, 2008.

[KS07] Gerhard Kramer and Shlomo Shamai (Shitz). Capacity for Classes of Broadcast Channels with Receiver Side Information. In *Proc. IEEE Inf. Theory Workshop*, pages 313–318, Tahoe City, CA, USA, September 2007.

[KW10a] Ashish Khisti and Gregory W. Wornell. Secure Transmission With Multiple Antennas–Part II: The MIMOME Wiretap Channel. *IEEE Trans. Inf. Theory*, 56(11):5515–5532, November 2010.

[KW10b] Ashish Khisti and Gregory W. Wornell. Secure Transmission With Multiple Antennas I: The MISOME Wiretap Channel. *IEEE Trans. Inf. Theory*, 56(7):3088–3104, July 2010.

[LC08] Sergey Loyka and Charalambos D. Charalambous. On the Capacity of a Class of MIMO Channels Subject to Normed Uncertainty. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2578–2582, Toronto, Canada, July 2008.

[Lia72] Henry Liao. *Multiple Access Channels*. PhD thesis, University of Hawaii, Honolulu, USA, 1972.

[LJS05] Peter Larsson, Niklas Johansson, and Kai-Erik Sunell. Coded Bi-directional Relaying. In *Proc. 5th Scandinavian Workshop on Ad Hoc Networks*, pages 851–855, Stockholm, Sweden, May 2005.

[LK09] Peng Liu and Il-Min Kim. Performance Analysis of Bidirectional Communication Protocols Based on Decode-and-Forward Relaying. *IEEE Trans. Commun.*, 58(9):2683–2696, September 2009.

[LKPS09] Yingbin Liang, Gerhard Kramer, H. Vincent Poor, and Shlomo Shamai (Shitz). Compound Wiretap Channels. *EURASIP J. Wireless Commun. Netw.*, Article ID 142374:1–13, 2009.

[LLL10] Hung D. Ly, Tie Liu, and Yingbin Liang. Multiple-Input Multiple-Output Gaussian Broadcast Channels With Common and Confidential Messages. *IEEE Trans. Inf. Theory*, 56(11):5477–5487, November 2010.

[LLPS10a] Ruoheng Liu, Tie Liu, H. Vincent Poor, and Shlomo Shamai (Shitz). MIMO Gaussian Broadcast Channels with Confidential and Common Messages. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2578–2582, Austin, TX, USA, June 2010.

[LLPS10b] Ruoheng Liu, Tie Liu, H. Vincent Poor, and Shlomo Shamai (Shitz). Multiple-Input Multiple-Output Gaussian Broadcast Channels With Confidential Messages. *IEEE Trans. Inf. Theory*, 56(9):4215–4227, September 2010.

[LMSY08] Ruoheng Liu, Ivana Marić, Predrag Spasojević, and Roy D. Yates. Discrete Memoryless Interference and Broadcast Channels With Confidential Messages: Secrecy Rate Regions. *IEEE Trans. Inf. Theory*, 54(6):2493–2507, June 2008.

[LN98] Amos Lapidoth and Prakash Narayan. Reliable Communication Under Channel Uncertainty. *IEEE Trans. Inf. Theory*, 44(6):2148–2177, October 1998.

[LP08] Yingbin Liang and H. Vincent. Poor. Multiple-Access Channels With Confidential Messages. *IEEE Trans. Inf. Theory*, 54(3):976–1002, March 2008.

[LP09] Ruoheng Liu and H. Vincent Poor. Secrecy Capacity Region of a Multiple-Antenna Gaussian Broadcast Channel With Confidential Messages . *IEEE Trans. Inf. Theory*, 55(3):1235–1249, March 2009.

[LPS08] Yingbin Liang, H. Vincent. Poor, and Shlomo Shamai (Shitz). Secure Communication Over Fading Channels. *IEEE Trans. Inf. Theory*, 54(6):2470–2492, June 2008.

[LPS09] Yingbin Liang, H. Vincent. Poor, and Shlomo Shamai (Shitz). Information Theoretic Security. *Foundations and Trends in Communications and Information Theory*, 5(4-5):355–580, 2009.

[LPV08] Tie Liu, Vinod Prabhakaran, and Sriram Vishwanath. The Secrecy Capacity of a Class of Parallel Gaussian Compound Wiretap Channels. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 116–120, Toronto, Canada, July 2008.

[LS09] Tie Liu and Shlomo Shamai (Shitz). A Note on the Secrecy Capacity of the Multiple-Antenna Wiretap Channel. *IEEE Trans. Inf. Theory*, 55(6):2547–2553, June 2009.

[LSPL10] Kyoung-Jae Lee, Hakjea Sung, Eunsung Park, and Inkyu Lee. Joint Optimization for One and Two-Way MIMO AF Multiple-Relay Systems. *IEEE Trans. Wireless Commun.*, 9(12):3671–3681, December 2010.

[LT10] Ruoheng Liu and Wade Trappe, editors. *Securing Wireless Communications at the Physical Layer*. Springer, 2010.

[LTXW09] Jianquan Liu, Meixia Tao, Youyun Xu, and Xiaodong Wang. Superimposed XOR: A New Physical Layer Network Coding Scheme for Two-Way Relay Channels. In *Proc. IEEE Global Commun. Conf.*, pages 1–6, Honolulu, HI, USA, December 2009.

[MDT06] Patrick Mitran, Natasha Devroye, and Vahid Tarokh. On Compound Channels With Side Information at the Transmitter. *IEEE Trans. Inf. Theory*, 52(4):1745–1755, April 2006.

[Mil51] John Milnor. Games Against Nature. *RAND Corporation*, pages 49–59, 1951.

[Moo65] Gordon E. Moore. Cramming more Components onto Integrated Circuits. *Electronics Magazine*, 38(8):114–117, April 1965.

[MS10] Amitav Mukherjee and A. Lee Swindlehurst. Securing Multi-Antenna Two-Way Relay Channels With Analog Network Coding Against Eavesdroppers. In *Proc. IEEE Signal Process. Adv. Wireless Commun.*, pages 1–5, Marrakech, Morocco, June 2010.

[MW00] Ueli M. Maurer and Stefan Wolf. Information-Theoretic Key Agreement: From Weak to Strong Secrecy for Free. In *EUROCRYPT 2000, Lecture Notes in Computer Science*, volume 1807, pages 351–368. Springer-Verlag, May 2000.

[MYK05] Ivana Marić, Roy D. Yates, and Gerhard Kramer. The Discrete Memoryless Compound Multiple Access Channel With Conferencing Encoders. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 407–410, Adelaide, Australia, September 2005.

[MYP11] Ninoslav Marina, Hideki Yagi, and H. Vincent Poor. Improved Rate-Equivocation Regions for Secure Cooperative Communication. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2832–2836, Saint Petersburg, Russia, August 2011.

[NCL10] Wooseok Nam, Sae-Young Chung, and Yong H. Lee. Capacity of the Gaussian Two-Way Relay Channel to Within $\frac{1}{2}$ Bit. *IEEE Trans. Inf. Theory*, 56(11):5488–5494, November 2010.

[NG11] Bobak Nazer and Michael Gastpar. Compute-and-Forward: Harnessing Interference through Structured Codes. *IEEE Trans. Inf. Theory*, 57(10):6463–6486, October 2011.

[Nit10] Sirin Nitinawarat. On the Deterministic Code Capacity of an Arbitrarily Varying Multiple-Access Channel Under List Decoding. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 290–294, Austin, TX, USA, June 2010.

[NQS09] Hien Quoc Ngo, Tony Q. S. Quek, and Hyundong Shin. Amplify-and-Forward Two-Way Relay Networks: Error Exponents and Resource Allocation. *IEEE Trans. Commun.*, 58(9):2653–2666, September 2009.

[OB06] Tobias J. Oechtering and Holger Boche. Optimal Resource Allocation for a Bidirectional Regenerative Half-duplex Relaying. In *Proc. Int. Symp. Inf. Theory Applications*, Seoul, Korea, October 2006.

[OB08a] Tobias J. Oechtering and Holger Boche. Optimal Time-Division for Bidirectional Relaying using Superposition Encoding. *IEEE Commun. Lett.*, 12(4):265–267, April 2008.

[OB08b] Tobias J. Oechtering and Holger Boche. Piggyback a Common Message on Half-Duplex Bidirectional Relaying. *IEEE Trans. Wireless Commun.*, 7(9):3397–3406, September 2008.

[OB08c] Tobias J. Oechtering and Holger Boche. Stability Region of an Optimized Bidirectional Regenerative Half-duplex Relaying Protocol. *IEEE Trans. Commun.*, 56(9):1519–1529, September 2008.

[ODS10] Tobias J. Oechtering, Hieu T. Do, and Mikael Skoglund. Achievable Rates for Embedded Bidirectional Relaying in a Cellular Downlink. In *Proc. IEEE Int. Conf. Commun.*, pages 1–5, Cape Town, South Africa, May 2010.

[OKJ10] Lawrence Ong, Christopher M. Kellett, and Sarah J. Johnson. Functional-Decode-Forward for the General Discrete Memoryless Two-Way Relay Channel. In *Proc. IEEE Int. Conf. Commun. Systems*, pages 351–355, Singapore, November 2010.

[OSBB08] Tobias J. Oechtering, Clemens Schnurr, Igor Bjelaković, and Holger Boche. Broadcast Capacity Region of Two-Phase Bidirectional Relaying. *IEEE Trans. Inf. Theory*, 54(1):454–458, January 2008.

[PCL03] Daniel Pérez Palomar, John M. Cioffi, and Miguel Angel Lagunas. Uniform Power Allocation in MIMO Channels: A Game-Theoretic Approach. *IEEE Trans. Inf. Theory*, 49(7):1707–1727, July 2003.

[PDT09] Etienne Perron, Suhas Diggavi, and Emre Telatar. A Multiple Access Approach for the Compound Wiretap Channel. In *Proc. IEEE Inf. Theory Workshop*, pages 11–15, Taormina, Italy, October 2009.

[Pin64] M. S. Pinsker. Information and Information Stability of Random Variables and Processes. *Holden-Day*, 1964.

[PKA09] Petar Popovski and Toshiaki Koike-Akino. *Coded Bidirectional Relaying in Wireless Networks*, chapter 11, pages 291–316. New Directions in Wireless Communications Research. Springer US, 2009.

[PPTH09] Steven W. Peters, Ali Y. Panah, Kien T. Truong, and Robert W. Heath. Relay Architectures for 3GPP LTE-Advanced. *EURASIP J. Wireless Commun. Netw.*, 2009, 2009.

[Pro00] John G. Proakis. *Digital Communications*. Mcgraw-Hill Higher Education, 4 edition, 2000.

[Rap02] Theodore S. Rappaport. *Wireless Communications*. Prentice Hall, 2 edition, 2002.

[RH10] Florian Roemer and Martin Haardt. Tensor-Based Channel Estimation and Iterative Refinements for Two-Way Relaying With Multiple Antennas and Spatial Reuse. *IEEE Trans. Signal Process.*, 58(11):5720–5735, November 2010.

[RPV09] Adnan Raja, Vinod M. Prabhakaran, and Pramod Viswanath. The Two-User Compound Interference Channel. *IEEE Trans. Inf. Theory*, 55(11):5100–5120, November 2009.

[RV68] W. L. Root and P. P. Varaiya. Capacity of Classes of Gaussian Channels. *SIAM J. Appl. Math*, 16(6):1350–1393, 1968.

[RW07] Boris Rankov and Armin Wittneben. Spectral Efficient Protocols for Half-Duplex Fading Relay Channels. *IEEE J. Sel. Areas Commun.*, 25(2):379–389, February 2007.

[SAKH11] Aydin Sezgin, A. Salman Avestimehr, M. Amin Khajehnejad, and Babak Hassibi. Divide-and-Conquer: Approaching the Capacity of the Two-Pair Bidirectional Gaussian Relay Network. *IEEE Trans. Inf. Theory*, 2011. accepted.

[SEP08] Lalitha Sankar, Elza Erkip, and H. Vincent Poor. Sum-Capacity of Ergodic Fading Interference and Compound Multiaccess Channels. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2712–2716, Toronto, Canada, July 2008.

[SGP+09] Osvaldo Simeone, Deniz Gündüz, H. Vincent Poor, Andrea Goldsmith, and Shlomo Shamai (Shitz). Compound Multiple-Access Channels With Partial Cooperation. *IEEE Trans. Inf. Theory*, 55(6):2425–2441, June 2009.

[Sha48] Claude E. Shannon. A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, 27:379–423, 623–656, July, October 1948.

[Sha61] Claude E. Shannon. Two-Way Communication Channels. *Proc. 4th Berkeley Symp. Math Stat. and Prob.*, 1:611–644, 1961.

[Shi96] Paul C. Shields. *The Ergodic Theory of Discrete Sample Paths*. American Mathematical Society, 1996.

[SOS07] Clemens Schnurr, Tobias J. Oechtering, and Sławomir Stańczak. Achievable Rates for the Restricted Half-Duplex Two-Way Relay Channel. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 1468–1472, Pacific Grove, CA, USA, November 2007.

[SP09] Brooke Shrader and Haim Permuter. Feedback Capacity of the Compound Channel. *IEEE Trans. Inf. Theory*, 55(8):3629–3644, August 2009.

[SWS09] Chi Wan Sung, Qi Wang, and Kenneth W. Shum. Capacity Region of the Linear Four-Node Half-Duplex Wireless Relay Network. *IEEE Commun. Lett.*, 13(4):224–226, April 2009.

[TGK11] Roy Timo, Alex Grant, and Gerhard Kramer. Rate-Distortion Functions for Source Coding with Complementary Side Information. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2988–2992, Saint Petersburg, Russia, August 2011.

[Tho87] Joy A. Thomas. Feedback Can at Most Double Gaussian Multiple Access Channel Capacity. *IEEE Trans. Inf. Theory*, 33(5):711–716, September 1987.

[Tia09] Chao Tian. Latent Capacity Region: A Case Study on Symmetric Broadcast With Common Messages. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1834–1838, Seoul, Korea, June 2009.

[TLSP11] Xiaojun Tang, Ruoheng Liu, Predrag Spasojević, and H. Vincent Poor. Interference Assisted Secret Communication. *IEEE Trans. Inf. Theory*, 57(5):3153–3167, 2011.

[Tun06] Ertem Tuncel. Slepian-Wolf Coding Over Broadcast Channels. *IEEE Trans. Inf. Theory*, 52(4):1469–1482, April 2006.

[TY08] Ender Tekin and Aylin Yener. The Gaussian Multiple Access Wire-Tap Channel. *IEEE Trans. Inf. Theory*, 54(12):5747–5755, December 2008.

[van71] Edward C. van der Meulen. Three-Terminal Communication Channels. *Adv. Appl. Prob.*, 3:120–154, September 1971.

[VG97] Mahesh K. Varanasi and Tommy Guess. Optimum Decision Feedback Multiuser Equalization with Successive Decoding Achieves the Total Capacity of the Gaussian Multiple- Access Channel. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 1405–1409, Pacific Grove, CA, USA, November 1997.

[WBBJ11] Moritz Wiese, Holger Boche, Igor Bjelaković, and Volker Jungnickel. The Compound Multiple Access Channel With Partially Cooperating Encoders . *IEEE Trans. Inf. Theory*, 57(5):3045–3066, May 2011.

[WCK05] Yunnan Wu, Philip Chou, and Sun-Yuan Kung. Information Exchange in Wireless Networks with Network Coding and Physical-Layer Broadcast. In *Proc. Conf. Inf. Sciences and Systems*, pages 1–6, Baltimore, MD, USA, March 2005.

[WES05] Ami Wiesel, Yonina C. Eldar, and Shlomo Shamai (Shitz). Beamforming Maximizes the Rank One Ricean MIMO Compound Capacity. In *Proc. IEEE Signal Process. Adv. Wireless Commun.*, pages 323–327, New York, NY, USA, June 2005.

[WES07] Ami Wiesel, Yonina C. Eldar, and Shlomo Shamai (Shitz). Optimization of the MIMO Compound Capacity. *IEEE Trans. Wireless Commun.*, 6(3):1094–1101, March 2007.

[WLS+09] Hanan Weingarten, Tie Liu, Shlomo Shamai (Shitz), Yossef Steinberg, and Pramod Viswanath. The Capacity Region of the Degraded Multiple-Input Multiple-Output Compound Broadcast Channel. *IEEE Trans. Inf. Theory*, 55(11):5011–5023, November 2009.

[WNPS10] Makesh Pravin Wilson, Krishna Narayanan, Henry D. Pfister, and Alex Sprintson. Joint Physical Layer Coding and Network Coding for Bidirectional Relaying. *IEEE Trans. Inf. Theory*, 56(11):5641–5654, November 2010.

[Wol60] Jacob Wolfowitz. Simultaneous Channels. *Arch. Rational Mech. Analysis*, 4(4):371–386, 1960.

[Wol78] Jacob Wolfowitz. *Coding Theorems of Information Theory*. Springer-Verlag, 3 edition, 1978.

[Woz58] John Wozencraft. List Decoding. *Quarterly Progress Report, Research Laboratory of Electronics, MIT*, 48:90–95, 1958.

[WSK07] Hanan Weingarten, Shlomo Shamai (Shitz), and Gerhard Kramer. On the Compound MIMO Broadcast Channel. In *Proc. Inf. Theory Appl.*, San Diego, CA, USA, January 2007.

[WSS06a] Hanan Weingarten, Yossef Steinberg, and Shlomo Shamai (Shitz). On the Capacity Region of the Multi-Antenna Broadcast Channel with Common Messages. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2195–2199, Seattle, WA, USA, July 2006.

[WSS06b] Hanan Weingarten, Yossef Steinberg, and Shlomo Shamai (Shitz). The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel. *IEEE Trans. Inf. Theory*, 52(9):3936–3964, September 2006.

[Wu07] Yunnan Wu. Broadcasting when Receivers Know Some Messages A Priori. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1141–1145, Nice, France, June 2007.

[Wyn75] Aaron D. Wyner. The Wire-Tap Channel. *Bell Syst. Tech. J.*, 54:1355–1387, October 1975.

[Xie07] Liang-Liang Xie. Network Coding and Random Binning for Multi-User Channels. In *Proc. Canadian Workshop on Inf. Theory*, pages 85–88, June 2007.

[XU10] Jianwei Xie and Sennur Ulukus. Real Interference Alignment for the K-User Gaussian Interference Compound Wiretap Channel. In *Proc. Allerton Conf. Commun., Control, Computing*, pages 1252–1257, Urbana-Champaign, IL, USA, September 2010.

[YLCZ05] Raymond W. Yeung, Shuo-Yen Robert Li, Ning Cai, and Zhen Zhang. Network Coding Theory Part I: Single Sources / Part II: Multiple Sources. *Foundations and Trends in Networking*, 2(4–5):241–381, 2005.

[YZGK10] Erhan Yilmaz, Randa Zakhour, David Gesbert, and Raymond Knopp. Multi-pair Two-way Relay Channel with Multiple Antenna Relay Station. In *Proc. IEEE Int. Conf. Commun.*, pages 1–5, Cape Town, South Africa, May 2010.

[ZKWB08] Jian Zhao, Marc Kuhn, Armin Wittneben, and Gerhard Bauch. Optimum Time-Division in MIMO Two-Way Decode-and-Forward Relaying Systems. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 1494–1500, Pacific Grove, CA, USA, October 2008.

[ZLCC09] Rui Zhang, Ying-Chang Liang, Chin Choy Chai, and Shuguang Cui. Optimal Beamforming for Two-Way Multi-Antenna Relay Channel with Analogue Network Coding. *IEEE J. Sel. Areas Commun.*, 27(5):699–712, June 2009.