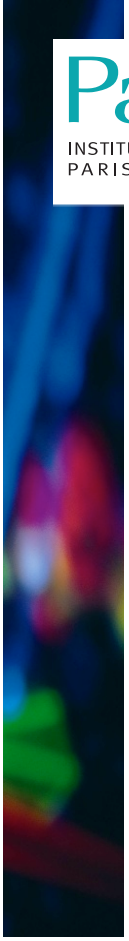
TELECOM ParisTech Spécialité " Electronique et Communications " présentée et soutenue publiquement par New Ways of Exploiting Feedback in Memoryless Broadcast Channels

Youlong Wu



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2014-ENST-00xx

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Electronique et Communications »

présentée et soutenue publiquement par

Youlong WU

le 31 Octobre 2014

New Ways of Exploiting Feedback in Memoryless Broadcast Channels

Directeur de thèse : **Michèle A. WIGGER**

Jury

Gerhard KRAMER, Professeur à l'Université Technique de Munich	Rapporteur
Yossef STEINBERG, Professeur à l'Institut Israélien de Technologie	Rapporteur
Petros ELIA, Maître de Conférences à Eurecom	Examinateur
Mari KOBAYASHI, Maître de Conférences à SUPELEC	Examinateur
Ghaya REKAYA-BEN OTHMAN, Professeur à Télécom ParisTech	Examinateur
Michèle A. WIGGER, Maître de Conférences à Télécom ParisTech	Examinateur

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

T H È S E

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Michèle Wigger. Three years ago, when I left my home country for the first time and started the PhD studies in France, I almost had to study everything from the beginning. The different languages, lives and research fields have posed immense difficulties that I had never thought to face. It was her endless patience and support that made me feel warm, confident and get engrossed in research. She always made time to discuss with me, no matter how busy her schedule and even when she was 9 months pregnant. She was a constant source of good ideas and has never hesitated to offer guidance and insightful comments. This dissertation would have never been possible without her invaluable aid. I am also wholeheartedly thankful for the many opportunities she has given me, such as attending special conferences, studying in the 2012 European School of Information Theory, and visiting Israel Institute of Technology as an exchange student in winter 2013. I feel privileged to have had the opportunity to have her as my "research mom".

My warmest thanks also goes to Professors Gerhard Kramer, Yossef Steinberg, Ghaya Rekaya, Mari Kobayashi and Petros Elia for being on my Ph.D. examination committees, and providing thoughtful feedback for me. I would like to thank Professor Yossef Steinberg with whom I had collaborated during my visit in Israel Institute of Technology. I really enjoyed the intriguing discussions with him and have always been inspired by his broad knowledge, profound expertise and unparalleled insight. I am also grateful to Professor Paolo Minero for the worthwhile collaborations. He initiated our first project and parts of this dissertation are joint works with him.

Also, I would like to thank my friends and colleagues: Selma Belhadj Amor, Tian Ban, Mohamed Essghair Chaibi, Arwa Ben Dhia, Asma Mejri , Seyed Hamed Mirghasemi and Julia Vinogradova, for all the pleasant discussions and fun we have had in the last

three years. I sincerely thank all the faculty and staff at the Telecom Paritech. Special thanks to Yvonne Bansimba, Florence Besnard, Chantal Cadiat and Bernard Cahen for helping me feel at home here.

I am grateful to Chinese Council Scholarship and the city of Paris for providing funding for the work done in this dissertation.

Last, but not the least, I would like to thank my wife and parents for their love and supporting me spiritually throughout my life. My wife, Yan Wang, gave up her job in Shanghai and moved to France with me. She is always behind me and gives her unconditional support. She has kept me healthy and energetic in these five years. Thanks to my parents for their countless sacrifices to raise me and give me the best education they can offer. For these three persons, I know that I can never fully repay them, but I will spend the rest of my life trying.

Dedication

To my family.

Abstract

In this dissertation we investigate the usefulness of feedback and propose new ways of exploiting rate-limited feedback for the memoryless broadcast channels (BC).

In the first part of the dissertation, we consider K-reciever Gaussian BC with only common message and feedback. We show that linear-feedback schemes with a message point, in the spirit of the Schalkwijk-Kailath scheme, are strictly suboptimal for this setup. To contrast this negative result, we describe a scheme for rate-limited feedback that uses the feedback in an intermittent way, which achieves all rates R up to capacity C with an L-th order exponential decay of the probability of error if the feedback rate $R_{\text{Fb}} \geq (L-1)R$, for some positive integer L.

In the second part, we study the two-receiver DMBC with private messages and rate-limited feedback. Two types of schemes based on block-Markov strategy and Marton's coding, have been proposed for this setup. In the first type of scheme, the transmitter simply relays the feedback messages obtained over the feedback links by encoding them into the Marton cloud center of the next-following block. With this type of scheme, we show that any positive feedback rate can strictly improve over the non-feedback capacity region for the class of *strictly essentially less-noisy* BCs, which we introduce in this dissertation. In our second type of scheme, the transmitter decodes all the feedback information and processes it with some local information before sending the result to the receivers. When the feedback rates are sufficiently large, then our scheme can recover all previously known capacity and degrees of freedom results for memoryless BCs with feedback.

Contents

Acknowledgements **i**

Dedication **iii**

Abstract **iv**

List of Figures **viii**

Notations **ix**

1 Introduction **1**

1.1 How Can Feedback Improve Communication Reliability? 2

1.2 What's the Capacity Region of Multiuser Channels with Feedback? . . . 3

1.2.1 Multiple Access Channels with Feedback 3

1.2.2 Broadcast Channels with Feedback 5

1.2.3 Interference Channels with Feedback 6

1.2.4 Relay Channels 8

1.3 Contributions and Outline 9

2 Reliability of the Gaussian Point-to-point Channel with Feedback **11**

2.1 Schalkwijk-Kailath Scheme 13

2.2 A Nonlinear Feedback Scheme 15

2.A Appendix: Proof of Existence of Codebook $\mathcal{C}_2$ 20

3 Broadcast Channel without Feedback **22**

3.1 Capacity for $R_1 = R_2 = 0$ 23

3.2 Capacity Region for $R_0 = 0$ 26
3.3 Achievable Regions for Broadcast Channels 29
3.3.1 Superposition Coding 29
3.3.2 Marton's Coding 31
3.4 Outer Bound on Broadcast Channel without Feedback 36
3.A Appendix: Proof of Remark 3.2 37

4 Previous Results on Memoryless Broadcast Channels with Feedback 40
4.1 Achievable Regions for Memoryless Broadcast Channels with Feedback . 40
4.1.1 Discrete Memoryless Broadcast Channels with Feedback 40
4.1.2 Memoryless Gaussian Broadcast Channels with Feedback 46
4.2 Outer Bound on Broadcast Channel with Feedback 48

5 Reliability of the Gaussian BC with Common Message and Feedback 49
5.1 Introduction 49
5.2 Setup 52
5.2.1 System Model and Capacity 52
5.2.2 Linear-Feedback Schemes with a Message Point 55
5.3 For comparison: Setup with Private Messages and Perfect Feedback . . 56
5.3.1 System Model and Capacity Region 56
5.3.2 Linear-Feedback Schemes with Message Points 58
5.4 Main Results 58
5.5 Proof of Proposition 5.1 61
5.6 Proof of Theorem 5.2: Coding Scheme Achieving L-th Order Exponential Decay 68
5.6.1 Code Construction 69
5.6.2 Transmission 70
5.6.3 Analysis 72
5.A Appendix: Proof of Proposition 5.2 76
5.B Appendix: Proof of Lemma 5.2 77
5.C Appendix: Existence of Code $\mathcal{C}_2, \ldots, \mathcal{C}_L$ with the Desired Properties . . 84

6 Coding Schemes for DMBCs with Private Messages and Rate-limited Feedback **87**

6.1 Introduction 87
6.2 Channel Model 89
6.3 Motivation: A Simple Scheme 91
6.4 New Achievable Regions and Usefulness of Feedback 94
6.4.1 Achievable Regions 94
6.4.2 Usefulness of Feedback 100
6.5 Examples 104
6.6 Coding Schemes 111
6.6.1 Coding Scheme 1A: Sliding-Window Decoding (Theorem 6.1) . . 111
6.6.2 Coding Scheme 1B: Backward Decoding (Theorem 6.2) 114
6.6.3 Coding Scheme 1C: Hybrid Sliding-Window Decoding and Backward Decoding (Theorem 6.3) 116
6.6.4 Coding Scheme 2: Encoder Processes Feedback-Info 117
6.7 Extension: Noisy Feedback 120
6.A Appendix: Analysis of Scheme 1A (Theorem 6.1) 121
6.B Appendix: Analysis of the Scheme 1B (Theorem 6.2) 127
6.C Appendix: Analysis of Scheme 2 (Theorem 6.4) 132
6.D Appendix: Proof of Theorem 6.5 138

7 Conclusion and Outlook **142**

References **144**

List of Figures

2.1 Gaussian point-to-point channel with feedback. 11
3.1 Two-receiver memoryless BC with private and common messages 22
4.1 Dueck's channel 41
5.1 K-receiver Gaussian BC with feedback and common message only. . . . 52
5.2 Upper bound (5.28) on the rates achievable with linear-feedback schemes with a message point in function of the number of receivers K. 60
5.3 Labeling of the transmission slots for our blocklength-$(n + 2K)$ scheme. 63
5.4 Transmissions considered at Receiver k and transmissions dedicated exclusively to Receiver k. 64
6.1 Broadcast channel with private messages and rate-limited feedback . . 90
6.2 $\mathcal{C}_{\text{NoFb}}$ and the achievable region in (6.36) are plotted for BSBCs with parameters $p_2 = 0.1$ and $p_1 \in \{0.2, 0.25, 0.3\}$ and for feedback rate $R_{\text{Fb},1} = 0.8$. 105
6.3 $\mathcal{C}_{\text{NoFb}}$ and the achievable regions in (6.40) and (6.45) are plotted for a BSC/BEC-BC when the BSC has parameter $p = 0.1$ and the BEC has parameter $e \in \{0.2, 0.7\}$. Notice that $0.2 < H_b(p) < 0.7$. The feedback rates $R_{\text{Fb},1} = R_{\text{Fb},2} = 0.8$. 108
6.4 $\mathcal{C}_{\text{NoFb}}$ and the achievable region in (6.47) are plotted for Gaussian BCs with parameters $P = 10$, $N_2 = 1$, $N_1 = 4$ and feedback rate $R_{\text{Fb},1} = 0.8$. 110

Notations

Sets, Scalars, Vectors, Matrices

Let $\mathbb{R}$ denote the set of real numbers and $\mathbb{Z}^+$ the set of positive integers. For a finite set $\mathcal{A}$, we denote by $|\mathcal{A}|$ its cardinality and by $\mathcal{A}^j$, for $j \in \mathbb{Z}^+$, its j-fold Cartesian product, $\mathcal{A}^j := \mathcal{A}_1 \times \ldots \times \mathcal{A}_j$. Given a set $\mathcal{S} \in \mathbb{R}^2$, we denote by $\mathrm{bd}(\mathcal{S})$ and $\mathrm{int}(\mathcal{S})$ the boundary and the interior of $\mathcal{S}$.

Given a real number x, $\lfloor x \rfloor$ is the integer part of x. We use the definitions $\bar{a} := (1-a)$ and $a * b := \bar{a}b + a\bar{b}$, for $a, b \in [0, 1]$. Also, $|\cdot|$ denotes the modulus operation for scalars.

Vectors are displayed in boldface, e.g., $\mathbf{X}$ and $\mathbf{x}$ for a random and deterministic vector. Given a positive integer n, let $\mathbf{1}_{[n]}$ denote the all-one tuple of length n, e.g., $\mathbf{1}_{[3]} = (1, 1, 1)$. Further, let $\|\cdot\|$ denote the norm operation for vectors.

For matrices we use the font A, and denote its Frobenius norm by $\|\mathsf{A}\|_F$.

Random Variables and Probability

We use capital letters to denote random variables and small letters for their realizations, e.g. X and x. For $j \in \mathbb{Z}^+$, we use the short hand notations X^j and x^j for the tuples $X^j := (X_1, \ldots, X_j)$ and $x^j := (x_1, \ldots, x_j)$.

$Z \sim \mathrm{Bern}(p)$ denotes that Z is a binary random variable taking values 0 and 1 with probabilities $1 - p$ and p. $X \sim \mathcal{N}(\mu, \sigma^2)$ denotes that X is a Gaussian random variable with mean μ and variance σ^2. We denote by $\mathcal{Q}(\cdot)$ the tail probability of the standard normal distribution. The abbreviation i.i.d. stands for *independent and identically distributed.*

Given a random variable Y, the expectation of Y is denoted by $\mathsf{E}[Y]$. The probability of an event $\mathcal{A}$ is denoted by $\Pr[A]$ and the conditional probability of $\mathcal{A}$ given an event $\mathcal{B}$ is denoted by $\Pr[\mathcal{A}|\mathcal{B}]$.

Common Functions

- $H(X)$: Shannon Entropy of the random variable X and is defined as

$$H(X) := \sum_{x \in \mathcal{X}} -P_X(x) \cdot \log_2(P_X(x)).$$

  Alternatively, we also write

$$H(P_X(x_1), P_X(x_2), \cdots, P_X(x_m)),$$

  to denote the Shannon entropy of a random variable X that takes values in $\mathcal{X} = \{x_1, \ldots, x_m\}$ according to the probability mass function $P_X(\cdot)$.
- $H_b(p)$: Binary Shannon Entropy of $p \in [0, 1]$ and is defined as

$$H_b(p) := -p \cdot \log_2(p) - (1-p) \cdot \log_2(1-p).$$

- $I(X;Y)$: Mutual Information between the random variables X and Y and is defined as

$$I(X;Y) := \sum_{(x,y) \in (\mathcal{X},\mathcal{Y})} P_{XY}(x,y) \cdot \log_2 \left(\frac{P_{XY}(x,y)}{P_X(x) \cdot P_Y(y)} \right).$$

- $\log(\cdot)$: denotes the natural logarithm.
- $\circ$: denotes function composition.
- $o(1)$: is the Landau symbol that stands for an arbitrary function that tends to 0 as $n \to \infty$.
- $\delta(\varepsilon)$: denotes a general nonnegative function of ε that tends to zero (arbitrarily slow) as $\varepsilon \to 0$.

Chapter 1

Introduction

The growing demands for reliable and high date rates have encouraged a huge amount of research on wireless communication systems with feedback, which allow the receivers to send signals back to the transmitter. Feedback channels, in fact, are present in many current commercial systems. A notable example are mobile cellular telephony systems, where the communication between the many mobiles and the only base station in a cell takes place in both directions.

In 1956 Shannon first showed that for memoryless point-to-point (single-user) channel, feedback cannot increase the capacity [1], even when the feedback is *perfect*, i.e. the feedback link is noise-free, delayless and of infinite rate. Nonetheless, feedback does improve the communication in terms of reducing the coding complexity and decreasing the probability of error [2]. Almost sixty years since the publication of Shannon's surprising result, a large number of works have been done, mainly to answer the following fundamental questions:

- How can feedback improve the communication reliability?
- What's the *capacity region* (i.e. the fundamental limits of reliable data rates) of *multiuser* channels with feedback?

In this dissertation we briefly review the works that are most relevant to our results in the following two sections.

1.1 How Can Feedback Improve Communication Reliability?

For Gaussian point-to-point channel without feedback, the probability of error decays at most exponentially in the blocklength n at any rate below capacity [3] (i.e. the probability of error is of the form $P_{\mathrm{e}}^{(n)} = \exp(-\Omega(n))$, where $\Omega(n)$ denotes a function that satisfies $\underline{\lim}_{n\to\infty} \frac{\Omega(n)}{n} > 0$). With the help of (perfect) feedback, a simple sequential linear scheme, proposed by Schalkwijk and Kailath [2], achieves capacity with a doubly exponential decay in the probability of error.

Notice that the Schalkwijk-Kailath scheme requires perfect feedback. When the feedback link is noisy, the improvement on the communication reliability drops dramatically [4, 5, 6]. In particular, when the feedback link is corrupted by additive white Gaussian noise, the Schalkwijk-Kailath scheme fails to achieve any positive rate [4].

In the presence of perfect feedback, the doubly exponential decay in error probability achieved by Schalkwijk-Kailath scheme is not the best one can obtain. In [7, 8] it is shown that perfect feedback allows for arbitrarily large super-exponential decay in error probability if the blocklength n is sufficiently large. Even if the feedback link is rate-limited (but noise-free and delayless), a nonlinear scheme [9] can achieve the capacity with L-th order exponential decay in the error probability when the feedback rate $R_{\mathrm{Fb}} > (L-1)R$, where R is the forward rate and L is any positive integer. That means, it achieves a probability of error of the form $P_{\mathrm{e}}^{(n)} = \exp(-\exp(\exp(\ldots\exp(\Omega(n)))))$, where there are L exponential terms. The Schalkwijk-Kailath scheme was later extended to many multiuser channels with (perfect) feedback [10, 11, 12, 13, 14, 15, 16, 17]. These variations also exhibit doubly exponential decay in the probability of error.

For discrete memoryless channel (DMC), the improvement on communication reliability afforded by feedback was studied in [18, 19, 20, 21, 22, 23, 24]. With perfect feedback, Burnashev [19] established a closed form expression for the reliability function which holds for all rates from zero to capacity. Subsequently, various two-phase coding schemes that achieve the optimal reliability function were given in [20, 21]. In [22] it is shown that for a binary symmetric channel (BSC) even if the encoder dose not know the crossover probability, the two-phase scheme can still achieve Burnashevs reliability function. Naghshvar, Javidi, and Wigger [23] proposed one-phase coding scheme

that achieves the optimal reliability function. In [24] communication reliability in the presence of Ack/Nack feedback was investigated.

1.2 What's the Capacity Region of Multiuser Channels with Feedback?

Contrary to the case of memoryless point-to-point channels with feedback, the capacity region of most multiuser channels with feedback is still unknown. Many previous works have been done aiming to find out whether or how feedback enlarges non-feedback capacity region. In this section we briefly review some results on various *memoryless* multiuser channels with feedback, including MAC, BC, interference channel (IC), and relay channel (RC) with feedback.

1.2.1 Multiple Access Channels with Feedback

Gaarder and Wolf [25] showed that (perfect) feedback can increase the non-feedback capacity region of certain discrete memoryless MAC. Subsequently, an achievable region for the general memoryless MAC with perfect feedback was proposed by Cover and Leung [26]. In [27] Willems has shown that the Cover-Leung region is tight for a certain class of discrete memoryless MACs, in which one of the two channel inputs is completely determined by the channel output and the other channel input. Bross and Lapidoth [28] derived an achievable region that strictly contains the Cover-Leung region for some channels. For the two-user memoryless Gaussian MAC, Ozarow [10] proposed a linear coding scheme based on the Schalkwijk-Kailath scheme [2], and showed that it is optimal, i.e. achieves the capacity region. For the $K \geq 3$-user Gaussian MAC, Kramer's scheme is optimal among a large class of schemes [12].

But so far, even with perfect feedback, no computable single-letter expression for the capacity region of general memoryless MACs with feedback is known. In [29, 30] Kramer presented a multi-letter expression based on "direct information". Permuter, Weissman, and Chen [31] later extended Kramer's results to finite-state MACs with feedback that may be an arbitrary time-invariant function of the channel output samples. Unfortunately, these multi-letter expressions are incomputable.

Memoryless MACs with imperfect feedback were considered in [32, 33, 34]. Carleial [32] derived an achievable region for the general MAC with generalized feedback signals and showed that the Cover-Leung region obtained for the perfect feedback setting, remains achievable for the perfect one-sided feedback, i.e. only one of the two transmitters knows perfect feedback signals whereas the other transmitter has no feedback at all. Shaviv and Steinberg [33] studied the general MAC with *rate-limited* feedback and derived an achievable region based on superposition coding, block-Markov coding, and coding with various degrees of side information at the feedback link, as in [35]. It is shown that this region coincides with the Cover-Leung region when feedback rate is large. Lapidoth and Wigger [34] considered the Gaussian MAC with noisy feedback and proposed achievable regions that exhibit the following two properties. 1) For all (finite) noise variances in the feedback links, the regions include rate points that lie outside the non-feedback capacity region, and 2) when the feedback-noise variances tend to zero, the regions coincide with Ozarow's region [10].

A different line of works has concentrated on the MACs with state [36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50]. In [37] and the references therein, the Gaussian MAC with perfect channel state information (CSI) at the transmitters (CSIT) and at the receiver (CSIR) was studied. In [36] the sum-rate capacity was given and later the capacity region was obtained in [37]. A relatively practical consideration where each transmitter knows its own CSI while the receiver has perfect CSI was investigated in [38, 39, 40]. In [41] Cemal and Steinberg considered the discrete memoryless MAC with partial CSIT, where the CSI is non-causally available at the transmitters in a compressed form. A single-letter capacity region was obtained for the two-user MAC when the CSI is available to the transmitter in a physically degraded version of that available to the other. This result was later extended by Jafar [42] to the case of casual CSI. In a survey paper [43], the fading MAC with perfect/partial/no CSIT was discussed. It is shown that with perfect CSIT and CSIR, the optimal power allocation leads to a substantial gain in capacity. In [44, 45] the fading MAC with distributed CSI, where the channel gains are only available at respective transmitters and receiver, was investigated. The *multiple-input multiple-output* (MIMO) fading MAC with CSI feedback was studied in [46, 47, 48, 49, 50].

Outer bounds on the capacity region of the MAC with feedback were derived by

Gastpar and Kramer [51] and Tandon and Ulukus [52] based on the idea of dependence-balance [53], which was first introduced for single-output two-way channels.

1.2.2 Broadcast Channels with Feedback

For most BCs in general, it is not known whether feedback can increase the capacity region. There are some exceptions. For example, for all physically degraded discrete memoryless broadcast channels (DMBCs) the capacity regions with and without feedback coincide [54]. The first simple example DMBC where (even rate-limited) feedback increases capacity was presented by Dueck [55]. His example and coding scheme were generalized by Shayevitz and Wigger [56] who proposed a general scheme and achievable region for DMBCs with generalized feedback. In the generalized feedback model, the feedback to the transmitter is modeled as an additional output of the DMBC that can depend on the input and the receivers' outputs in an arbitrary manner. It has recently been shown [57] that the Shayevitz-Wigger scheme for generalized feedback includes as special cases the two-user schemes by Wang [58], by Georgiadis and Tassiulas [59], and by Maddah-Ali and Tse [60], which achieve the capacity region and the degrees of freedom (DoF) region of their respective channels.

Other achievable regions for general DMBCs with perfect or noisy feedback have been proposed by Kramer [29] and by Venkataramanan and Pradhan [61]. Comparing the general achievable regions in [29, 56, 61] to each other is hard because of their complex form which involves several auxiliary random variables.

A different line of works has concentrated on the memoryless Gaussian BCs [13, 14, 16, 62, 63, 64, 65, 66]. In [14], a coding scheme was proposed which can achieve the full-cooperation bound for all noise correlations $-1 < \rho_z <$ in the high signal-to-noise ratio (SNR) limit. When the feedback links are noisy, the gains afforded by feedback are bounded, unless the feedback noise decays to zero sufficiently fast with SNR. The asymptotic high SNR sum-capacity for the cases $\rho_z \in \{+1, -1\}$ was also investigated in [14]. The best achievable region when the noises at the two receivers are independent was given in [16, 17] (see also [13, 64] for special cases) and is based on a MAC-BC duality approach.

In [66] Li and Goldsmith investigated the K-receiver Gaussian fading BC assuming perfect CSIT and CSIR, and proposed optimal resource allocation strategies for code

division, time division, and frequency division. The MIMO Gaussian fading BC with rate-limited feedback was studied in [67, 68, 69, 70, 71, 72, 73, 74, 75], where the receivers are assumed to have perfect knowledge of the channel matrix, and use quantization or analog scaling to send the it back to the transmitter via the rate-limited feedback link. Jindal [67] proved that to achieve the full multiplexing gain, the feedback rate per receiver must increase linearly with the SNR. In [68] Caire *et al.* analyzed and compared the quantized and analog CSI feedback schemes under various assumptions. Kobayashi [69] proposed a useful guideline to optimize the overall system throughput. Lapidoth, Shamai, and Wigger [71] showed that the uncertainty of CSI at the transmitter greatly reduces the SNR throughput even if the receivers know perfect CSI. Weingarten, Shamai, and Kramer [73] proposed inner and outer bounds on the DoF for finite-state compound BC with two receivers. Gou, Jafar, and Wang showed that for a few cases the inner bound is tight. In [60, 76, 77, 78, 79, 80, 81, 82] the case of delayed (stale) CSI at the transmitter was considered, i.e. there is a delay between the time measuring channel state at the receiver and the time observing CSI at the transmitter. It is shown that delayed CSI can still increase the capacity region.

An outer bound on the capacity region of the BC with feedback was given by Ozarow and Leung [62], by assuming one of the two receivers observes the channel output of the other receiver. This converts the BC to physically degraded BC and thus the capacity region with feedback and without feedback coincide.

A more detailed introduction to memoryless BC with feedback is provided in Chapter 4.

1.2.3 Interference Channels with Feedback

ICs with perfect feedback were investigated in [83, 84, 85, 86, 14]. Jiang *et al.* [83] derived an achievable region for discrete memoryless IC with feedback, based on binning and block-Markov coding. In [84, 85], Kramer proposed feedback strategies for the Gaussian IC. Suh and Tse [86] considered the two-user Gaussian IC with feedback and characterized the capacity region to within 2 bits/s/Hz and the symmetric capacity to within 1 bit/s/Hz. They also showed that feedback can provide *unbounded* gain in ICs at high SNR, i.e. the gap between the feedback and non-feedback capacity can be arbitrarily large for certain channel parameters. Sahai *et al.* [87] proved that for

deterministic IC, only one perfect feedback link from one receiver to either of the two transmitters is sufficient to achieve the capacity region of the setting with both perfect feedback links. Gastpar *et al.* [14] considered the two-user Gaussian IC with noiseless one-sided feedback, where each of the two transmitter obtains feedback only from its intended receiver. It is shown that feedback can provide unbounded gain for this setting, approximately doubling the sum-rate capacity in the high SNR regime.

More realistic scenarios where feedback cost is taken into consideration were studied in [88, 89, 90, 91]. Vahid, Suh and Avestimehr [88] studied the two-user IC with rate-limited feedback under three different models: the El Gamal-Costa deterministic model [92], the linear deterministic model [93] and the Gaussian model, and it is shown that one bit of feedback can provide at most one bit of capacity increase. In [90] the K-user Gaussian IC with feedback was studied and it is shown that feedback increases the DoF for some classes of scalar ICs.

MIMO Gaussian IC with CSI feedback was considered in [94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107]. In [94], Jafar and Fakhereddin showed that zero-forcing scheme is sufficient to achieve the DoF region for the two-user MIMO IC. In [97] Wang, Gou, and Jafar studied the three-user Gaussian IC and solved issue of feasibility of linear interference alignment introduced by Yetis [98]. In [99, 100, 101], it is shown that the DoF achieved with perfect CSIT remains achievable when the CSIT is imperfect but increases at sufficiently fast rate with SNR. Recently, inspired by the Maddah-Ali&Tse's surprising result which shows that even delayed CSIT can increase the DoF for MIMO Gaussian BC, the impact of delayed CSIT on MIMO Gaussian IC was investigated in [102, 103, 104, 105, 106, 107].

Using the idea of dependence-balance, Tandon and Ulukus [52] derived an outer bound for the general IC with feedback. Gastpar and Kramer [51] applied the same idea to obtain an outer-bound on the usefulness of noisy feedback for the IC and showed that feedback gain decreases roughly when the feedback-noise variance is larger than the noise on the forward link. More outer bounds for various specific ICs with feedback were established in [86, 88, 90, 91].

1.2.4 Relay Channels

The relay channel model was first introduced by Van der Meulen [108], which is in fact a single-user communication channel where a relay helps the communication between the transmitter and the receiver. In [109], Cover and El Gamal proposed two coding strategies for the discrete memoryless relay channel: the decode-and-forward scheme and the compress-and-forward scheme. A generalized compress-and-forward strategy for

The relay channel with feedback has first been considered by Cover and Gamal [109]. It is shown that for the setting with feedback from the receiver to the relay, the channel is physically degraded and therefore decode-and-forward strategy achieves the capacity. However, for the settings with partial feedback either from the receiver to the transmitter, from the relay to the transmitter, or both (but without receiver-relay feedback), the capacity is still not known except for the semi-deterministic case [110, 92] and for the physically degraded case.

In [111] Gabbai and Bross considered the setting with partial feedback from the receiver to the transmitter as well as partial feedback from the relay to the transmitter. Achievable rates were derived for the general Gaussian and the Z relay channels and were shown to be strictly larger than the best known achievable rates without feedback.

Bross and Wigger [112] investigated the discrete memoryless relay channel and the Gaussian memoryless relay channel with partial feedback from the receiver to the transmitter. For discrete memoryless relay channels, an achievable rate was calculated by combining the ideas of restricted decoding used in [113], the nested binning used in [111], and the generalized coding strategy for the relay channel in [109]. For the Gaussian relay channel, another achievable rate was derived based on the Schalkwijk-Kailath scheme [2]. It is shown that for some channels both achievable rates are strictly larger than all best known achievable rates.

An upper bound for relay channels with partial feedback from the relay to the transmitter was established in [110]. This outer bound includes the cut-set upper bound [109] on the capacity of the one-way relay channel, and was shown to be tight for the semi-deterministic relay channel with partial feedback from the relay to the transmitter.

The classic relay channel in [108, 109] can be extended to more general models by introducing multiple nodes such as transmitters, relays and receivers [114, 115, 116,

117]. When adding multiple transmitters, the model turns to multiple access relay channel (MARC) [114, 118]. In [119] Hou, Koetter, and Kramer studied the MARC with perfect feedback and derived inner bounds and outer bound for feedback from the relay to the transmitters. The MARC with generalized feedback was studied in [120]. When adding multiple transmitters and relays, it results in broadcast relay channel (BRC) [121, 122], the work on BRC with perfect/limited feedback can be found in [123, 124, 125, 126, 127]. When adding multiple transmitters, relays and and receivers, this model is called interference channel with relays (ICR) [116, 117]. The ICR with feedback was investigated in [128, 129, 130] and the references therein.

1.3 Contributions and Outline

In this dissertation we focus on memoryless BCs with feedback, for which we investigate the usefulness of feedback and propose new ways of exploiting feedback to help the communication, i.e. to increase the capacity or to improve the communication reliability.

The dissertation is organized as follows.

In Chapter 2 we review the reliability results on memoryless Gaussian point-to-point channel with feedback, where the Schalkwijk-Kailath scheme and a nonlinear feedback scheme proposed by Mirghaderi, Goldsmith and Weissman are present.

In Chapter 3 we review the capacity results on some classes of memoryless BCs without feedback. We also describe superposition coding and Marton's coding, which lead to two important inner bounds for general BCs. These two schemes are also used as building stones in our schemes proposed in Chapter 6 for memoryless BCs with feedback. In Chapter 3 we also present the best known outer bound proposed by Nair and El Gamal.

In Chapter 4 we review several achievable regions for DMBC and introduce several linear-feedback coding schemes for memoryless Gaussian BC. At the end of Chapter 4, we present two well-known outer bounds: cut-set outer bound and the Ozarow-Leung outer bound.

In Chapter 5 we consider the $K \geq 2$-user memoryless Gaussian BCs with feedback and common message only. We show that the type of Schalkwijk-Kailath coding scheme, which performs well for point-to-point channels, Gaussian MACs and Gaussian BCs

with private messages, is strictly suboptimal for this setup. Even with perfect feedback, the largest rate achieved by this type of linear-feedback scheme is strictly smaller than capacity, which is the same with and without feedback. As a consequence, for this setup, linear-feedback schemes also fail to achieve double-exponential decay of the probability of error for rates close to capacity. In contrast, we present a coding scheme with rate-limited feedback which is inspired by the nonlinear scheme presented in Chapter 2. When the feedback rate $R_{\text{fb}} > (L-1)R$, where R is the forward rate and L is any positive integer, our intermittent-feedback scheme can achieve the capacity with a L-th order exponential decay in the probability of error.

In Chapter 6 we study the two-receiver memoryless BCs with private messages and feedback. For this setup we propose two types of coding schemes with rate-limited feedback from one or two receivers. Our first type of scheme strictly improves over the non-feedback capacity region for the class of *strictly essentially less-noisy* BCs, for any positive feedback rate and even when there is only feedback from the weaker receiver. Examples of essentially strictly less-noisy BCs are the binary symmetric BCs or the binary erasure BCs with unequal cross-over probabilities or unequal erasure probabilities to the two receivers. Previous to our work, feedback was known to increase capacity only for a few very specific memoryless BCs. Our second type of scheme can recover all previously known capacity and DoF results for memoryless BCs with feedback when the feedback rates are sufficiently large. This includes in particular the result by Wang [58] and by Georgiadis and Tassiulas [59] for the binary erasure BC when all erasures are known to both receivers, the results by Shayevitz and Wigger [56] and by Chia, Kim, and El Gamal [57] on variations of the Blackwell DMBC, and the result by Maddah-Ali and Tse [60] on memoryless fading BCs with completely stale state information. In fact, as the feedback-rates tend to infinity our scheme improves over a special case of the Shayevitz-Wigger scheme which is known to recover the mentioned results.

Finally in Chapter 7 we conclude this dissertation and discuss some possible future directions.

Chapter 2

Reliability of the Gaussian Point-to-point Channel with Feedback

Consider a memoryless Gaussian channel with feedback, depicted in Fig. 2.1.

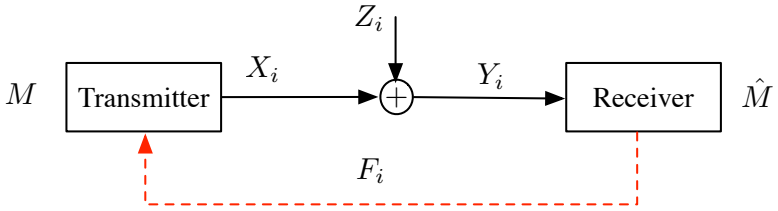


Figure 2.1: Gaussian point-to-point channel with feedback.

The transmitter wishes to communicate a message $M \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$ over the discrete-time memoryless Gaussian noise channel.

The cannel output is

$$Y_i = X_i + Z_i, \text{ for } i \in \{1, \ldots, n\} \tag{2.1}$$

where $\{Z_i\}$ is an i.i.d sequence with each sample $Z_i \sim \mathcal{N}(0, 1)$. The transmitter is comprised of a sequence of encoding functions $\{f_i^{(n)}\}_{i=1}^n$ that is used to produce the

channel inputs as

$$X_i = f_i^{(n)}(M, F_1, \ldots, F_{i-1}), \quad \text{for } i \in \{1, \ldots, n\} \tag{2.2}$$

where F_i takes values in feedback alphabet set $\mathcal{F}_i$ and denotes the signal sent back by the receiver at the end of time slot i. The channel input X_i is restricted to an expected average power constraint:

$$\sum_{i=1}^{n} \mathsf{E}\left[X_i^2\right] \leq nP. \tag{2.3}$$

The receiver is comprised of a decoding function $g^{(n)}$ used to produce a guess of message M

$$\hat{M} = g^{(n)}(Y^n) \tag{2.4}$$

where $\hat{M} \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$.

A rate is called achievable if for every blocklength n, there exists an encoding function and a decoding function such that

$$P_e^n := \Pr\left[\hat{M} \neq M\right]$$

tends to 0 as the blocklength n tends to infinity.

The capacity of this channel is the supremum over all achievable rate, which is same as the non-feedback capacity: $C = 1/2 \log_2(1 + P)$.

Although feedback can not increase the capacity of point-to-point channel, it does improve the reliability, i.e. reduce the smallest possible error probability in the communication at a given blocklengh n.

In this chapter, we briefly review the results on reliability of memoryless Gaussian point-to-point channels with feedback. In Section 2.1, we describe the prominent linear Schalkwijk-Kailath scheme [2], which can achieve the capacity with a doubly exponential decay in error probability as a function of blocklength n. In Section 2.2, a nonlinear scheme [9] with rate-limited feedback is present, which uses the feedback in an intermittent way: only in few time slots the receiver sends feedback signals. This

nonlinear feedback scheme can also achieve the capacity with doubly exponential decay in the probability of error, but under a milder constraint–only requiring that the average feedback rate satisfies $R_{\mathrm{Fb}} > R$.

2.1 Schalkwijk-Kailath Scheme

Assume that the feedback link is perfect (noise-free, delayless and of infinite capacity), then we have $F_i = Y_i$ for $i \in \{1, \ldots, n\}$. Here we review the Schalkwijk-Kailath scheme [2], see also Section 17.1.1 in [131].

1) Codebook generation: Divide the interval $[-\sqrt{P}, \sqrt{P}]$ into $\lfloor 2^{nR} \rfloor$ small subintervals with equal length $\Delta = 2\sqrt{P}/\lfloor 2^{nR} \rfloor$. Convert each message m into a real number $\theta(m)$ that denotes the midpoint of the m-th subinterval.

2) Encoding: Given $M = m$, initially, the transmitter sends

$$X_0 = \theta(m),$$

and the receiver observes $Y_0 = X_0 + Z_0$. With perfect feedback, the transmitter can learn the noise Z_0. Then at time $i = 1$, it sends $X_1 = \alpha_1 Z_0$ with $\alpha_1 = \sqrt{P}$, which satisfies the average power constraint. Subsequently, at time $i \in \{1, \ldots, n\}$, the encoder sends

$$X_i = \alpha_i (Z_0 - \hat{Z}_0(Y^{i-1}))$$

where $\hat{Z}_0(Y^{i-1})$ is the minimum mean square error (MMSE) estimate of Z_0 given Y^{i-1} and α_i is chosen to satisfy $\mathsf{E}\left[X_i^2\right] = P$ for $i \in \{1, \ldots, n\}$.

3) Decoding: After observing Y^n, the receiver estimates $\hat{\theta}_n$ of $\theta(m)$ by taking

$$\begin{aligned} \hat{\theta}_n &= Y_0 - \hat{Z}_0(Y^{n-1}) \\ &= \theta(m) + Z_0 - \hat{Z}_0(Y^{n-1}) \end{aligned} \tag{2.5}$$

and declares $\hat{m}$ as the message sent if $\theta(\hat{m})$ is the nearest message point to $\hat{\theta}_n$.

3) Analysis of the probability of error: By the orthogonality principle of MMSE estimate and by the joint Gaussianity, we observe that $\hat{Z}(Y^{i-1})$ is linear in Y^{i-1}, and (X_i, Z_i, Y_i) are independent of Y^{i-1}. Furthermore, since the scaling factor α_i, for $i \in$

$\{1, \ldots, n\}$, is chosen so that $\mathsf{E}\left[X_i^2\right] = P$, the channel outputs Y^n are i.i.d with $Y_i \sim \mathcal{N}(0, P+1)$.

Now we look at the mutual information $I(Z_0; Y^n)$ in two ways. On one hand,

$$\begin{aligned}
I(Z_0; Y^n) &= h(Y^n) - H(Y^n|Z_0) \\
&= \sum_{i=1}^{n} \big(h(Y_i|Y^{i-1}) - h(Y_i|Z_0, Y^{i-1})\big) \\
&= \sum_{i=1}^{n} \big(h(Y_i|Y^{i-1}) - h(X_i + Z_i|Z_0, Y^{i-1})\big) \\
&= \sum_{i=1}^{n} \big(h(Y_i|Y^{i-1}) - h(Z_i|Z_0, Y^{i-1})\big) \\
&= \sum_{i=1}^{n} \big(h(Y_i|Y^{i-1}) - h(Z_i)\big) \\
&= \sum_{i=1}^{n} \big(h(Y_i) - h(Z_i)\big) \\
&= \frac{n}{2} \log_2(1+P) \\
&= nC. \qquad (2.6)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
I(Z_0; Y^n) &= h(Z_0) - h(Z_0|Y^n) \\
&= \frac{1}{2} \log_2 \frac{1}{\mathrm{var}(Z_0|Y^n)}. \qquad (2.7)
\end{aligned}$$

From (2.6) and (2.7), we have

$$\mathrm{var}(Z_0|Y^n) = 2^{-2nC}.$$

Thus, given (2.5), we obtain that $\hat{\theta}_n \sim \mathcal{N}(\theta(m), 2^{-2nC})$. Notice that the receiver performs the nearest neighbor decoding, which implies that the decoder makes an error only if $\hat{\theta}_n$ is closer to a different message point not equal to $\theta(m)$, i.e. $|\hat{\theta}_n - \theta(m)| > \Delta/2$. The probability of error is thus upper bounded by

$$P_e^{(n)} \leq 2\mathcal{Q}(2^{n(C-R)}\sqrt{P}). \qquad (2.8)$$

Recall that for the stand Gaussian cumulative density function: $Q(x) \leq (1/\sqrt{2\pi})e^{-x^2/2}$. Therefore, and if $R < C$

$$P_e^{(n)} \leq \sqrt{\frac{2}{\pi}}\exp(-\frac{2^{n(C-R)}\sqrt{P}}{2}),$$

the probability of error decays doubly exponentially to 0 in blocklength n.

2.2 A Nonlinear Feedback Scheme

Assume the feedback link is noise-free, delayless but is rate-limited by R_{Fb}. It is shown in [9] that a super-exponential decay in the probability of error as a function of the blocklengh n is achievable when R_{Fb} is sufficiently large. For completeness, in the following we describe their scheme that achieves capacity C with doubly exponential decay in the probability of error.

Fix a positive rate $R < C$ and assume that

$$R_{\text{Fb}} \geq R. \tag{2.9}$$

Also, fix a large blocklength n and small numbers $\epsilon, \delta > 0$ such that

$$R < C(1-\delta) \tag{2.10}$$

and

$$(1-\epsilon)^{-1} < 1+\delta. \tag{2.11}$$

Define

$$n_1 := (1-\epsilon)n. \tag{2.12}$$

Notice that by (2.11) and (2.12),

$$\frac{n}{n_1} < 1+\delta. \tag{2.13}$$

The coding scheme takes place in two phases. After each phase $l \in \{1,2\}$, the receiver

makes a temporary guess $\hat{M}_l$ of message M. Define the probability of error after phase $l \in \{1, 2\}$:

$$P_{e,l}^{(n)} := \Pr\left[\hat{M}_l \neq M\right]. \tag{2.14}$$

1) Code Construction: Construct a codebook $\mathcal{C}_1$ that

- is of blocklength n_1,
- is of rate $R_{\text{phase},1} = \frac{n}{n_1}R$,
- satisfies an expected average block-power constraint P, and
- when using a non-feedback coding rule, it achieves probability of error γ_1 not exceeding

$$\gamma_1 \leq e^{-n(\zeta - o(1))} \tag{2.15}$$

for some $\zeta > 0$.

Notice that such a code exists because, by (2.10) and (2.13), the rate of the code $\frac{n}{n_1}R < C(1-\delta^2)$. Construct a codebook $\mathcal{C}_2$ that:

- is of blocklength $\epsilon n - 1$,
- is of rate $R_{\text{phase},2} := \frac{R}{\epsilon - 1/n}$,
- satisfies an expected average block-power constraint P/γ_1,
- when using a non-feedback coding rule, it achieves probability of error γ_2 not exceeding

$$\gamma_2 \leq \exp(-\exp(\Omega(n))). \tag{2.16}$$

The existence of codebook $\mathcal{C}_2$ is given in Appendix 2.A.

2) Transmission: Transmission takes place in two phases. In the first phase, which occupies channel uses $1, \ldots, n_1$, the transmitter sends the codeword in $\mathcal{C}_1$ corresponding to message M.

After observing the channel outputs $Y_1, \ldots, Y_{n_1}$, the receiver makes a temporary decision $\hat{M}_1$ about M. It then sends this temporary decision $\hat{M}_1$ to the transmitter over the feedback link. By $R_{\text{Fb}} > R$, message $\hat{M}_1$ can be perfectly known at the transmitter.

If in phase 1 the decoding result is wrong, i.e.

$$(\hat{M}_1 \neq M) \tag{2.17}$$

then in channel use $n_1 + 1$ the transmitter sends an error signal to indicate an error:

$$X_{n_1+1} = \sqrt{P/\gamma_1}. \tag{2.18}$$

During the remaining channel uses $i = n_1 + 2, \ldots, n$, it then retransmits the message M by sending the codeword from $\mathcal{C}_2$ that corresponds to M.

On the other hand, if the receiver's temporary decisions was correct,

$$\hat{M}_1 = M, \tag{2.19}$$

then the transmitter remains silent during the entire phase 2:

$$X_i = 0, \qquad i = n_1 + 1, \ldots, n. \tag{2.20}$$

In this case, no power is consumed in phase 2.

The receiver first detects whether the transmitter sent an error signal in channel use $n_1 + 1$. Depending on the output of this detection, they either stick to their temporary decision in phase 1 or make a new decision based on the transmissions in phase 2. Specifically, if

$$Y_{n_1+1} < T \tag{2.21}$$

where

$$T := \frac{\sqrt{P/\gamma_1}}{2}, \tag{2.22}$$

then the receiver decides that its decision $\hat{M}_1$ in phase 1 was correct and keeps it as its temporary guess of the message M:

$$\hat{M} = \hat{M}_1. \tag{2.23}$$

If instead,

$$Y_{n_1+1} \geq T, \tag{2.24}$$

the receiver then decides that its temporary decision $\hat{M}_1$ was wrong and discards it. It then produces a new guess $\hat{M}_2$ by decoding the code $\mathcal{C}_2$ based on the outputs $Y_{n_1+2}, \ldots, Y_n$. In this case, the receiver produces the final guess as

$$\hat{M} = \hat{M}_2. \tag{2.25}$$

3) Analysis:

An error occurs in the communication if

$$(\hat{M} \neq M). \tag{2.26}$$

We next analyze the probability of error and we bound the consumed power. The analysis relies on the following events. Define the events:

- ϵ_1: The receiver's decision in phase 1 is wrong:

$$\hat{M}_1 \neq M. \tag{2.27}$$

- ϵ_T: The receiver observes

$$Y_{n_1+1} < T. \tag{2.28}$$

- ϵ_2: Decoding message M based on the outputs $Y_{n_1+2}, \ldots, Y_n$ using codebook $\mathcal{C}_2$ results in an error:

$$\hat{M}_2 \neq M. \tag{2.29}$$

Define also the events:

$\mathcal{E}_1$: The receiver's decision in phase 1 is correct, and the receiver obtains an error signal in channel use $n_1 + 1$:

$$\big(\epsilon_1\big)^c \cap \big(\epsilon_T\big)^c. \tag{2.30}$$

$\mathcal{E}_2$: The receiver's decision in phase 1 is wrong, but the receiver obtains no error signal in channel use $n_1 + 1$:

$$\epsilon_1 \cap \epsilon_T. \tag{2.31}$$

$\mathcal{E}_3$: The receiver's decision in phase 1 is incorrect, the receiver observes $Y_{n_1+1} \geq T$ and errs when decoding M based on the outputs $Y_{n_1+2}, \ldots, Y_n$

$$\epsilon_1 \cap \left(\epsilon_T\right)^c \cap \epsilon_2. \tag{2.32}$$

The probability $P_e^{(n)}$ is included in the union of the events $(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3)$, and thus, by the union bound,

$$P_e^{(n)} \leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2] + \Pr[\mathcal{E}_3]\,. \tag{2.33}$$

We bound each summand in (2.33) individually, starting with $\Pr[\mathcal{E}_1]$. By (2.30), we have

$$\begin{aligned} \Pr[\mathcal{E}_1] &= \Pr\left[\left(\epsilon_1\right)^c \cap \left(\epsilon_T\right)^c\right] \\ &\leq \Pr\left[\left(\epsilon_T\right)^c | \left(\epsilon_1\right)^c\right] \\ &= \mathcal{Q}(T) \end{aligned} \tag{2.34}$$

where the first inequality follows by Bayes' rule and because a probability cannot exceed 1; and the last equality because in the event $\left(\epsilon_1\right)^c$, we have $X_{n_1+1} = 0$ and thus $Y_{n_1+1} \sim \mathcal{N}(0,1)$.

Next, by (2.31) and similar arguments as before, we obtain,

$$\begin{aligned} \Pr[\mathcal{E}_2] &= \Pr[\epsilon_1 \cap \epsilon_T] \\ &\leq \sum_{k=1}^{K} \Pr[\epsilon_T | \epsilon_1] \\ &= \mathcal{Q}(T). \end{aligned} \tag{2.35}$$

Finally, by (2.32) and similar arguments as before,

$$\begin{aligned} \Pr[\mathcal{E}_3] &= \Pr\left[\epsilon_1 \cap \left(\epsilon_T\right)^c \cap \epsilon_2\right] \\ &\leq \Pr\left[\left(\epsilon_T\right)^c \cap \epsilon_2 | \epsilon_1\right] \\ &\leq \Pr[\epsilon_2 | \epsilon_1] \\ &= \gamma_2 \end{aligned} \tag{2.36}$$

where the last inequality follows by the definition of γ_2.

In view of (2.33)–(2.36),

$$\begin{aligned} P_e^{(n)} &\leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2] + \Pr[\mathcal{E}_3] \\ &\leq 2\mathcal{Q}(T) + \gamma_2. \end{aligned} \tag{2.37}$$

Given the definitions of γ_1, γ_2 and T in (2.15), (2.16) and (2.22), respectively, and by $\mathcal{Q}(x) \leq (1/\sqrt{2\pi})e^{-x^2/2}$, (2.37) implies that probability of error decays doubly exponentially in the blocklength n.

Now consider the consumed expected average block-power. By the definition in (2.14), we have

$$P_{e,1}^{(n)} \leq \gamma_1. \tag{2.38}$$

Furthermore, since in phase 2 we consume power P/γ_1 in the event (2.17) and power 0 in the event (2.19),

$$\frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{n} X_i^2\right] \leq \frac{1}{n}\Big(P(1-\epsilon)n + P_{e,1}^{(n)}\frac{P}{\gamma_1}\epsilon n\Big) \leq P. \tag{2.39}$$

Also note that the average feedback rate is R that meets the constraint on the feedback link.

2.A Appendix: Proof of Existence of Codebook $\mathcal{C}_2$

By [3], for all rates

$$\tilde{R} < \frac{1}{2}\log_2 \frac{2 + \sqrt{\tilde{P}^2/\sigma^4 + 4}}{4}, \tag{2.40}$$

and for sufficiently large n there exists a blocklength-$\tilde{n}$, rate-$\tilde{R}$ non-feedback coding scheme for the memoryless Gaussian point-to-point channel with noise variance σ^2, with expected average block-power no larger than $\tilde{P}$ and with probability of error P_e

satisfying

$$P_e \leq e^{-\tilde{n}(E(\tilde{R},\tilde{P}/\sigma^2)-\epsilon')} \tag{2.41}$$

for some fixed $\epsilon' > 0$ and

$$E(\tilde{R},\tilde{P}) = \frac{\tilde{P}}{4\sigma^2}\left(1-\sqrt{1-2^{-2\tilde{R}}}\right). \tag{2.42}$$

Now apply this statement to $\tilde{R} = R_{\text{phase},2} = \frac{R}{\epsilon - 1/n}$, $\sigma^2 = 1$, $\tilde{P} = P/\gamma_1$ and $\tilde{n} = \epsilon n - 1$. Since for sufficiently large n and for all rates

$$R_{\text{phase},2} < \frac{1}{2}\log_2 \frac{2+\sqrt{P^2/\gamma_1^2+4}}{4}, \tag{2.43}$$

we conclude by (2.40–2.42) that there exists a code $\mathcal{C}_2$ of rate-$R_{\text{phase},2}$, block-power P/γ_1, blocklength $(\epsilon n - 1)$ and probability of error γ_2 satisfying

$$\begin{aligned}\gamma_2 &\leq e^{-\left(\epsilon n - 1\right)\left(\frac{P}{4\gamma_1}\left(1-\sqrt{1-2^{-2\frac{R}{\epsilon-1/n}}}\right)-\epsilon'\right)} \\ &= \exp(-\exp(\Omega(n))).\end{aligned} \tag{2.44}$$

Chapter 3

Broadcast Channel without Feedback

In this chapter we review some previous results on the capacity region of memoryless BC without feedback.

Consider a two-receiver memoryless BC with private and common messages depicted in Figure 3.1. The setup is characterized by the input alphabet $\mathcal{X}$, the output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ and a probability transition function $P_{Y_1Y_2|X}(y_1, y_2|x)$. If at time i, for $i \in \{1, \ldots, n\}$, the transmitter sends the channel input $x_i \in \mathcal{X}$, then Receiver $k \in \{1, 2\}$ observes the output $Y_{k,i} \in \mathcal{Y}_k$. The BC is said to be memoryless if $P_{Y_1^n Y_2^n|X^n}(y_1^n, y_2^n|x^n) = \prod_{t=1}^n P_{Y_1Y_2|X}(y_{1,i}, y_{2,i}|x_i)$.

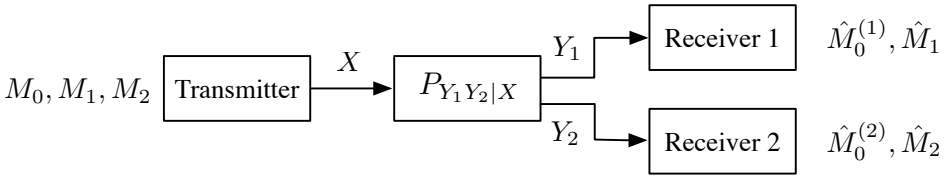


Figure 3.1: Two-receiver memoryless BC with private and common messages

The goal of the communication is that the transmitter conveys a common message $M_0 \in \{1, \ldots, \lfloor 2^{nR_0} \rfloor\}$ to both receivers, and two private messages $M_1 \in \{1, \ldots, \lfloor 2^{nR_1} \rfloor\}$ and $M_2 \in \{1, \ldots, \lfloor 2^{nR_2} \rfloor\}$, to Receiver 1 and 2, respectively. Each M_k, $k \in \{0, 1, 2\}$, is independently and uniformly distributed over the set $\mathcal{M}_k := \{1, \ldots, \lfloor 2^{nR_k} \rfloor\}$, where R_0

denotes the common rate of transmission and R_k, $k \in \{1,2\}$ denotes the private rate of transmission of Receiver i.

The encoder maps the messages (M_0, M_1, M_2) to a sequence $x^n \in \mathcal{X}^n$:

$$X^n = f^{(n)}(M_0, M_1, M_2) \tag{3.1}$$

where the encoding function $f^{(n)}$ is of form $\mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n$ and the Receiver $k \in \{1,2\}$ uses channel outputs y_k^n to estimate $(\hat{M}_0^{(k)}, \hat{M}_k)$ as a guess of messages (M_0, M_k):

$$(\hat{M}_0^{(k)}, \hat{M}_k) = g_k^{(n)}(Y_k^N). \tag{3.2}$$

where the decoding function $g_k^{(n)}$ is of form $\mathcal{Y}_k^n \to \mathcal{M}_0 \times \mathcal{M}_k$.

A rate region (R_0, R_1, R_2) is called achievable if for every blocklength n, there exists an encoding function $f^{(n)}$ and two decoding functions $g_1^{(n)}, g_2^{(n)}$ such that the error probability

$$P_e^{(n)} := \Pr\Big[(\hat{M}_0^{(1)}, \hat{M}_0^{(2)}, \hat{M}_1, \hat{M}_2) \neq (M_0, M_0, M_1, M_2)\Big] \tag{3.3}$$

tends to zero as the blocklength n tends to infinity. The closure of the set of achievable rate tuple (R_0, R_1, R_2) is called the *capacity region* and is denoted by $\mathcal{C}_{\text{NoFb}}$.

3.1 Capacity for $R_1 = R_2 = 0$

For the case of $R_1 = R_2 = 0$, where the transmitter only conveys a common message to the receivers, the capacity is

$$R_0 = \max_{P_X} \min\{I(X;Y_1), I(X;Y_2)\}. \tag{3.4}$$

In the following we recall the proof for the capacity for $R_1 = R_2 = 0$, see also [132].

Achievability

1) Codebook generation: Randomly and independently generate $\lfloor 2^{nR_0} \rfloor$ sequences $x^n(m_0)$. Each sequence is drawn according to the product distribution $\prod_{i=1}^n P_X(x_i)$, where x_i

denotes the i-th entry of $x^n(m_0)$. All codebooks are revealed to the transmitter and receivers.

2) Encoding: Given that $M_0 = m_0$, the transmitter sends $x^n(m_0)$.

3) Decoding at Receivers: We describe the operations performed at Receiver 1. Receiver 2 estimates $\hat{m}_0^{(2)}$ of messages M_0 in an analogous way.

Given that Receiver 1 observes the sequence y_1^n, it looks for an index $\hat{m}_0^{(1)}$ such that

$$\left(x^n(\hat{m}_0^{(1)}), y_1^n\right) \in \mathcal{T}_\varepsilon^n(P_{XY_1}).$$

If there is exactly one index $\hat{m}_0^{(1)}$ that satisfies the above condition, Receiver 1 chooses this index. If there are multiple such indices, it chooses one of them uniformly at random. Otherwise it chooses an index uniformly at random over the entire set $\{1, \ldots, \lfloor 2^{nR_0} \rfloor\}$.

4) Analysis: We analyze the average error probability over the random message, codebooks, and channel realizations. To simplify exposition we therefore assume that $M_0 = 1$. Under this assumption, an error occurs if, and only if,

$$(\hat{M}_0^{(1)}, \hat{M}_0^{(2)}) \neq (1, 1).$$

For $k \in \{1, 2\}$, define the following events.

- Let $\mathcal{E}_{0,k}$ be the event that

$$\left(X^n(1), Y_k^n\right) \notin \mathcal{T}_\varepsilon^n(P_{XY_k}).$$

  Since the channel is memoryless, by the law of large numbers, $\Pr[\mathcal{E}_{0,k}]$ tends to 0 as $n \to \infty$.

- Let $\mathcal{E}_{1,k}$ be the event that there is an index $\hat{m}_0^{(k)} \neq 1$ such that

$$\left(X^n(\hat{m}_0^{(k)}), Y_k^n\right) \in \mathcal{T}_\varepsilon^n(P_{XY_k}).$$

  By the Packing lemma [131], $\Pr[\mathcal{E}_{1,k}|\mathcal{E}_{0,k}]$ tends to 0 as $n \to \infty$ if

$$R_0 \leq I(X; Y_k) - \delta(\varepsilon) \tag{3.5}$$

An error in the communication occurs only if one or more of the the above events

happens. Thus we conclude that for a suitable $\delta(\varepsilon)$ which tends to 0 as $\varepsilon \to 0$, if

$$R_0 \leq I(X;Y_1) - \delta(\varepsilon) \tag{3.6}$$

$$R_0 \leq I(X;Y_2) - \delta(\varepsilon) \tag{3.7}$$

then the average error probability for this setup tends to 0 as n tends to infinity. Taking $\varepsilon \to 0$, we get the achievable region as shown in (3.4). Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(n)}$ tends to 0 as $n \to \infty$ when the mentioned conditions are satisfied.

Converse

Note that

$$\begin{aligned}
nR_0 &= H(M_0) \\
&\leq I(M_0; \hat{M}_0^{(1)}) + n\epsilon_n \\
&\leq I(X^n; Y_1^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} \big(H(Y_{1,i}|Y_1^{i-1}) - H(Y_1^n|X^n) \big) + n\epsilon_n \\
&= \sum_{i=1}^{n} \big(H(Y_{1,i}|Y_1^{i-1}) - H(Y_{1,i}|X_i) \big) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \big(H(Y_{1,i}) - H(Y_{1,i}|X_i) \big) + n\epsilon_n \\
&\leq nI(\bar{X}; \bar{Y}_1) + n\epsilon_n
\end{aligned} \tag{3.8}$$

where the first inequality follows by Fano's inequality and ϵ_n is a positive function that tends to 0 as $n \to \infty$; the second inequality follows by the data processing inequality; the last inequality follows by the convexity of mutual information and by setting

$$P_{\bar{X}\bar{Y}_1\bar{Y}_2}(x, y_1, y_2) = \Big(\frac{1}{n} \sum_{i=1}^{n} P_{X_i}(x) \Big) P_{Y_1Y_2|X}(y_1, y_2|x) \tag{3.9}$$

for all appropriate x, y_1, y_2. Similarly, we have $nR_0 \leq nI(\bar{X}; \bar{Y}_2)$. Thus, we have

$$R_0 = \max_{P_X} \min\{I(X; Y_1), I(X; Y_2)\}.$$

3.2 Capacity Region for $R_0 = 0$

The capacity region of DMBCs with $R_0 = 0$ is in general unknown. There are some exceptions. For the following classes of broadcast channels, where one of the two receivers is stronger than the other receiver in some sense, the capacity region is known and can be achieved by *superposition coding*:

- *stochastically or physically degraded DMBCs [133]*
- *less noisy DMBCs [134]*
- *more capable DMBCs [134].*
- *essentially less noisy DMBCs [135]*
- *essentially more capable DMBCs [135]*

Definition 1 (From [133]). *A DMBC is called* physically degraded *if*

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

More generally, a DMBC is called stochastically degraded *if there exists a distribution* $p'(y_2|y_1)$ *such that*

$$p(y_2|x) = \sum_{y_1} p(y_1|x)p'(y_2|y_1),$$

The capacity region of stochastically or physically degraded DMBCs is given by [136, 137], which is the closure set of all nonnegative rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq I(U; Y_1) \\ R_1 + R_2 &\leq I(U; Y_1) + I(X; Y_2|U) \end{aligned} \tag{3.10}$$

for some probability mass function (pmf) P_{UX} such that $U - X - (Y_1, Y_2)$ forms a Markov chain.

Definition 2 (From [134]). *A DMBC is called* less noisy *if* $I(U; Y_2) \geq I(U; Y_1)$ *holds for all* P_{UX}.

The capacity region of a less noisy BC has the same expression as that of a degraded BC [134], see (3.10).

Definition 3 (From [134]). *A DMBC is called* more capable *if* $I(X;Y_2) \geq I(X;Y_1)$ *holds for all* P_X.

The capacity region of a more capable BC is the closure set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U;Y_1) \tag{3.11}$$

$$R_1 + R_2 \leq I(U;Y_1) + I(X;Y_2|U) \tag{3.12}$$

$$R_1 + R_2 \leq I(X;Y_2) \tag{3.13}$$

for some pmf P_{UX} such that $U - X - (Y_1, Y_2)$ forms a Markov chain [134].

Before defining essentially less noisy/more capable DMBCs, we need to first introduce a new term *sufficient class* of pmfs.

Definition 4 (From [135]). *A subset* $\mathcal{P}_{\mathcal{X}}$ *of all pmfs on the input alphabet* $\mathcal{X}$ *is said to be a* sufficient class of pmfs *for a DMBC if the following holds: Given any joint pmf* P_{UVX} *there exists a joint pmf* P'_{UVX} *that satisfies*

$$\begin{aligned} P'_X(x) &\in \mathcal{P}_{\mathcal{X}} \\ I_P(U;Y_1) &\leq I_{P'}(U;Y_1) \\ I_P(V;Y_2) &\leq I_{P'}(V;Y_2) \\ I_P(U;Y_1) + I_P(X;Y_2|U) &\leq I_{P'}(U;Y_1) + I_{P'}(X;Y_2|U) \\ I_P(V;Y_2) + I_P(X;Y_1|V) &\leq I_{P'}(V;Y_2) + I_{P'}(X;Y_1|V) \end{aligned} \tag{3.14}$$

where the notations I_P *and* $I_{P'}$ *indicate that the mutual informations are computed assuming that* $(U,V,X) \sim P_{UVX}$ *and* $(U,V,X) \sim P'_{UVX}$ *and* $P'_X(x)$ *is the marginal obtained from* P'_{UVX}.

Now we are ready to define essentially less noisy/more capable DMBCs.

Definition 5 (From [135]). *A DMBC is called* essentially less noisy *if there exists a sufficient class of pmfs* $\mathcal{P}_{\mathcal{X}}$ *such that whenever* $P_X \in \mathcal{P}_{\mathcal{X}}$, *then for all random variables*

U that form the Markov chain $U - X - (Y_1, Y_2)$,

$$I(U; Y_1) \leq I(U; Y_2). \tag{3.15}$$

The capacity region [135] of an essentially less-noisy BC is the closure set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U; Y_1) \tag{3.16}$$

$$R_1 + R_2 \leq I(U; Y_1) + I(X; Y_2|U) \tag{3.17}$$

for pmfs P_{UX} such that $U - X - (Y_1, Y_2)$ forms a Markov chain and $P_X \in \mathcal{P}$.

Definition 6 (From [135])**.** *A DMBC is called* essentially more capable *if there exists a sufficient class of pmfs $\mathcal{P}_{\mathcal{X}}$ such that whenever $P_X \in \mathcal{P}_{\mathcal{X}}$, then for all random variables (U, X) that form the Markov chain $U - X - (Y_1, Y_2)$,*

$$I(X; Y_1|U) \leq I(X; Y_2|U). \tag{3.18}$$

The capacity region [135] of an essentially more capable BC is the closure set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U; Y_1) \tag{3.19}$$

$$R_1 + R_2 \leq I(U; Y_1) + I(X; Y_2|U) \tag{3.20}$$

$$R_1 + R_2 \leq I(X; Y_2) \tag{3.21}$$

for some pmf P_{UX} such that $U - X - (Y_1, Y_2)$ forms a Markov chain and $P_X \in \mathcal{P}$.

Remark 3.1. *The relationship among these various classes of BCs is established in [135], [138]:*

- *degraded $\subsetneq$ less-noisy $\subsetneq$ more capable.*
- *less noisy $\subsetneq$ essentially less noisy.*
- *essentially less-noisy $\nsubseteq$ more capable.*
- *more capable $\nsubseteq$ essentially less noisy.*
- *more capable $\subsetneq$ essentially more capable.*

- *essentially more capable* $\not\subseteq$ *essentially less noisy.*
- *essentially less noisy* $\not\subseteq$ *essentially more capable.*

3.3 Achievable Regions for Broadcast Channels

3.3.1 Superposition Coding

Superposition coding is optimal for the classes of BCs mentioned in Section 3.2. In the following we describe the superposition coding scheme (see also [139, 132]) that achieves rate region $\mathcal{R}^{(1)}_{\text{SuperPos}}$, which is defined as the set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U;Y_1) \tag{3.22a}$$

$$R_2 \leq I(X;Y_2|U) \tag{3.22b}$$

$$R_1 + R_2 \leq I(X;Y_2) \tag{3.22c}$$

for some pmf P_{UX}. The superposition coding region $\mathcal{R}^{(2)}_{\text{SuperPos}}$, which is defined similarly to $\mathcal{R}^{(1)}_{\text{SuperPos}}$ but where indices 1 and 2 need to be exchanged, is achieved if in the following scheme message M_1 and M_2 are exchanged as well as Receiver 1 and 2.

1) Codebook generation: Fix a pmf P_{UX}. Randomly and independently generate $\lfloor 2^{nR_1} \rfloor$ cloud center codewords: $u^n(m_1)$, for $m_1 \in \{1, \ldots, \lfloor 2^{nR_1} \rfloor\}$. Each codeword is drawn according to the product distribution $\prod_{i=1}^n P_U(u_i)$, where u_i denotes the i-th entry of $u^n(m_1)$.

For each codeword $u^n(m_1)$, randomly generate $\lfloor 2^{nR_2} \rfloor$ satellite codewords $x^n(m_2|m_1)$, for $m_2 \in \{1, \ldots, \lfloor 2^{nR_2} \rfloor\}$. Each codeword $x^n(m_2|m_1)$ is drawn according to the product distribution $\prod_{i=1}^n P_{X|U}(x_i|u_i)$, where x_i denotes the i-th entry of $x^n(m_2|m_1)$. All codebooks are revealed to the transmitter and receivers.

2) Encoding: Given that $M_1 = m_1$, $M_2 = m_2$, the transmitter sends $x^n(m_2|m_1)$.

3) Decoding at Receivers: After observing the channel outputs y_1^n, Receiver 1 looks for an index $\hat{m}_1^{(1)}$ such that

$$\left(u^n(\hat{m}_1^{(1)}), y_1^n\right) \in \mathcal{T}_\varepsilon^n(P_{UY_1}).$$

If there is exactly one index $\hat{m}_1^{(1)}$ that satisfies the above condition, Receiver 1 chooses this index. If there are multiple such indices, it chooses one of them uniformly at random. Otherwise it chooses an index uniformly at random over the entire set $\{1, \ldots, \lfloor 2^{nR_1} \rfloor\}$.

Given observing channel outputs y_2^n, Receiver 2 looks for the pair $(\hat{m}_1^{(2)}, \hat{m}_2)$ such that

$$\left(u^n(\hat{m}_1^{(2)}), x^n(\hat{m}_2|\hat{m}_1^{(2)}), y_2^n\right) \in \mathcal{T}_\varepsilon^n(P_{UXY_2}).$$

If there is exactly one pair $(\hat{m}_1^{(2)}, \hat{m}_2)$ that satisfies the above condition, Receiver 2 chooses this pair. If there are multiple such pair, it chooses one of them uniformly at random. Otherwise it chooses a pair uniformly at random over the entire set $\{1, \ldots, \lfloor 2^{nR_1} \rfloor\} \times \{1, \ldots, \lfloor 2^{nR_2} \rfloor\}$.

4) Analysis: We analyze the average error probability of the superposition coding scheme over the random messages, codebooks, and channel realizations. To simplify exposition we assume that $M_1 = M_2 = 1$. Under this assumption, an error occurs if, and only if,

$$(\hat{M}_1^{(1)}, \hat{M}_1^{(2)}, \hat{M}_2) \neq (1, 1, 1).$$

Define the following events.

- Let $\mathcal{E}_0$ be the event that

  $$\left(U^n(1), X^n(1|1), Y_1^n, Y_2^n\right) \in \mathcal{T}_{\varepsilon/16}^n(P_{UXY_1Y_2}).$$

  Since the channel is memoryless, by the law of large numbers, $\Pr[\mathcal{E}_0]$ tends to 0 as $n \to \infty$.
- Let $\mathcal{E}_1$ be the event that there is an index $\hat{m}_2 \neq 1$ such that

  $$\left(U^n(1), X^n(\hat{m}_2|1), Y_2^n\right) \in \mathcal{T}_\varepsilon^n(P_{UXY_2}).$$

  By the Packing Lemma, $\Pr[\mathcal{E}_1|\mathcal{E}_0^c]$ tends to 0 as $n \to \infty$ if

  $$R_2 \leq I(X; Y_2|U) - \delta(\varepsilon), \tag{3.23}$$

- Let $\mathcal{E}_2$ be the event that there is a pair $(\hat{m}_1^{(2)}, \hat{m}_2) \neq (1,1)$ such that

$$\left(U^n(\hat{m}_1^{(2)}), X^n(\hat{m}_2|\hat{m}_1^{(2)}), Y_2^n\right) \in \mathcal{T}_\varepsilon^n(P_{UXY_2}).$$

  By the Packing Lemma, $\Pr[\mathcal{E}_2|\mathcal{E}_0^c]$ tends to 0 as $n \to \infty$ if

$$R_1 + R_2 \leq I(X;Y_2) - \delta(\varepsilon), \tag{3.24}$$

- Let $\mathcal{E}_3$ be the event that there is an index $\hat{m}_1^{(1)} \neq 1$ such that

$$\left(U^n(\hat{m}_1^{(2)}), Y_1^n\right) \in \mathcal{T}_\varepsilon^n(P_{UY_1}).$$

  By the Packing Lemma $\Pr[\mathcal{E}_3|\mathcal{E}_0^c]$ tends to 0 as $n \to \infty$ if

$$R_1 \leq I(U;Y_1) - \delta(\varepsilon), \tag{3.25}$$

An error in the communication occurs only if one or more of the the above events happens. Thus we conclude that for a suitable $\delta(\varepsilon)$ which tends to 0 as $\varepsilon \to 0$, if

$$R_1 \leq I(U;Y_1) - \delta(\varepsilon) \tag{3.26}$$

$$R_2 \leq I(X;Y_2|U) - \delta(\varepsilon) \tag{3.27}$$

$$R_1 + R_2 \leq I(X;Y_2) - \delta(\varepsilon), \tag{3.28}$$

then the average error probability for this setup tends to 0 as n tends to infinity. Taking $\varepsilon \to 0$, we get the achievable region as shown in (3.22). Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(n)}$ tends to 0 as $n \to \infty$ when the mentioned conditions are satisfied.

3.3.2 Marton's Coding

The best known inner bound on the capacity of DMBC without feedback is Marton's region [140], $\mathcal{R}_{\text{Marton,general}}$, which is defined as the set of all nonnegative rate tuples

(R_0, R_1, R_2) satisfying

$$R_0 + R_1 \le I(U_0, U_1; Y_1) \tag{3.29a}$$

$$R_0 + R_2 \le I(U_0, U_2; Y_2) \tag{3.29b}$$

$$R_0 + R_1 + R_2 \le I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0) \tag{3.29c}$$

$$R_0 + R_1 + R_2 \le I(U_0, U_2; Y_2) + I(U_1; Y_1|U_0) - I(U_1; U_2|U_0) \tag{3.29d}$$

$$2R_0 + R_1 + R_2 \le I(U_0, U_1; Y_1) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0) \tag{3.29e}$$

for some pmf $P_{U_0U_1U_2}$ and a function $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$ such that $X = f(U_0, U_1, U_2)$.

Remark 3.2. *To evaluate the region $\mathcal{R}_{\text{Marton,general}}$, it suffices to consider distributions $P_{U_0U_1U_2X}$ for which one of the following conditions holds [141, 142]:*

- $I(U_0; Y_1) = I(U_0; Y_2)$*;*
- $I(U_0; Y_1) < I(U_0; Y_2)$ *and* $U_1 = \text{const}$*;*
- $I(U_0; Y_1) > I(U_0; Y_2)$ *and* $U_2 = \text{const}$*.*

Proof. The proof is given in [141, 142]. For convenience to the reader, we reprove the statement in Appendix 3.A. □

When $R_0 = 0$, the general Marton's region $\mathcal{R}_{\text{Marton,general}}$ specializes to $\mathcal{R}_{\text{Marton}}$, which is the set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \le I(U_0, U_1; Y_1) \tag{3.30a}$$

$$R_2 \le I(U_0, U_2; Y_2) \tag{3.30b}$$

$$R_1 + R_2 \le I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0) \tag{3.30c}$$

$$R_1 + R_2 \le I(U_0, U_2; Y_2) + I(U_1; Y_1|U_0) - I(U_1; U_2|U_0) \tag{3.30d}$$

for some pmf $P_{U_0U_1U_2}$ and a function $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$ such that $X = f(U_0, U_1, U_2)$.

Remark 3.3. *The region $\mathcal{R}_{\text{Marton}}$ specializes to the superposition coding region, $\mathcal{R}^{(1)}_{\text{SuperPos}}$ when choosing $U_1 = const.$ and $X = U_2$ and is specialized to $\mathcal{R}^{(2)}_{\text{SuperPos}}$ when choosing $U_2 = const.$ and $X = U_1$ in Marton's constraints* (3.30).

We describe the Marton scheme achiving $\mathcal{R}_{\text{Marton,general}}$ for a DMBC $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, P_{Y_1Y_2|X})$. Choose nonnegative rates R_1', R_2', auxiliary finite alphabets $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2$, a function $f^{(n)}$

of the form $f^{(n)}$: $\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and a pmf $P_{U_0U_1U_2}$. Define $\mathcal{T}_k := \{1, \ldots, \lfloor 2^{nR'_k} \rfloor\}$, for $k \in \{0, 1, 2\}$. Split the private messages into "common" and "private" parts: $M_k = (M_{c,k}, M_{p,k})$, for $k \in \{1, 2\}$. The submessages $M_{c,k}$ and $M_{p,k}$ are independently and uniformly distributed over the sets $\mathcal{M}_{c,k} := \{1, \ldots, \lfloor 2^{nR_{c,k}} \rfloor\}$ and $\mathcal{M}_{p,k} := \{1, \ldots, \lfloor 2^{nR_{p,k}} \rfloor\}$, respectively, where $R_{p,k}, R_{c,k} > 0$ and so that $R_k = R_{p,k} + R_{c,k}$[1] . Let $\mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $R_c := (R_0 + R_{c,1} + R_{c,2})$.

1) Codebook generation: Randomly and independently generate $\lfloor 2^{nR_c} \rfloor$ sequences $u_0^n(m_0, m_{c,1}, m_{c,2})$. Each sequence is drawn according to the distribution $\prod_{i=1}^n P_{U_0}(u_{0,i})$, where $u_{0,i}$ denotes the i-th entry of $u_0^n(m_0, m_{c,1}, m_{c,2})$.

For $k \in \{1, 2\}$ and each tuple $(m_0, m_{c,1}, m_{c,2})$ randomly generate $\lfloor 2^{n(R_{p,k}+R'_k)} \rfloor$ sequences $u_k^n(m_{p,k}, t_k | m_0, m_{c,1}, m_{c,2})$, for $m_{p,k} \in \mathcal{M}_{p,k}$ and $t_k \in \mathcal{T}_k$. Each codeword $u_k^n(m_{p,k}, t_k | m_0, m_{c,1}, m_{c,2})$ is drawn according to the distribution $\prod_{t=1}^n P_{U_k|U_0}(u_{k,i}|u_{0,i})$, where $u_{k,i}$ denotes the i-th entry of $u_k^n\big(m_{p,k}, t_k | m_0, m_{c,1}, m_{c,2}\big)$.

All codebooks are revealed to the transmitter and receivers.

2) Encoding: Assume that $M_0 = m_0$, $M_{c,k} = m_{c,k}$ and $M_{p,k} = m_{p,k}$, for $k \in \{1, 2\}$. Define $\mathbf{m}_c := (m_0, m_{c,1}, m_{c,2})$. The transmitter looks for a pair $(t_1, t_2) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\big(u_0^n(\mathbf{m}_c), u_1^n(m_{p,1}, t_1 | \mathbf{m}_c), u_2^n(m_{p,2}, t_2 | \mathbf{m}_c)\big) \in \mathcal{T}_{\varepsilon'/16}^n(P_{U_0U_1U_2}). \tag{3.31}$$

If there is exactly one pair (t_1, t_2) that satisfies the above condition, the transmitter chooses this pair. If there are multiple such pairs, it chooses one of them uniformly at random. Otherwise it chooses a pair (t_1, t_2) uniformly at random over the entire set $\mathcal{T}_1 \times \mathcal{T}_2$. The transmitter then sends the inputs $x^n = (x_1, \ldots, x_n)$, where

$$x_i = f^{(n)}(u_{0,i}, u_{1,i}, u_{2,i}), \qquad i \in \{1, \ldots, n\}, \tag{3.32}$$

and $u_{0,i}$, $u_{1,i}$, $u_{2,i}$ denote the i-th symbols of the chosen Marton codewords $u_0^n(\mathbf{m}_c)$, $u_1^n(m_{p,1}, t_1 | \mathbf{m}_c)$ and $u_2^n(m_{p,2}, t_2 | \mathbf{m}_c)$, respectively.

3) Decoding at Receivers: We describe the operations performed at Receiver 1.

[1] Due to the floor operations, R_1 and R_2 here do not exactly represent the transmission rates of messages M_1 and M_2. In the limit $n \to \infty$, which is our case of interest, R_1 and R_2 however approach these transmission rates. Therefore, we neglect this technicality in the following.

Receiver 2 estimates $(\hat{m}_0^{(2)}, \hat{m}_2)$ of messages (m_0, m_2) in an analogous way. Given that Receiver 1 observes the sequence y_1^n, it looks for all the tuples $(\hat{\mathbf{m}}_c^{(1)}, \hat{m}_{p.1}, \hat{t}_1)$ such that

$$\left(u_0^n(\hat{\mathbf{m}}_c^{(1)}), u_1^n(\hat{m}_{p,1}, \hat{t}_1|\hat{\mathbf{m}}_c^{(1)}), y_1^n\right) \in \mathcal{T}_\varepsilon^n(P_{U_0U_1Y_1}).$$

4) Analysis of Marton's Scheme: We analyze the average error probability of Marton's scheme over the random messages, codebooks, and channel realizations, see also [140, 141, 56]. To simplify exposition we assume that $M_0 = M_{c,k} = M_{p,k} = T_k = 1$ for all $k \in \{1, 2\}$. Under this assumption, an error occurs if, and only if,

$$(\hat{\mathbf{M}}_c^{(1)}, \hat{\mathbf{M}}_c^{(2)}, \hat{M}_{p,1}, \hat{M}_{p,2}) \neq (\mathbf{1}_{[3]}, \mathbf{1}_{[3]}, 1, 1).$$

Define the following events.

- Let $\mathcal{E}_0$ be the event that there is no pair $(t_1, t_2) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\left(U_0^n(\mathbf{1}_{[3]}), U_1^n(1, t_1|\mathbf{1}_{[3]}), U_2^n(1, t_2|\mathbf{1}_{[3]})\right) \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2}).$$

  By the Covering Lemma [131], $\Pr[\mathcal{E}_0]$ tends to 0 as $n \to \infty$ if

$$R_1' + R_2' \geq I(U_1; U_2|U_0) + \delta(\varepsilon). \tag{3.33}$$

- Let $\mathcal{E}_{1,k}$ be the event that

$$\left(U_0^n(\mathbf{1}_{[3]}), U_k^n(1, 1|\mathbf{1}_{[3]}), Y_k^n\right) \notin \mathcal{T}_\varepsilon^n(P_{U_0U_kY_k}).$$

  Since the channel is memoryless, by the law of large numbers, $\Pr[\mathcal{E}_{1,k}|\mathcal{E}_0^c]$ tends to 0 as $n \to \infty$.
- Let $\mathcal{E}_{2,k}$ be the event that there exists $\hat{\mathbf{m}}_c^{(k)} \neq \mathbf{1}_{[3]}$ that satisfies

$$\left(U_0^n(\hat{\mathbf{m}}_c^{(k)}), U_k(1, 1|\hat{\mathbf{m}}_c^{(k)}), Y_k^n\right) \in \mathcal{T}_\varepsilon^n(P_{U_0U_kY_k}).$$

  By the Packing Lemma, $\Pr\left[\mathcal{E}_{2,k}|\mathcal{E}_{1,k}^c\right]$ tends to 0 as $n \to \infty$, if

$$R_0 + R_{c,k} + R_{c,k} \leq I(U_0, U_k; Y_k) - \delta(\varepsilon). \tag{3.34}$$

- Let $\mathcal{E}_{3,k}$ be the event that there is a pair $(\hat{m}_{p,k}, \hat{t}_k) \neq (1,1)$ that satisfies

$$\left(U_0^n(\mathbf{1}_{[3]}), U_k^n(\hat{m}_{p,k}, \hat{t}_k|\mathbf{1}_{[3]}), Y_k^n\right) \in \mathcal{T}_\varepsilon^n(P_{U_0U_kY_k}).$$

  By the Packing Lemma, $\Pr\left[\mathcal{E}_{3,k}|\mathcal{E}_{1,k}^c\right]$ tends to 0 as $n \to \infty$, if

$$R_{p,k} + R'_k \leq I(U_k; Y_k|U_0) - \delta(\varepsilon). \tag{3.35}$$

- Let $\mathcal{E}_{4,k}$ be the event that there is a tuple $\hat{\mathbf{m}}_c^{(k)} \neq \mathbf{1}_{[3]}$ and $(\hat{m}_{p,k}, \hat{t}_k) \neq (1,1)$ that satisfies

$$\left(U_0^n(\hat{\mathbf{m}}_c^{(k)}), U_k^n(\hat{m}_{p,k}, \hat{t}_k|\hat{\mathbf{m}}_c^{(k)}), Y_k^n\right) \in \mathcal{T}_\varepsilon^n(P_{U_0U_kY_k}).$$

  By the Packing Lemma, $\Pr\left[\mathcal{E}_{4,k}|\mathcal{E}_{1,k}^c\right]$ tends to 0 as $n \to \infty$, if

$$R_0 + R_{c,1} + R_{c,2} + R_{p,k} + R'_k \leq I(U_0, U_k; Y_k) - \delta(\varepsilon). \tag{3.36}$$

An error in the communication occurs only if one or more of the above events happens. Then we conclude that if for $k \in \{1,2\}$

$$R'_1 + R'_2 \geq I(U_1; U_2|U_0) + \delta(\varepsilon) \tag{3.37}$$

$$R_{p,k} + R'_k \leq I(U_k; Y_k|U_0) - \delta(\varepsilon) \tag{3.38}$$

$$R_0 + R_{c,1} + R_{c,2} + R_{p,k} + R'_k \leq I(U_0, U_k; Y_k) - \delta(\varepsilon). \tag{3.39}$$

then the average error probability of Marton's scheme tends to 0 as n tends to infinity.

By Fourier-Motzkin elimination, we conclude that whenever

$$I(U_1; Y_1|U_0) + I(U_2; Y_2|U_0) \geq I(U_1; U_2|U_0) \tag{3.40}$$

then for every rate tuple (R_0, R_1, R_2) satisfying

$$R_0 + R_1 \leq I(U_0, U_1; Y_1) - \delta(\varepsilon) \tag{3.41a}$$

$$R_0 + R_2 \leq I(U_0, U_2; Y_2) - \delta(\varepsilon) \tag{3.41b}$$

$$R_0 + R_1 + R_2 \leq I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0) - \delta(\varepsilon) \tag{3.41c}$$

$$R_0 + R_1 + R_2 \leq I(U_0, U_2; Y_2) + I(U_1; Y_1|U_0) - I(U_1; U_2|U_0) - \delta(\varepsilon) \tag{3.41d}$$

$$2R_0 + R_1 + R_2 \leq I(U_0, U_1; Y_1) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0) - \delta(\varepsilon) \tag{3.41e}$$

for a suitable $\delta(\varepsilon)$ which tends to 0 as $\varepsilon \to 0$, there exists a choice of nonnegative rates $R_1, R_2, R_0, R_1', R_2'$ such that $R_k = R_{p,k} + R_{c,k}$, for $k \in 1, 2$ and (3.33)–(3.39) hold. Notice that we can ignore the rate constraint (3.40) because the rate region achieved by any random variables (U_0, U_1, U_2) violating (3.40) can be strictly enlarged by choosing $U_0' = (U_0, U_1, U_2)$ and $U_1' = U_2' = \text{const.}$. The new choice (U_0', U_1', U_2') satisfies (3.40). Thus we conclude that the rate region in (3.29) is achievable also when (3.40) is not satisfied. Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(n)}$ tends to 0 as $n \to \infty$ when the mentioned conditions are satisfied.

3.4 Outer Bound on Broadcast Channel without Feedback

In [143], Liang, Kramer and Shamai proposed the New-Jersey outer bound based on the ideas of [122, 141, 144]. This outer bound was later simplified by Nair [145], which can be written as the closure set of all nonnegative rate tuples such that

$$R_0 \leq \min\{I(U_0; Y_1), I(U_0; Y_2)\} \tag{3.42a}$$

$$R_0 + R_1 \leq I(U_1; Y_1|U_0) + \min\{I(U_0; Y_1), I(U_0; Y_2)\} \tag{3.42b}$$

$$R_0 + R_2 \leq I(U_2; Y_2|U_0) + \min\{I(U_0; Y_1), I(U_0; Y_2)\} \tag{3.42c}$$

$$R_0 + R_1 + R_2 \leq I(U_1; Y_1|U_0, U_2) + I(U_2; Y_2|U_0) + \min\{I(U_0; Y_1), I(U_0; Y_2)\} \tag{3.42d}$$

$$R_0 + R_1 + R_2 \leq I(U_2; Y_2|U_0, U_1) + I(U_1; Y_1|U_0) + \min\{I(U_0; Y_1), I(U_0; Y_2)\} \tag{3.42e}$$

for some pmf $P_{U_1 U_1 U_2}$ and a function $X = f(U_0, U_1, U_2)$.

For the BC with no common message, i.e. with $R_0 = 0$, the New-Jersey outer bound reduces to the following outer bound [144, 145] that is the closure set of all nonnegative rate pairs such that

$$R_1 \leq I(U_1; Y_1) \tag{3.43a}$$

$$R_2 \leq I(U_2; Y_2) \tag{3.43b}$$

$$R_1 + R_2 \leq I(U_1; Y_1) + I(X; Y_2|U_1) \tag{3.43c}$$

$$R_1 + R_2 \leq I(U_2; Y_2) + I(X; Y_1|U_2) \tag{3.43d}$$

for some pmf $P_{U_1U_2}$ and a function $X = f(U_1, U_2)$. In [146] Geng, *et al.* presented an example of a product BC which shows that this outer bound is not tight in general.

3.A Appendix: Proof of Remark 3.2

Fix a distribution $P_{U_0U_1U_2X}$. We prove that there exists a distribution $P_{U_0'U_1'U_2'X'}$ that satisfies one of the three conditions in Remark 3.2 and so that the rate region defined by Marton's constraints (3.29) and distribution $P_{U_0'U_1'U_2'X'}$ contains the rate region defined by Marton's constraints (3.29) and distribution $P_{U_0U_1U_2X}$.

We assume without loss of generality that $I(U_0; Y_1) \leq I(U_0; Y_2)$, and we separately treat the two cases

- $I(U_0, U_1; Y_1) \leq I(U_0, U_1; Y_2)$
- $I(U_0, U_1; Y_1) > I(U_0, U_1; Y_2)$.

For the first case, $I(U_0, U_1; Y_1) \leq I(U_0, U_1; Y_2)$, let $U_0' = (U_0, U_1)$, $U_1' = \text{const.}$, $U_2' = U_2$ and $X' = X$. Evaluating Marton's constraints (3.29) for the auxiliaries (U_0', U_1', U_2', X') results in

$$R_0 + R_1 \leq I(U_0, U_1; Y_1) \tag{3.44a}$$

$$R_0 + R_1 + R_2 \leq I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0, U_1) \tag{3.44b}$$

$$R_0 + R_1 + R_2 \leq I(U_0, U_1, U_2; Y_2) \tag{3.44c}$$

Note that the constraint (3.44c) is redundant in view of (3.44b) because $I(U_0, U_1; Y_1) \leq I(U_0, U_1; Y_2)$.

We show that the first two constraints in (3.44) are no tighter than Marton's constraints in (3.29), which proves the desired result for the first case. In fact, the constraint (3.44a) coincides with Marton's constraint (3.29a). The constraint in (3.44b) is looser than Marton's constraint (3.29c),

$$\begin{aligned}
&I(U_0,U_1;Y_1)+I(U_2;Y_2|U_0,U_1)\\
&\quad=I(U_0,U_1;Y_1)+H(U_2|U_0,U_1)-H(U_2|U_0,U_1,Y_2)\\
&\quad\geq I(U_0,U_1;Y_1)+H(U_2|U_0,U_1)-H(U_2|U_0,Y_2)\\
&\quad=I(U_0,U_1;Y_1)+H(U_2|U_0,U_1)-H(U_2|U_0,Y_2)\\
&\qquad+H(U_2|U_0)-H(U_2|U_0)\\
&\quad=I(U_0,U_1;Y_1)+I(U_2;Y_2|U_0)-I(U_1;U_2|U_0).
\end{aligned}$$

We now treat the second case $I(U_0,U_1;Y_1) > I(U_0,U_1;Y_2)$. Since $I(U_0;Y_1) < I(U_0;Y_2)$ by assumption and by the continuity of mutual information, there exists a deterministic function f such that

$$I(U_0,f(U_1);Y_1)=I(U_0,f(U_1);Y_2) \tag{3.45}$$

Let now $U_0'=(U_0,f(U_1))$, $U_1'=U_1$, $U_2'=U_2$ and $X'=X$. For this choice of auxiliaries, Marton's constraints (3.29) result in:

$$R_0+R_1\leq I(U_0,f(U_1),U_1;Y_1) \tag{3.46a}$$

$$R_0+R_2\leq I(U_0,f(U_1),U_2;Y_2) \tag{3.46b}$$

$$\begin{aligned}
R_0+R_1+R_2\leq\ & I(U_0,f(U_1),U_1;Y_1)+I(U_2;Y_2|U_0,f(U_1))\\
&-I(U_1;U_2|U_0,f(U_1))
\end{aligned} \tag{3.46c}$$

$$\begin{aligned}
R_0+R_1+R_2\leq\ & I(U_1;Y_1|U_0,f(U_1))+I(U_0,f(U_1),U_2;Y_2)\\
&-I(U_1;U_2|U_0,f(U_1))
\end{aligned} \tag{3.46d}$$

$$\begin{aligned}
2R_0+R_1+R_2\leq\ & I(U_0,f(U_1),U_1;Y_1)+I(U_0,f(U_1),U_2;Y_2)\\
&-I(U_1;U_2|U_0,f(U_1))
\end{aligned} \tag{3.46e}$$

Note that the constraints (3.46c) and (3.46d) coincide because $I(U_0,f(U_1);Y_1)=I(U_0,f(U_1);Y_2)$.

We again show that these constraints are no tighter than Marton's constraints in (3.29), which proves the desired result also for this second case and concludes the proof. The constraint (3.46a) coincides with Marton's constraint (3.29a):

$$I(U_0, f(U_1), U_1; Y_1) = I(U_0, U_1; Y_1).$$

The constraint (3.46b) is looser than constraint (3.29b):

$$I(U_0, f(U_1), U_2; Y_2) \geq I(U_0, U_2; Y_2).$$

The constraints in (3.46c) is looser than Marton's constraint (3.29c):

$$\begin{aligned}
&I(U_0, f(U_1), U_1; Y_1) - I(U_1; U_2|U_0, f(U_1)) + I(U_2; Y_2|U_0, f(U_1)) \\
&\quad = I(U_0, U_1; Y_1) - H(U_2|U_0, f(U_1)) + H(U_2|U_0, U_1) \\
&\qquad + H(U_2|U_0, f(U_1)) - H(U_2|U_0, f(U_1), Y_2) \\
&\quad = I(U_0, U_1; Y_1) + H(U_2|U_0, U_1) - H(U_2|U_0, f(U_1), Y_2) \\
&\qquad + H(U_2|U_0) - H(U_2|U_0) \\
&\quad \geq I(U_0, U_1; Y_1) + H(U_2|U_0, U_1) - H(U_2|U_0, Y_2) \\
&\qquad + H(U_2|U_0) - H(U_2|U_0) \\
&\quad = I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0). \qquad (3.47)
\end{aligned}$$

The constraints in (3.46e) is looser than Marton's constraint (3.29e):

$$\begin{aligned}
&I(U_0, f(U_1), U_1; Y_1) + I(U_0, f(U_1), U_2; Y_2) - I(U_1; U_2|U_0, f(U_1)) \\
&\quad \geq I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0) + I(U_0, f(U_1); Y_2) \\
&\quad = I(U_0, U_1; Y_1) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0) + I(U_0, f(U_1); Y_2) - I(U_0; Y_2) \\
&\quad \geq I(U_0, U_1; Y_1) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0)
\end{aligned}$$

where the first inequality follows from the inequality (3.47).

Chapter 4

Previous Results on Memoryless Broadcast Channels with Feedback

In this chapter, we review some previous works most relevant to ours on achievable regions and outer bound for memoryless BC with feedback.

4.1 Achievable Regions for Memoryless Broadcast Channels with Feedback

4.1.1 Discrete Memoryless Broadcast Channels with Feedback

Consider Dueck's channel depicted in Fig. 4.1. The channel outputs are

$$Y_1 = (X_0, X_1 \oplus Z), \quad Y_2 = (X_0, X_2 \oplus Z) \tag{4.1}$$

where $X = (X_0, X_1, X_2)$ is the channel input with $X_0, X_1, X_2 \in \{0, 1\}$ and Z is the noise with $Z \sim \text{Bern}(1/2)$. It is easy to show that for this channel the maximum sum-rate is $R_1 + R_2 = 1$. In the presence of perfect feedback, Dueck [55] showed that the rate pair $(R_1, R_2) = (1, 1)$ is achievable by a simple scheme. To prove the achievability, it suffices to show that the transmitter can perfectly convey two i.i.d. Bern$(1/2)$ sequences X_1^n and

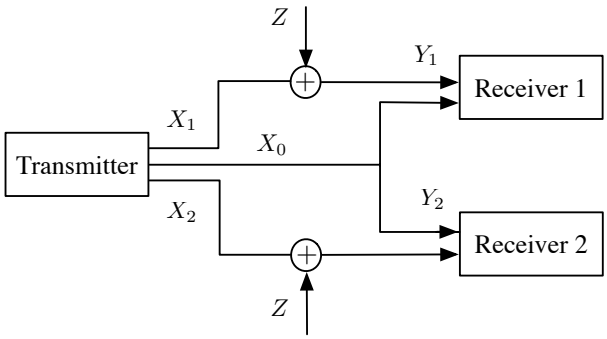


Figure 4.1: Dueck's channel

X_2^n to Receiver 1 an 2, respectively, in $n+1$ channel uses. The proof is as follows. In the first channel use, the transmitter sends $(0, X_{1,1}, X_{2,1})$ over the channel. Upon obtaining the outputs $Y_{1,1}$ and $Y_{2,1}$, the receivers send them back to the transmitter. Since the feedback links are noiseless, the transmitter can recover the noise Z_1 in the first channel use. In the second channel use, the transmitter sends $(Z_1, X_{1,2}, X_{2,2})$. In view of the outputs $Y_{k,1} = (0, X_{k,1} \oplus Z_1)$ and $Y_{k,2} = (Z_1, X_{k,2} \oplus Z_2)$, Receiver $k \in \{1, 2\}$ can recover $X_{k,1}$ perfectly. After $n + 1$ transmissions, Receiver k finally recovers Z^n and therefore can determine its intended sequence X_k^n perfectly. Thus, the rate $(R_1, R_2) = (1, 1)$ is achievable as $n \to \infty$. Dueck's channel is the first example demonstrating that feedback can increase the capacity region of BC.

In [11] Kramer applied the idea of directed information to establish a multi-letter achievable region for the DMBC with feedback. It is shown that when evaluating this achievable region for a BSBC with specific parameters, perfect feedback increases the non-feedback capacity.

Recall that in Dueck's example, the transmitter at each time broadcasts common information about the channel of previous time (obtained through feedback). This common information helps the receivers recover previous messages at a rate higher than that without feedback. Inspired by this idea, Shayevitz and Wigger proposed a scheme [56], based on block-Markov strategy, Marton's coding and Gray-Wyner coding [147], for the two-receiver DMBC with generalized feedback.

In the Shayevitz-Wigger scheme, the transmitter in each block uses Marton's coding

to send fresh messages together with *refinement* information. This refinement information is about the messages sent in the previous block and is generated using a generalization of Gray-Wyner coding. The receivers perform backward decoding where the refinement information decoded in a block is used to decode the refinement information and the messages sent in the previous block.

Given a two-receiver DMBC with generalized feedback that consists of an input alphabet $\mathcal{X}$, two output alphabets $(\mathcal{Y}_1, \mathcal{Y}_2)$, a feedback alphabet $\mathcal{F}$, and a conditional pmf $P_{Y_1Y_2F|X}$, the Shayevitz-Wigger region is characterized by the set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \le I(U_0, U_1; Y_1, V_1|Q) - I(U_0, U_1, U_2, F; V_0, V_1|Q, Y_1) \tag{4.2a}$$

$$R_2 \le I(U_0, U_2; Y_2, V_2|Q) - I(U_0, U_1, U_2, F; V_0, V_2|Q, Y_2) \tag{4.2b}$$

$$\begin{aligned} R_1 + R_2 \le\ & I(U_1; Y_1, V_1|Q, U_0) + I(U_2; Y_2, V_2|Q, U_0) + \min_{k\in\{1,2\}} I(U_0; Y_k, V_k|Q) \\ & - \max_{k\in\{1,2\}} I(U_0, U_1, U_2, F; V_0|Q, Y_k) - I(U_0, U_1, U_2, F; V_1|Q, V_0, Y_1) \\ & - I(U_0, U_1, U_2, F; V_2|Q, V_0, Y_2) - I(U_1; U_2|Q, U_0) \end{aligned} \tag{4.2c}$$

$$\begin{aligned} R_1 + R_2 \le\ & I(U_0, U_1; Y_1, V_1|Q) + I(U_0, U_2; Y_2, V_2|Q) - I(U_1; U_2|Q, U_0) \\ & - I(U_0, U_1, U_2, F; V_0, V_1|Q, Y_1) - I(U_0, U_1, U_2, F; V_0, V_2|Q, Y_2) \end{aligned} \tag{4.2d}$$

for some pmf $P_Q P_{U_0U_1U_2|Q} P_{V_0V_1V_2|U_0U_1U_2Y_1Y_2Q}$ and some function $f\colon \mathcal{Q}\times\mathcal{U}_0\times\mathcal{U}_1\times\mathcal{U}_2 \to \mathcal{X}$, where $X = f(U_0, U_1, U_2, Q)$.

This region is generally difficult to evaluate due to the presence of seven auxiliary random variables in the rate constraints. Recently, Kim, Chia, and El Gamal [57] studied the more general Shayevitz-Wigger region for generalized feedback [56], which differs from the above region only in that in some places the outputs Y_1 or Y_2 have to be replaced by the generalized feedback output F. In particular, they proposed two choices for the auxiliary random variables for the Shayevitz-Wigger region with generalized feedback and presented simplified expressions for the maximum sum-rates that these choices achieve for symmetric state-dependent DMBCs with state known at both receivers and where the generalized feedback equals the delayed state sequence.

Their first choice is given by

$$Q = \begin{cases} 0 & \text{w. p. } 1-2p \\ 1 & \text{w. p. } p \\ 2 & \text{w. p. } p \end{cases}, \tag{4.3a}$$

$$V_0 = V_1 = V_2 = \begin{cases} \emptyset & \text{if } Q = 0 \\ Y_1 & \text{if } Q = 1 \\ Y_2 & \text{if } Q = 2 \end{cases}, \tag{4.3b}$$

and

$$X = \begin{cases} U_0 & \text{if } Q = 0 \\ U_1 & \text{if } Q = 1 \\ U_2 & \text{if } Q = 2 \end{cases} \tag{4.3c}$$

for joint pmf $P_{U_0U_1U_2} = P_{U_0}P_{U_1}P_{U_2}$. This choice essentially results in a coded time-sharing scheme. Their second choice is

$$Q = \begin{cases} 1 & \text{w. p. } 1/2 \\ 2 & \text{w. p. } 1/2 \end{cases}, \tag{4.4a}$$

$$V_0 = V_1 = V_2 = \begin{cases} Y_1 & \text{if } Q = 1 \\ Y_2 & \text{if } Q = 2 \end{cases}, \tag{4.4b}$$

and

$$X = \begin{cases} U_1 & \text{if } Q = 1 \\ U_2 & \text{if } Q = 2 \end{cases} \tag{4.4c}$$

for some joint pmf $P_{U_0U_1U_2} = P_{U_0}P_{U_1|U_0}P_{U_2|U_0}$. This choice results in a randomized superposition coding scheme.

The deterministic model of the two-user i.i.d. fading BC with delayed CSI at the transmitter considered by Maddah-Ali and Tse [60], and the two-user BEBCs where *all* erasure events are known at all receivers considered by Wang [58] and by Georgiadis

and Tassiulas [59][1] belong to this class of DMBCs. In fact, it is shown in [57] that the capacity-achieving schemes in [60], [58], and [59] are special cases of a simplified version—without block-Markov coding—of the Shayevitz-Wigger scheme for generalized feedback specialized to the choice of parameters in (4.3).

On a more general note, also various results for two-user multi-input single-output (MISO) Gaussian fading BCs with delayed CSI at the transmitter [60], [77], [80], [79], [81] and [82] are related to the Shayevitz-Wigger scheme, if this latter is extended to continuous-valued channels. Maddah-Ali and Tse [60] studied the K-user MISO Gaussian fading BC where the transmitter learns the CSI with one unit delay. They presented a coding scheme for general stationary and ergodic fading processes, and proved that in some special cases it achieves the optimal DoF region. This is in particular the case when the fadings are i.i.d. and independent across users. For i.i.d. fading sequences (even when correlated across users), the setup studied by Maddah-Ali and Tse can be modeled as a (continuous-valued) memoryless BC, and their scheme is a special case of a simplified Shayevitz-Wigger scheme without block-Markov coding [56], when this latter is naturally extended to continuous-valued channels.

Yang, Kobayashi, Gesbert, and Yi [79] studied the more general setup where the transmitter also obtains imperfect (rate-limited) CSI. Specifically, they modified and improved Maddah-Ali&Tse's scheme to apply to this more general setup, and showed that, under some mild assumptions, their improved scheme achieves the optimal DoF region for arbitrary stationary and ergodic fading processes and all qualities of current CSI at the transmitter. In this sense, they could bridge the gap between Maddah-Ali&Tse's no current-CSI result and the standard perfect current-CSI DoF result where zero-forcing is DoF optimal. The new components in the Yang-Kobayashi-Gesbert-Yi scheme are: a clever power allocation strategy; the fact that fresh information is sent in all blocks; precoding of transmit signals; and transmission of quantized versions of interferences as common message.

Chen and Elia [77] considered an even more general setup where the transmitter accumulates CSI about each fading sample over time. In their model there is thus current CSI as well as various levels of predicted and delayed CSI. Chen and Elia

[1] Notice that the main contribution of [58, 59] is not the capacity-achieving schemes for the described two-user BEBC, but its extension to arbitrary number of users $K \geq 2$.

proposed a coding scheme that exploits current and delayed CSI, and they showed that with a careful choice of parameters, under some mild assumptions, their scheme achieves the optimal DoF region. Interestingly this optimal DoF region depends only on the current CSI and the asymptotic acquisition rate of delayed CSI. Given these two quantities the DoF region is in particular independent of the predicted CSI. Their results also illustrate the tradeoff between the required current and asymptotic delayed CSI to achieve a certain performance. Compared to the Yang-Kobayashi-Gesbert-Yi scheme, the Chen&Elia scheme shows as new components: block-Markov strategy and a more sophisticated power- and rate-allocation policy.

In the special case without predicted or current CSI and when all delayed CSI is acquired with only one unit delay, the setups considered by Yang, Kobayashi, Gesbert, and Yi and by Chen and Elia can be modeled as (continuous-valued) memoryless BCs with generalized feedback. Under these assumptions the Yang-Kobayashi-Gesbert-Yi scheme and the Chen&Elia scheme specialize to simplified versions of the Shayevitz-Wigger scheme for generalized feedback. In fact, the specialized Yang-Kobayashi-Gesbert-Yi scheme [79] then corresponds to the Shayevitz-Wigger scheme without block-Markov strategy, with successive decoding instead of the more performant joint decoding, and with the following choice of auxiliaries $Q = \{1, 2\}$; $X = U_0 + U_1 + U_2$; $U_0 - U_1 - U_2$ form a Markov chain; $U_0 = \text{const.}$ when $Q = 1$ and U_0 arbitrary when $Q = 2$; $V_0 = V_1 = V_2 = (\hat{\eta}_1, \hat{\eta}_2)$ where $(\hat{\eta}_1, \hat{\eta}_2)$ are defined in [79]. The specialized Chen&Elia scheme [77] corresponds to the Shayevitz-Wigger scheme but with only successive decoding and with the following choice of auxiliaries: $X = U_0 + U_1 + U_2$; $U_0 - U_1 - U_2$ form a Markov chain; and $V_0 = V_1 = V_2 = (\bar{\bar{l}}^{(1)}, \bar{\bar{l}}^{(2)})$ where $(\bar{\bar{l}}^{(1)}, \bar{\bar{l}}^{(2)})$ are defined in (54) in [77] and need to be specialized to the assumption of no current CSI.

Recently, another achievable region was proposed independently by Venkataramanan and Pradhan [61]. The scheme present is based on block-Markov strategy and Marton's coding. More specifically, in their scheme, the transmitter in each block uses Marton's coding to encode fresh messages into random variables U_1 and U_2 intended for Receiver 1 and 2, respectively. In next block, the transmitter sends refinement information using new random variables (A, B, C), where (A, C) is intended for Receiver 1 and (B, C) is for Receiver 2. This scheme is similar to Shayevitz-Wigger's scheme but here it applies joint source-channel coding and forward decoding.

Comparing the three general regions–Kramer's region, the Shayevitz-Wigger region and the Venkataramanan-Pradhan region, with each other is difficult due to their complex form which involves several auxiliary random variables.

For the two-receiver Gaussian BC with perfect feedback, Ozarow and Leung [62] proposed a scheme based on the Schalkwijk-Kailath scheme. It is shown that the Ozarow-Leung scheme can enlarge the non-feedback capacity region and improve the communication reliability. Their scheme is summarized as follows. In the initialization phase (occupying two channel uses), two private messages intended for Receiver 1 and 2, respectively, are mapped into real numbers and sent individually over the channel. Upon obtaining the channel outputs, each receiver makes temporary guess of its desired message. After the initialization phase, the transmitter iteratively sends the errors of the MMSE of the transmitted messages given the previous channel outputs obtained through feedback. When receiving a linear combination of these errors together with a correlated noise term, the receivers estimate the noise occurred in the initialization phase and update the guess of their desired messages. This scheme was later extended by Kramer to more than two receivers case [11]. Note that the Ozarow-Leung scheme and Kramer's scheme both belong to linear-feedback scheme and exhibit doubly exponential decay in the probability of error.

4.1.2 Memoryless Gaussian Broadcast Channels with Feedback

For the $K \geq 2$-receiver Gaussian BC with perfect feedback, Ardestanizadeh, Minero, and Franceschetti [13] proposed a coding scheme based on the theory of linear quadratic Gaussian (LQG) optimal control. Similar to the Schalkwijk-Kailath scheme and Kramer's scheme, in this LQG scheme the transmitter iteratively sends refinement of the estimate of the messages, but here the estimate is based on a stabilizing control instead of MMSE. The performance of this LQG scheme depends on the noise correlation at the receivers. When the noises at the receivers are independent, this scheme recovers the region proposed by Elia [64] and strictly improves over the Ozarow-Leung scheme and Kramer's scheme. When the noises are correlated, the *prelog* of the sum capacity is determined by the rank of noise correlation matrix and can be strictly larger than 1. Specifically, for correlation matrix of rank L, the prelog is at most $(K-L+1)$. Conversely, there exists a noise correlation matrix of rank $L \leq K$ for which the LQG coding scheme achieves

the prelog $(K-L+1)$. This generalizes a result by Wigger and Gastpar [148], to the case of more than two receivers.

More recently, Selma, Yossef, and Wigger [16, 17] investigated the duality between multi-antenna Gaussian MAC and BC with perfect feedback. For these two channels, a class of linear-feedback coding schemes was proposed that achieves the linear-feedback capacity region, i.e. the set of all rate pairs achieved by linear-feedback schemes. This scheme achieves the best known achievable region and includes as special cases the schemes by Ardestanizadeh *et al.* [13] and by Elia [64].

Given a two-receiver single-antenna Gaussian BC that is characterized by input-power constraint P, independent noises $Z_1 \sim \mathcal{N}(0,1)$, $Z_2 \sim \mathcal{N}(0,1)$, and channel coefficients h_1, h_2, from [16, 17], it is known that the linear-feedback capacity region is given by

$$\mathcal{R}_{\text{linfb}}(h_1, h_2, P) = \bigcup_{\rho \in [0,1]} \mathcal{R}^*_{\text{linfb}}(h_1, h_2, P, \rho) \tag{4.5a}$$

where for each $\rho \in [0,1]$, $\mathcal{R}^*_{\text{linfb}}(h_1, h_2, P, \rho)$ is the set of all nonnegative rate-pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log_2 \left(1 + h_1^2 \alpha P (1-\rho^2)\right) \tag{4.5b}$$

$$R_2 \leq \frac{1}{2} \log_2 \left(1 + h_2^2 (1-\alpha) P (1-\rho^2)\right) \tag{4.5c}$$

$$R_1 + R_2 \leq \frac{1}{2} \log_2 \left(1 + h_1^2 \alpha P + h_2^2 (1-\alpha) P + 2\sqrt{\alpha(1-\alpha)} h_1 h_2 P \rho\right) \tag{4.5d}$$

for arbitrary α with $0 \leq \alpha \leq 1$.

4.2 Outer Bound on Broadcast Channel with Feedback

Cut-set Outer Bound

Applying the Theorem 14.10.1 in [139], the cut-set outer bound $\mathcal{R}_{\text{outer,cut}}$ can be obtained as the closure set of all nonnegative rate pairs such that

$$R_1 \leq I(X;Y_1) \tag{4.6a}$$

$$R_2 \leq I(X;Y_2) \tag{4.6b}$$

$$R_1 + R_2 \leq I(X;Y_1,Y_2) \tag{4.6c}$$

for some pmf $P_X P_{Y_1Y_2|X}$.

Ozarow-Leung Outer Bound

A simple outer bound on the capacity with output feedback is given in [62]. It equals the capacity region $\mathcal{C}_{\text{Enh}}^{(1)}$ of an enhanced DMBC where the outputs Y_1^n are also revealed to Receiver 2. Notice that this enhanced DMBC is physically degraded and thus, with and without feedback, its capacity region is given by the set of all nonnegative rate pairs (R_1, R_2) that satisfy

$$R_1 \leq I(U;Y_1) \tag{4.7a}$$

$$R_2 \leq I(X;Y_1,Y_2|U) \tag{4.7b}$$

for some pmf $P_{UX} P_{Y_1Y_2|X}$.

Exchanging everywhere in the previous paragraph indices 1 and 2, we can define a similar enhanced capacity region $\mathcal{C}_{\text{Enh}}^{(2)}$, which is also an outer bound to $\mathcal{C}_{\text{Fb}}$. The intersection $\mathcal{C}_{\text{Enh}}^{(1)} \cap \mathcal{C}_{\text{Enh}}^{(2)}$ yields an even tighter outerbound [58, 59].

Chapter 5

Reliability of the Gaussian BC with Common Message and Feedback

5.1 Introduction

In this chapter we show that linear-feedback schemes with a message point, in the spirit of Schalkwijk-Kailath's scheme, are strictly suboptimal for the K-user memoryless Gaussian BC with common message and fail to achieve capacity. As a consequence, for this setup, linear-feedback schemes also fail to achieve double-exponential decay of the probability of error for rates close to capacity. To our knowledge, this is the first example of a memoryless Gaussian network with perfect feedback, where linear-feedback schemes with message points are shown to be strictly suboptimal. In all previously studied networks with perfect feedback, they attained the optimal performance or the best so far performance. (In case of noisy feedback, they are known to perform badly even in the memoryless Gaussian point-to-point channel [149].)

In the asymptotic scenario of infinitely many receivers $K \to \infty$, the performance of linear-feedback schemes with a message point completely collapses: the largest rate that is achievable with these schemes tends to 0 as $K \to \infty$. This latter result holds under some mild assumptions regarding the variances of the noises experienced at the

receivers, which are for example met when all the noise variances are equal. Notice that, in contrast, the capacity of the K-user Gaussian BC with common message does not tend to 0 as $K \to \infty$ when e.g., all the noise variances are equal. In this case, the capacity does not depend on K, because it is simply given by the point-to-point capacity to the receiver with the largest noise variance.

That the performance of linear-feedback schemes with a common message point degenerates with increasing number of users K is intuitively explained as follows. At each time instant, the transmitter sends a linear combination of the message point and past noise symbols. Resending the noise symbols previously experienced at some Receiver k can be beneficial for this Receiver k because it allows it to mitigate the noise corrupting previous outputs. However, resending these noise symbols is of no benefit for all other Receivers $k' \neq k$ and only harms them. Therefore, the more receivers there are, the more noise symbols the transmitter sends in each channel use that are useless for a given Receiver k. Our result hinges upon the independence of the noises at difference receivers. In the case of correlated noises a noise symbol can be beneficial to multiple receivers. In the extreme case where all noises are identical, for instance, the BC degenerates to a point-to-point channel and Schalkwijk-Kailath's scheme is capacity achieving.

For the memoryless Gaussian point-to-point channel [2] and MAC [10], the (sum-)capacity achieving linear-feedback schemes with message points transmit in each channel use a scaled version of the linear minimum mean square estimation LMMSE errors of the message points given the previous channel outputs. The same strategy is however strictly suboptimal—even among the class of linear-feedback schemes with message points—when sending private messages over a Gaussian BC [13]. It is unknown whether LMMSE estimates are optimal among linear-feedback schemes when sending a common message over the Gaussian BC.

In our proof that any linear-feedback scheme with a message point cannot achieve the capacity of the Gaussian BC with common message, the following proposition is key: For any sequence of linear-feedback schemes with a *common* message point that achieves rate $R > 0$, one can construct a sequence of linear-feedback schemes that achieves the rate tuple $R_1 = \ldots = R_K = R$ when sending K *private* message points with a linear-feedback scheme. This proposition shows that the class of linear-feedback

schemes with message points cannot take advantage of the fact that all the $K \geq 2$ receivers are interested in the same message.

To contrast the bad performance of linear-feedback schemes, we present a coding scheme that exploits the feedback in a intermittent way (only in few time slots the receivers send feedback signals) and that achieves double-exponential decay of the probability of error for all rates up to capacity. In our scheme it suffices to have rate-limited feedback with feedback rate R_{Fb} no smaller than the forward rate R. If the feedback rate $R_{\text{Fb}} < R$ then, even for the setup with only one receiver, the probability of error can decay only exponentially in the blocklength [9]. This implies immediately that also for the $K \geq 2$ receivers BC with common message no double-exponential decay in the probability of error is achievable when $R_{\text{Fb}} < R$. When the feedback rate $R_{\text{Fb}} > (L-1)R$, for some positive integer L, then our intermittent-feedback scheme can achieve an L-th order exponential decay in the probability of error. That means, it achieves a probability of error of the form $P_{\text{e}}^{(n)} = \exp(-\exp(\exp(\ldots \exp(\Omega(n)))))$, where there are L exponential terms and where $\Omega(n)$ denotes a function that satisfies $\underline{\lim}_{n\to\infty} \frac{\Omega(n)}{n} > 0$.

Our intermittent-feedback scheme is inspired by the scheme in [9] for the memoryless Gaussian point-to-point channel with rate-limited feedback. Also the schemes in [150] and [151] for the memoryless Gaussian point-to-point channel with perfect feedback are related. In fact, in our scheme communication takes place in L phases. In the first phase, the transmitter uses a Gaussian code of power P to send the common message to the K Receivers. The transmission in phase $l \in \{2, \ldots, L\}$ depends on the feedback signals. After each phases $l \in \{1, \ldots, L-1\}$ each Receiver k feeds back a temporary guess of the message. Now, if one receiver's temporary guesses after phase $(l-1)$ is wrong, then in phase l the transmitter resends the common message using a new code. If all receivers' temporary guesses after phase $(l-1)$ were correct, in phase l the transmitter sends the all-zero sequence. In this latter case, no power is consumed in phase l. The receivers' final guess is their temporary guess after phase L.

The fact that the described scheme can achieve an L-th order decay of the probability of error, roughly follows from the following inductive argument. Assume that the probability of the event "one of the receivers' guesses is wrong after phase l", for $l \in \{1, \ldots, L-1\}$, has an l-th order exponential decay in the blocklength. Then, when

sending the common message in phase $l+1$, the transmitter can use power that is l-th order exponentially large in the blocklength without violating the *expected* average blockpower constraint. With such a code, in turn, the probability that after phase $l+1$ one of the receivers has a wrong guess can have an $(l+1)$-th order exponential decay in the blocklength.

The rest of the chapter is organized as follows. Section 5.2 describes the Gaussian BC with common message and defines the class of linear-feedback schemes with a message point. Section 5.3 introduces the Gaussian BC with private messages and defines the class of linear-feedback schemes with private message points. Section 5.4 presents our main results. Finally, Sections 5.5 and 5.6 contain the proofs of our Theorems 5.1 and 5.2.

5.2 Setup

5.2.1 System Model and Capacity

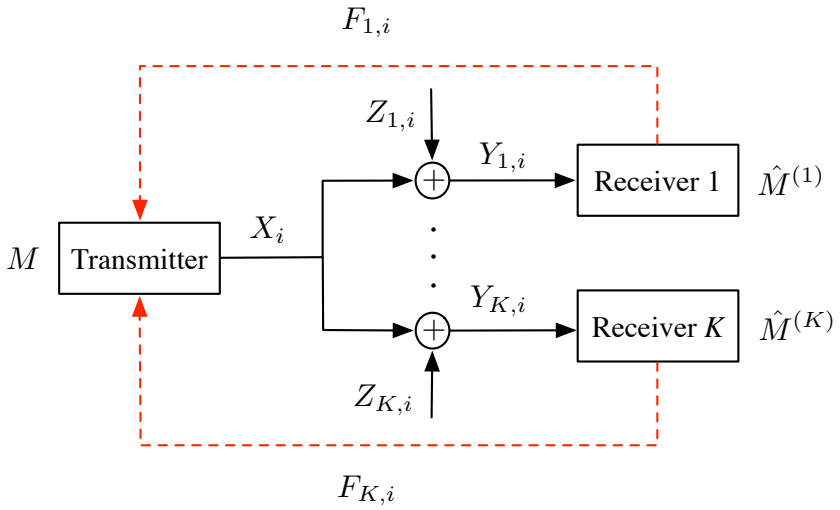


Figure 5.1: K-receiver Gaussian BC with feedback and common message only.

We consider the $K \geq 2$-receiver Gaussian BC with common message and feedback depicted in Figure 5.1. Specifically, if X_i denotes the transmitter's channel input at

time-i, the channel output at Receiver $k \in \mathcal{K} := \{1, \ldots, K\}$ is

$$Y_{k,i} = X_i + Z_{k,i} \tag{5.1}$$

where $\{Z_{k,i}\}_{i=1}^n$ models the additive noise at Receiver k. The sequence of noises $\{(Z_{1,i}, \ldots, Z_{K,i})\}_{i=1}^n$ is a sequence of i.i.d. centered Gaussian vectors, each of diagonal covariance matrix

$$\mathsf{K}_z = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_K^2 \end{pmatrix}. \tag{5.2}$$

Without loss of generality, we assume that

$$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_K^2. \tag{5.3}$$

The transmitter wishes to convey a common message M to all receivers, where M is uniformly distributed over the message set $\mathcal{M} := \{1, ..., \lfloor 2^{nR} \rfloor\}$ independent of the noise sequences $\{Z_{1,i}\}_{i=1}^n, \ldots, \{Z_{K,i}\}_{i=1}^n$. Here, n denotes the blocklength and $R > 0$ the rate of transmission. It is assumed that the transmitter has either rate-limited or perfect feedback from all receivers. That means, after each channel use $i \in \{1, \ldots, n\}$, each Receiver $k \in \mathcal{K}$ feeds back a signal $F_{k,i} \in \mathcal{F}_{k,i}$ to the transmitter. The feedback alphabet $\mathcal{F}_{k,i}$ is a design parameter of the scheme. In the case of rate-limited feedback, the signals from Receiver k have to satisfy:

$$\sum_{i=1}^{n} H(F_{k,i}) \leq nR_{\mathrm{Fb}}, \qquad k \in \mathcal{K} \tag{5.4}$$

where R_{Fb} denotes the symmetric feedback rate. In the case of perfect feedback, we have no constraint on the feedback signals $\{F_{k,i}\}_{i=1}^n$, and it is thus optimal to choose $\mathcal{F}_{k,i} = \mathbb{R}$ and

$$F_{k,i} = Y_{k,i}, \tag{5.5}$$

because in this way any processing that can be done at the receivers can also be done at the transmitter.

An encoding strategy is comprised of a sequence of encoding functions $\{f_i^{(n)}\}_{i=1}^n$ of

the form

$$f_i^{(n)} \colon \mathcal{M} \times \mathcal{F}_1^{i-1} \times \ldots \times \mathcal{F}_K^{i-1} \to \mathbb{R} \tag{5.6}$$

that is used to produce the channel inputs as

$$X_i = f_i^{(n)}(M, F_1^{i-1}, \ldots, F_K^{i-1}), \qquad i \in \{1, \ldots, n\}. \tag{5.7}$$

We impose an expected average block-power constraint P on the channel input sequence:

$$\frac{1}{n} \mathsf{E}\left[\sum_{i=1}^{n} X_i^2\right] \leq P. \tag{5.8}$$

Each Receiver $k \in \mathcal{K}$ decodes the message M by means of a decoding function $g_k^{(n)}$ of the form

$$g_k^{(n)} \colon \mathbb{R}^n \to \mathcal{M}. \tag{5.9}$$

That means, Receiver k produces as its guess

$$\hat{M}^{(k)} = g_k^{(n)}(Y_k^n). \tag{5.10}$$

An error occurs in the communication if

$$(\hat{M}^{(k)} \neq M), \tag{5.11}$$

for some $k \in \mathcal{K}$. Thus, the average probability of error is

$$P_e^{(n)} := \Pr\left[\bigcup_{k \in \mathcal{K}} \left(\hat{M}^{(k)} \neq M\right)\right]. \tag{5.12}$$

We say that *a rate $R > 0$ is achievable* for the described setup if for every $\epsilon > 0$ there exists a sequence of encoding and decoding functions $\left\{\{f_i^{(n)}\}_{i=1}^n, \{g_k^{(n)}\}_{k=1}^K\right\}_{n=1}^\infty$ as in (5.6) and (5.9) and satisfying the power constraint (5.8) such that for sufficiently large blocklengths n the probability of error $P_e^{(n)} < \epsilon$. The supremum of all achievable rates is called the capacity. The capacity is the same in the case of perfect feedback, of rate-limited feedback (irrespective of the feedback rate R_{Fb}), and without feedback.

We denote it by C and by assumption (5.3) it is given by

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_1^2}\right). \tag{5.13}$$

Our main interest in this chapter is in the speed of decay of the probability of error at rates $R < C$.

Definition 7. *Given a positive integer L, we say that the L-th order exponential decay in the probability of error is achievable at a given rate $R < C$, if there exists a sequence of schemes of rate R such that their probabilities of error $\{P_e^{(n)}\}_{n=1}^{\infty}$ satisfy*

$$\varliminf_{n\to\infty} \frac{1}{n}\log\log\ldots\log(-\log P_e^{(n)}) > 0, \tag{5.14}$$

where the number of logarithms in (5.14) *is L.*

5.2.2 Linear-Feedback Schemes with a Message Point

When considering perfect feedback, we will be interested in the class of coding schemes where the feedback is only used in a linear fashion. Specifically, we say that a scheme is a linear-feedback scheme with a message point, if the sequence of encoding functions $\{f_i^{(n)}\}_{i=1}^{n}$ is of the form

$$f_i^{(n)} = \Phi^{(n)} \circ L_i^{(n)} \tag{5.15}$$

with

$$\Phi^{(n)} \colon M \mapsto \Theta^{(n)} \in \mathbb{R} \tag{5.16a}$$

$$L_i^{(n)} \colon (\Theta^{(n)}, Y_1^{i-1}, \ldots, Y_K^{i-1}) \mapsto X_i \tag{5.16b}$$

where $\Phi^{(n)}$ is an arbitrary function on the respective domains and $L_i^{(n)}$ is a *linear mapping* on the respective domains. There is no constraint on the decoding functions $g_1^{(n)}, \ldots, g_K^{(n)}$.

By the definition of a linear-feedback coding scheme in (5.16), for each blocklength n, if we define $\mathbf{X} = (X_1, \ldots, X_n)^\mathsf{T}$, $\mathbf{Y}_k = (Y_{k,1}, \ldots, Y_{k,n})^\mathsf{T}$, and $\mathbf{Z}_k = (Z_{k,1}, \ldots, Z_{k,n})^\mathsf{T}$,

for $k \in \mathcal{K}$, the channel inputs can be written as:

$$\mathbf{X} = \Theta^{(n)} \cdot \mathbf{d}^{(n)} + \sum_{k=1}^{K} \mathsf{A}_k^{(n)} \mathbf{Z}_k, \tag{5.17}$$

for some n-dimensional vector $\mathbf{d}^{(n)}$ and n-by-n strictly lower-triangular matrices $\mathsf{A}_1^{(n)}$, $\ldots, \mathsf{A}_K^{(n)}$. (The lower-triangularity of $\mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}$ ensures that the feedback is used in a strictly causal fashion.) Thus, for a given blocklength n, a linear-feedback scheme is described by the tuple

$$\Phi^{(n)}, \mathbf{d}^{(n)}, \mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}, g_1^{(n)}, \ldots, g_K^{(n)}. \tag{5.18}$$

It satisfies the average block-power constraint (5.8) whenever

$$\sum_{k=1}^{K} \|\mathsf{A}_k^{(n)}\|_F^2 \sigma_k^2 + \|\mathbf{d}^{(n)}\|^2 \mathsf{E}\left[|\Theta^{(n)}|^2\right] \leq nP. \tag{5.19}$$

The supremum of all rates that are achievable with a sequence of linear-feedback schemes with a message point is denoted by $C^{(\mathrm{Lin})}$.

5.3 For comparison: Setup with Private Messages and Perfect Feedback

5.3.1 System Model and Capacity Region

For comparison, we also discuss the scenario where the transmitter wishes to communicate a private message M_k to each Receiver $k \in \mathcal{K}$ over the Gaussian BC in Figure 5.1. The messages $M_1, \ldots, M_K$ are assumed independent of each other and of the noise sequences $\{Z_{1,i}\}_{i=1}^n, \ldots, \{Z_{K,i}\}_{i=1}^n$ and each M_k is uniformly distributed over the set $\mathcal{M}_k := \{1, \ldots, \lfloor 2^{nR_k} \rfloor\}$. For this setup we restrict attention to perfect feedback. Thus, here the channel inputs are produced as

$$X_i = f_{\mathrm{priv},i}^{(n)}(M_1, \ldots, M_K, Y_1^{i-1}, \ldots, Y_K^{i-1}),\ i \in \{1, \ldots, n\}. \tag{5.20}$$

Receiver k produces the guess

$$\hat{M}_k \;=\; g^{(n)}_{\text{priv},k}(Y^n_k) \tag{5.21}$$

where the sequence of decoding function $\{g^{(n)}_{\text{priv},k}\}_{k=1}^K$ is of the form

$$g^{(n)}_{\text{priv},k}\colon \mathbb{R}^n \to \{1,\ldots,\lfloor 2^{nR_k}\rfloor\}, \tag{5.22}$$

A rate tuple $(R_1,\ldots,R_K)$ is said to be achievable if for every blocklength n there exists a set of n encoding functions as in (5.20) satisfying the power constraint (5.8) and a set of K decoding functions as in (5.22) such that the probability of decoding error tends to 0 as the blocklength n tends to infinity, i.e.,

$$\lim_{n\to\infty} \Pr\left[(M_1,\ldots,M_K) \neq (\hat{M}_1,\ldots,\hat{M}_K)\right] = 0.$$

The closure of the set of all achievable rate tuples $(R_1,\ldots,R_K)$ is called the capacity region. We denote it $\mathcal{C}_{\text{private}}$. This capacity region is unknown to date. (The sum-capacity in the high-SNR asymptotic regime is derived in [14].) Achievable regions were presented in [13, 14, 15]; the tightest known outer bound on capacity for $K = 2$ users was presented in [10] based on the idea of revealing one of the output sequences to the other receiver. This idea generalizes to $K \geq 2$ users, and leads to the following outer bound [11, 152]:

Lemma 5.1. *If the rate tuple $(R_1,\ldots,R_K)$ lies in $\mathcal{C}_{\text{private}}$, then there exist coefficients $\alpha_1,\ldots,\alpha_K$ in the closed interval $[0,1]$ such that for each $k \in \mathcal{K}$,*

$$R_k \leq \frac{1}{2}\log_2\left(1 + \frac{\alpha_k P}{(1-\alpha_1-\ldots-\alpha_k)P + N_k}\right) \tag{5.23}$$

where

$$N_k = \left(\sum_{k'=1}^{k} \frac{1}{\sigma^2_{k'}}\right)^{-1}, \qquad k \in \mathcal{K}. \tag{5.24}$$

Proof. Let a genie reveal each output sequence Y^n_k to Receivers $k+1,\ldots,K$. The resulting BC is physically degraded, and thus its capacity is the same as without feedback [54] and known. Evaluating this capacity region readily gives the outer bound in

the lemma. □

5.3.2 Linear-Feedback Schemes with Message Points

A *linear-feedback scheme with message points* for this setup with independent messages consists of a sequence of K decoding functions as in (5.22) and of a sequence of encoding functions $\{f_{\text{priv},i}^{(n)}\}_{i=1}^{n}$ of the form

$$f_{\text{priv},i}^{(n)} = \Phi_{\text{priv}}^{(n)} \circ L_{\text{priv},i}^{(n)} \tag{5.25}$$

with

$$\Phi_{\text{priv}}^{(n)} \colon \begin{pmatrix} M_1 \\ \vdots \\ M_K \end{pmatrix} \mapsto \mathbf{\Theta} := \begin{pmatrix} \Theta_1 \\ \vdots \\ \Theta_K \end{pmatrix} \in \mathbb{R}^K \tag{5.26a}$$

$$L_{\text{priv},i}^{(n)} \colon (\mathbf{\Theta}, Y_1^{i-1}, \ldots, Y_K^{i-1}) \mapsto X_i \tag{5.26b}$$

where $\Phi_{\text{priv}}^{(n)}$ is an arbitrary function on the respective domains and $L_{\text{priv},i}^{(n)}$ is a *linear mapping* on the respective domains.

We denote the closure of the set of rate tuples $(R_1, \ldots, R_K)$ that are achievable with a linear-feedback scheme with message points by $\mathcal{C}_{\text{private}}^{(\text{Lin})}$. This region is unknown to date.

5.4 Main Results

The main question we wish to answer is whether for the Gaussian BC with common message a super-exponential decay in the probability of error is achievable for all rates $R < C$. We first show that the class of linear-feedback schemes with message point fails in achieving this goal even with perfect feedback, because it does not achieve capacity (Theorem 5.1 and Corollary 5.1). As the number of receivers K increases, the largest rate that is achievable with linear-feedback schemes with a message point vanishes (Proposition 5.2). However, as we show then, a super-exponential decay in the probability of error is still possible by means of an intermittent feedback scheme similar to [9] (Theorem 5.2).

Proposition 5.1. *If a sequence of linear-feedback schemes with a message point achieves a* common rate $R > 0$*, then there exists a sequence of linear-feedback schemes with message points that achieves the* private rates $(R, \ldots, R) \in \mathbb{R}^K$*:*

$$0 < R \leq C^{(\mathrm{Lin})} \quad \Longrightarrow \quad (R, \ldots, R) \in \mathcal{C}^{(\mathrm{Lin})}_{\mathrm{private}}. \tag{5.27}$$

Proof. See Section 5.5. □

Proposition 5.1 and the upper bound in Lemma 5.1 yield the following result:

Theorem 5.1. *We have:*

$$C^{(\mathrm{Lin})} \leq \frac{1}{2} \log_2 \left(1 + \frac{\alpha_1^\star P}{(1 - \alpha_1^\star)P + \sigma_1^2} \right) \tag{5.28}$$

where $\alpha_1^\star$ *lies in the open interval* $(0, 1)$ *and is such that there exist* $\alpha_2^\star, \ldots, \alpha_K^\star \in (0, 1)$ *that satisfy*

$$\alpha_1^\star + \alpha_2^\star + \ldots + \alpha_K^\star = 1 \tag{5.29a}$$

and for $k \in \{2, \ldots, K\}$*:*

$$\begin{aligned} &\frac{1}{2} \log_2 \left(1 + \frac{\alpha_k^\star P}{(1 - \alpha_1^\star - \alpha_2^\star - \ldots - \alpha_k^\star)P + N_k} \right) \\ &\qquad = \frac{1}{2} \log_2 \left(1 + \frac{\alpha_1^\star P}{(1 - \alpha_1^\star)P + \sigma_1^2} \right) \end{aligned} \tag{5.29b}$$

where the noise variances $\{N_k\}_{k=1}^K$ *are defined in* (5.24)*.*

Since $\alpha_1^\star$ is strictly smaller than 1, irrespective of K and the noise variances $\sigma_1^2, \ldots, \sigma_K^2$, we obtain the following corollary.

Corollary 5.1. *Linear-feedback schemes with a message point cannot achieve the capacity of the Gaussian BC with common message:*

$$C^{(\mathrm{Lin})} < C \tag{5.30}$$

where the inequality is strict.

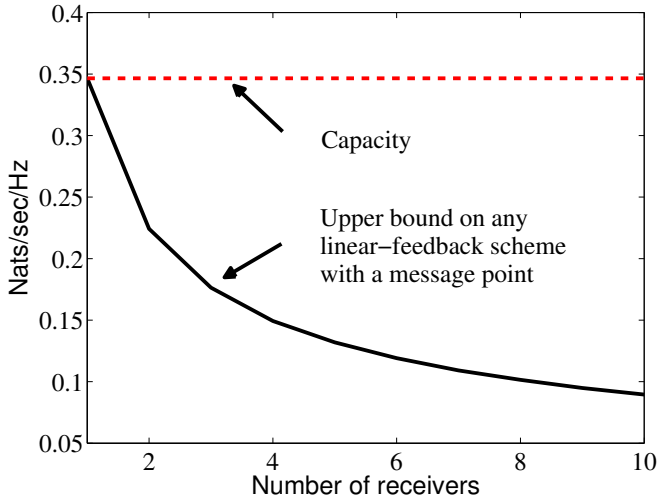


Figure 5.2: Upper bound (5.28) on the rates achievable with linear-feedback schemes with a message point in function of the number of receivers K.

Proposition 5.2. *If the noise variances* $\{\sigma_k^2\}_{k=1}^K$ *are such that*

$$\sum_{k=1}^{\infty} N_k = \infty, \tag{5.31}$$

then

$$\lim_{K\to\infty} C^{(\mathrm{Lin})} = 0. \tag{5.32}$$

Proof. See Appendix 5.A. □

In Figure 5.2 we plot the upper bond on $C^{(\mathrm{Lin})}$ shown in (5.28), Theorem 5.1, as a function of the number of receivers K, which have all the same noise variance $\sigma_1^2 = \ldots = \sigma_K^2 = 1$. As we observe, this upper bound, and thus also $C^{(\mathrm{Lin})}$, tends to 0 as K tends to infinity

Theorem 5.2. *For any positive rate* $R < C$*, if the feedback rate*

$$R_{\mathrm{Fb}} \geq (L-1)R, \tag{5.33}$$

for some positive integer L*, then it is possible to achieve an* L*-th order exponential decay of the probability of error in the blocklength.*

Proof. See Section 5.6. □

5.5 Proof of Proposition 5.1

Let $\delta > 0$ be a small real number. Fix a sequence of rate-$R > 0$, power-$(P - \delta)$ linear-feedback schemes that sends a common message point over the Gaussian BC with probability of error $P_{\mathrm{e}}^{(n)}$ tending to 0 as $n \to \infty$. For each $n \in \mathbb{Z}^+$, let

$$\Phi^{(n)}, \mathbf{d}^{(n)}, \mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}, g_1^{(n)}, \ldots, g_K^{(n)} \tag{5.34}$$

denote the parameters of the blocklength-n scheme, which satisfy the power constraint

$$\mathsf{E}\left[|\Theta^{(n)}|^2\right] \cdot \|\mathbf{d}^{(n)}\|^2 + \sum_{k=1}^{K} \left\|\mathsf{A}_k^{(n)}\right\|_F^2 \le n(P - \delta) \tag{5.35}$$

where $\Theta^{(n)} = \Phi^{(n)}(M)$.

We have the following lemma.

Lemma 5.2. *For each blocklength n, there exist n-dimensional row-vectors $\mathbf{v}_1^{(n)}, \ldots, \mathbf{v}_K^{(n)}$ of unit norms,*

$$\|\mathbf{v}_1^{(n)}\|^2 = \cdots = \|\mathbf{v}_K^{(n)}\|^2 = 1, \tag{5.36}$$

and K indices $j_1^{(n)}, \ldots, j_K^{(n)} \in \{1, \ldots, n\}$ such that for each $k \in \mathcal{K}$ the following three limits holds:

1.

$$R \le \varliminf_{n\to\infty} -\frac{1}{2n} \log_2 c_k^{(n)} \tag{5.37}$$

where

$$c_k^{(n)} := \sigma_k^2 \left\|\mathbf{v}_k^{(n)}\left(\mathsf{I} + \mathsf{A}_k^{(n)}\right)\right\|^2 + \sum_{k' \in \mathcal{K}\setminus\{k\}} \sigma_{k'}^2 \left\|\mathbf{v}_{k'}^{(n)} \mathsf{A}_{k'}^{(n)}\right\|^2; \tag{5.38}$$

2.

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left[\left(X^{(n)}_{j_k^{(n)}}\right)^2\right] = 0 \tag{5.39}$$

where for $i \in \{1,\ldots,n\}$, $X_i^{(n)}$ denotes the i-th channel input of the blocklength-n scheme; and

3.

$$\lim_{n\to\infty} \frac{1}{2n} \log_2\left(|v^{(n)}_{k,j_k^{(n)}}|\right) = 0 \tag{5.40}$$

where for $i \in \{1,\ldots,n\}$, $v_{k,i}^{(n)}$ denotes the i-th component of the vector $\mathbf{v}_k^{(n)}$.

Proof. See Appendix 5.B. □

Remark 5.1. *In the statement of the above lemma, the vector $\mathbf{v}_k^{(n)}$ is a scaled version of the LMMSE filter of the the input given observations $Y_{k,1},\ldots,Y_{k,n}$, and $c_k^{(n)}$ represents the volume of uncertainty about the message point at receiver k (hence R is bounded by its rate of decay). The last two claims of Lemma 5.2 hinge upon the fact that the channel input is power limited and therefore there must exists channel inputs that use less or equal than average power.*

In the following, let for each $n \in \mathbb{Z}^+$, $\mathbf{v}_1^{(n)},\ldots,\mathbf{v}_K^{(n)}$ be n-dimensional unit-norm row-vectors and $j_1^{(n)},\ldots,j_K^{(n)}$ be positive integers satisfying the limits (5.37), (5.39), and (5.40).

We now construct a sequence of linear-feedback schemes with message points that can send K independent messages $M_1,\ldots,M_K$ to Receivers $1,\ldots,K$ at rates

$$R_k \geq \Big(\varliminf_{n\to\infty} -\frac{1}{2n}\log_2 c_k^{(n)}\Big) - \epsilon, \qquad k \in \mathcal{K}, \tag{5.41}$$

for an arbitrary small $\epsilon > 0$ with: 1) a probability of error that tends to 0 as the blocklength tends to infinity and 2) with an average blockpower that is no larger than P when the blocklength is sufficiently large. By (5.37), since $\delta, \epsilon > 0$ can be chosen arbitrary small, and since $C^{(\mathrm{Lin})}$ is continuous in the power P (Remark 5.2 ahead) and is defined as a supremum, the result in Proposition 5.1 will follow.

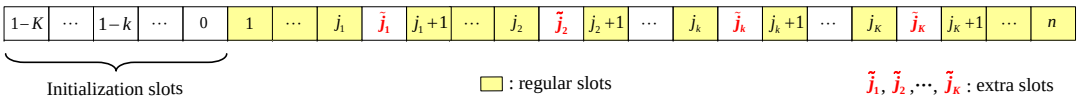


Figure 5.3: Labeling of the transmission slots for our blocklength-$(n+2K)$ scheme.

We describe our scheme for blocklength-$(n+2K)$, for some fixed $n \in \mathbb{Z}^+$. Our scheme is based on the parameters $\mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}$ in (5.34), on the vectors $\mathbf{v}_1^{(n)}, \ldots, \mathbf{v}_K^{(n)}$, and on the indices $j_1^{(n)}, \ldots, j_K^{(n)}$ where $\mathbf{v}_k^{(n)}$ and $j_k^{(n)}$, $k \in \mathcal{K}$ are defined in Lemma 5.2. For ease of notation, when describing our scheme in the following, we drop the superscript (n), i.e., we write

$$\mathsf{A}_1, \ldots, \mathsf{A}_K, \ \mathbf{v}_1, \ldots, \mathbf{v}_K, \ \text{ and } \ j_1, \ldots, j_K.$$

We also assume that

$$j_1 \le j_2 \le \ldots \le j_K. \tag{5.42}$$

(If this is not the case, we simply relabel the receivers.) Also, to further simplify the description of the linear-feedback coding and the decoding, we rename the $n+2K$ transmission slots as depicted in Figure 5.3. Transmission starts at slot $1-K$ and ends at slot n; also, after each slot j_k, for $k \in \mathcal{K}$, we introduce an additional slot $\tilde{j}_k$. We call the slots $1-K, \ldots, 0$ the *initialization slots*, the slots $\tilde{j}_1, \ldots, \tilde{j}_K$ the *extra slots*, and the remaining slots $1, 2, 3, \ldots, n$ the *regular slots*.

In our scheme, the message points $\{\Theta_k\}_{k=1}^K$ are constructed as in the Ozarow-Leung scheme [62]:

$$\Theta_k := 1/2 - \frac{M_k - 1}{\lfloor 2^{(n+2K)R_k} \rfloor}, \qquad k \in \mathcal{K}. \tag{5.43}$$

These messages are sent during the initialization phase. Specifically, in the initialization slots $i = 1-K, \ldots, 0$, the transmitter sends the K message points $\Theta_1, \ldots, \Theta_K$:

$$X_{1-k} = \sqrt{\frac{P}{\mathsf{Var}(\Theta_k)}}\Theta_k, \qquad k \in \mathcal{K}. \tag{5.44}$$

In the regular slots $i = 1, \ldots, n$, the transmitter sends the same inputs as in the scheme with common message described by the parameters in (5.34), but without the component from the message point and where for each $k \in \mathcal{K}$ the noise sample Z_{k,j_k} is replaced by $Z_{k,\tilde{j}_k}$. Thus, defining the n-length vector of regular inputs $\mathbf{X} \triangleq (X_1, X_2, X_3, \ldots, X_n)^\mathsf{T}$,

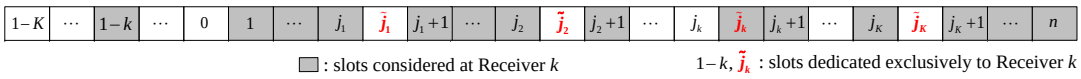


Figure 5.4: Transmissions considered at Receiver k and transmissions dedicated exclusively to Receiver k.

we have

$$\mathbf{X} = \sum_{k=1}^{K} \mathsf{A}_k \tilde{\mathbf{Z}}_k \tag{5.45}$$

where for $k \in \mathcal{K}$,

$$\tilde{\mathbf{Z}}_k := (Z_{k,1}, Z_{k,2}, \ldots, Z_{k,j_k-1}, Z_{k,\tilde{j}_k}, Z_{k,j_k+1}, \ldots, Z_{k,n})^{\mathsf{T}} \tag{5.46}$$

denotes the n-length noise vector experienced at Receiver k during the regular slots $1, \ldots, j_k - 1$, the extra slot $\tilde{j}_k$, and the regular slots $j_k + 1, \ldots, n$.

Since for each $k \in \mathcal{K}$, the extra slot $\tilde{j}_k$ precedes all regular slots $j_k + 1, \ldots, n$, the strict lower-triangularity of the matrices $\mathsf{A}_1, \ldots, \mathsf{A}_K$ ensures that in (5.45) the feedback is used in a strictly causal way.

In each extra slot $\tilde{j}_k$, for $k \in \mathcal{K}$, the transmitter sends the regular input X_{j_k}, but now with the noise sample $Z_{k,1-k}$,

$$X_{\tilde{j}_k} = X_{j_k} + Z_{k,1-k}. \tag{5.47}$$

The noise sample $Z_{k,1-k}$ is of interest to Receiver k (and only to Receiver k) because from this noise sample and $Y_{k,1-k}$ one can recover Θ_k, see (5.44). Therefore—as described shortly—in the decoding, Receiver k considers the extra output $Y_{k,\tilde{j}_k}$ which contains $Z_{k,1-k}$ whereas all other receivers $k' \neq k$ instead consider the regular outputs Y_{k',j_k} which do not have the $Z_{k,1-k}$-component.

The decoding is similar as in the Ozarow-Leung scheme. However, here, each Receiver $k \in \mathcal{K}$ only considers the initialization output $Y_{k,1-k}$, the regular outputs $Y_{k,1}, \ldots, Y_{k,j_k-1}, Y_{k,j_k+1}, \ldots, Y_{k,K}$ and the extra output $Y_{k,\tilde{j}_k}$, see also Figure 5.4. Specifically, Receiver k forms the n-length vector

$$\tilde{\mathbf{Y}}_k := \big(Y_{k,1}, \ldots, Y_{k,j_k-1}, Y_{k,\tilde{j}_k}, Y_{k,j_k+1}, \ldots, Y_{k,n}\big)^{\mathsf{T}}, \tag{5.48}$$

and produces the LMMSE estimate $\hat{Z}_{k,1-k}$ of the noise $Z_{k,1-k}$ based on the vector $\tilde{\mathbf{Y}}_k$. It then forms

$$\hat{\Theta}_k = \sqrt{\frac{\mathrm{Var}(\Theta_k)}{P}} \left(Y_{k,1-k} - \hat{Z}_{k,1-k} \right). \tag{5.49}$$

and performs nearest neighbor decoding to decode its desired message M_k based on $\hat{\Theta}_k$.

We now analyze the described scheme. The expected blockpower of our scheme is:

$$\begin{aligned} &\sum_{i=1-K}^{0} \mathsf{E}\left[|X_i|^2\right] + \sum_{i=1}^{n} \mathsf{E}\left[|X_i|^2\right] + \sum_{k=1}^{K} \mathsf{E}\left[\left|X_{\tilde{j}_k}\right|\right] \\ &\qquad \le KP + n(P-\delta) + \sum_{k=1}^{K} \mathsf{E}\left[|X_{j_k}|^2\right] + \sum_{k=1}^{K} \sigma_k^2 \end{aligned} \tag{5.50}$$

where the inequality follows from (5.44), (5.45), and (5.47), and from (5.35), which assures that the regular inputs $X_1, \ldots, X_n$ are block-power constrained to $n(P-\delta)$. Further, since the indices $j_1, \ldots, j_K$ satisfy Assumption (5.39),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{K} \mathsf{E}\left[|X_{j_k}|^2\right] = 0, \tag{5.51}$$

and thus for sufficiently large n the proposed scheme for independent messages is average blockpower constrained to P.

We analyze the probability of error. Notice that

$$\hat{\Theta}_k = \Theta_k + E_k \tag{5.52}$$

where

$$E_k := \sqrt{\frac{\mathrm{Var}(\Theta_k)}{P}} \left(Z_{k,1-k} - \hat{Z}_{k,1-k} \right) \tag{5.53}$$

is zero-mean Gaussian of variance

$$\mathsf{Var}(E_k) = \frac{\mathsf{Var}(\Theta_k)}{P} \sigma_k^2 2^{-2I\left(Z_{k,1-k};\tilde{\mathbf{Y}}_k\right)}. \tag{5.54}$$

Equation (5.54) is justified by

$$\begin{aligned} I\big(Z_{k,1-k}\ ;\ \tilde{\mathbf{Y}}_k\big) & \\ &= h(Z_{k,1-k}) - h\left(Z_{k,1-k}\middle|\tilde{\mathbf{Y}}_k\right) \\ &= \frac{1}{2}\log_2\left(\frac{\sigma_k^2}{\mathsf{Var}\Big(Z_{k,1-k}-\hat{Z}_{k,1-k}\Big)}\right) \end{aligned} \tag{5.55}$$

where the last equality follows because $Z_{k,1-k}$ and $\tilde{\mathbf{Y}}_k$ are jointly Gaussian, and thus the LMMSE estimation error $Z_{k,1-k} - \hat{Z}_{k,1-k}$ is independent of the observations $\tilde{\mathbf{Y}}_k$.

The nearest neighbor decoding rule is successful if $|E_k|$ is smaller than half the distance between any two message points. Since E_k is Gaussian and independent of the message point, the probability of this happening is

$$\begin{aligned} \Pr\Big[\hat{M}_k \neq M_k\Big] &\leq \Pr\left[|E_k| \geq \frac{1}{2\cdot\lfloor 2^{(n+2K)R_k}\rfloor}\right] \\ &= 2\mathcal{Q}\left(\frac{2^{I(Z_{k,1-k};\tilde{\mathbf{Y}}_k)}}{2\cdot\lfloor 2^{(n+2K)R_k}\rfloor}\cdot\frac{P}{\mathsf{Var}(\Theta_k)\,\sigma_k^2}\right). \end{aligned}$$

We conclude that the probability of error tends to 0, double-exponentially, whenever

$$R_k < \varliminf_{n\to\infty}\frac{1}{n}I(Z_{k,1-k};\tilde{\mathbf{Y}}_k). \tag{5.56}$$

Notice that the vector $\tilde{\mathbf{Y}}_k$ as defined in (5.48), satisfies

$$\tilde{\mathbf{Y}}_k = \sum_{k'\in\mathcal{K}\setminus\{k\}}\mathsf{A}_{k'}\tilde{\mathbf{Z}}_{k'} + (\mathsf{I}+\mathsf{A}_k)\tilde{\mathbf{Z}}_k + \mathbf{e}_{j_k}Z_{k,1-k} \tag{5.57}$$

where for each $i \in \{1,\ldots,n\}$ the vector $\mathbf{e}_i$ is the n-length unit-norm vector with all zero entries except at position i where the entry is 1. Thus, by the data processing

inequality,

$$\begin{aligned}
&I(Z_{k,1-k};\tilde{\mathbf{Y}}_k)\\
&\geq I(Z_{k,1-k};\mathbf{v}_k^{\mathsf{T}}\tilde{\mathbf{Y}}_k)\\
&=\frac{1}{2}\log_2\left(1+\frac{|v_{k,j_k}|^2}{\sigma_k^2\|\mathbf{v}_k(\mathsf{I}+\mathsf{A}_k)\|^2+\sum\limits_{k'\in\mathcal{K}\setminus\{k\}}\sigma_{k'}^2\|\mathbf{v}_{k'}\mathsf{A}_{k'}\|^2}\right)\\
&=\frac{1}{2}\log_2\left(1+\frac{|v_{k,j_k}|^2}{c_k}\right) \qquad (5.58)
\end{aligned}$$

where the first equality follows by (5.57) and the joint Gaussianity of all involved random variables and the second equality follows by the definition of c_k in (5.38).

Combining (5.56) and (5.58), we obtain that the probability $\Pr\left[\hat{M}_k \neq M_k\right]$ tends to 0 as $n \to \infty$ whenever

$$R_k < \varliminf_{n\to\infty} \frac{1}{2n}\log_2\left(1+\frac{|v_{k,j_k}|^2}{c_k}\right). \qquad (5.59)$$

(Recall that the quantities j_k, c_k, and v_{k,j_k} depend on n, but here we do not show this dependence for readability.)

Further, by the converse in (5.37),

$$\begin{aligned}
0 < R &\leq \varliminf_{n\to\infty} \frac{-1}{2n}\log_2 c_k\\
&= \varliminf_{n\to\infty} \frac{1}{2n}\log_2 \frac{|v_{k,j_k}|^2}{c_k} \qquad (5.60)\\
&= \varliminf_{n\to\infty} \frac{1}{2n}\log_2\left(1+\frac{|v_{k,j_k}|^2}{c_k}\right) \qquad (5.61)
\end{aligned}$$

where the first equality holds by Condition (5.40) and the second equality holds because (5.60) implies that the ratio $\frac{|v_{k,j_k}|^2}{c_k}$ tends to infinity with n.

Combining (5.59) with (5.61) establishes that for arbitrary $\epsilon > 0$ there exists a rate tuple $(R_1, \ldots, R_K)$ satisfying (5.41) such that the described scheme with independent messages achieves probability of error that tends to 0 as the blocklength tends to infinity.

Remark 5.2. *In the spirit of the scheme for private messages described above, one can construct a linear-feedback scheme with a common message point that has arbitrary*

small probability of error whenever

$$R < \overline{\lim_{n\to\infty}} -\frac{1}{2n}\log_2 c_k, \qquad k \in \mathcal{K}.$$

Combined with the converse in (5.37)*, this gives a (multi-letter) characterization of* $C^{(\mathrm{Lin})}$*. Based on this multi-letter characterization one can show the continuity of* $C^{(\mathrm{Lin})}$ *in the transmit-power constraint* P*.*

5.6 Proof of Theorem 5.2: Coding Scheme Achieving L-th Order Exponential Decay

The scheme is based on the scheme in [9], see also [150], [151]. Fix a positive rate $R < C$ and a positive integer L. Assume that

$$R_{\mathrm{Fb}} \geq R(L-1). \tag{5.62}$$

Also, fix a large blocklength n and small numbers $\epsilon, \delta > 0$ such that

$$R < C(1-\delta) \tag{5.63}$$

and

$$(1-\epsilon)^{-1} < 1+\delta. \tag{5.64}$$

Define

$$n_1 := (1-\epsilon)n \tag{5.65}$$

and for $l \in \{2, \ldots, L\}$

$$n_l := n_1 + \frac{\epsilon n}{L-1}(l-1). \tag{5.66}$$

Notice that by (5.64) and (5.65),

$$\frac{n}{n_1} < 1+\delta. \tag{5.67}$$

The coding scheme takes place in L phases. After each phase $l \in \{1, \ldots, L\}$, each Receiver $k \in \mathcal{K}$ makes a temporary guess $\hat{M}_l^{(k)}$ of message M. The final guess is the guess after phase L:

$$\hat{M}^{(k)} = \hat{M}_L^{(k)}, \tag{5.68}$$

Define the probability of error after phase $l \in \{1, \ldots, L\}$:

$$P_{e,l}^{(n)} := \Pr\left[\bigcup_{k \in \mathcal{K}} \hat{M}_l^{(k)} \neq M\right] \tag{5.69}$$

and thus

$$P_e^{(n)} = P_{e,L}^{(n)}. \tag{5.70}$$

5.6.1 Code Construction

We construct a codebook $\mathcal{C}_1$ that

- is of blocklength n_1,
- is of rate $R_{\text{phase},1} = \frac{n}{n_1} R$,
- satisfies an expected average block-power constraint P, and
- when used to send a common message over the Gaussian BC in (5.1) and combined with an optimal decoding rule, it achieves probability of error ρ_1 not exceeding

$$\rho_1 \leq e^{-n(\zeta - o(1))} \tag{5.71}$$

  for some $\zeta > 0$.

Notice that such a code exists because, by (5.63) and (5.67), the rate of the code $\frac{n}{n_1} R < C(1 - \delta^2)$, and because the error exponent of the BC with common message without feedback is positive for all rates below capacity.[1]

Let

$$\gamma_1 := \rho_1. \tag{5.72}$$

[1] The positiveness of the error exponent for the Gaussian BC with common message and without feedback follows from the fact that without feedback the probability of error for the Gaussian BC with common messages is at most K times the probability of error to the weakest receiver.

For l from 2 to L, do the following.

Construct a codebook $\mathcal{C}_l$ that:

- is of blocklength $\frac{\epsilon n}{L-1} - 1$,
- is of rate $R_{\text{phase},l} := \frac{R(L-1)}{\epsilon-(L-1)/n}$,
- satisfies an expected average block-power constraint P/γ_{l-1},
- when used to send a common message over the Gaussian BC in (5.1) and combined with an optimal decoding rule, it achieves probability of error ρ_l not exceeding

$$\rho_l \leq \exp(-\underbrace{\exp \circ \ldots \circ \exp}_{l-1 \text{ times}}(\Omega(n))). \tag{5.73}$$

Define

$$\gamma_l := \rho_l + 2 \sum_{k \in \mathcal{K}} \mathcal{Q}\left(\frac{\sqrt{P/\gamma_{l-1}}}{2\sigma_k}\right). \tag{5.74}$$

(As shown in Section 5.6.3 ahead, γ_l upper bounds $P_{e,l}^{(n)}$ defined in (5.69).) By (5.73) and (5.74), inductively one can show that

$$\gamma_l \leq \exp(-\underbrace{\exp \circ \ldots \circ \exp}_{l-1 \text{ times}}(\Omega(n))). \tag{5.75}$$

In Appendix 5.C, we prove that such codes $\mathcal{C}_2, \ldots, \mathcal{C}_L$ exist.

5.6.2 Transmission

Transmission takes place in L phases.

First phase with channel uses $i = 1, \ldots, n_1$

During the first n_1 channel uses, the transmitter sends the codeword in $\mathcal{C}_1$ corresponding to message M.

After observing the channel outputs $Y_k^{n_1}$, Receiver $k \in \mathcal{K}$ makes a temporary decision $\hat{M}_1^{(k)}$ about M. It then sends this temporary decision $\hat{M}_1^{(k)}$ to the transmitter over the feedback channel:

$$F_{k,n_1} = \hat{M}_1^{(k)}. \tag{5.76}$$

All previous feedback signals from Receiver k are deterministically 0.

Phase $l \in \{2, \ldots, L\}$ with channel uses $i \in \{n_{l-1}+1, \ldots, n_l\}$

The communication in phase l depends on the receivers' temporary decisions ($\hat{M}_{l-1}^{(1)}, \ldots, \hat{M}_{l-1}^{(K)}$) after the previous phase $(l-1)$. These decisions have been communicated to the transmitter over the respective feedback links.

If in phase $(l-1)$ at least one of the receivers made an incorrect decision,

$$(\hat{M}_{l-1}^{(k)} \neq M), \qquad \text{for some } k \in \mathcal{K}, \tag{5.77}$$

then in channel use $n_{l-1}+1$ the transmitter sends an error signal to indicate an error:

$$X_{n_l+1} = \sqrt{P/\gamma_{l-1}}. \tag{5.78}$$

During the remaining channel uses $i = n_{l-1}+2, \ldots, n_l$ it then retransmits the message M by sending the codeword from $\mathcal{C}_l$ that corresponds to M.

On the other hand, if all receivers' temporary decisions to the phase $(l-1)$ were correct,

$$\hat{M}_{l-1}^{(1)} = \hat{M}_{l-1}^{(2)} = \ldots = \hat{M}_{l-1}^{(K)} = M, \tag{5.79}$$

then the transmitter sends 0 during the entire phase l:

$$X_i = 0, \qquad i = n_{l-1}+1, \ldots, n_l. \tag{5.80}$$

In this case, no power is consumed in phase l.

The receivers first detect whether the transmitter sent an error signal in channel use $n_{l-1}+1$. Depending on the output of this detection, they either stick to their temporary decision in phase $(l-1)$ or make a new decision based on the transmissions in phase l. Specifically, if

$$Y_{k,n_{l-1}+1} < T_{l-1} \tag{5.81}$$

where

$$T_{l-1} := \frac{\sqrt{P/\gamma_{l-1}}}{2}, \tag{5.82}$$

then Receiver $k \in \mathcal{K}$ decides that its decision $\hat{M}_{l-1}^{(k)}$ in phase $(l-1)$ was correct and keeps it as its temporary guess of the message M:

$$\hat{M}_l^{(k)} = \hat{M}_{l-1}^{(k)}. \tag{5.83}$$

If instead,

$$Y_{k,n_{l-1}+1} \geq T_{l-1}, \tag{5.84}$$

Receiver k decides that its temporary decision $\hat{M}_{l-1}^{(k)}$ was wrong and discards it. It then produces a new guess $\hat{M}_l^{(k)}$ by decoding the code $\mathcal{C}_l$ based on the outputs $Y_{k,n_{l-1}+2}, \ldots,$ Y_{k,n_l}.

After each phase $l \in \{2, \ldots, L-1\}$, each Receiver $k \in \mathcal{K}$ feeds back to the transmitter its temporary guess $\hat{M}_l^{(k)}$:

$$F_{k,n_l} = \hat{M}_l^{(k)}. \tag{5.85}$$

All other feedback signals $F_{k,n_{l-1}+1}, \ldots, F_{k,n_l-1}$ in phase l are deterministically 0.

After L transmission phases, Receiver k's final guess is

$$\hat{M}^{(k)} = \hat{M}_L^{(k)}. \tag{5.86}$$

Thus, an error occurs in the communication if

$$(\hat{M}_L^{(k)} \neq M), \quad \text{for some } k \in \mathcal{K}. \tag{5.87}$$

5.6.3 Analysis

In view of (5.62), by (5.76) and (5.85), and because all other feedback signals are deterministically 0, our scheme satisfies the feedback rate constraint in (5.4).

We next analyze the probability of error and we bound the consumed power. These analysis rely on the following events. For each $k \in \mathcal{K}$ and $l \in \{1, \ldots, L\}$ define the events:

- $\epsilon_l^{(k)}$: Receiver k's decision in phase l is wrong:

$$\hat{M}_l^{(k)} \neq M; \tag{5.88}$$

- $\epsilon_{T,l}^{(k)}$: Receiver k observes
$$Y_{k,n_l+1} < T_l; \tag{5.89}$$
- $\epsilon_{\rho,l}^{(k)}$: Decoding message M based on Receiver k's phase-l outputs $Y_{k,n_{l-1}+2}, \ldots, Y_{k,n_l}$ using codebook $\mathcal{C}_l$ results in an error.

Define also the events:

$\mathcal{E}_{1,l}$: All receivers' decisions in phase $(l-1)$ are correct, and at least one Receiver $k \in \mathcal{K}$ obtains an error signal in channel use $n_{l-1}+1$:
$$\left(\bigcap_{k\in\mathcal{K}} \left(\epsilon_{l-1}^{(k)}\right)^c\right) \cap \left(\bigcup_{k\in\mathcal{K}} \left(\epsilon_{T,l-1}^{(k)}\right)^c\right). \tag{5.90}$$

$\mathcal{E}_{2,l}$: At least one Receiver $k \in \mathcal{K}$ makes an incorrect decision in phase $(l-1)$ but obtains no error signal in channel use $n_{l-1}+1$:
$$\bigcup_{k\in\mathcal{K}} \left(\epsilon_{l-1}^{(k)} \cap \epsilon_{T,l-1}^{(k)}\right). \tag{5.91}$$

$\mathcal{E}_{3,l}$: At least one Receiver $k \in \mathcal{K}$ makes an incorrect temporary decision in phase $(l-1)$, and at least one Receiver $k' \in \mathcal{K}$ observes $Y_{k',n_{l-1}+1} \geq T_{l-1}$ and errs when decoding M based on its phase-l outputs $Y_{k',n_{l-1}+2}, \ldots, Y_{k',n_l}$:
$$\left(\bigcup_{k\in\mathcal{K}} \epsilon_{l-1}^{(k)}\right) \cap \left(\bigcup_{k'\in\mathcal{K}} \left(\left(\epsilon_{T,l}^{(k')}\right)^c \cap \epsilon_{\rho,l}^{(k')}\right)\right). \tag{5.92}$$

For each $l \in \{1, \ldots, L\}$, the probability $P_{e,l}^{(n)}$ is included in the union of the events $(\mathcal{E}_{1,l} \cup \mathcal{E}_{2,l} \cup \mathcal{E}_{3,l})$, and thus, by the union bound,
$$P_{e,l}^{(n)} \leq \Pr[\mathcal{E}_{1,l}] + \Pr[\mathcal{E}_{2,l}] + \Pr[\mathcal{E}_{3,l}]\,. \tag{5.93}$$

In particular, by (5.70) and (5.93), the probability of error of our scheme
$$P_e^{(n)} \leq \Pr[\mathcal{E}_{1,L}] + \Pr[\epsilon_{2,L}] + \Pr[\epsilon_{3,L}]\,. \tag{5.94}$$

We bound each summand in (5.94) individually, starting with $\Pr[\mathcal{E}_{1,L}]$. By (5.90),

we have

$$\begin{aligned}\Pr[\mathcal{E}_{1,L}] &= \Pr\left[\left(\bigcap_{k\in\mathcal{K}}\left(\epsilon_{L-1}^{(k)}\right)^c\right)\cap\left(\bigcup_{k\in\mathcal{K}}\left(\epsilon_{T,L-1}^{(k)}\right)^c\right)\right]\\ &\leq \Pr\left[\bigcup_{k\in\mathcal{K}}\left(\epsilon_{T,L-1}^{(k)}\right)^c\Big|\bigcap_{k\in\mathcal{K}}\left(\epsilon_{L-1}^{(k)}\right)^c\right]\\ &\leq \sum_{k=1}^{K}\Pr\left[\left(\epsilon_{T,L-1}^{(k)}\right)^c\Big|\bigcap_{k\in\mathcal{K}}\left(\epsilon_{L-1}^{(k)}\right)^c\right]\\ &= \sum_{k=1}^{K}\mathcal{Q}\left(\frac{T_{L-1}}{\sigma_k}\right)\end{aligned} \tag{5.95}$$

where the first inequality follows by Bayes' rule and because a probability cannot exceed 1; the second inequality by the union bound; and the last equality because in the event $\left(\bigcap_{k\in\mathcal{K}}(\epsilon_{L-1}^{(k)})^c\right)$, we have $X_{n_{L-1}+1}=0$ and thus $Y_{k,n_{L-1}+1}\sim\mathcal{N}(0,\sigma_k^2)$.

Next, by (5.91) and similar arguments as before, we obtain,

$$\begin{aligned}\Pr[\mathcal{E}_{2,L}] &= \Pr\left[\bigcup_{k\in\mathcal{K}}\left(\epsilon_{L-1}^{(k)}\cap\epsilon_{T,L-1}^{(k)}\right)\right]\\ &\leq \sum_{k=1}^{K}\Pr\left[\epsilon_{L-1}^{(k)}\cap\epsilon_{T,L-1}^{(k)}\right]\\ &\leq \sum_{k=1}^{K}\Pr\left[\epsilon_{T,L-1}^{(k)}\big|\epsilon_{L-1}^{(k)}\right]\\ &= \sum_{k=1}^{K}\mathcal{Q}\left(\frac{T_{L-1}}{\sigma_k}\right).\end{aligned} \tag{5.96}$$

Finally, by (5.92) and similar arguments as before,

$$\begin{aligned}\Pr[\mathcal{E}_{3,L}] &= \Pr\left[\left(\bigcup_{k\in\mathcal{K}}\epsilon_{L-1}^{(k)}\right)\cap\left(\bigcup_{k'\in\mathcal{K}}\left(\left(\epsilon_{T,L}^{(k')}\right)^c\cap\epsilon_{\rho,L}^{(k')}\right)\right)\right]\\ &\le \Pr\left[\bigcup_{k'\in\mathcal{K}}\left(\left(\epsilon_{T,L}^{(k')}\right)^c\cap\epsilon_{\rho,L}^{(k')}\right)\Big|\bigcup_{k\in\mathcal{K}}\epsilon_{L-1}^{(k)}\right]\\ &\le \Pr\left[\bigcup_{k'\in\mathcal{K}}\epsilon_{\rho,L}^{(k')}\Big|\bigcup_{k\in\mathcal{K}}\epsilon_{L-1}^{(k)}\right]\\ &\le \rho_L \end{aligned} \tag{5.97}$$

where the last inequality follows by the definition of ρ_L.

In view of (5.82) and (5.94)–(5.97),

$$\begin{aligned}P_e^{(n)} &\le \Pr[\mathcal{E}_{1,L}]+\Pr[\mathcal{E}_{2,L}]+\Pr[\mathcal{E}_{3,L}]\\ &\le \rho_L + 2\sum_{k\in\mathcal{K}}\mathcal{Q}\left(\frac{\sqrt{P/\gamma_{L-1}}}{2\sigma_k}\right)\\ &= \gamma_L \end{aligned} \tag{5.98}$$

where the equality follows by the definition of γ_L in (5.74). Combining this with the L-th order exponential decay of γ_L, see (5.75), we obtain

$$\underline{\lim}_{n\to\infty} -\frac{1}{n}\underbrace{\log\log\ldots\log}_{L-1\,\text{times}}(-\log P_e^{(n)}) > 0, \tag{5.99}$$

Now consider the consumed expected average block-power. Similarly to (5.98), we can show that for $l\in\{1,\ldots,L-1\}$,

$$P_{e,l}^{(n)} \le \gamma_l. \tag{5.100}$$

Since in each phase $l\in\{2,\ldots,L\}$ we consume power P/γ_{l-1} in the event (5.77) and power 0 in the event (5.79), by the definition in (5.69),

$$\frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{n}X_i^2\right] \le \frac{1}{n}\Big(P(1-\epsilon)n+\sum_{l=2}^{L}P_{e,l-1}^{(n)}\frac{P}{\gamma_{l-1}}\frac{\epsilon n}{L-1}\Big) \le P \tag{5.101}$$

where the second inequality follows from (5.100).

This completes the proof of Theorem 5.2.

5.A Appendix: Proof of Proposition 5.2

We show that under assumption (5.31),

$$\lim_{K\to\infty} \alpha_1^\star = 0, \tag{5.102}$$

which implies (5.32).

Notice that (5.29b) implies for $k \in \{1,\ldots,K-1\}$:

$$\frac{\alpha_K^* P}{N_K} = \frac{\alpha_k^* P}{(1-\alpha_1^\star - \alpha_2^* .... - \alpha_k^*)P + N_k}. \tag{5.103}$$

Since for each k, the term $(1-\alpha_1^\star - \alpha_2^* - \ldots - \alpha_k^*)$ is nonnegative,

$$\alpha_k^* \;\geq\; \frac{N_k}{N_K}\alpha_K^*, \qquad k \in \{1,\ldots,K-1\}. \tag{5.104}$$

Thus, by (5.29a),

$$1 = \sum_{k=1}^K \alpha_k^* \geq \sum_{k=1}^K \frac{N_k}{N_K}\alpha_K^*$$

and

$$\alpha_K^* \leq \frac{N_K}{\sum_{k=1}^K N_k}.$$

We conclude that, for every finite positive integer K,

$$R_K \;\leq\; \frac{1}{2}\log_2\left(1+\frac{P}{\sum_{k=1}^K N_k}\right),$$

and under Assumption (5.31), in the limit as $K\to\infty$,

$$\lim_{K\to\infty} R_K = 0.$$

5.B Appendix: Proof of Lemma 5.2

We first prove the converse (5.37). Fix a blocklength n. By Fano's inequality, for each $k \in \mathcal{K}$,

$$\begin{aligned} nR &= H(M^{(n)}) \\ &\leq I\left(M^{(n)}; Y_{k,1}^{(n)}, \ldots, Y_{k,n}^{(n)}\right) + \epsilon(n) \\ &\leq I\left(\Theta^{(n)}; Y_{k,1}^{(n)}, \ldots, Y_{k,n}^{(n)}\right) + \epsilon(n) \\ &\overset{(a)}{\leq} I\left(\bar{\Theta}^{(n)}; \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}\right) + \epsilon(n) \end{aligned} \tag{5.105}$$

where $\frac{\epsilon(n)}{n} \to 0$ as $n \to \infty$ and where we defined the tuple $(\bar{\Theta}^{(n)}, \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)})$ to be jointly Gaussian with the same covariance matrix as the tuple $(\Theta^{(n)}; Y_{k,1}^{(n)}, \ldots, Y_{k,n}^{(n)})$. Inequality (a) holds because the Gaussian distribution maximizes differential entropy under a covariance constraint.

Now, since $\bar{\Theta}^{(n)}, \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}$ are jointly Gaussian, there exists a linear combination $\sum_{i=1}^{n} v_{k,i}^{(n)} \bar{Y}_{k,i}^{(n)}$ such that

$$I\left(\bar{\Theta}^{(n)}; \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}\right) = I\left(\bar{\Theta}^{(n)}; \sum_{i=1}^{n} v_{k,i}^{(n)} \bar{Y}_{k,i}^{(n)}\right). \tag{5.106}$$

(In fact, the linear combination is simply the LMMSE-estimate of $\bar{\Theta}^{(n)}$ based on $\bar{Y}_{k,1}^{(n)}$,..., $\bar{Y}_{k,n}^{(n)}$.) Defining the n-dimensional row-vector $\mathbf{v}_k^{(n)} = \left(v_{k,1}^{(n)}, \ldots, v_{k,n}^{(n)}\right)$, in view of (5.106), we have

$$I\left(\bar{\Theta}^{(n)}; \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}\right) = \frac{1}{2} \log_2 \left(1 + \frac{\left(\mathbf{v}_k^{(n)} \mathbf{d}^{(n)}\right)^2 \mathsf{Var}\left(\Theta^{(n)}\right)}{c_k^{(n)}}\right) \tag{5.107}$$

where $c_k^{(n)}$ is as defined in (5.38).

Notice that the right-hand side of (5.107) does not depend on the norm of $\mathbf{v}_k^{(n)}$ (as long as it is non-zero) but only on the direction. Therefore, without loss of generality, we can assume that

$$\|\mathbf{v}_k^{(n)}\|^2 = 1. \tag{5.108}$$

By (5.105) and (5.107), we conclude that for each $k \in \mathcal{K}$, there exists a unit-norm vector $\mathbf{v}_k^{(n)}$ such that

$$R \leq \varliminf_{n\to\infty} \frac{1}{2n} \log_2 \left(1+\frac{\left(\mathbf{v}_k^{(n)}\mathbf{d}^{(n)}\right)^2 \mathsf{Var}\left(\Theta^{(n)}\right)}{c_k^{(n)}}\right). \tag{5.109}$$

Since by assumption $R > 0$, (5.109) implies that the ratio $(\mathbf{v}_k^{(n)}\mathbf{d}^{(n)})^2\mathsf{Var}\left(\Theta^{(n)}\right)/c_k^{(n)}$ tends to infinity and thus

$$R \leq \varliminf_{n\to\infty} \frac{1}{2n} \log_2 \left(\frac{\left(\mathbf{v}_k^{(n)}\mathbf{d}^{(n)}\right)^2 \mathsf{Var}\left(\Theta^{(n)}\right)}{c_k^{(n)}}\right). \tag{5.110}$$

Now, consider the average block-power constraint (5.35). Since the trace of a positive semidefinite matrix is non-negative and $\mathsf{Var}\left(\Theta^{(n)}\right) \leq \mathsf{E}\left[|\Theta^{(n)}|^2\right]$, by (5.35), for each $n \in \mathbb{Z}^+$:

$$\|\mathbf{d}^{(n)}\|^2 \mathsf{E}\left[|\Theta^{(n)}|^2\right] \leq n(P-\delta). \tag{5.111}$$

Since $\|\mathbf{v}_k^{(n)}\| = 1$, (5.108), by the Cauchy-Schwarz Inequality,

$$\left(\mathbf{v}_k^{(n)}\mathbf{d}^{(n)}\right)^2 \mathsf{Var}\left(\Theta^{(n)}\right) \leq n(P-\delta) \tag{5.112}$$

and as a consequence

$$\varliminf_{n\to\infty} \frac{1}{2n} \log_2 \left(\left(\mathbf{v}_k^{(n)}\mathbf{d}^{(n)}\right)^2 \mathsf{Var}\left(\Theta^{(n)}\right)\right) \leq 0. \tag{5.113}$$

Combining this with (5.110), proves the desired inequality (5.37).

The proof of Inequalities (5.39) and (5.40) relies on Lemmas 5.3 and 5.4 at the end of this appendix. Notice that the monotonicity of the log-function and the nonnegativity of the norm combined with (5.37) imply that for each $k \in \mathcal{K}$,

$$R \leq \varliminf_{n\to\infty} -\frac{1}{2n} \log_2 \left\|\mathbf{v}_k^{(n)}\left(\mathsf{I}+\mathsf{A}_k^{(n)}\right)\right\|^2, \tag{5.114}$$

where recall that we assumed $R > 0$.

Define for each $k \in \mathcal{K}$ and positive integer n the set

$$\mathcal{S}_k^{(n)} := \left\{ i \in \{1, \ldots, n\} \colon v_{k,i}^{(n)} > n^{-2\log_2 n} \right\}. \tag{5.115}$$

By Lemma 5.3 and Inequality (5.114), the cardinality of each set $\mathcal{S}_k^{(n)}$ is unbounded,

$$|\mathcal{S}_k^{(n)}| \to \infty \qquad \text{as} \qquad n \to \infty, \qquad k \in \mathcal{K}. \tag{5.116}$$

Applying now Lemma 5.4 to $p = P - \delta$, to

$$\pi_i^{(n)} = \mathsf{E}\left[\left(X_i^{(n)} \right)^2 \right], \tag{5.117}$$

and to $\mathcal{T}^{(n)} = \mathcal{S}_k^{(n)}$ implies that for each $k \in \mathcal{K}$ there exists a sequences of indices $\{j_k^{(n)} \in \mathcal{S}_k^{(n)}\}_{n=1}^{\infty}$ that satisfies (5.39). Since every sequence of indices $\{i^{(n)} \in \mathcal{S}_k^{(n)}\}_{n=1}^{\infty}$ also satisfies (5.40), this concludes the proof of the lemma.

Lemma 5.3. *For each $n \in \mathbb{Z}^+$, let $\mathsf{A}^{(n)}$ be a strictly lower-triangular n-by-n matrix and $\mathbf{v}^{(n)}$ an n-dimensional row-vector. Let $a_{i,j}^{(n)}$ denote the row-i, column-j entry of $\mathsf{A}^{(n)}$ and $v_i^{(n)}$ denote the i-th entry of $\mathbf{v}^{(n)}$. Assume that the elements $a_{i,j}^{(n)}$ are bounded as*

$$|a_{i,j}^{(n)}|^2 \leq np \tag{5.118}$$

for some real number $p > 0$, and that the inequality

$$\varliminf_{n\to\infty} -\frac{1}{2n} \log_2 \|\mathbf{v}^{(n)}(\mathsf{I} + \mathsf{A}^{(n)})\|^2 \geq \Gamma \tag{5.119}$$

holds for some real number $\Gamma > 0$. Then, for each $\epsilon \in (0, \Gamma)$ and for all sufficiently large n the following implication holds: If

$$|v_j^{(n)}| > 2^{-n(\Gamma-\epsilon)} \tag{5.120a}$$

for some index $j \in \{1, \ldots, n\}$, then there must exist an index $i \in \{j+1, \ldots, n\}$ such that

$$|v_i^{(n)}| \geq \frac{|v_j^{(n)}| - 2^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}}\sqrt{p}}. \tag{5.120b}$$

If moreover, the vectors $\{\mathbf{v}^{(n)}\}_{n=1}^{\infty}$ *are of unit norm, then the cardinality of the set*

$$\mathcal{S}^{(n)} := \left\{ j \in \{1, \dots, n\} : |v_j^{(n)}| > n^{-2\log_2(n)} \right\} \tag{5.121}$$

is unbounded in n*.*

Proof. Fix $\epsilon \in (0, \Gamma)$ and let n be sufficiently large so that

$$-\frac{1}{2n} \log_2 \left\| \mathbf{v}^{(n)} (\mathsf{I} + \mathsf{A}^{(n)}) \right\|^2 \geq \Gamma - \epsilon. \tag{5.122}$$

This is possible by (5.119).

Since $\mathsf{A}^{(n)}$ is strictly lower-triangular,

$$\begin{aligned} \|\mathbf{v}^{(n)}(\mathsf{I} + \mathsf{A}^{(n)})\|^2 &= \sum_{j=1}^{n} (v_j^{(n)} + \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)})^2 \\ &\geq \sum_{j=1}^{n} \left(|v_j^{(n)}| - \Big| \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)} \Big| \right)^2 \\ &\geq \left(|v_j^{(n)}| - \Big| \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)} \Big| \right)^2 \end{aligned}$$

and by (5.122) and the monotonicity of the log-function, for all $j \in \{1, \dots, n\}$:

$$-\frac{1}{2n} \log_2 \left(|v_j^{(n)}| - \Big| \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)} \Big| \right)^2 \geq \Gamma - \epsilon.$$

Thus,

$$|v_j^{(n)}| \leq \Big| \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)} \Big| + 2^{-n(\Gamma - \epsilon)}$$

and by (5.118):

$$\begin{aligned} |v_j^{(n)}| - 2^{-n(\Gamma-\epsilon)} &\leq \Big| \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)} \Big| \\ &\leq \sum_{i=j+1}^{n} |v_i^{(n)}| \sqrt{np}. \end{aligned} \tag{5.123}$$

If $|v_j^{(n)}| \leq 2^{-n(\Gamma-\epsilon)}$, then the sum on the right-hand side of (5.123) can be empty, i.e.,

$j = n$. However, if

$$|v_j^{(n)}| > 2^{-n(\Gamma-\epsilon)}, \tag{5.124}$$

then the sum needs to have at least one term. Indeed, if (5.124) holds and $i < n$, there must exist an index $i \in \{j+1, ..., n\}$ such that

$$\frac{1}{n}\Big(|v_j^{(n)}| - 2^{-n(\Gamma-\epsilon)}\Big) \leq |v_i^{(n)}|\sqrt{np}, \tag{5.125}$$

which is equivalent to the desired bound (5.120b).

We now prove the second part of the lemma, i.e., the unboundedness of the cardinalities of the sets $\mathcal{S}^{(n)}$, where we assume that the vectors $\{\mathbf{v}^{(n)}\}$ are of unit norm. In the following, let n be sufficiently large so that the first part of the lemma, Implication (5.120), holds and so that

$$\frac{1}{\sqrt{n}} > \frac{1}{n^{2\log_2(n)}} > 2^{-n(\Gamma-\epsilon)} \tag{5.126}$$

and for every $\ell \in \{1, \ldots, \log_2(n)\}$

$$\begin{aligned} &\frac{1}{n^{(3\ell+1)/2}p^{\ell/2}} - 2^{-n(\Gamma-\epsilon)}n^{-3/2}p^{-1/2}\frac{1-n^{-3\ell/2}p^{-\ell/2}}{1-n^{-3/2}p^{-1/2}} \\ &\quad > \frac{1}{n^{2\log_2(n)}} \end{aligned} \tag{5.127}$$

Since $\|\mathbf{v}^{(n)}\|^2 = 1$, for each n, there must exist an index $i_0^{(n)} \in \{1, \ldots, n\}$ such that

$$|v_{i_0^{(n)}}^{(n)}| \geq \frac{1}{\sqrt{n}}, \tag{5.128}$$

and by (5.126)

$$|v_{i_0^{(n)}}^{(n)}| > n^{2\log_2(n)} > 2^{-n(\Gamma-\epsilon)}. \tag{5.129}$$

We conclude by (5.120) that there exists an index $i_1^{(n)} \in \{i_0^{(n)}+1,\ldots,n\}$ satisfying

$$\begin{aligned} |v^{(n)}_{i_1^{(n)}}| &\geq \frac{|v^{(n)}_{i_0^{(n)}}| - 2^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}}\sqrt{p}} \\ &\geq \frac{1}{n^2\sqrt{p}} - \frac{2^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}}\sqrt{p}} \end{aligned} \tag{5.130}$$

where the inequality follows from (5.128). By (5.126) and (5.127), (applied for $\ell = 1$), Inequality (5.130) implies that

$$|v^{(n)}_{i_1^{(n)}}| > 2^{-n(\Gamma-\epsilon)}, \tag{5.131}$$

and consequently, by (5.120), there exists an index $i_2^{(n)} \in \{i_1^{(n)}+1,\ldots,n\}$ satisfying

$$|v^{(n)}_{i_2^{(n)}}| \geq \frac{|v^{(n)}_{i_1^{(n)}}| - 2^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}}\sqrt{p}} \tag{5.132}$$

$$\geq \frac{1}{n^{7/2}p} - \frac{2^{-n(\Gamma-\epsilon)}}{n^3 p} - \frac{2^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}}\sqrt{p}} \tag{5.133}$$

$$> 2^{-n(\Gamma-\epsilon)}, \tag{5.134}$$

where the last inequality follows by (5.126) and (5.127) (applied for $\ell = 2$).

Repeating these arguments iteratively, we conclude that it is possible to find indices $1 \leq i_0^{(n)} < i_1^{(n)} < \ldots < i_{\log_2(n)}^{(n)} < n$ such that for each $\ell \in \{1,\ldots,\log_2(n)\}$:

$$\begin{aligned} |v^{(n)}_{i_\ell^{(n)}}| &\geq \frac{1}{n^{3(\ell+1)/2}p^{\ell/2}} - 2^{-n(\Gamma-\epsilon)}\sum_{j=1}^{\ell}\left(n^{-3/2}p^{-1/2}\right)^j \\ &= \frac{1}{n^{3(\ell+1)/2}p^{\ell/2}} \\ &\quad -2^{-n(\Gamma-\epsilon)}n^{-3/2}p^{-1/2}\frac{1-n^{-3\ell/2}p^{-\ell/2}}{1-n^{-3/2}p^{-1/2}} \qquad (5.135) \\ &> \frac{1}{n^{2\log_2(n)}} \qquad (5.136) \\ &> 2^{-(\Gamma-\epsilon)} \qquad (5.137) \end{aligned}$$

where the last two inequalities follow from (5.126) and (5.127). This proves that for sufficiently large n the cardinality of the set $\mathcal{S}^{(n)}$ as defined in (5.121) is at least $\log_2(n)$ and thus unbounded in n. □

Lemma 5.4. *For each positive integer n, let $(\pi_1^{(n)}, \ldots, \pi_n^{(n)})$ be a tuple of nonnegative real numbers that satisfy*

$$\frac{1}{n}\sum_{i=1}^{n} \pi_i^{(n)} \leq p \tag{5.138}$$

for some real number $p > 0$, and let $\mathcal{T}^{(n)}$ be a subset of the indices from 1 to n,

$$\mathcal{T}^{(n)} \subseteq \{1, \ldots, n\}, \tag{5.139}$$

that satisfies

$$|\mathcal{T}^{(n)}| \to \infty \quad \text{as} \quad n \to \infty. \tag{5.140}$$

Then, there exists a sequence of indices $\left\{i^{(n)} \in \mathcal{T}^{(n)}\right\}_{n=1}^{\infty}$ such that

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\pi_{i^{(n)}}^{(n)} = 0. \tag{5.141}$$

Proof. Since all numbers $\pi_i^{(n)}$ are nonnegative, for every sequence of indices $\{i^{(n)} \in \mathcal{T}^{(n)}\}_{n=1}^{\infty}$,

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\pi_{i^{(n)}}^{(n)} \geq 0. \tag{5.142}$$

We thus have to prove that there exists at least one sequence of indices $\{i^{(n)} \in \mathcal{T}^{(n)}\}_{n=1}^{\infty}$ that satisfies

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\pi_{i^{(n)}}^{(n)} \leq 0. \tag{5.143}$$

We prove this by contradiction. Assume that for each sequence of indices $\{i^{(n)} \in \mathcal{T}^{(n)}\}_{n=1}^{\infty}$

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\pi_{i^{(n)}}^{(n)} > 0. \tag{5.144}$$

Define for each $n \in \mathbb{Z}^+$

$$\pi_{\min}^{(n)} := \min_{i\in\mathcal{T}^{(n)}} \pi_i^{(n)}, \tag{5.145}$$

and define the limit

$$\delta_{\min} := \overline{\lim_{n\to\infty}} \frac{1}{n}\pi_{\min}^{(n)}, \tag{5.146}$$

which by Assumption (5.144) is strictly positive,

$$\delta_{\min} > 0. \tag{5.147}$$

Now, since all the terms $\pi_i^{(n)}$ are nonnegative:

$$\frac{1}{n}\sum_{i=1}^{n}\pi_i^{(n)} \geq \frac{1}{n}\sum_{i\in\mathcal{T}^{(n)}}\pi_i^{(n)} \geq \frac{1}{n}\pi_{\min}^{(n)}|\mathcal{T}^{(n)}|, \tag{5.148}$$

where the second inequality follows by the definition in (5.145). By (5.146) and (5.147) and by the undboundedness of the cardinality of the sets $\mathcal{T}^{(n)}$, we conclude that the sum in (5.148) is unbounded in n, which contradicts Assumption (5.138) and thus concludes our proof. □

5.C Appendix: Existence of Code $\mathcal{C}_2, \ldots, \mathcal{C}_L$ with the Desired Properties

The proof is by induction: for each $\ell \in \{2, \ldots, L\}$, when proving the existence of the desired $\mathcal{C}_\ell$, we assume that

$$\gamma_{l-1} \leq \exp(-\underbrace{\exp\circ\ldots\circ\exp}_{l-2 \text{ times}}(\Omega(n))). \tag{5.149}$$

For $l = 2$, Inequality (5.149) follows from (5.71).

By [3], for all rates

$$\tilde{R} < \frac{1}{2}\log_2\frac{2+\sqrt{\tilde{P}^2/\sigma^4+4}}{4},$$

and for sufficiently large n there exists a blocklength-$\tilde{n}$, rate-$\tilde{R}$ non-feedback coding scheme for the memoryless Gaussian point-to-point channel with noise variance σ^2, with expected average block-power no larger than $\tilde{P}$ and with probability of error P_e

satisfying

$$P_e \leq e^{-\tilde{n}(E(\tilde{R},\tilde{P}/\sigma^2)-\epsilon')} \tag{5.150}$$

for some fixed $\epsilon' > 0$ and

$$E(\tilde{R},\tilde{P}) = \frac{\tilde{P}}{4\sigma^2}\left(1-\sqrt{1-2^{-2\tilde{R}}}\right). \tag{5.151}$$

Since the probability of error of a non-feedback code over the Gaussian BC with common message is at most K times the probability of error to the weakest receiver, we conclude that for all $\tilde{P} > 0$ and

$$0 < \tilde{R} < \frac{1}{2}\log_2 \frac{2+\sqrt{\tilde{P}^2/\sigma_1^4+4}}{4}, \tag{5.152a}$$

there exists a rate-$\tilde{R}$ code with power $\tilde{P}$ and blocklength $\tilde{n}$ that for the Gaussian BC with common message achieves probability of error

$$P_e^{(\mathrm{BC})} \leq K e^{-\tilde{n}\left(\frac{P}{4\sigma_1^2}\left(1-\sqrt{1-2^{-2\tilde{R}}}\right)-\epsilon'\right)}. \tag{5.152b}$$

Now apply this statement to $\tilde{R} = R_{\mathrm{phase},l}$, $\tilde{P} = P/\gamma_{l-1}$ and $\tilde{n} = \frac{\epsilon n}{L-1} - 1$. Since for sufficiently large n, by (5.149),

$$R_{\mathrm{phase},l} < \frac{1}{2}\log_2 \frac{2+\sqrt{\frac{P^2}{\gamma_{l-1}^2\sigma_1^4}+4}}{4}, \tag{5.153}$$

we conclude by (5.152) that there exists a code $\mathcal{C}_l$ of rate-$R_{\mathrm{phase},l}$, block-power P/γ_{l-1}, blocklength $\frac{\epsilon n}{L-1} - 1$ and probability of error ρ_l satisfying

$$\begin{aligned}\rho_l &\leq K e^{-\left(\frac{\epsilon n}{L-1}-1\right)\left(\frac{P}{4\gamma_{l-1}\sigma_1^2}\left(1-\sqrt{1-2^{-2\frac{R(L-1)}{\epsilon-(L-1)/n}}}\right)-\epsilon'\right)} \\ &\leq \exp(-\underbrace{\exp\circ\ldots\circ\exp}_{l-1 \text{ times}}(\Omega(n)))\end{aligned} \tag{5.154}$$

where the inequality follows again by (5.149).

By the definition of γ_l in (5.74), Inequalities (5.154) and (5.149) also yield:

$$\gamma_l \leq \exp(-\underbrace{\exp \circ \ldots \circ \exp}_{l-1 \text{ times}}(\Omega(n))). \tag{5.155}$$

Chapter 6

Coding Schemes for DMBCs with Private Messages and Rate-limited Feedback

6.1 Introduction

In this chapter, we present two types of coding schemes for DMBCs with rate-limited feedback. Our schemes use a block-Markov strategy where in each block they apply Marton coding [140], which to date is the best known coding scheme without feedback. The messages sent over the feedback links are simply compression information that describe the channel outputs that the receivers observed during a block.

In our first type of scheme, (Schemes 1A–1C), the encoder transmits exactly these compression informations as part of the cloud center of the Marton code employed in the next block. Thus, here, the encoder only *relays* the feedback messages from one receiver to the other. Each receiver can hence reconstruct a compressed version of the other receiver's outputs and apply a modified Marton decoding to these compressed outputs and its own observed outputs. The Marton decoding is modified to account for the fact that each receiver already knows a part of the message sent in the cloud center—namely the compression information it had generated itself after the previous block. As we will see, the decoding can be performed as well as if the part of the

cloud-center message known at a receiver was not present. In this sense, in the cloud center we are sending information that is useful to one of the two receivers without disturbing the other receiver, or in other words, without occupying the other receiver's resources. For asymmetric setups where one of the two receivers is stronger than the other, e.g., less noisy, this implies that we can send the compression message, and thus the information about the other receiver's outputs, to the stronger receiver without harming the performance of the weaker receiver. This allows in particular to improve over Marton's original non-feedback scheme.

We discuss the described coding strategy when the two receivers apply backward decoding (Scheme 1A), when they apply sliding-window decoding (Scheme 1B), and when one receiver applies backward decoding and the other sliding-window decoding (Scheme 1C).

Our coding strategy is reminiscent of the compress-and-forward relay strategy [109] or the noisy network coding for general networks [153, 154] in the sense that the two receivers compress their channel outputs and send these compression indices over the feedback links. However, in our schemes, we use Marton coding since our transmitter has to send two independent private messages to the two receivers (we could treat them as a big common message, but this would perform poorly). Moreover, whereas in noisy network coding the transmitter where to generate new compression indices that describe its observed feedback outputs, in our schemes the transmitter *decode-and-forward (or relays)* the compression messages that were sent over the feedback links. Thus, in our schemes the transmitter sends compression indices that describe the outputs observed at the two receivers.

Our schemes are particularly beneficial for the class of *strictly essentially less-noisy* DMBCs, which we define in this chapter and which represents a subclass of Nair's essentially less-noisy DMBCs [135]. Our class includes the BSBC and the *binary erasure BCs* (BEBC) with unequal cross-over probabilities or unequal erasure probabilities at the two receivers, and the *binary symmetric channel/binary erasure channel BC* (BSC/BEC-BC) for a large range of parameters. For strictly essentially less-noisy DMBCs Marton coding is known to achieve capacity [135]. For this class of DMBCs, our schemes improve strictly over the non-feedback capacity region no matter how small but positive the feedback rates are and even when there is feedback only from the weaker receiver.

In fact, for most of these channels our scheme can improve over all boundary points $(R_1 > 0, R_2 > 0)$ of the non-feedback capacity region. The described schemes also improve over the non-feedback capacity region of the BSC/BEC-BC when the DMBC is more capable [155], unless the BSC and BEC have same capacities.

Thus, unlike for previous schemes, with our new schemes we can easily show that feedback increases the capacity region for a large set of DMBCs.

We present a fourth scheme, Scheme 2, where the encoder uses the feedback messages to *reconstruct* compressed versions of the channel outputs, and then *processes* these compressed signals together with the previously sent codewords to generate update (compression) information intended to both receivers. This update information is sent as part of the cloud center of the Marton code employed in the next-following block. This scheme is reminiscent of the Shayevitz-Wigger scheme [56] but for rate-limited feedback. Moreover, in our Scheme 2 here, the update information is sent only in the cloud center and using a joint source-channel code, whereas in the Shayevitz-Wigger scheme parts of it are also sent in the satellite codewords but using only a separate source-channel code.

Since here the update information is sent using a joint source-channel code, in the limit as the feedback rates increase, the region achieved with our Scheme 2 improves over the region achieved by the Shayevitz-Wigger scheme when this latter is restricted to send all the update information in the cloud center. Notice that this represents a prominent special case of the Shayevitz-Wigger scheme which subsumes the schemes by Wang [58], by Georgiadis and Tassiulas [59], by Maddah-Ali and Tse [60], and also the schemes in [77], [79], [81] and [82] when these are specialized to memoryless BCs and to delayed state-information only.

All our results hold also with noisy feedback when the receivers can code over the feedback links.

6.2 Channel Model

Communication takes place over a DMBC with rate-limited feedback, see Figure 6.1. The setup is characterized by the finite input alphabet $\mathcal{X}$, the finite output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, the channel law $P_{Y_1Y_2|X}$, and nonnegative feedback rates $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$. If at

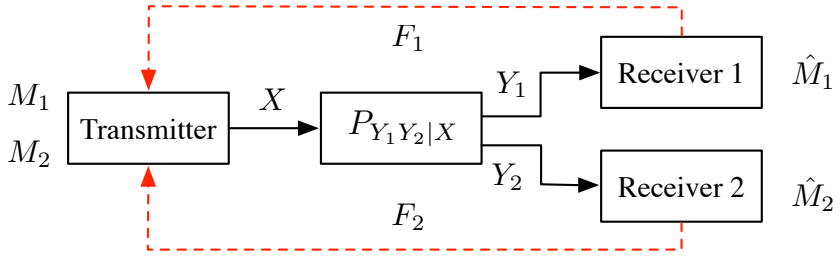


Figure 6.1: Broadcast channel with private messages and rate-limited feedback

discrete-time i the transmitter sends the channel input $x_i \in \mathcal{X}$, then Receiver $k \in \{1,2\}$ observes the output $Y_{k,i} \in \mathcal{Y}_k$, where the pair $(Y_{1,i}, Y_{2,i}) \sim P_{Y_1Y_2|X}(\cdot,\cdot|x_i)$. Also, after observing $Y_{k,i}$, Receiver k can send a feedback signal $F_{k,i} \in \mathcal{F}_{k,i}$ to the transmitter, where $\mathcal{F}_{k,i}$ denotes the finite alphabet of $F_{k,i}$ and is a design parameter of a scheme. The feedback link between the transmitter and Receiver k is assumed to be instantaneous and noise-free but rate-limited to $R_{\mathrm{Fb},k}$ bits on average. Thus, if the transmission takes place over a total blocklength N, then

$$|\mathcal{F}_{k,1}| \times \cdots \times |\mathcal{F}_{k,N}| \leq \lfloor 2^{NR_{\mathrm{Fb},k}} \rfloor, \quad k \in \{1,2\}. \tag{6.1a}$$

The goal of the communication is that the transmitter conveys two independent private messages $M_1 \in \{1,\ldots,\lfloor 2^{NR_1}\rfloor\}$ and $M_2 \in \{1,\ldots,\lfloor 2^{NR_2}\rfloor\}$, to Receiver 1 and 2, respectively. Each M_k, for $k \in \{1,2\}$, is uniformly distributed over the set $\mathcal{M}_k := \{1,\ldots,\lfloor 2^{NR_k}\rfloor\}$, where R_k denotes the private rate of transmission of Receiver k.

The transmitter is comprised of a sequence of encoding functions $\{f_i^{(N)}\}_{i=1}^N$ of the form $f_i^{(N)} : \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{F}_{1,1} \times \cdots \times \mathcal{F}_{1,i-1} \times \mathcal{F}_{2,1} \times \cdots \times \mathcal{F}_{2,i-1} \to \mathcal{X}$ that is used to produce the channel inputs as

$$X_i = f_i^{(N)}\big(M_1, M_2, F_1^{i-1}, F_2^{i-1}\big), \qquad i \in \{1,\ldots,N\}. \tag{6.2}$$

Receiver $k \in \{1,2\}$ is comprised of a sequence of feedback-encoding functions $\{\psi_{k,i}^{(N)}\}_{i=1}^N$ of the form $\psi_{k,i}^{(N)} : \mathcal{Y}_k^i \to \mathcal{F}_{k,i}$ that is used to produce the symbols

$$F_{k,i} = \psi_{k,i}^{(N)}(Y_{k,1},\ldots,Y_{k,i}), \qquad i \in \{1,\ldots,N\}, \tag{6.3}$$

sent over the feedback link, and of a decoding function $g_k^{(N)} : \mathcal{Y}_k^N \to \mathcal{M}_k$ used to produce

a guess of Message M_k:

$$\hat{M}_k = g_k^{(N)}(Y_k^N). \tag{6.4}$$

A rate region (R_1, R_2) with averaged feedback rates $R_{\text{Fb},1}$, $R_{\text{Fb},2}$ is called achievable if for every blocklength N, there exists a set encoding functions $\left\{f_i^{(N)}\right\}_{i=1}^N$ and for $k = \{1, 2\}$ there exists a set of decoding functions $g_k^{(N)}$, feedback alphabets $\{\mathcal{F}_{k,i}\}_{i=1}^N$ satisfying (6.1), and feedback-encoding functions $\left\{\psi_{k,i}^{(N)}\right\}_{i=1}^N$ such that the error probability

$$P_e^{(N)} := \Pr\left[M_1 \neq \hat{M}_1 \text{ or } M_2 \neq \hat{M}_2\right] \tag{6.5}$$

tends to zero as the blocklength N tends to infinity. The closure of the set of achievable rate pairs (R_1, R_2) is called the *feedback capacity region* and is denoted by $\mathcal{C}_{\text{Fb}}$.

In the special case $R_{\text{Fb},1} = R_{\text{Fb},2} = 0$ the feedback signals are constant and the setup is equivalent to a setup without feedback. We denote the capacity region for this setup $\mathcal{C}_{\text{NoFb}}$.

6.3 Motivation: A Simple Scheme

We sketch a simple scheme that motivates our schemes in Section 6.6. We assume there is only feedback from Receiver 1, i.e., $R_{\text{Fb},1} > 0$ and $R_{\text{Fb},2} = 0$.

We apply block-Markov coding with $B + 1$ blocks of length n, where in each block we use superposition coding (without feedback) to send fresh messages $M_{1,b}$ and $M_{2,b}$. Message $M_{1,b}$ is sent in the cloud center U_b^n and $M_{2,b}$ in the satellite codeword X_b^n. Thus, the scheme is expected to perform well when the following gap is nonnegative:

$$\Gamma := I(U; Y_2) - I(U; Y_1) \geq 0. \tag{6.6}$$

(This is for example the case in a BSBC when the cross-over probability to Receiver 2 is no larger than the cross-over probability to Receiver 1.)

After each block, both Receivers 1 and 2 decode the cloud center codeword U_b^n by means of joint typicality decoding. By the Packing Lemma, this is possible whenever

$$R_1 \leq I(U; Y_1) \tag{6.7}$$

$$R_1 \leq I(U; Y_2), \tag{6.8}$$

where here, by (6.6), the second constraint is not active. We notice that when

$$\Gamma > 0 \tag{6.9}$$

Receiver 2 would be able to decode the cloud center even if—besides $M_{1,b}$—it also encoded an extra message of rate not exceeding Γ. Of course, we cannot just add an arbitrary rate-Γ message to the cloud center, because this would make it impossible for Receiver 1 to decode this larger cloud center. Instead, we shall add a rate-Γ message that is known to Receiver 1. If in the typicality check Receiver 1 only considers the candidate codewords for the cloud center that correspond to the correct value of this extra message, then the decoding at Receiver 1 performs as well as if the additional message was not present. Thus, if the additional message is known at Receiver 1, it does not disturb its decoding.

With rate-limited feedback, we can identify a suitable additional message to send in the cloud center of block b: the feedback message $M_{\text{Fb},1,b-1}$ that Receiver 1 had fed back after the previous block $b-1$. In fact, as we describe shortly, in our scheme Receiver 1 only feeds back a message at the end of each block.

The transmitter thus simply relays the information it received over the feedback link to the other receiver. In this sense, the feedback link and part of the cloud center can be seen as an independent communication pipe from Receiver 1 to Receiver 2, where the pipe is rate-limited to

$$\min\{\Gamma, R_{\text{Fb},1}\}. \tag{6.10}$$

In our scheme, we use this pipe to send a compressed version of the channel outputs observed at Receiver 1 to Receiver 2. Specifically, the feedback message $M_{\text{Fb},1,b-1}$ sent after block $b-1$ is a Wyner-Ziv message that compresses outputs $Y_{1,b-1}^n$ while taking into account that the reconstructor has side-information $Y_{2,b-1}^n, U_{b-1}^n$. The rate-required for this Wyner-Ziv message is

$$\tilde{R}_1 > I(\tilde{Y}_1; Y_1 | Y_2, U). \tag{6.11}$$

and, in order to satisfy the feedback-rate constraint, it also has to satisfy

$$\tilde{R}_1 < R_{\mathrm{Fb},1}. \tag{6.12}$$

After decoding the additional message $M_{\mathrm{Fb},1,b-1}$, which is transmitted in the cloud center of block b, Receiver 2 first reconstructs a compressed version of Receiver 1's outputs $\tilde{Y}_{1,b-1}^n$. It then uses this reconstruction to decode its intended Message $M_{2,b-1}$ based on the tuple $(\tilde{Y}_{1,b-1}^n, Y_{2,b-1}^n, U_{b-1}^n)$. This is possible, with arbitrary small probability of error, if

$$R_2 \leq I(X; \tilde{Y}_1, Y_2|U). \tag{6.13}$$

Combining now constraints (6.7), (6.11), (6.12), and (6.13), we conclude that our scheme achieves all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U; Y_1) \tag{6.14a}$$

$$R_2 \leq I(X; \tilde{Y}_1, Y_2|U) = I(X; Y_2|U) + I(X; \tilde{Y}_1|U, Y_2) \tag{6.14b}$$

for some pmf $P_{UX}P_{\tilde{Y}_1|U,Y_1}$ that satisfies

$$I(\tilde{Y}_1; Y_1|Y_2, U) \leq \min\{\Gamma, R_{\mathrm{Fb},1}\}. \tag{6.15}$$

The left-hand side of (6.15) gives the minimum rate required for a Wyner-Ziv code that compresses $Y_{1,b-1}^n$ given that the reconstructor has side-information $Y_{2,b-1}^n$ and U_{b-1}^n.

Comparing constraints (6.14) to the superposition coding constraints in (3.22), we see that the constraints here are strictly looser whenever $I(X; \tilde{Y}_1|U, Y_2) > 0$. Or in other words, whenever observing a compressed version of Receiver 1's outputs improves the decoding at Receiver 2.

What is remarkable about this scheme is that when $\Gamma > 0$, there is no cost in conveying the compressed version of Receiver 1's outputs to Receiver 2. It is as if there were free resources in the communication from the transmitter to Receiver 2, which the feedback allows to exploit. Without feedback, the resources cannot be exploited because the transmitter cannot identify a messages that is known at Receiver 1 and useful at Receiver 2.

6.4 New Achievable Regions and Usefulness of Feedback

6.4.1 Achievable Regions

The following achievable regions are based on the coding schemes in Section 6.6. These coding schemes are motivated by the scheme sketched in the previous section, but use the more general Marton coding instead of superposition coding and exploit the feedback from both receivers.

In our first scheme 1A (Section 6.6.1), the receivers apply sliding-window decoding. The scheme achieves the region in the following Theorem 6.1.

Theorem 6.1 (Sliding-Window Decoding)**.** *The capacity region* $\mathcal{C}_{\mathrm{Fb}}$ *includes the set* $\mathcal{R}_{\mathrm{relay,sw}}$[1] *of all nonnegative rate pairs* (R_1, R_2) *that satisfy*

$$R_1 \le I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; U_0, Y_2|Y_1, Q) \tag{6.16a}$$

$$R_1 \le I(U_0; Y_2|Q) + I(U_1; Y_1, \tilde{Y}_2|U_0, Q) - \Delta_2 - I(\tilde{Y}_1; Y_1|U_0, U_2, Y_2, Q) \tag{6.16b}$$

$$R_2 \le I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; U_0, Y_1|Y_2, Q) \tag{6.16c}$$

$$R_2 \le I(U_0; Y_1|Q) + I(U_2; Y_2, \tilde{Y}_1|U_0, Q) - \Delta_1 - I(\tilde{Y}_2; Y_2|U_0, U_1, Y_1, Q) \tag{6.16d}$$

$$\begin{aligned} R_1 + R_2 \le\ & I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; U_0, Y_2|Y_1, Q) \\ & + I(U_2; Y_2, \tilde{Y}_1|U_0, Q) - \Delta_1 - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.16e}$$

$$\begin{aligned} R_1 + R_2 \le\ & I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; U_0, Y_1|Y_2, Q) \\ & + I(U_1; Y_1, \tilde{Y}_2|U_0, Q) - \Delta_2 - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.16f}$$

$$\begin{aligned} R_1 + R_2 \le\ & I(U_0, U_1; Y_1, \tilde{Y}_2|Q) + I(U_0, U_2; Y_2, \tilde{Y}_1|Q) \\ & - I(\tilde{Y}_2; U_0, Y_2|Y_1, Q) - I(\tilde{Y}_1; U_0, Y_1|Y_2, Q) - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.16g}$$

where

$$\Delta_1 := \max\{0, I(\tilde{Y}_1; Y_1|U_0, Y_2, Q) - R_{\mathrm{Fb},1}\}$$
$$\Delta_2 := \max\{0, I(\tilde{Y}_2; Y_2|U_0, Y_1, Q) - R_{\mathrm{Fb},2}\}$$

for some pmf $P_Q P_{U_0U_1U_2|Q} P_{\tilde{Y}_1|Y_1U_0Q} P_{\tilde{Y}_2|Y_2U_0Q}$ *and some function* $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to$

[1] *The subscript "relay" indicates that the transmitter simply relays the feedback information and the subscript "sw" indicates that sliding-window decoding is applied.*

$\mathcal{X}$ *such that*

$$I(U_1; Y_1, \tilde{Y}_2|U_0, Q) - \Delta_2 \geq 0 \tag{6.18a}$$

$$I(U_2; Y_2, \tilde{Y}_1|U_0, Q) - \Delta_1 \geq 0. \tag{6.18b}$$

$$I(\tilde{Y}_1; Y_1|U_0, U_2, Y_2, Q) \leq \min\{I(U_0; Y_2|Q), R_{\text{Fb},1}\} \tag{6.18c}$$

$$I(\tilde{Y}_2; Y_2|U_0, U_1, Y_1, Q) \leq \min\{I(U_0; Y_1|Q), R_{\text{Fb},2}\} \tag{6.18d}$$

where $X = f(U_0, U_1, U_2, Q)$.

Proof. See Section 6.6.1. □

For sufficiently large feedback rates $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$ (in particular for $R_{\text{Fb},1} \geq |\mathcal{Y}_1|$ and $R_{\text{Fb},2} \geq |\mathcal{Y}_2|$), the terms Δ_1 and Δ_2 as defined in (6.17) are 0.

In our second scheme 1B (Section 6.6.2), the receivers apply backward decoding. This way, for each block, they can jointly decode the cloud center and their intended satellite codewords. In this scheme, the Wyner-Ziv compression cannot be superpositioned on the cloud center because the receivers have not yet decoded this latter when compressing their channel outputs at the end of each block. The following Theorem 6.2 presents the achievable region for this second scheme.

Theorem 6.2 (Backward Decoding)**.** *The capacity region* $\mathcal{C}_{\text{Fb}}$ *includes the set* $\mathcal{R}_{\text{relay,bw}}$[2] *of all nonnegative rate pairs* (R_1, R_2) *that satisfy*

$$R_1 \leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \tag{6.19a}$$

$$R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) \tag{6.19b}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \\ & + I(U_2; Y_2, \tilde{Y}_1|U_0, Q) - \Delta_1 - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.19c}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) \\ & + I(U_1; Y_1, \tilde{Y}_2|U_0, Q) - \Delta_2 - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.19d}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \\ & + I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.19e}$$

[2] *The subscript "bw" stands for backward decoding.*

for some pmf $P_Q P_{U_0U_1U_2|Q} P_{\tilde{Y}_1|Y_1Q} P_{\tilde{Y}_2|Y_2Q}$ *and some function* $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$ *such that*

$$I(\tilde{Y}_1; Y_1|U_0, U_2, Y_2, Q) \leq R_{\text{Fb},1} \tag{6.20a}$$

$$I(\tilde{Y}_2; Y_2|U_0, U_1, Y_1, Q) \leq R_{\text{Fb},2} \tag{6.20b}$$

where $X = f(U_0, U_1, U_2, Q)$.

Proof. See Section 6.6.2. □

Setting $\tilde{Y}_1 = \tilde{Y}_2 = \text{const.}$, i.e., both receivers do not send any feedback, the region $\mathcal{R}_{\text{relay,bw}}$ specializes to $\mathcal{R}_{\text{Marton}}$.

Remark 6.1. *Constraints* (6.19) *and* (6.20) *are looser than constraints* (6.16) *and* (6.18)*, respectively. But in Theorem 6.2 we have the conditional pmfs* $P_{\tilde{Y}_1|Y_1}$ *and* $P_{\tilde{Y}_2|Y_2}$ *whereas in Theorem 6.1 we allow for more general pmfs* $P_{\tilde{Y}_1|Y_1,U_0}$ *and* $P_{\tilde{Y}_2|Y_2,U_0}$*. It is thus not clear in general which of the achievable regions in Theorems 6.1 or 6.2 is larger.*

Remark 6.2. *Consider the Shayevitz-Wigger region* (4.2) *restricted to the choice of auxiliaries*

$$V_1 = V_2 = V_0 = (f_1(Y_1, Q), f_2(Y_2, Q)) \tag{6.21}$$

for two deterministic functions f_1 *and* f_2*. (Notice that Kim, Chia, and El Gamal's choice of auxiliaries* (4.3) *or* (4.4) *is of this form.) Our new achievable region* $\mathcal{R}_{\text{relay,bw}}$ *improves over this restricted Shayevitz-Wigger region whenever the feedback rates* $R_{\text{Fb},1}$*,* $R_{\text{Fb},2}$ *are sufficiently large so that in our new region we can choose*

$$\tilde{Y}_1 = f_1(Y_1, Q) \quad \text{and} \quad \tilde{Y}_2 = f_2(Y_2, Q) \tag{6.22}$$

and so that $\Delta_1 = \Delta_2 = 0$.

In fact, for the choices (6.21) *and* (6.22) *the rate constraints in* (6.19a)*,* (6.19b)*, and* (6.19e) *characterizing our new region coincide with the rate constraints* (4.2a)–(4.2b) *which characterize the Shayevitz-Wigger region. Moreover, the combination of the two sum-rate constraints* (6.19c) *and* (6.19d) *is looser than the sum-rate constraint* (4.2c)*,*

because the former involves a "$\min_{k=\{1,2\}}\{a_k - b_k\}$*-term" whereas the latter involves the smaller "*$\min_{k\in\{1,2\}} a_k - \max_{k\in\{1,2\}} b_k$*-term", for* $a_k, b_k \geq 0$.

Our third scheme 1C (Section 6.6.3) is a mixture of the first two: Receiver 1 behaves as in the first scheme and Receiver 2 as in the second scheme. This is particularly interesting when there is no feedback from Receiver 2, $R_{\mathrm{Fb},2} = 0$, and when Marton's scheme specializes to superposition coding with no satellite codeword for Receiver 1. Theorem 6.3 presents the region achieved by this third scheme with Marton coding and Corollary 6.1 with superposition coding.

Theorem 6.3 (Hybrid Sliding-Window and Backward Decoding)**.** *Even for* $R_{\mathrm{Fb},2} = 0$*, the capacity region* $\mathcal{C}_{\mathrm{Fb}}$ *includes the set* $\mathcal{R}^{(1)}_{\mathrm{relay,hb}}$[3] *of all nonnegative rate pairs* (R_1, R_2) *that satisfy*

$$R_1 \leq I(U_0, U_1; Y_1|Q) \tag{6.23a}$$

$$R_2 \leq I(U_0, U_2; \tilde{Y}_1, Y_2|Q) - I(\tilde{Y}_1; U_0, U_1, U_2, Y_1|Y_2, Q) \tag{6.23b}$$

$$\begin{aligned} R_1+R_2 \leq\ & I(U_0, U_1; Y_1|Q) + I(U_2; Y_2, \tilde{Y}_1|U_0, Q) \\ & -\Delta_1 - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.23c}$$

$$\begin{aligned} R_1+R_2 \leq\ & I(U_1; Y_1|U_0, Q) + I(U_0, U_2; \tilde{Y}_1, Y_2|Q) \\ & -I(\tilde{Y}_1; U_0,U_1,U_2,Y_1|Y_2,Q) - I(U_1; U_2|U_0,Q) \end{aligned} \tag{6.23d}$$

for some pmf $P_Q P_{U_0U_1U_2|Q} P_{\tilde{Y}_1|Y_1U_0Q}$ *and some function* $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$ *such that*

$$I(\tilde{Y}_1; U_1,Y_1|U_0, U_2, Y_2, Q) \leq R_{\mathrm{Fb},1}. \tag{6.24}$$

The capacity region $\mathcal{C}_{\mathrm{Fb}}$ *also includes the region* $\mathcal{R}^{(2)}_{\mathrm{relay,hb}}$ *which is obtained by exchanging indices* 1 *and* 2 *in the above definition of* $\mathcal{R}^{(1)}_{\mathrm{relay,hb}}$.

Proof. See Section 6.6.3. □

If superposition coding is used instead of Marton coding and only one of the two receivers sends feedback, Theorem 6.3 reduces to the following corollary.

[3] *The subscript "hb" stands for hybrid decoding.*

Corollary 6.1. *The capacity region $\mathcal{C}_{\mathrm{Fb}}$ includes the set $\mathcal{R}^{(1)}_{\mathrm{relay,sp}}$*[4] *of all nonnegative rate pairs (R_1, R_2) that satisfy*

$$R_1 \leq I(U; Y_1|Q) \tag{6.25a}$$

$$R_1 + R_2 \leq I(U; Y_1|Q) + I(X; Y_2, \tilde{Y}_1|U, Q) \tag{6.25b}$$

$$R_1 + R_2 \leq I(X; Y_2|Q) - I(\tilde{Y}_1; Y_1|U, Y_2, Q) \tag{6.25c}$$

for some pmf $P_Q P_{UX|Q} P_{\tilde{Y}_1|Y_1UQ}$ such that

$$I(\tilde{Y}_1; Y_1|U, Y_2, Q) \leq R_{\mathrm{Fb},1}. \tag{6.26}$$

The capacity region $\mathcal{C}_{\mathrm{Fb}}$ also includes the region $\mathcal{R}^{(2)}_{\mathrm{relay,sp}}$ which is obtained by exchanging indices 1 *and* 2 *in the above definition of $\mathcal{R}^{(1)}_{\mathrm{relay,sp}}$.*

Proof. Let $\tilde{Y}_2 = U_1 = \text{const.}$, $U = U_0$ and $X = U_2$. Constraint (6.23a) then specializes to (6.25a) and constraint (6.23b) is redundant compared to constraint (6.23d). Observe that constraints (6.23d) and (6.24) are looser than constraints (6.25c) and (6.26), respectively. Also, by (6.26), constraint (6.23c) reduces to (6.25b). Thus the capacity region $\mathcal{C}_{\mathrm{Fb}}$ includes the region $\mathcal{R}^{(1)}_{\mathrm{relay,sp}}$. Similar arguments hold for $\mathcal{R}^{(2)}_{\mathrm{relay,sp}}$. □

Remark 6.3. *The region $\mathcal{R}^{(1)}_{\mathrm{relay,hb}}$ contains the regions in Theorems 6.1 and 6.2 when these latter are specialized to $U_1 =$const., $U_2 = X$, and $R_{\mathrm{Fb},2} = 0$.*

In our first three schemes 1A–1C the transmitter simply relays the compression information it received over each of the feedback links to the other receiver, as is the case also for our motivating scheme in the previous section 6.3. Alternatively, the transmitter can also use this feedback information to first reconstruct the compressed versions of the channel outputs and then compress them jointly with the Marton codewords. The indices resulting from this latter compression are then sent to the two receivers. The following Theorem 6.4 presents the region achieved by this fourth scheme 2.

Theorem 6.4. *The capacity region $\mathcal{C}_{\mathrm{Fb}}$ includes the set $\mathcal{R}_{\mathrm{proc.}}$*[5] *of all nonnegative rate*

[4] *The subscript "sp" stands for superposition coding.*

[5] *The subscript "proc." indicates that the transmitter processes the feedback information it receives.*

pairs (R_1, R_2) *that satisfy*

$$\begin{aligned}
R_1 &\le I(U_0, U_1; Y_1, \tilde{Y}_1, V|Q) - I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1, Q) \\
R_2 &\le I(U_0, U_2; Y_2, \tilde{Y}_2, V|Q) - I(V; U_0, U_1, U_2, \tilde{Y}_1|\tilde{Y}_2, Y_2, Q) \\
R_1+R_2 &\le I(U_0, U_1; Y_1, \tilde{Y}_1, V|Q)+I(U_2; Y_2, \tilde{Y}_2, V|U_0, Q) \\
&\quad -I(V; U_0,U_1,U_2,\tilde{Y}_2|\tilde{Y}_1,Y_1,Q)-I(U_1; U_2|U_0,Q) \\
R_1+R_2 &\le I(U_0, U_2; Y_2, \tilde{Y}_2, V|Q)+I(U_1; Y_1, \tilde{Y}_1, V|U_0, Q) \\
&\quad -I(V; U_0,U_1,U_2,\tilde{Y}_1|\tilde{Y}_2,Y_2,Q)-I(U_1; U_2|U_0,Q) \\
R_1+R_2 &\le I(U_0, U_1; Y_1, \tilde{Y}_1, V|Q) + I(U_0, U_2; Y_2, \tilde{Y}_2, V|Q) \\
&\quad -I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1, Q) \\
&\quad -I(V; U_0,U_1,U_2,\tilde{Y}_1|\tilde{Y}_2,Y_2,Q)-I(U_1; U_2|U_0,Q)
\end{aligned}$$

for some pmf $P_Q P_{U_0U_1U_2|Q} P_{\tilde{Y}_1|Y_1Q} P_{\tilde{Y}_2|Y_2Q} P_{V|U_0U_1U_2\tilde{Y}_1\tilde{Y}_2}$ *and some function* $f\colon \mathcal{X} \to \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q}$ *where the feedback-rates have to satisfy*

$$I(Y_1; \tilde{Y}_1|U_0, U_1, U_2, \tilde{Y}_2, Q) \le R_{\mathrm{Fb},1} \tag{6.28a}$$

$$I(Y_2; \tilde{Y}_2|U_0, U_1, U_2, \tilde{Y}_1, Q) \le R_{\mathrm{Fb},2} \tag{6.28b}$$

$$I(Y_1, Y_2; \tilde{Y}_1, \tilde{Y}_2|U_0, U_1, U_2, Q) \le R_{\mathrm{Fb},1} + R_{\mathrm{Fb},2}. \tag{6.28c}$$

and where $X = f(U_0, U_1, U_2, Q)$.

Proof. See Section 6.6.4. □

When the feedback rates $R_{\mathrm{Fb},1}, R_{\mathrm{Fb},2}$ are sufficiently large, we can choose $\tilde{Y}_k = Y_k$ for $k \in \{1, 2\}$.

Corollary 6.2. *In the limit $R_{\mathrm{Fb},1}, R_{\mathrm{Fb},2} \to \infty$, $\mathcal{C}_{\mathrm{Fb}}$ includes the set $\mathcal{R}_{\mathrm{proc.}}^{\infty}$ of all nonnegative rate pairs (R_1, R_2) that satisfy*

$$R_1 \le I(U_0, U_1; Y_1, V|Q) - I(V; U_0, U_1, U_2, Y_2|Y_1, Q) \tag{6.29a}$$

$$R_2 \le I(U_0, U_2; Y_2, V|Q) - I(V; U_0, U_1, U_2, Y_1|Y_2, Q) \tag{6.29b}$$

$$\begin{aligned} R_1 + R_2 \le\ & I(U_1; Y_1, V|U_0, Q) + I(U_2; Y_2, V|U_0, Q) \\ & - I(U_1; U_2|U_0, Q) + \min_{k\in\{1,2\}} \{I(U_0; Y_k, V|Q) \\ & - I(V; U_0, U_1, U_2, Y_1, Y_2|Y_k, Q)\} \end{aligned} \tag{6.29c, 6.29d}$$

$$\begin{aligned} R_1 + R_2 \le\ & I(U_0, U_1; Y_1, V|Q) - I(V; U_0, U_1, U_2, Y_2|Y_1, Q) \\ & + I(U_0, U_2; Y_2, V|Q) - I(V; U_0, U_1, U_2, Y_1|Y_2, Q) \\ & - I(U_1; U_2|U_0, Q) \end{aligned} \tag{6.29e}$$

for some pmf $P_Q P_{U_0U_1U_2|Q} P_{V|U_0U_1U_2Y_1Y_2}$ and some function $f\colon \mathcal{X} \to \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q}$, where $X = f(U_0, U_1, U_2, Q)$.

Remark 6.4. *The region $\mathcal{R}_{\mathrm{proc.}}^{\infty}$ improves over the Shayevitz-Wigger region for output feedback when this latter is specialized to the choice $V_1 = V_2 = V_0$. Observe that except for the sum-rate constraints* (6.29d) *and* (4.2c)*, all other rate constraints defining $\mathcal{R}_{\mathrm{proc.}}$ and the Shayevitz-Wigger region coincide when the latter are specialized to $V_1 = V_2 = V_0$. Since $\min_{k=\{1,2\}}\{a_k - b_k\} \ge \min_{k\in\{1,2\}} a_k - \max_{k\in\{1,2\}} b_k$ holds for any nonnegative $\{a_k, b_k\}_{k=1}^2$, we conclude that the rate region $\mathcal{R}_{\mathrm{proc.}}$ contains the Shayevitz-Wigger region specialized to the choice $V_1 = V_2 = V_0$. As proved in [57], our region $\mathcal{R}_{\mathrm{proc.}}^{\infty}$ thus also recovers the two-user capacity result in [58, 59] and the degrees of freedom achievability result in [60].*

6.4.2 Usefulness of Feedback

Our third scheme 1C (which leads to Theorem 6.3) can be used to prove the following result on the usefulness of rate-limited feedback for DMBCs. (Similar results can be shown based on our other proposed schemes.)

Theorem 6.5. *Fix a DMBC. Consider random variables $(U_0^{(\mathrm{M})}, U_1^{(\mathrm{M})}, U_2^{(\mathrm{M})}, X^{(\mathrm{M})})$ such*

that

$$\Gamma^{(\mathrm{M})} := I(U_0^{(\mathrm{M})}; Y_2^{(\mathrm{M})}) - I(U_0^{(\mathrm{M})}; Y_1^{(\mathrm{M})}) > 0. \tag{6.30}$$

Let the rate pair $(R_1^{(\mathrm{M})}, R_2^{(\mathrm{M})})$ *satisfy Marton's constraints* (3.30) *when evaluated for* $(U_0^{(\mathrm{M})}, U_1^{(\mathrm{M})}, U_2^{(\mathrm{M})}, X^{(\mathrm{M})})$ *where constraint* (3.30b) *has to hold with strict inequality.*

Also, let $(R_1^{(\mathrm{Enh})}, R_2^{(\mathrm{Enh})})$ *be a rate pair in the capacity region* $\mathcal{C}_{\mathrm{Enh}}^{(1)}$ *of the enhanced DMBC.*

If the feedback-rate from Receiver 1 is positive, $R_{\mathrm{Fb},1} > 0$*, then for all sufficiently small* $\gamma \in (0,1)$*, the rate pair* (R_1, R_2)*,*

$$R_1 = (1-\gamma)R_1^{(\mathrm{M})} + \gamma R_1^{(\mathrm{Enh})} \tag{6.31a}$$

$$R_2 = (1-\gamma)R_2^{(\mathrm{M})} + \gamma R_2^{(\mathrm{Enh})} \tag{6.31b}$$

lies in $\mathcal{R}_{\mathrm{relay,hb}}^{(1)}$*,*

$$(R_1, R_2) \in \mathcal{R}_{\mathrm{relay,hb}}^{(1)}, \tag{6.32}$$

and is thus achievable.

An analogous statement holds when indices 1 *and* 2 *are exchanged.*

Proof. See Appendix 6.D. □

The following remark elaborates on the condition of the theorem that a rate pair satisfies constraint (3.30b) with strict inequality.

Remark 6.5. *For given random variables* $U_0^{(\mathrm{M})}, U_1^{(\mathrm{M})}, U_2^{(\mathrm{M})}, X^{(\mathrm{M})}$ *Marton's region, i.e., the rate region defined by constraints* (3.30)*, is either a pentagon (both single-rate constraints as well as at least one of the sum-rates are active), a quadrilateral (only the two single-rate constraints are active), or a triangle (only one single-rate constraint and at least one of the sum-rate constraints are active).*

In the case of superposition coding with $U_1^{(\mathrm{M})} =$*const. and* $U_2^{(\mathrm{M})} = X^{(\mathrm{M})}$ *and when condition* (6.30) *holds, then the region is a quadrilateral and the only active constraints are* (3.30a) *and* (3.30c)*. Thus, in this case, constraint* (3.30b) *holds with strict inequality for all rate pairs in this region.*

Whenever the region defined by Marton's constraints (3.30) *is a pentagon, then the*

only rate pair in this pentagon that satisfies constraint (3.30b) *with equality is the dominant corner point of maximum R_2-rate.*

Corollary 6.3. *Assume $R_{\mathrm{Fb},1} > 0$. If there exists a rate pair $(R_1^{(\mathrm{M})}, R_2^{(\mathrm{M})})$ that satisfies the conditions in Theorem 6.5 and that lies on the boundary of $\mathcal{R}_{\mathrm{Marton}}$ but strictly in the interior of $\mathcal{C}_{\mathrm{Enh}}^{(1)}$, then*

$$\mathcal{R}_{\mathrm{Marton}} \subsetneq \mathcal{C}_{\mathrm{Fb}}. \tag{6.33}$$

If for the considered DMBC moreover $\mathcal{R}_{\mathrm{Marton}} = \mathcal{C}_{\mathrm{NoFb}}$,

$$\mathcal{C}_{\mathrm{NoFb}} \subsetneq \mathcal{C}_{\mathrm{Fb}}. \tag{6.34}$$

Proof. Inclusion (6.34) follows from (6.33). We show (6.33). Since $(R_1^{(\mathrm{M})}, R_2^{(\mathrm{M})})$ is in the interior of $\mathcal{C}_{\mathrm{Enh}}^{(1)}$, there exists a rate pair $(R_1^{(\mathrm{Enh})}, R_2^{(\mathrm{Enh})}) \in \mathcal{C}_{\mathrm{Enh}}^{(1)}$ with $R_1^{(\mathrm{Enh})} > R_1^{(\mathrm{M})}$ and $R_2^{(\mathrm{Enh})} > R_2^{(\mathrm{M})}$. Now, since $(R_1^{(\mathrm{M})}, R_2^{(\mathrm{M})})$ lies on the boundary of $\mathcal{R}_{\mathrm{Marton}}$, the rate pair in (6.31) must lie outside $\mathcal{R}_{\mathrm{Marton}}$ for any $\gamma \in (0,1)$. By Theorem 6.5, Equation (6.32), this rate pair is achievable with rate-limited feedback for all $\gamma \in (0,1)$ that are sufficiently close to 0. □

For many DMBCs such as the BSBC or the BEBC with unequal cross-over probabilities or unequal erasure probabilities to the two receivers, or the BSC/BEC-BC where the two channels have different capacities, the conditions of Corollary 6.3 can easily be checked. Thus, our corollary immediately shows that for these DMBCs rate-limited feedback strictly increases capacity. (See also Examples 6.1 and 6.2

For the BSBC and the BEBC, Theorem 6.5 can even be used to show that all the boundary points $(R_1 > 0, R_2 > 0)$ of $\mathcal{C}_{\mathrm{NoFb}}$ can be improved with rate-limited feedback, see the following Corollary 6.4, the paragraph thereafter, and Example 6.1 in the next Section.

More generally speaking, Corollary 6.3 is particularly interesting in view of the following class of BCs. We introduce the new term *strictly essentially less-noisy.*

Definition 8 (Strictly Essentially Less-Noisy)**.** *The definition of a* strictly essentially less-noisy *DMBC coincides with the definition of an essentially less-noisy DMBC except that inequality* (3.15) *needs to be strict whenever $I(U;Y_1) > 0$.*

The BSBC and the BEBC with different cross-over probabilities or different erasure probabilities at the two receivers are strictly essentially less-noisy.

Corollary 6.4. *Consider a DMBC where Y_2 is strictly essentially less-noisy than Y_1. Assume $R_{\mathrm{Fb},1} > 0$. We have:*

1. *If a rate pair (R_1, R_2) lies on the boundary of $\mathcal{C}_{\mathrm{NoFb}}$ but in the interior of $\mathcal{C}_{\mathrm{Enh}}^{(1)}$, then (R_1, R_2) lies in the interior of $\mathcal{C}_{\mathrm{Fb}}$, i.e., with rate-limited feedback one can improve over this rate pair.*
2. *If $\mathcal{C}_{\mathrm{NoFb}}$ does not coincide with $\mathcal{C}_{\mathrm{Enh}}^{(1)}$, then $\mathcal{C}_{\mathrm{NoFb}}$ is also a strict subset of $\mathcal{C}_{\mathrm{Fb}}$, i.e., feedback strictly improves capacity.*

Analogous statements hold if indices 1 *and* 2 *are exchanged.*

As mentioned, all BSBCs and BEBCs with unequal cross-over probabilities or unequal erasure probabilities to the two receivers are strictly essentially less-noisy. Also, for these BCs $\mathcal{C}_{\mathrm{NoFb}}$ has no common boundary points $(R_1 > 0, R_2 > 0)$ with the sets $\mathcal{C}_{\mathrm{Enh}}^{(1)}$ or $\mathcal{C}_{\mathrm{Enh}}^{(2)}$ unless the BC is physically degraded. Thus, for these BCs the corollary implies that, unless the BC is physically degraded, rate-limited feedback improves all boundary points $(R_1 > 0, R_2 > 0)$ of $\mathcal{C}_{\mathrm{NoFb}}$ whenever $R_{\mathrm{Fb},1}, R_{\mathrm{Fb},2} > 0$.

Notice that when a DMBC is physically degraded in the sense that output Y_1 is a degraded version of Y_2, then $\mathcal{C}_{\mathrm{NoFb}} = \mathcal{C}_{\mathrm{Enh}}^{(1)}$. Of course (even infinite-rate) feedback does not increase the capacity of physically degraded DMBCs [54].

Proof of Corollary 6.4. 2.) follows from 1.) We prove 1.) For strictly essentially less-noisy DMBCs, $\mathcal{C}_{\mathrm{NoFb}}$ is achieved by superposition coding. Thus, $\mathcal{R}_{\mathrm{Marton}} = \mathcal{C}_{\mathrm{NoFb}}$ and in the evaluation of Marton's region one can restrict attention to auxiliaries of the form U_1 =const. and $U_2 = X$. By the definition of strictly essentially-less noisy, when evaluating Marton's region we can further restrict attention to auxiliary random variables that satisfy (6.30). Thus, by Remark 6.5, any boundary point of $\mathcal{R}_{\mathrm{Marton}}$ satisfies the conditions of Theorem 6.5. Repeating the proof steps for Corollary 6.3, we can prove that these boundary points cannot be boundary points of $\mathcal{C}_{\mathrm{Fb}}$ whenever they lie in the interior of $\mathcal{C}_{\mathrm{Enh}}^{(1)}$. □

6.5 Examples

Example 6.1. *Consider the BSBC with input X and outputs Y_1 and Y_2 described by:*

$$Y_1 = X \oplus Z_1, \quad Y_2 = X \oplus Z_2, \tag{6.35a}$$

for $Z_1 \sim Bern(p_1)$ and $Z_2 \sim Bern(p_2)$ independent noises. Let $Q = const.$, $U \sim Bern(1/2)$, $W_1 \sim Bern(\beta_1)$ and $W_2 \sim Bern(\beta_2)$, for $\beta_1, \beta_2 \in [0, 1/2]$, where U, W_1, W_2 are mutually independent. Also set $X = U \oplus W_1$, and $\tilde{Y}_1 = Y_1 \oplus W_2$. Then

$$I(U; Y_1) = 1 - H_b(\beta_1 * p_1), \quad I(X; Y_2) = 1 - H_b(p_2),$$

and

$$\begin{aligned} I(X; \tilde{Y}_1, Y_2|U) &= H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) - H_b(p_2) - H_b(\beta_2 * p_1) \\ I(\tilde{Y}_1; Y_1|Y_2, U) &= H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) - H_b(\beta_1 * p_2) - H_b(\beta_2) \end{aligned}$$

where

$$\begin{aligned} \alpha_1 &= (p_1 * \beta_2)p_2\beta_1 + (1 - p_1 * \beta_2)\bar{p_2}\bar{\beta_1} \\ \alpha_2 &= (p_1 * \beta_2)\bar{p_2}\beta_1 + (1 - p_1 * \beta_2)p_2\bar{\beta_1} \\ \alpha_3 &= (p_1 * \beta_2)\bar{p_2}\bar{\beta_1} + (1 - p_1 * \beta_2)p_2\beta_1 \\ \alpha_4 &= (p_1 * \beta_2)p_2\bar{\beta_1} + (1 - p_1 * \beta_2)\bar{p_2}\beta_1. \end{aligned}$$

For this choice, the region defined by the constraints in Corollary 6.1 evaluates to:

$$R_1 \leq 1 - H_b(\beta_1 * p_1) \tag{6.36a}$$

$$\begin{aligned} R_1 + R_2 \leq 1 &- H_b(\beta_1 * p_1) + H(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ &- H_b(p_2) - H_b(\beta_2 * p_1) \end{aligned} \tag{6.36b}$$

$$\begin{aligned} R_1 + R_2 \leq 1 &- H_b(p_2) - H(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ &+ H_b(\beta_1 * p_2) + H_b(\beta_2) \end{aligned} \tag{6.36c}$$

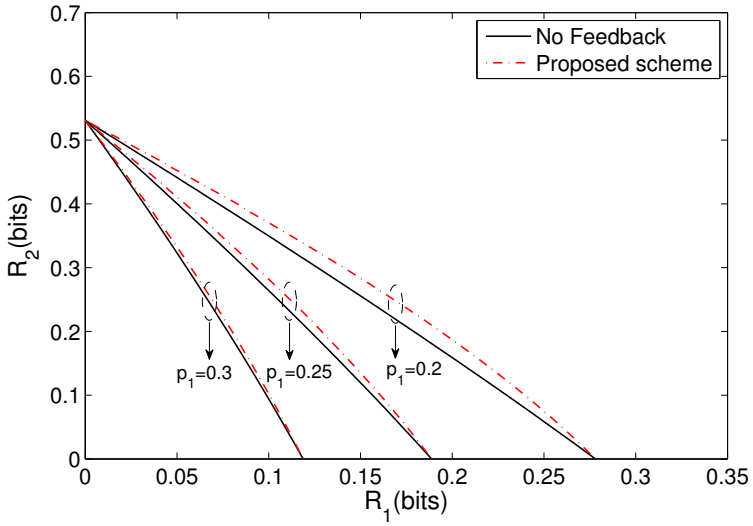


Figure 6.2: $\mathcal{C}_{\text{NoFb}}$ and the achievable region in (6.36) are plotted for BSBCs with parameters $p_2 = 0.1$ and $p_1 \in \{0.2, 0.25, 0.3\}$ and for feedback rate $R_{\text{Fb},1} = 0.8$.

for some $\beta_1, \beta_2 \in [0, 1/2]$ *satisfying*

$$H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) - H_b(\beta_1 * p_2) - H_b(\beta_2) \leq R_{\text{Fb},1} \tag{6.37}$$

and where $H(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ *denotes the entropy of a quaternary random variable with probability masses* $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

The region is plotted in Figure 6.2 against the non-feedback capacity region $\mathcal{C}_{\text{NoFb}}$.

Example 6.2. *Consider a DMBC where the channel from* X *to* Y_1 *is a BSC with crossover probability* $p \in (0, 1/2)$*, and the channel from* X *to* Y_2 *is an independent BEC with erasure probability* $e \in (0, 1)$*. We show that our feedback regions* $\mathcal{R}^{(1)}_{\text{relay,sp}}$ *and* $\mathcal{R}^{(2)}_{\text{relay,sp}}$ *improve over a large part of the boundary points of* $\mathcal{C}_{\text{NoFb}}$ *for all values of* e, p *unless* $H_b(p) = e$*, no matter how small* $R_{\text{Fb},1}, R_{\text{Fb},2} > 0$.

We distinguish different parameter ranges of our channel.

- $0 < e < H_b(p)$*: In this case, the non-feedback capacity region* $\mathcal{C}_{\text{NoFb}}$ *[135] is*

formed by the set of rate pairs (R_1, R_2) *that for some* $s \in [0, 1/2]$ *satisfy*

$$R_1 \leq 1 - H_b(s * p), \tag{6.38a}$$

$$R_2 \leq (1 - e)H_b(s), \tag{6.38b}$$

$$R_1 + R_2 \leq 1 - e. \tag{6.38c}$$

We specialize the region $\mathcal{R}^{(1)}_{\text{relay,sp}}$ *to the following choices. Let* $Q = const.$, $U \sim \text{Bern}(1/2)$, $X = U \oplus V$, *where* $V \sim \text{Bern}(s)$ *independent of* U, *and* $\tilde{Y}_1 = Y_1$ *with probability* $\gamma \in (0, 1)$ *and* $\tilde{Y}_1 = ?$ *with probability* $1 - \gamma$, *where*

$$\gamma \leq \frac{R_{\text{Fb},1}}{(1 - e)H_b(p) + eH_b(s * p)}. \tag{6.39}$$

Condition (6.39) *assures that the described choice satisfies* (6.26). *Then,*

$$I(U; Y_1) = 1 - H_b(s * p), \quad I(X; Y_2) = 1 - e,$$

and

$$I(X; \tilde{Y}_1, Y_2|U) = \gamma e(H_b(s * p) - H_b(p)) + (1 - e)H_b(s)$$
$$I(\tilde{Y}_1; Y_1|Y_2, U) = \gamma(H_b(p)(1 - e) + eH_b(s * p)).$$

When $R_{\text{Fb},1} > 0$, *by Corollary 6.1, all rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq 1 - H_b(s * p) \tag{6.40a}$$

$$R_1 + R_2 \leq 1 - H_b(s * p) + (1 - e)H_b(s) + \gamma e(H_b(s * p) - H_b(p)) \tag{6.40b}$$

$$R_1 + R_2 \leq 1 - e - \gamma(H_b(p)(1 - e) + eH_b(s * p)) \tag{6.40c}$$

are achievable for any $\gamma \in (0, 1)$ *satisfying* (6.39).
As shown in [135], the points (R_1, R_2) *of the form*

$$(1 - H_b(s * p), (1 - e)H_b(s)), \qquad s \in (0, s_0), \tag{6.41}$$

are all on the dominant boundary of $\mathcal{C}_{\text{NoFb}}$*, where* $s_0 \in (0, 1/2)$ *is the unique solution to*

$$1 - H_b(s_0 * p) + (1 - e)H_b(s_0) = 1 - e. \tag{6.42}$$

For these boundary points, only the single-rate constraints (6.38a) *and* (6.38b) *are active, but not* (6.38c)*. Thus, comparing* (6.41) *to our feedback region* (6.40)*, we can conclude that by choosing* γ *sufficiently small, all boundary points* (6.41) *lie strictly in the interior of our feedback region* $\mathcal{R}^{(1)}_{\text{relay,sp}}$ *when* $R_{\text{Fb},1} > 0$*.*

- $0 < H_b(p) < e < 1$*: The non-feedback capacity region* $\mathcal{C}_{\text{NoFb}}$ *equals the time-sharing region given by the union of all rate pairs* (R_1, R_2) *that for some* $\alpha \in [0, 1]$ *satisfy*

$$R_1 \le \alpha(1 - H_b(p)) \tag{6.43a}$$

$$R_2 \le (1 - \alpha)(1 - e). \tag{6.43b}$$

We specialize the region $\mathcal{R}^{(2)}_{\text{relay,sp}}$ *to the following choices:* $Q \sim \text{Bern}(\alpha)$*; if* $Q = 0$ *then* $U \sim \text{Bern}(1/2)$*,* $X = U$*, and* $\tilde{Y}_2 = const.$*; if* $Q = 1$ *then* $U = const.$*,* $X \sim \text{Bern}(1/2)$*, and* $\tilde{Y}_2 = Y_2$ *with probability* $\gamma \in (0, 1)$ *and* $\tilde{Y}_2 = ?$ *with probability* $1 - \gamma$*, where in order to satisfy the average feedback rate constraint,*

$$\gamma \le \frac{R_{\text{Fb},2}}{\alpha((1 - e)H_b(p) + H_b(e))}. \tag{6.44}$$

When $R_{\text{Fb},2} > 0$*, by Theorem 6.3, all rate pairs* (R_1, R_2) *satisfying*

$$R_1 \le \alpha(1 - H_b(p)) + \alpha(1 - e)\gamma H_b(p) \tag{6.45a}$$

$$R_1 + R_2 \le (1 - \alpha)(1 - e) + \alpha(1 - H_b(p)) + \alpha(1 - e)\gamma H_b(p) \tag{6.45b}$$

$$R_1 + R_2 \le (1 - H_b(p)) - (1 - \alpha)\gamma H_b(e). \tag{6.45c}$$

are achievable for any $\gamma \in (0, 1)$ *satisfying* (6.44)*.*
Since here $1 - H_b(p) > 1 - e$*, for small* $\gamma > 0$ *the feedback region in* (6.45) *improves over* $\mathcal{C}_{\text{NoFb}}$ *given in* (6.43)*. In fact,* (6.45) *improves over all boundary points* $(R_1 > 0, R_2 > 0)$ *of* $\mathcal{C}_{\text{NoFb}}$*.*

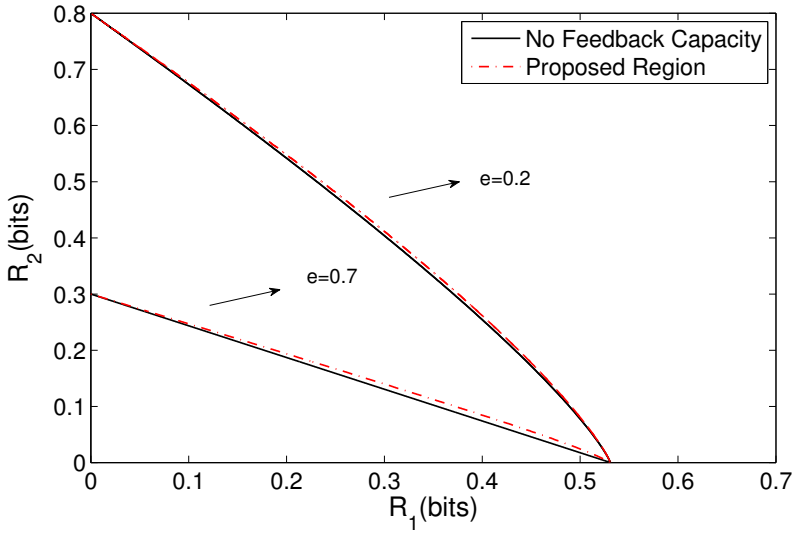


Figure 6.3: $\mathcal{C}_{\text{NoFb}}$ and the achievable regions in (6.40) and (6.45) are plotted for a BSC/BEC-BC when the BSC has parameter $p = 0.1$ and the BEC has parameter $e \in \{0.2, 0.7\}$. Notice that $0.2 < H_b(p) < 0.7$. The feedback rates $R_{\text{Fb},1} = R_{\text{Fb},2} = 0.8$.

Remark 6.6. *The BSC/BEC-BC in Example 6.2, is particularly interesting, because depending on the values of the parameters e and p, the BC is either degraded, less noisy, more capable, or essentially less-noisy [135]. We conclude that our feedback regions* $\mathcal{R}^{(1)}_{\text{relay,sp}}$ *and* $\mathcal{R}^{(2)}_{\text{relay,sp}}$ *can improve over the non-feedback capacity regions for all these classes of BCs even with only one feedback link that is of arbitrary small, but positive rate.*

We plotted our regions (6.40) *and* (6.45) *versus the non-feedback capacity region in Figure 6.3 for* $p = 0.1$ *and* $e = 0.2$ *or* $e = 0.7$. *In the first case the DMBC is more capable and in the second case it is essentially less-noisy.*

In the next example we consider the Gaussian BC with independent noises. We evaluate the region defined by the constraints of Corollary 6.1 for a set of jointly Gaussian distributions on the input and the auxiliary random variables. A rigorous proof that our achievability result in Corollary 6.1 holds also for the Gaussian BC and Gaussian random variables is omitted for brevity.

Example 6.3. *Consider the* Gaussian broadcast channel

$$Y_1 = X + Z_1 \tag{6.46a}$$

$$Y_2 = X + Z_2 \tag{6.46b}$$

where $Z_1 \sim \mathcal{N}(0, N_1)$ *and* $Z_2 \sim \mathcal{N}(0, N_2)$ *are independent noises. Assume an average transmission power* P*, and* $0 < N_2 < N_1 < P$*.*

Let $Q = const.$*,* $U \sim \mathcal{N}(0, \bar{\alpha}P)$*,* $W_1 \sim \mathcal{N}(0, \alpha P)$ *and* $W_2 \sim \mathcal{N}(0, \beta)$*, for* $\alpha \in [0, 1], \beta > 0$*, where* U, W_1, W_2 *are mutually independent. Set* $X = U + W_1$*,* $\tilde{Y}_1 = Y_1 + W_2$*, then*

$$I(U; Y_1) = C\Big(\frac{\bar{\alpha}P}{\alpha P + N_1}\Big), \quad I(X; Y_2) = C\Big(\frac{P}{N_2}\Big),$$

and

$$I(X; Y_2, \tilde{Y}_1 | U) = C\Big(\frac{\alpha P}{N_2}\Big) + C\Big(\frac{\alpha P N_2}{(\alpha P + N_2)(N_1 + \beta)}\Big)$$
$$I(\tilde{Y}_1; Y_1 | Y_2, U) = C\Big(\frac{\alpha P(N_1 + N_2) + N_1 N_2}{\beta(N_2 + \alpha P)}\Big).$$

For these choices, the region defined by the constraints in Corollary 6.1 evaluates to:

$$R_1 \le C\Big(\frac{\bar{\alpha}P}{\alpha P + N_1}\Big) \tag{6.47a}$$

$$R_1 + R_2 \le C\Big(\frac{\bar{\alpha}P}{\alpha P + N_1}\Big) + C\Big(\frac{\alpha P}{N_2}\Big) + C\Big(\frac{\alpha P N_2}{(\alpha P + N_2)(N_1 + \beta)}\Big) \tag{6.47b}$$

$$R_1 + R_2 \le C\Big(\frac{P}{N_2}\Big) - C\Big(\frac{\alpha P(N_1 + N_2) + N_1 N_2}{\beta(N_2 + \alpha P)}\Big) \tag{6.47c}$$

for some $\alpha \in [0, 1]$ *and* $\beta > 0$ *satisfying*

$$C\Big(\frac{\alpha P(N_1 + N_2) + N_1 N_2}{\beta(N_2 + \alpha P)}\Big) \le R_{\mathrm{Fb},1}. \tag{6.48}$$

Here, we use $C(x) := \frac{1}{2}\log_2(1 + x)$*, for any* $x \ge 0$*.*

The region is plotted in Figure 6.4 against the non-feedback capacity region $\mathcal{C}_{\mathrm{NoFb}}$

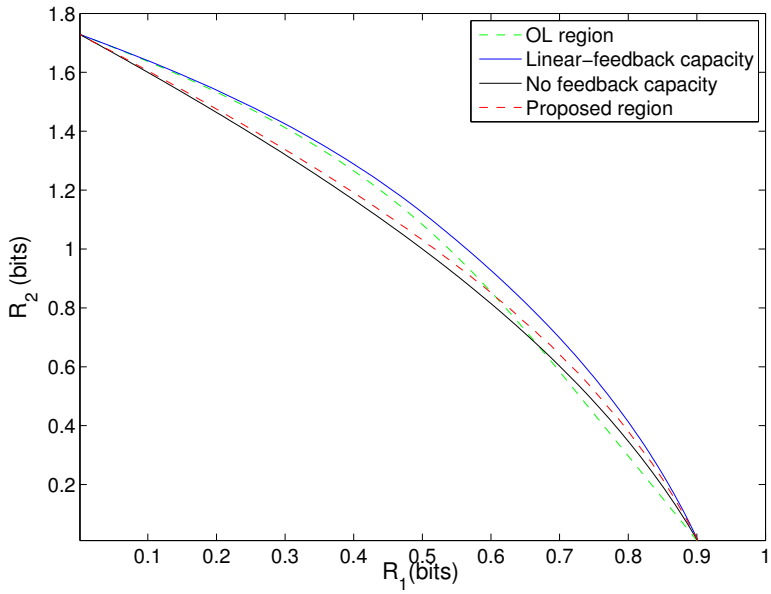


Figure 6.4: $\mathcal{C}_{\text{NoFb}}$ and the achievable region in (6.47) are plotted for Gaussian BCs with parameters $P = 10$, $N_2 = 1$, $N_1 = 4$ and feedback rate $R_{\text{Fb},1} = 0.8$.

and the region achieved by Ozarow-Leung coding scheme [62], and the linear-feedback capacity region [16, 17].

Example 6.4. *(Blackwell Channel with State [57]) We consider the Blackwell DMBC with random state. The state is described by a random variable* $S \sim \text{Bern}(1/2)$*, which is also part of the outputs. That means Receiver 1's output is* $Y_1 = (Y_1^*, S)$ *and Receiver 2's output is* $Y_2 = (Y_2^*, S)$*. If* $S = 0$ *then the BC to* Y_1^* *and* Y_2^* *is a reversed Blackwell channel:*

$$Y_1^* = \begin{cases} 0 & X = 0 \\ 1 & X = 1, 2 \end{cases} \qquad Y_2^* = \begin{cases} 0 & X = 0, 2 \\ 1 & X = 1. \end{cases} \tag{6.49}$$

If $S = 1$*, then the BC to* Y_1^* *and* Y_2^* *is a standard Blackwell channel:*

$$Y_1^* = \begin{cases} 0 & X = 0, 2 \\ 1 & X = 1 \end{cases} \qquad Y_2^* = \begin{cases} 0 & X = 0 \\ 1 & X = 1, 2. \end{cases} \tag{6.50}$$

For this BC, the non-feedback capacity region is achieved by time-sharing and the maximum sum-rate is 1. In [57] it was shown that the Shayevitz-Wigger scheme with choices of auxiliary random variables as in (4.3) *and* (4.4) *achieves the rate pairs* $(0.5958, 0.5958)$ *and* $(0.6103, 0.6103)$*, respectively. By Remark 6.2, we obtain that the proposed scheme pertaining to Theorem 6.2 can enlarge the non-feedback capacity of this BC. Notice that for this setup,* $I(U;Y_2) - I(U;Y_1) = 0$ *holds for all* P_{UX}*, which means the statement above holds even when one of the receivers is not "stronger" than the other.*

6.6 Coding Schemes

6.6.1 Coding Scheme 1A: Sliding-Window Decoding (Theorem 6.1)

For simplicity, we only describe the scheme for Q =const. A general Q can be introduced by coded time-sharing [131, Section 4.5.3]. That means all the codebooks need to be superpositioned on a P_Q-i.i.d. random vector Q^n that is revealed to transmitter and receivers, and this Q^n sequence needs to be included in all the joint-typicality checks.

Choose nonnegative rates $R_1', R_2', \tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2$, auxiliary finite alphabets $\mathcal{U}_0$, $\mathcal{U}_1$, $\mathcal{U}_2$, $\tilde{\mathcal{Y}}_1$, $\tilde{\mathcal{Y}}_2$, a function $f^{(n)}$ of the form $f^{(n)}$: $\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and pmfs $P_{U_0U_1U_2}$, $P_{\tilde{Y}_1|U_0Y_1}$, $P_{\tilde{Y}_2|U_0Y_2}$. Transmission takes place over $B+1$ consecutive blocks, with length n for each block. We denote the n-length blocks of inputs and outputs in block b by x_b^n, $y_{1,b}^n$ and $y_{2,b}^n$.

Define $\mathcal{J}_k := \{1, \ldots, \lfloor 2^{n\hat{R}_k} \rfloor\}$, $\mathcal{T}_k := \{1, \ldots, \lfloor 2^{nR_k'} \rfloor\}$, and $\mathcal{L}_k := \{1, \ldots, \lfloor 2^{n\tilde{R}_k} \rfloor\}$, for $k \in \{1,2\}$. The messages are in product form: $M_k = (M_{k,1}, \ldots, M_{k,B})$, $k \in \{1,2\}$, with $M_{k,b} = (M_{c,k,b}, M_{p,k,b})$ for $b \in \{1, \ldots, B\}$. The submessages $M_{c,k,b}$, and $M_{p,k,b}$ are uniformly distributed over the sets $\mathcal{M}_{c,k} := \{1, \ldots, \lfloor 2^{nR_{c,k}} \rfloor\}$ and $\mathcal{M}_{p,k} := \{1, \ldots, \lfloor 2^{nR_{p,k}} \rfloor\}$, respectively, where $R_{p,k}, R_{c,k} > 0$ and so that $R_k = R_{p,k} + R_{c,k}$[6] . Let $R_c := (R_{c,1} + R_{c,2} + \tilde{R}_1 + \tilde{R}_2)$.

[6] Due to the floor operations and since transmission takes place over $B+1$ blocks whereas the messages M_1 and M_2 are split into only B submessages, R_1 and R_2 here do not exactly represent the transmission rates of messages M_1 and M_2. In the limit $n \to \infty$ and $B \to \infty$, which is our case of interest, R_1 and R_2 however approach these transmission rates. Therefore, we neglect this technicality in the following.

1) Codebook generation: For each block $b \in \{1, \ldots, B+1\}$, randomly and independently generate $\lfloor 2^{nR_c} \rfloor$ sequences $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $\mathbf{m}_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $l_{k,b-1} \in \mathcal{L}_k$, for $k \in \{1,2\}$. (We use vector notation for $\mathbf{m}_{c,b}$ to emphasize that it represents a pair of indices.) Each sequence $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ is drawn according to the product distribution $\prod_{i=1}^n P_{U_0}(u_{0,b,i})$, where $u_{0,b,i}$ denotes the i-th entry of $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$.

For $k \in \{1,2\}$ and each $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ randomly and conditionally independently generate $\lfloor 2^{n(R_{p,k}+R'_k)} \rfloor$ sequences $u_{k,b}^n(m_{p,k,b}, t_{k,b} | \mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $m_{p,k,b} \in \mathcal{M}_{p,k}$ and $t_{k,b} \in \mathcal{T}_k$, where each $u_{k,b}^n(m_{p,k,b}, t_{k,b} | \mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ is drawn according to the product distribution $\prod_{i=1}^n P_{U_k|U_0}(u_{k,b,i}|u_{0,b,i})$, where $u_{k,b,i}$ denotes the i-th entry of $u_{k,b}^n\big(m_{p,k,b}, t_{k,b} | \mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$.

Similarly, for $k \in \{1,2\}$ and each tuple $(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_1 \times \mathcal{L}_2$ randomly generate $\lfloor 2^{n(\tilde{R}_k + \hat{R}_k)} \rfloor$ sequences $\tilde{y}_{k,b}^n\big(l_{k,b}, j_{k,b} | \mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$, for $l_{k,b} \in \mathcal{L}_k$ and $j_{k,b} \in \mathcal{J}_k$, by drawing each $\tilde{y}_{k,b}^n\big(l_{k,b}, j_{k,b} | \mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$ according to the product distribution $\prod_{i=1}^n P_{\tilde{Y}_k|U_0Y_k}(\tilde{y}_{k,b,i}|u_{0,b,i})$ where $\tilde{y}_{k,b,i}$ denotes the i-th entry of $\tilde{y}_{k,b}^n$.

All codebooks are revealed to transmitter and receivers.

2) Encoding: We describe the encoding for a fixed block $b \in \{1, \ldots, B+1\}$. Assume that $M_{c,k,b} = m_{c,k,b}$, $M_{p,k,b} = m_{p,k,b}$, for $k \in \{1,2\}$ and that the feedback messages sent after block $b-1$ are $L_{1,b-1} = l_{1,b-1}$ and $L_{2,b-1} = l_{2,b-1}$. Define $\mathbf{m}_{c,b} := (m_{c,1,b}, m_{c,2,b})$. To simplify notation, let $l_{k,0} = m_{c,k,B+1} = m_{p,k,B+1} = 1$, for $k \in \{1,2\}$ and $\mathbf{m}_{c,B+1} = (1,1)$.

The transmitter looks for a pair $(t_{1,b}, t_{2,b}) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\big(u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}), u_{1,b}^n(m_{p,1,b}, t_{1,b} | \mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}), \\ u_{2,b}^n(m_{p,2,b}, t_{2,b} | \mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})\big) \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2}). \tag{6.51}$$

If there is exactly one pair $(t_{1,b}, t_{2,b})$ that satisfies the above condition, the transmitter chooses this pair. If there are multiple such pairs, it chooses one of them uniformly at random. Otherwise it chooses a pair $(t_{1,b}, t_{2,b})$ uniformly at random over the entire set $\mathcal{T}_1 \times \mathcal{T}_2$. In block b the transmitter then sends the inputs $x_b^n = (x_{b,1}, \ldots, x_{b,n})$, where

$$x_{b,i} = f(u_{0,b,i}, u_{1,b,i}, u_{2,b,i}), \qquad i \in \{1, \ldots, n\}, \tag{6.52}$$

and $u_{0,b,i}$ and $u_{k,b,i}$, for $k = \{1,2\}$, denote the i-th symbols of the chosen Marton codewords $u^n_{0,b}(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ and $u^n_{k,b}(m_{p,k,b}, t_{k,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, respectively.

3) Decoding and Generation of Feedback Messages at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

After each block $b \in \{1,\dots,B+1\}$, and after observing the outputs $y^n_{1,b}$, Receiver 1 looks for a pair of indices $(\hat{\mathbf{m}}^{(1)}_{c,b}, \hat{l}_{2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_2$ that satisfies

$$\big(u^n_{0,b}(\hat{\mathbf{m}}^{(1)}_{c,b}, l_{1,b-1}, \hat{l}_{2,b-1}), y^n_{1,b}\big) \in \mathcal{T}^n_{\varepsilon/8}(P_{U_0Y_1}).$$

Notice that Receiver 1 already knows $l_{1,b-1}$ because it has created it itself after the previous block $b-1$. If there are multiple such pairs, the receiver chooses one of them at random. If there is no such pair, then it chooses $(\hat{\mathbf{m}}^{(1)}_{c,b}, \hat{l}_{2,b-1})$ randomly over the set $\mathcal{M}_c \times \mathcal{L}_2$.

After decoding the cloud center in block b, Receiver 1 then looks for a tuple $(\hat{j}_{2,b-1}, \hat{m}_{p,1,b-1}, \hat{t}_{1,b-1}) \in \mathcal{J}_2 \times \mathcal{M}_{p,1} \times \mathcal{T}_1$ that satisfies

$$\begin{aligned}\big(&u^n_{0,b-1}(\hat{\mathbf{m}}^{(1)}_{c,b-1}, l_{1,b-2}, \hat{l}_{2,b-2}), u^n_{1,b-1}(\hat{m}_{p,1,b-1}, \hat{t}_{1,b-1}|\hat{\mathbf{m}}^{(1)}_{c,b-1}, l_{1,b-2}, \hat{l}_{2,b-2}),\\ &\tilde{y}^n_{2,b-1}(\hat{l}_{2,b-1}, \hat{j}_{2,b-1}|\hat{\mathbf{m}}^{(1)}_{c,b-1}, l_{1,b-2}, \hat{l}_{2,b-2}), y^n_{1,b-1}\big) \in \mathcal{T}^n_{\varepsilon}(P_{U_0U_1Y_1\tilde{Y}_2}).\end{aligned}$$

It further looks for a pair $(l_{1,b}, j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$\begin{aligned}(&\tilde{y}^n_{1,b}(l_{1,b}, j_{1,b}|\hat{\mathbf{m}}^{(1)}_{c,b}, l_{1,b-1}, \hat{l}_{2,b-1}),\\ &\qquad u^n_{0,b}(\hat{\mathbf{m}}^{(1)}_{c,b}, l_{1,b-1}, \hat{l}_{2,b-1}), y^n_{1,b}) \in \mathcal{T}^n_{\varepsilon/4}(P_{Y_1U_0\tilde{Y}_1})\end{aligned}$$

and sends the index $l_{1,b}$ over the feedback link. If there is more than one such pair $(l_{1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $l_{1,b}$ that it sends over the feedback link uniformly at random over $\mathcal{L}_1$. The receivers thus only send a feedback message at the end of each block $1,\dots,B$.

After decoding Block $B+1$, Receiver 1 produces the estimation $\hat{m}_1$ of m_1 by $\hat{m}_1 = (\hat{m}_{1,1}, \dots, \hat{m}_{1,B})$ as its guess, where $\hat{m}_{1,b} = (\hat{m}^{(1)}_{c,1,b}, \hat{m}_{p,1,b})$, for $b \in \{1,\dots,B\}$, and $\hat{m}^{(1)}_{c,1,b}$ denotes the first component of $\hat{\mathbf{m}}^{(1)}_{c,b}$.

5) Analysis: See Appendix 6.A.

6.6.2 Coding Scheme 1B: Backward Decoding (Theorem 6.2)

For simplicity, we describe the scheme without the coded time-sharing random variable Q, i.e., for Q =const.

Choose nonnegative rates $R'_1, R'_2, \tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2$, auxiliary finite alphabets $\mathcal{U}_0$, $\mathcal{U}_1$, $\mathcal{U}_2$, $\tilde{\mathcal{Y}}_1$, $\tilde{\mathcal{Y}}_2$, a function $f^{(n)}$ of the form $f^{(n)}$: $\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and pmfs $P_{U_0U_1U_2}$, $P_{\tilde{Y}_1|Y_1}$, $P_{\tilde{Y}_2|Y_2}$. Transmission takes place over $B+1$ consecutive blocks, with length n for each block. We denote the n-length blocks of inputs and outputs in block b by x_b^n, $y_{1,b}^n$ and $y_{2,b}^n$.

Define $\mathcal{J}_k := \{1, \ldots, \lfloor 2^{n\hat{R}_k} \rfloor\}$, $\mathcal{T}_k := \{1, \ldots, \lfloor 2^{nR'_k} \rfloor\}$, and $\mathcal{L}_k := \{1, \ldots, \lfloor 2^{n\tilde{R}_k} \rfloor\}$, for $k \in \{1, 2\}$. The messages are in product form: $M_k = (M_{k,1}, \ldots, M_{k,B})$, $k \in \{1, 2\}$, with $M_{k,b} = (M_{c,k,b}, M_{p,k,b})$ for $b \in \{1, \ldots, B\}$. The submessages $M_{c,k,b}$, and $M_{p,k,b}$ are uniformly distributed over the sets $\mathcal{M}_{c,k} := \{1, \ldots, \lfloor 2^{nR_{c,k}} \rfloor\}$ and $\mathcal{M}_{p,k} := \{1, \ldots, \lfloor 2^{nR_{p,k}} \rfloor\}$, respectively, where $R_{p,k}, R_{c,k} > 0$ and so that $R_k = R_{p,k} + R_{c,k}$. Let $R_c := (R_{c,1} + R_{c,2} + \tilde{R}_1 + \tilde{R}_2)$.

1) Codebook generation: For each block $b \in \{1, \ldots, B+1\}$, randomly and independently generate $\lfloor 2^{nR_c} \rfloor$ sequences $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $\mathbf{m}_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $l_{k,b-1} \in \mathcal{L}_k$, for $k \in \{1, 2\}$. Each sequence $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ is drawn according to the product distribution $\prod_{i=1}^n P_{U_0}(u_{0,b,i})$, where $u_{0,b,i}$ denotes the i-th entry of $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$.

For $k \in \{1, 2\}$ and each tuple $(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ randomly generate $\lfloor 2^{n(R_{p,k}+R'_k)} \rfloor$ sequences $u_{k,b}^n(m_{p,k,b}, t_{k,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $m_{p,k,b} \in \mathcal{M}_{p,k}$ and $t_{k,b} \in \mathcal{T}_k$ by randomly drawing each codeword $u_{k,b}^n(m_{p,k,b}, t_{k,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ according to the product distribution $\prod_{i=1}^n P_{U_k|U_0}(u_{k,b,i}|u_{0,b,i})$, where $u_{k,b,i}$ denotes the i-th entry of $u_{k,b}^n$.

Also, for $k \in \{1, 2\}$, generate $\lfloor 2^{n(\tilde{R}_k+\hat{R}_k)} \rfloor$ sequences $\tilde{y}_{k,b}^n(l_{k,b}, j_{k,b})$, for $l_{k,b} \in \mathcal{L}_k$ and $j_{k,b} \in \mathcal{J}_k$, by drawing all the entries independently according to the same distribution $P_{\tilde{Y}_k}$.

All codebooks are revealed to transmitter and receivers.

2) Encoding: We describe the encoding for a fixed block $b \in \{1, \ldots, B+1\}$. Assume that $M_{c,k,b} = m_{c,k,b}$, $M_{p,k,b} = m_{p,k,b}$, for $k \in \{1, 2\}$, and that the feedback messages sent after block $b-1$ are $L_{1,b-1} = l_{1,b-1}$ and $L_{2,b-1} = l_{2,b-1}$. Define $\mathbf{m}_{c,b} := (m_{c,1,b}, m_{c,2,b})$. To simplify notation, let $l_{k,0} = m_{c,k,B+1} = m_{p,k,B+1} = 1$, for $k \in \{1, 2\}$ and $m_{c,B+1} = (1, 1)$.

The transmitter looks for a pair $(t_{1,b}, t_{2,b}) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\big(u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}), u_{1,b}^n(m_{p,1,b}, t_{1,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}), \\ u_{2,b}^n(m_{p,2,b}, t_{2,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})\big) \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2}). \tag{6.53}$$

If there is exactly one pair $(t_{1,b}, t_{2,b})$ that satisfies the above condition, the transmitter chooses this pair. If there are multiple such pairs, it chooses one of them uniformly at random. Otherwise it chooses a pair $(t_{1,b}, t_{2,b})$ uniformly at random over the entire set $\mathcal{T}_1 \times \mathcal{T}_2$. In block b the transmitter then sends the inputs $x_b^n = (x_{b,1}, \ldots, x_{b,n})$, where

$$x_{b,i} = f^{(n)}(u_{0,b,i}, u_{1,b,i}, u_{2,b,i}), \qquad i \in \{1, \ldots, n\}, \tag{6.54}$$

and $u_{0,b,i}$, $u_{1,b,i}$, $u_{2,b,i}$ denote the i-th symbols of the chosen Marton codewords $u_{0,b}^n$, $u_{1,b}^n$, and $u_{2,b}^n$.

3) Generation of Feedback Messages at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

After each block $b \in \{1, \ldots, B\}$, and after observing the outputs $y_{1,b}^n$, Receiver 1 looks for a pair $(l_{1,b}, j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$(\tilde{y}_{1,b}^n(l_{1,b}, j_{1,b}), y_{1,b}^n) \in \mathcal{T}_{\varepsilon/4}^n(P_{Y_1\tilde{Y}_1}) \tag{6.55}$$

and sends the index $l_{1,b}$ over the feedback link. If there is more than one such pair $(l_{1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $l_{1,b}$ that it sends over the feedback link uniformly at random over $\mathcal{L}_1$.

In our scheme the receivers thus only send a feedback message at the end of each block $1, \ldots, B$.

4) Decoding at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

The receivers apply backward decoding and thus start decoding only after the transmission terminates. Then, for each block $b \in \{1, \ldots, B+1\}$, starting with the last block $B+1$, Receiver 1 performs the following operations. From the previous decoding step in block $b+1$, it already knows the feedback message $l_{2,b}$. Moreover, it also knows its own feedback messages $l_{1,b-1}$ and $l_{1,b}$ because it has created them itself, see point *3)*.

Now, when observing $y_{1,b}^n$, Receiver 1 looks for a tuple $(\hat{j}_{2,b}, \hat{\mathbf{m}}_{c,b}^{(1)}, \hat{l}_{2,b-1}, \hat{m}_{p,1,b}, \hat{t}_{1,b}) \in \mathcal{J}_2 \times \mathcal{M}_c \times \mathcal{L}_2 \times \mathcal{M}_{p,1} \times \mathcal{T}_1$ that satisfies

$$\big(u_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, l_{1,b-1}, \hat{l}_{2,b-1}), u_{1,b}^n(\hat{m}_{p,1,b}, \hat{t}_{1,b} | \hat{\mathbf{m}}_{c,b}^{(1)}, l_{1,b-1}, \hat{l}_{2,b-1}), \\ \tilde{y}_{2,b}^n(l_{2,b}, \hat{j}_{2,b}), y_{1,b}^n\big) \in \mathcal{T}_\varepsilon^n(P_{U_0 U_1 Y_1 \tilde{Y}_2}).$$

After decoding Block 1, Receiver 1 produces the product message $\hat{m}_1 = (\hat{m}_{1,1}, \ldots, \hat{m}_{1,B})$ as its guess, where $\hat{m}_{1,b} = (\hat{m}_{c,1,b}^{(1)}, \hat{m}_{p,1,b})$, for $b \in \{1, \ldots, B\}$, and $\hat{m}_{c,1,b}^{(1)}$ denotes the first component of $\hat{\mathbf{m}}_{c,b}^{(1)}$.

5) Analysis: See Appendix 6.B.

6.6.3 Coding Scheme 1C: Hybrid Sliding-Window Decoding and Backward Decoding (Theorem 6.3)

For simplicity, we only describe the scheme achieving region $\mathcal{R}_{\text{relay,hb}}^{(1)}$ for Q =const. A scheme achieving region $\mathcal{R}_{\text{relay,hb}}^{(2)}$ is obtained if in the following description indices 1 and 2 are exchanged.

1) Codebook generation: The codebooks are generated as in Scheme 1A, described in point 1) in Section 6.6.1, but where $\tilde{R}_2 = \hat{R}_2 = 0$.

2) Encoding: The transmitter performs the same encoding procedure as in Section 6.6.1, but where $l_{2,b-1} = 1$ is constant for each block $b \in \{1, \ldots, B+1\}$.

3) Receiver 1: In each block $b \in \{1, \ldots, B+1\}$, Receiver 1 first simultaneously decodes the cloud center and its satellite. Specifically, Receiver 1 looks for a tuple $(\hat{\mathbf{m}}_{c,b-1}, \hat{m}_{p,1,b-1}, \hat{t}_{1,b-1}) \in \mathcal{M}_c \times \mathcal{M}_{p,1} \times \mathcal{T}_1$ that satisfies

$$\big(u_{0,b-1}^n(\hat{\mathbf{m}}_{c,b-1}, l_{1,b-2}, 1), y_{1,b-1}^n, \\ u_{1,b-1}^n(\hat{m}_{p,1,b-1}, \hat{t}_{1,b-1} | \hat{\mathbf{m}}_{c,b-1}, l_{1,b-2}, 1)\big) \in \mathcal{T}_\varepsilon^n(P_{U_0 U_1 Y_1}).$$

It further compresses the outputs $y_{1,b}^n$ and sends the feedback message $l_{1,b}$ over the feedback link as in Scheme 1A, see point 3) in Section 6.6.1.

4) Receiver 2: Receiver 2 performs backward decoding as in Scheme 1B, see point 4) in Section 6.6.2.

5) Analysis: Similar to the analysis of the schemes 1A and 1B in presented in

appendices 6.A and 6.B. Details are omitted.

6.6.4 Coding Scheme 2: Encoder Processes Feedback-Info

The scheme described in this subsection differs from the previous scheme in that in each block b, after receiving the feedback messages $M_{\text{Fb},1,b}, M_{\text{Fb},2,b}$, the encoder first reconstructs the compressed versions of the channel outputs, $\tilde{Y}_{1,b}^n$ and $\tilde{Y}_{2,b}^n$, and then newly compresses the quintuple consisting of $\tilde{Y}_{1,b}^n$ and $\tilde{Y}_{2,b}^n$ and the Marton codewords $U_{0,b}^n$, $U_{1,b}^n$, $U_{2,b}^n$ that it had sent during block b. This new compression information is then sent to the two receivers in the next-following block $b+1$ as part of the cloud center of Marton's code.

Decoding at the receivers is based on backward decoding. For each block b, each receiver $k \in \{1,2\}$ uses its observed outputs $Y_{k,b}^n$ to simultaneously reconstruct the encoder's compressed signal and decode its intended messages sent in block b.

For simplicity, we only describe the scheme for Q =const.

Choose nonnegative rates $R_1', R_2', \tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2, \tilde{R}_v$, auxiliary finite alphabets $\mathcal{U}_0$, $\mathcal{U}_1, \mathcal{U}_2, \tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2, \mathcal{V}$, a function $f^{(n)}$ of the form $f^{(n)}\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and pmfs $P_{U_0U_1U_2}$, $P_{\tilde{Y}_1|Y_1}$, $P_{\tilde{Y}_2|Y_2}$, and $P_{V|U_0U_1U_2\tilde{Y}_1\tilde{Y}_2}$. Transmission takes place over $B+1$ consecutive blocks, with length n for each block. We denote the n-length blocks of channel inputs and outputs in block b by x_b^n, $y_{1,b}^n$ and $y_{2,b}^n$.

Define $\mathcal{J}_k := \{1, \ldots, \lfloor 2^{n\hat{R}_k} \rfloor\}$, $\mathcal{T}_k := \{1, \ldots, \lfloor 2^{nR_k'} \rfloor\}$, and $\mathcal{L}_k := \{1, \ldots, \lfloor 2^{n\tilde{R}_k} \rfloor\}$, for $k \in \{1,2\}$, and $\mathcal{N} := \{1, \ldots, \lfloor 2^{n\tilde{R}_v} \rfloor\}$ The messages are in product form: $M_k = (M_{k,1}, \ldots, M_{k,B})$, $k \in \{1,2\}$, with $M_{k,b} = (M_{c,k,b}, M_{p,k,b})$ for $b \in \{1, \ldots, B\}$. The sub-messages $M_{c,k,b}$, and $M_{p,k,b}$ are uniformly distributed over the sets $\mathcal{M}_{c,k} := \{1, \ldots, \lfloor 2^{nR_{c,k}} \rfloor\}$ and $\mathcal{M}_{p,k} := \{1, \ldots, \lfloor 2^{nR_{p,k}} \rfloor\}$, respectively, where $R_{p,k}, R_{c,k} > 0$ and so that $R_k = R_{p,k} + R_{c,k}$. Let $R_c := (R_{c,1} + R_{c,2} + \tilde{R}_v)$.

1) Codebook generation: For each block $b \in \{1, \ldots, B+1\}$, randomly and independently generate $\lfloor 2^{nR_c} \rfloor$ sequences $u_{0,b}^n(\mathbf{m}_{c,b}, n_{b-1})$, for $\mathbf{m}_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $n_{b-1} \in \mathcal{N}$. Each sequence $u_{0,b}^n(\mathbf{m}_{c,b}, n_{b-1})$ is drawn according to the product distribution $\prod_{i=1}^n P_{U_0}(u_{0,b,i})$, where $u_{0,b,i}$ denotes the i-th entry of $u_{0,b}^n(\mathbf{m}_{c,b}, n_{b-1})$.

For $k \in \{1,2\}$ and each pair $(\mathbf{m}_{c,b}, n_{b-1})$ randomly generate $\lfloor 2^{n(R_{p,k}+R_k')} \rfloor$ sequences $u_{k,b}^n(m_{p,k,b}, t_{k,b} | \mathbf{m}_{c,b}, n_{b-1})$, for $m_{p,k,b} \in \mathcal{M}_{p,k}$ and $t_{k,b} \in \mathcal{T}_k$, by drawing each codeword $u_{k,b}^n(m_{p,k,b}, t_{k,b} | \mathbf{m}_{c,b}, n_{b-1})$ according to the product distribution $\prod_{i=1}^n P_{U_k|U_0}(u_{k,b,i} | u_{0,b,i})$,

where $u_{k,b,i}$ denotes the i-th entry of $u^n_{k,b}\big(m_{p,k,b}, t_{k,b}|\mathbf{m}_{c,b}, n_{b-1}\big)$.

Also, for $k \in \{1,2\}$, generate $\lfloor 2^{n(\tilde{R}_k+\hat{R}_k)}\rfloor$ sequences $\tilde{y}^n_{k,b}(l_{k,b}, j_{k,b})$, for $l_{k,b} \in \mathcal{L}_k$ and $j_{k,b} \in \mathcal{J}_k$ by drawing all the entries independently according to the same distribution $P_{\tilde{Y}_k}$;

Finally, for each $n_{b-1} \in \mathcal{N}$, generate $\lfloor 2^{nR_v}\rfloor$ sequences $v^n_b(n_b|n_{b-1})$, for $n_b \in \mathcal{N}$ by drawing all entries independently according to the same distribution P_V.

All codebooks are revealed to transmitter and receivers.

2) Encoding: We describe the encoding for a fixed block $b \in \{1,\ldots,B+1\}$. Assume that in this block we wish to send messages $M_{c,k,b} = m_{c,k,b}$, $M_{p,k,b} = m_{p,k,b}$, for $k \in \{1,2\}$, and define $\mathbf{m}_{c,b} := (m_{c,1,b}, m_{c,2,b})$. To simplify notation, let $l_{k,0} = m_{c,k,B+1} = m_{p,k,B+1} = 1$, for $k \in \{1,2\}$, and also $n_{-1} = n_0 = 1$.

The first step in the encoding is to reconstruct the compressed outputs pertaining to the previous block $\tilde{Y}^n_{1,b-1}$ and $\tilde{Y}^n_{2,b-1}$. Assume that after block $b-1$ the transmitter received the feedback messages $L_{1,b-1} = l_{1,b-1}$ and $L_{2,b-1} = l_{2,b-1}$, and that in this previous block it had produced the Marton codewords $u^n_{0,b-1} := u^n_{0,b-1}(\mathbf{m}_{c,b-1}, n_{b-2})$, $u^n_{1,b-1} := u^n_{1,b-1}(m_{p,1,b-1}, t_{1,b-1}|\mathbf{m}_{c,b-1}, n_{b-2})$, and $u^n_{2,b-1} := u^n_{2,b-1}(m_{p,2,b-1}, t_{2,b-1}|\mathbf{m}_{c,b-1}, n_{b-2})$. The transmitter then looks for a pair $(\hat{j}_{1,b-1}, \hat{j}_{2,b-1}) \in \mathcal{J}_1 \times \mathcal{J}_2$ that satisfies

$$\big(u^n_{0,b-1}, u^n_{1,b-1}, u^n_{2,b-1}, \tilde{y}^n_{1,b-1}(l_{1,b-1}, \hat{j}_{1,b-1}), \\ \tilde{y}^n_{2,b-1}(l_{2,b-1}, \hat{j}_{2,b-1})\big) \in \mathcal{T}_{\varepsilon/4}(P_{U_0U_1U_2\tilde{Y}_1,\tilde{Y}_2}).$$

In a second step the encoder produces the new compression information pertaining to block $b-1$, which it then sends to the receivers during block b. To this end, it looks for an index $\hat{n}_{b-1} \in \mathcal{N}$ that satisfies

$$\big(u^n_{0,b-1}, u^n_{1,b-1}, u^n_{2,b-1}, \tilde{y}^n_{1,b-1}(l_{1,b-1}, \hat{j}_{1,b-1}), \\ \tilde{y}^n_{2,b-1}(l_{2,b-1}, \hat{j}_{2,b-1}), v^n_{b-1}(\hat{n}_{b-1}|n_{b-2})\big) \in \mathcal{T}_{\varepsilon/2}(P_{U_0U_1U_2\tilde{Y}_1,\tilde{Y}_2V}).$$

The transmitter now sends the fresh data and the compression message $\hat{n}_{b-1}$ over the

channel: It thus looks for a pair $(t_{1,b}, t_{2,b}) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\big(u_{0,b}^n(\mathbf{m}_{c,b}, \hat{n}_{b-1}), u_{1,b}^n(m_{p,1,b}, t_{1,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1}), \\ u_{2,b}^n(m_{p,2,b}, t_{2,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1})\big) \in \mathcal{T}_{\varepsilon/64}^n(P_{U_0U_1U_2}).$$

If there is exactly one pair $(t_{1,b}, t_{2,b})$ that satisfies the above condition, the transmitter chooses this pair. If there are multiple such pairs, it chooses one of them uniformly at random. Otherwise it chooses a pair $(t_{1,b}, t_{2,b})$ uniformly at random over the entire set $\mathcal{T}_1 \times \mathcal{T}_2$. In block b the transmitter then sends the inputs $x_b^n = (x_{b,1}, \ldots, x_{b,n})$, where

$$x_{b,i} = f^{(n)}(u_{0,b,i}, u_{1,b,i}, u_{2,b,i}), \qquad i \in \{1, \ldots, n\}. \tag{6.56}$$

and $u_{0,b,i}$, $u_{1,b,i}$, $u_{2,b,i}$ denote the i-th symbols of the chosen Marton codewords $u_{0,b}^n(\mathbf{m}_{c,b}, \hat{n}_{b-1})$, $u_{1,b}^n(m_{p,1,b}, t_{1,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1})$, and $u_{2,b}^n(m_{p,2,b}, t_{2,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1})$.

3) Generation of Feedback Messages at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

After each block $b \in \{1, \ldots, B\}$, and after observing the outputs $y_{1,b}^n$, Receiver 1 looks for a pair of indices $(l_{1,b}, j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$(\tilde{y}_{1,b}^n(l_{1,b}, j_{1,b}), y_{1,b}^n) \in \mathcal{T}_{\varepsilon/16}^n(P_{Y_1\tilde{Y}_1}) \tag{6.57}$$

and sends the index $l_{1,b}$ over the feedback link. If there is more than one such pair $(l_{1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $l_{1,b}$ sent over the feedback link uniformly at random over $\mathcal{L}_1$.

In our scheme the receivers thus only send a feedback message at the end of each block.

4) Decoding at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

The receivers apply backward decoding, so they wait until the end of the transmission. Then, for each block $b \in \{1, \ldots, B+1\}$, starting with the last block $B+1$, Receiver 1 performs the following operations. From the previous decoding step in block $b+1$, it already knows the compression index n_b. Now, when observing $y_{1,b}^n$, Receiver 1

looks for a tuple $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{m}_{p,1,b}, \hat{t}_{1,b}, \hat{n}_{b-1}) \in \mathcal{M}_c \times \mathcal{M}_{p,1} \times \mathcal{T}_1 \times \mathcal{N}$ that satisfies

$$\big(u_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}), u_{1,b}^n(\hat{m}_{p,1,b}, \hat{t}_{1,b}|\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}), \\ v_b^n(n_b|\hat{n}_{b-1}), y_{1,b}^n, \tilde{y}_{1,b}^n(l_{1,b}, j_{1,b})\big) \in \mathcal{T}_\varepsilon^n(P_{U_0U_1VY_1\tilde{Y}_1}),$$

where recall that Receiver 1 knows the indices $l_{1,b}$ and $j_{1,b}$ because it has constructed them itself under *3)*.

After the decoding Block 1, Receiver 1 produces the product message $\hat{m}_1 = (\hat{m}_{1,1}, \ldots, \hat{m}_{1,B})$ as its guess, where $\hat{m}_{1,b} = (\hat{m}_{c,1,b}^{(1)}, \hat{m}_{p,1,b})$, for $b \in \{1, \ldots, B\}$, and $\hat{m}_{c,1,b}^{(1)}$ denotes the first component of $\hat{m}_{c,1,b}^{(1)}$.

5) Analysis: See Appendix 6.C.

6.7 Extension: Noisy Feedback

Our results also apply to the related setup where the two feedback links are noisy channels of capacities $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$ and where *the decoders can code over their feedback links*. The following three modifications to our coding schemes suffice to ensure that our achievable regions remain valid:

- We time-share two instances of our coding schemes: one scheme operates during the odd blocks of the BC and occupies the even blocks on the feedback links; the other scheme operates during the even blocks of the BC and occupies the odd blocks on the feedback links.
- Instead of sending after each block an uncoded feedback message over the feedback links, the receivers encode them using a capacity-achieving code for their feedback links and send these codewords during the next block.
- After each block, the transmitter first decodes the messages sent over the feedback links during this block, and then uses the decoded feedback-messages in the same way as it used them in the original scheme.

Let $\mathcal{E}_{\text{Fb},k,b}$, for $k = 1, 2$, denote the event that during Block b there is an error in the feedback communication from Receiver k to the transmitter, and let $\mathcal{E}$ denote the

event that $\hat{M}_1 \neq M_1$ or $\hat{M}_2 \neq M_2$. Then,

$$\begin{aligned}
&\Pr[\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2] \\
&\quad \leq \Pr\left[\mathcal{E} \cup \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},1,b}\right) \cup \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},2,b}\right)\right] \\
&\quad \leq \Pr\left[\mathcal{E}\Big| \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},1,b}\right)^c \cap \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},2,b}\right)^c\right] \\
&\qquad + \Pr\left[\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},1,b}\right] + \Pr\left[\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},2,b}\right] \\
&\quad \leq \Pr\left[\mathcal{E}\Big| \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},1,b}\right)^c \cap \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},2,b}\right)^c\right] \\
&\qquad + \sum_{b=1}^{B} \Pr[\mathcal{E}_{\mathrm{Fb},1,b}] + \Pr[\mathcal{E}_{\mathrm{Fb},2,b}]\,. \qquad (6.58)
\end{aligned}$$

Since we use capacity-achieving codes on the feedback links, the probabilities $\Pr[\mathcal{E}_{\mathrm{Fb},1,b}]$ and $\Pr[\mathcal{E}_{\mathrm{Fb},2,b}]$ vanish as the blocklength increases. When the feedback communications in all the blocks are error-free, then the probability of error in the setup with noisy feedback is no larger than that in the setup with noise-free feedback. Thus, under the corresponding rate constraints, also the probability $\Pr\left[\mathcal{E}\Big| \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},1,b}\right)^c \cap \left(\bigcup_{b=1}^{B} \mathcal{E}_{\mathrm{Fb},2,b}\right)^c\right]$ vanishes as the blocklength increases. Combining all these observations proves that the rate regions in Theorems 6.1–6.4 are achievable also in a setup with noisy feedback if the receivers can code over the feedback links.

6.A Appendix: Analysis of Scheme 1A (Theorem 6.1)

By the symmetry of our code construction, the probability of error does not depend on the realizations of $M_{c,k,b}$, $M_{p,k,b}$, $T_{k,b}$, $J_{k,b}$, $L_{k,b}$, for $k \in \{1,2\}$ and $b \in \{1,\ldots,B\}$. To simplify exposition we therefore assume that $M_{c,k,b} = M_{p,k,b} = T_{k,b} = J_{k,b} = L_{k,b} = 1$ for all $k \in \{1,2\}$ and $b \in \{1,\ldots,B\}$. Under this assumption, an error occurs if, and only if, for some $b \in \{1,\ldots,B\}$,

$$(\hat{M}_{p,1,b}, \hat{M}_{p,2,b}, \hat{M}^{(1)}_{c,1,b}, \hat{M}^{(2)}_{c,2,b}) \neq (1,1,1,1).$$

For each $b \in \{1, \ldots, B\}$, let $\mathcal{E}_b$ denote the event that in our coding scheme at least one of the following holds for $k \in \{1, 2\}$:

- $\hat{J}_{k,b-1} \neq 1$;
- $\hat{T}_{k,b-1} \neq 1$;
- $\hat{L}_{k,b-1} \neq 1$;
- $\hat{M}_{p,k,b-1} \neq 1$;
- $\hat{\mathbf{M}}_{c,b}^{(k)} \neq (1, 1)$;
- There is no pair $(t_{1,b}, t_{2,b}) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, t_{1,b}|\mathbf{1}_{[4]}), U_{2,b}^n(1, t_{2,b}|\mathbf{1}_{[4]})\big) \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2})$$

- $\big(U_{0,b-1}^n(\mathbf{1}_{[4]}), U_{1,b-1}^n(1, 1|\mathbf{1}_{[4]}), U_{2,b-1}^n(1, 1, |\mathbf{1}_{[4]}), Y_{1,b-1}^n, Y_{2,b-1}^n\big) \notin \mathcal{T}_{\varepsilon/12}^n(P_{U_0U_1U_2Y_1Y_2})$
- There is no pair $(l_{k,b}, j_{k,b}) \in \mathcal{L}_k \times \mathcal{J}_k$ that satisfies

$$\big(\tilde{Y}_{k,b}^n(l_{k,b}, j_{k,b}|\mathbf{1}_{[4]}), U_{0,b}^n(\mathbf{1}_{[4]}), Y_{k,b}^n\big) \in \mathcal{T}_{\varepsilon/4}^n(P_{\tilde{Y}_k U_0 Y_k}).$$

Then,

$$P_e^{(N)} \leq \Pr\left[\bigcup_{b=1}^{B+1} \mathcal{E}_b\right] \leq \sum_{b=2}^{B+1} \Pr\big[\mathcal{E}_b|\mathcal{E}_{b-1}^c\big] + \Pr[\mathcal{E}_1]. \tag{6.59}$$

In the following we analyze the probabilities of these events averaged over the random code construction. In particular, we shall identify conditions such that for each $b \in \{2, \ldots, B+1\}$, the probability $\Pr\big[\mathcal{E}_b|\mathcal{E}_{b-1}^c\big]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\mathcal{E}_1] \to 0$ as $n \to \infty$. Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(N)}$ tends to 0 as $N \to \infty$ when the mentioned conditions are satisfied.

Fix $b \in \{2, \ldots, B+1\}$ and $\varepsilon > 0$, and define the following events.

- Let $\mathcal{E}_{0,b}$ be the event that there is no pair $(t_{1,b}, t_{2,b}) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, t_{1,b}|\mathbf{1}_{[4]}), U_{2,b}^n(1, t_{2,b}|\mathbf{1}_{[4]})\big) \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2}).$$

By the Covering Lemma, $\Pr[\mathcal{E}_{0,b}]$ tends to 0 as $n \to \infty$ if

$$R_1' + R_2' \geq I(U_1; U_2|U_0) + \delta(\varepsilon), \tag{6.60}$$

where throughout this section $\delta(\varepsilon)$ stands for some function that tends to 0 as $\varepsilon \to 0$.

- Let $\mathcal{E}_{1,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1,1|\mathbf{1}_{[4]}), U_{2,b}^n(1,1,|\mathbf{1}_{[4]}), Y_{1,b}^n, Y_{2,b}^n\big) \notin \mathcal{T}_{\varepsilon/12}^n(P_{U_0U_1U_2Y_1Y_2}).$$

  Since the channel is memoryless, by the law of large numbers, $\Pr\Big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\Big]$ tends to 0 as $n \to \infty$.

- Let $\mathcal{E}_{2,1,b}$ be the event that there is no tuple $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{l}_{2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_2$ that is not equal to $(\mathbf{1}_{[2]}, 1)$ and that satisfies

$$\big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, 1, \hat{l}_{2,b-1}), Y_{1,b}^n\big) \in \mathcal{T}_{\varepsilon/8}^n(P_{U_0Y_1}).$$

  By the Packing Lemma, $\Pr\Big[\mathcal{E}_{2,1,b}|\mathcal{E}_{1,b}^c\Big]$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_2 + R_{c,1} + R_{c,2} \leq I(U_0; Y_1) + \delta(\varepsilon). \tag{6.61}$$

- Let $\mathcal{E}_{2,2,b}$ be the event that there is no tuple $(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{l}_{1,b-1}) \in \mathcal{M}_c \times \mathcal{L}_1$ with $(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{l}_{1,b-1})$ not equal to $(\mathbf{1}_{[2]}, 1)$ that satisfies

$$\big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{l}_{1,b-1}, 1), Y_{2,b}^n\big) \in \mathcal{T}_{\varepsilon/8}^n(P_{U_0Y_2}).$$

  By the Packing Lemma, $\Pr\Big[\mathcal{E}_{2,2,b}|\mathcal{E}_{1,b}^c\Big]$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_1 + R_{c,1} + R_{c,2} \leq I(U_0; Y_2) + \delta(\varepsilon). \tag{6.62}$$

- Let $\mathcal{E}_{3,1,b}$ be the event that

$$\big(U_{0,b-1}^n(\mathbf{1}_{[4]}),\, U_{1,b-1}^n(1,1|\mathbf{1}_{[4]}), \tilde{Y}_{2,b-1}^n(1,1), Y_{1,b-1}^n\big) \notin \mathcal{T}_{\varepsilon/2}^n(P_{U_0U_1\tilde{Y}_2Y_1}).$$

  By the Markov Lemma, $\Pr\big[\mathcal{E}_{3,1,b}|\mathcal{E}_{b-1}^c\big]$ tends to 0 as $n \to \infty$.

- Let $\mathcal{E}_{3,2,b}$ be the event that

$$\big(U^n_{0,b-1}(\mathbf{1}_{[4]}),\, U^n_{2,b-1}(1,1|\mathbf{1}_{[4]}), \tilde{Y}^n_{1,b-1}(1,1), Y^n_{2,b-1}\big) \notin \mathcal{T}^n_{\varepsilon/2}(P_{U_0U_2\tilde{Y}_1Y_2}).$$

  By the Markov Lemma, $\Pr\big[\mathcal{E}_{3,2,b}|\mathcal{E}^c_{b-1}\big]$ tends to 0 as $n \to \infty$.

- Let $\mathcal{E}_{4,1,b}$ be the event that there exists a tuple $(\hat{m}_{p,1,b-1}, \hat{t}_{1,b-1}, \hat{j}_{2,b-1}) \in \mathcal{M}_{p,1} \times \mathcal{T}_1 \times \mathcal{J}_2$ not equal to the all-one tuple and that satisfies

$$\begin{aligned}\big(U^n_{0,b-1}(\mathbf{1}_{[4]}), U^n_{1,b-1}(\hat{m}_{p,1,b-1}, \hat{t}_{1,b-1}|\mathbf{1}_{[4]}),\\ \tilde{Y}^n_{2,b-1}(1, \hat{j}_{2,b-1}|\mathbf{1}_{[4]}), Y^n_{1,b-1}\big) \in \mathcal{T}^n_{\varepsilon}(P_{U_0U_1\tilde{Y}_2Y_1}).\end{aligned}$$

  By the Packing Lemma, $\Pr\Big[\mathcal{E}_{4,1,b}|\mathcal{E}^c_{3,1,b}\Big]$ tends to zero as $n \to \infty$, if

$$\hat{R}_2 \le I(\tilde{Y}_2; U_1, Y_1|U_0) - \delta(\varepsilon) \tag{6.63}$$

$$R_{p,1} + R'_1 \le I(U_1; Y_1, \tilde{Y}_2|U_0) - \delta(\varepsilon) \tag{6.64}$$

$$R_{p,1} + R'_1 + \hat{R}_2 \le I(U_1; Y_1, \tilde{Y}_2|U_0) + I(\tilde{Y}_2; Y_1|U_0) - \delta(\varepsilon). \tag{6.65}$$

- Let $\mathcal{E}_{4,2,b}$ be the event that there exists a tuple $(\hat{m}_{p,2,b-1}, \hat{t}_{2,b-1}, \hat{j}_{1,b-1}) \in \mathcal{M}_{p,2} \times \mathcal{T}_2 \times \mathcal{J}_1$ not equal to the all-one tuple and that satisfies

$$\begin{aligned}\big(U^n_{0,b-1}(\mathbf{1}_{[4]}), U^n_{2,b-1}(\hat{m}_{p,2,b-1}, \hat{t}_{2,b-1}|\mathbf{1}_{[4]}),\\ \tilde{Y}^n_{1,b-1}(1, \hat{j}_{1,b-1}|\mathbf{1}_{[4]}), Y^n_{2,b-1}\big) \in \mathcal{T}^n_{\varepsilon}(P_{U_0U_2\tilde{Y}_1Y_2}).\end{aligned}$$

  By the Packing Lemma, $\Pr\Big[\mathcal{E}_{4,2,b}|\mathcal{E}^c_{3,2,b}\Big]$ tends to zero as $n \to \infty$, if

$$\hat{R}_1 \le I(\tilde{Y}_1; U_2, Y_2|U_0) - \delta(\varepsilon) \tag{6.66}$$

$$R_{p,2} + R'_2 \le I(U_2; Y_2, \tilde{Y}_1|U_0) - \delta(\varepsilon) \tag{6.67}$$

$$R_{p,2} + R'_2 + \hat{R}_1 \le I(U_2; Y_2, \tilde{Y}_1|U_0) + I(\tilde{Y}_1; Y_2|U_0) - \delta(\varepsilon). \tag{6.68}$$

- For $k \in \{1, 2\}$, let $\mathcal{E}_{5,k,b}$ be the event that there is no pair $(l_{k,b}, j_{k,b}) \in \mathcal{L}_k \times \mathcal{J}_k$ that satisfies

$$\big(\tilde{Y}^n_{k,b}(l_{k,b}, j_{k,b}|\mathbf{1}_{[4]}), U^n_{0,b}(\mathbf{1}_{[4]}), Y^n_{k,b}\big) \in \mathcal{T}^n_{\varepsilon/4}(P_{\tilde{Y}_kU_0Y_k}).$$

By the Covering Lemma, $\Pr\Big[\mathcal{E}_{5,k,b}|\mathcal{E}_{1,b}^c\Big]$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_k + \hat{R}_k \geq I(\tilde{Y}_k; Y_k|U_0) + \delta(\varepsilon). \tag{6.69}$$

Whenever the event $\mathcal{E}_{b-1}^c$ occurs but none of the events $\{\mathcal{E}_{0,b}, \mathcal{E}_{1,b}, \mathcal{E}_{2,1,b}, \mathcal{E}_{2,2,b}, \mathcal{E}_{3,1,b}, \mathcal{E}_{3,2,b}, \mathcal{E}_{4,1,b}, \mathcal{E}_{4,2,b}, \mathcal{E}_{5,1,b}, \mathcal{E}_{5,2,b}\}$ above, then $\mathcal{E}_b^c$. Therefore,

$$\begin{aligned}
&\Pr\big[\mathcal{E}_b|\mathcal{E}_{b-1}^c\big] \\
&\leq \Pr\Big[\mathcal{E}_{0,b} \cup \mathcal{E}_{1,b} \cup \bigcup_{k=1}^{2} \big(\mathcal{E}_{2,k,b} \cup \mathcal{E}_{3,k,b} \cup \mathcal{E}_{4,k,b} \cup \mathcal{E}_{5,k,b}\big)\Big|\mathcal{E}_{b-1}^c\Big] \\
&\leq \Pr\big[\mathcal{E}_{0,b}|\mathcal{E}_{b-1}^c\big] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c, \mathcal{E}_{b-1}^c\big] \\
&\qquad + \sum_{k=1}^{2} \Big(\Pr\big[\mathcal{E}_{2,k,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{b-1}^c\big] + \Pr\big[\mathcal{E}_{3,k,b}|\mathcal{E}_{b-1}^c\big] \\
&\qquad + \Pr\big[\mathcal{E}_{4,k,b}|\mathcal{E}_{3,k,b}^c, \mathcal{E}_{b-1}^c\big] + \Pr\big[\mathcal{E}_{5,k,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{b-1}^c\big] \Big) \\
&= \Pr[\mathcal{E}_{0,b}] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\big] \\
&\qquad + \sum_{k=1}^{2} \Big(\Pr\big[\mathcal{E}_{2,k,b}|\mathcal{E}_{1,b}^c\big] + \Pr\big[\mathcal{E}_{3,k,b}|\mathcal{E}_{b-1}^c\big] \\
&\qquad\qquad + \Pr\big[\mathcal{E}_{4,k,b}|\mathcal{E}_{3,k,b}^c\big] + \Pr\big[\mathcal{E}_{5,k,b}|\mathcal{E}_{1,b}^c\big] \Big).
\end{aligned}$$

The last equality holds because the channel is memoryless and the codebooks employed in blocks $b-1$ and b are drawn independently. As explained in the previous paragraphs, the remaining terms in the last three lines tend to 0 as $n \to \infty$, if constraints (6.60)–(6.69) are satisfied. Thus, by (6.59) and (6.70) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if constraints (6.60)–(6.69) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error can be made to tend

to 0 as $n \to \infty$ whenever

$$R'_1 + R'_2 > I(U_1; U_2|U_0) \tag{6.70a}$$
$$\tilde{R}_2 + R_{c,1} + R_{c,2} < I(U_0; Y_1) \tag{6.70b}$$
$$\tilde{R}_1 + R_{c,1} + R_{c,2} < I(U_0; Y_2) \tag{6.70c}$$
$$\hat{R}_1 < I(\tilde{Y}_1; U_2, Y_2|U_0) \tag{6.70d}$$
$$\hat{R}_2 < I(\tilde{Y}_2; U_1, Y_1|U_0) \tag{6.70e}$$
$$R_{p,1} + R'_1 < I(U_1; Y_1, \tilde{Y}_2|U_0) \tag{6.70f}$$
$$R_{p,2} + R'_2 > I(U_2; Y_2, \tilde{Y}_1|U_0) \tag{6.70g}$$
$$R_{p,1} + R'_1 + \hat{R}_2 < I(U_1; Y_1, \tilde{Y}_2|U_0) + I(\tilde{Y}_2; Y_1|U_0) \tag{6.70h}$$
$$R_{p,2} + R'_2 + \hat{R}_1 < I(U_2; Y_2, \tilde{Y}_1|U_0) + I(\tilde{Y}_1; Y_2|U_0) \tag{6.70i}$$
$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1|U_0) \tag{6.70j}$$
$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2|U_0). \tag{6.70k}$$

Moreover, the feedback-rate constraints (6.1) impose that:

$$\tilde{R}_1 \leq R_{\mathrm{Fb},1} \tag{6.70l}$$
$$\tilde{R}_2 \leq R_{\mathrm{Fb},2}. \tag{6.70m}$$

Applying the Fourier-Motzkin elimination algorithm to these constraints, we obtain the desired result in Theorem 6.1 with the additional constraint that

$$\begin{aligned} &I(U_1; Y_1, \tilde{Y}_2|U_0) + I(U_2; Y_2, \tilde{Y}_1|U_0) \\ &\qquad -\Delta_1 - \Delta_2 - I(U_1; U_2|U_0) \geq 0 \end{aligned} \tag{6.71}$$

Notice that we can ignore constraint (6.71) because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (6.71), the region defined by the constraints in Theorem 6.1 is contained in the time-sharing region.

6.B Appendix: Analysis of the Scheme 1B (Theorem 6.2)

An error occurs whenever

$$\hat{M}_{1,b} \neq M_{1,b} \text{ or } \hat{M}_{2,b} \neq M_{2,b}, \text{ for some } b \in \{1, \ldots, B\}.$$

For each $b \in \{1, \ldots, B+1\}$, let $\mathcal{E}_b$ denote the event that in our coding scheme at least one of the following holds for $k \in \{1, 2\}$:

$$\hat{J}_{k,b} \neq J_{k,b} \tag{6.72}$$

$$\hat{T}_{k,b} \neq T_{k,b} \tag{6.73}$$

$$\hat{L}_{k,b-1} \neq L_{k,b-1} \tag{6.74}$$

$$\hat{M}_{p,k,b} \neq M_{p,k,b} \tag{6.75}$$

$$\hat{\mathbf{M}}_{c,b}^{(k)} \neq \mathbf{M}_{c,b}^{(k)} \tag{6.76}$$

Then,

$$P_e^{(N)} \leq \Pr\left[\bigcup_{b=1}^{B+1} \mathcal{E}_b\right] \leq \sum_{b=1}^{B} \Pr\left[\mathcal{E}_b | \mathcal{E}_{b+1}^c\right] + \Pr[\mathcal{E}_{B+1}]\,. \tag{6.77}$$

In the following we analyze the probabilities of these events averaged over the random code construction. In particular, we shall identify conditions such that for each $b \in \{1, \ldots, B\}$, the probability $\Pr\left[\mathcal{E}_b | \mathcal{E}_{b+1}^c\right]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\mathcal{E}_{B+1}] \to 0$ as $n \to \infty$. Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(N)}$ tends to 0 as $N \to \infty$ when the mentioned conditions are satisfied.

Fix $b \in \{1, \ldots, B\}$ and $\varepsilon > 0$. By the symmetry of our code construction, the probability $\Pr\left[\mathcal{E}_b | \mathcal{E}_{b+1}^c\right]$ does not depend on the realization of $M_{c,k,b}$, $M_{p,k,b}$, $T_{k,b}$, $J_{k,b}$, $L_{k,b}$, $L_{k,b-1}$, for $k \in \{1, 2\}$. To simplify exposition we therefore assume that $M_{c,k,b} = M_{p,k,b} = T_{k,b} = J_{k,b} = L_{k,b} = L_{k,b-1} = 1$.

Define the following events.

- Let $\mathcal{E}_{0,b}$ be the event that there is no pair $(t_{1,b}, t_{2,b}) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\big(U_{0,b}(\mathbf{1}_{[4]}), U_{1,b}^n(1, t_{1,b}|\mathbf{1}_{[4]}), U_{2,b}^n(1, t_{2,b}|\mathbf{1}_{[4]})\big) \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2}).$$

  By the Covering Lemma, $\Pr[\mathcal{E}_{0,b}]$ tends to 0 as $n \to \infty$, if

$$R_1' + R_2' \geq I(U_1; U_2|U_0) + \delta(\varepsilon), \tag{6.78}$$

  where throughout this section $\delta(\varepsilon)$ stands for some function that tends to 0 as $\varepsilon \to 0$.
- Let $\varepsilon_{1,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, 1|\mathbf{1}_{[4]}, U_{2,b}^n(1, 1, |\mathbf{1}_{[4]}), Y_{1,b}^n, Y_{2,b}^n\big) \notin \mathcal{T}_{\varepsilon/8}^n(P_{U_0U_1Y_2Y_1Y_2}).$$

  Since the channel is memoryless, according to the law of large numbers, $\Pr\Big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\Big]$ tends to 0 as $n \to \infty$.
- For $k \in \{1, 2\}$, let $\mathcal{E}_{2,k,b}$ be the event that there is no pair $(l_{k,b}, j_{k,b}) \in \mathcal{L}_k \times \mathcal{J}_k$ that satisfies

$$\big(\tilde{Y}_{k,b}^n(l_{k,b}, j_{k,b}), Y_{k,b}^n\big) \in \mathcal{T}_{\varepsilon/4}^n(P_{\tilde{Y}_kY_k}).$$

  By the Covering Lemma, $\Pr\Big[\mathcal{E}_{2,k,b}|\mathcal{E}_{1,b}^c\Big]$ tends to 0 as $n \to \infty$ if

$$\tilde{R}_k + \hat{R}_k \geq I(\tilde{Y}_k; Y_k) + \delta(\varepsilon). \tag{6.79}$$

- Let $\mathcal{E}_{3,1,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, 1|\mathbf{1}_{[4]}), \tilde{Y}_{2,b}^n(1, 1), Y_{1,b}^n\big) \notin \mathcal{T}_{3\varepsilon/4}^n(P_{U_0U_1\tilde{Y}_2Y_1}).$$

  By the Markov Lemma, $\Pr\Big[\mathcal{E}_{3,1,b}|\mathcal{E}_{2,2,b}^c, \mathcal{E}_{1,b}^c\Big]$ tends to 0 as $n \to \infty$.
- Let $\mathcal{E}_{3,2,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{2,b}^n(1, 1|\mathbf{1}_{[4]}), \tilde{Y}_{1,b}^n(1, 1), Y_{2,b}^n\big) \notin \mathcal{T}_{3\varepsilon/4}^n(P_{U_0U_2\tilde{Y}_1Y_2}).$$

By the Markov Lemma, $\Pr\left[\mathcal{E}_{3,2,b}|\mathcal{E}_{2,1,b}^c,\mathcal{E}_{1,b}^c\right]$ tends to 0 as $n\to\infty$.

- Let $\mathcal{E}_{4,1,b}$ be the event that there exists a tuple $(\hat{j}_{2,b},\hat{\mathbf{m}}_{c,b}^{(1)},\hat{l}_{2,b-1},\hat{m}_{p,1,b},\hat{t}_{1,b})\in \mathcal{J}_2\times\mathcal{M}_c\times\mathcal{L}_2\times\mathcal{M}_{p,1}\times\mathcal{T}_1$ not equal to the all-one tuple $(1,\mathbf{1}_{[2]},1,1,1)$ and that satisfies

$$\Big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)},1,\hat{l}_{2,b-1}),U_{1,b}^n(\hat{m}_{p,1,b},\hat{t}_{1,b}|\hat{\mathbf{m}}_{c,b}^{(1)},1,\hat{l}_{2,b-1}),\\ \tilde{Y}_{2,b}^n(1,\hat{j}_{2,b}),Y_{1,b}^n\Big)\in\mathcal{T}_\varepsilon^n(P_{U_0U_1\tilde{Y}_2Y_1}).$$

By the Packing Lemma, we conclude that $\Pr\left[\mathcal{E}_{4,1,b}|\mathcal{E}_{3,1,b}^c\right]$ tends to zero as $n\to\infty$ if

$$\begin{aligned}
\hat{R}_2 &\le I(U_0,U_1,Y_1;\tilde{Y}_2|U_0)-\delta(\varepsilon)\\
R_{p,1}+R_1' &\le I(U_1;Y_1,\tilde{Y}_2|U_0)-\delta(\varepsilon)\\
R_1+R_{c,2}+\tilde{R}_2+R_1' &\le I(U_0,U_1;Y_1,\tilde{Y}_2)-\delta(\varepsilon)\\
R_1+R_{c,2}+\tilde{R}_2+R_1'+\hat{R}_2 &\le I(U_0,U_1;Y_1,\tilde{Y}_2)+I(Y_1;\tilde{Y}_2)-\delta(\varepsilon)\\
R_{p,1}+R_1'+\hat{R}_2 &\le I(U_1;Y_1,\tilde{Y}_2|U_0)+I(\tilde{Y}_2;Y_1,U_0)-\delta(\varepsilon).
\end{aligned}\tag{6.80}$$

- Let $\mathcal{E}_{4,2,b}$ be the event that there exists a tuple $(\hat{j}_{1,b},\hat{\mathbf{m}}_{c,b}^{(2)},\hat{l}_{1,b-1},\hat{m}_{p,2,b},\hat{t}_{2,b})\in \mathcal{J}_1\times\mathcal{M}_c\times\mathcal{L}_1\times\mathcal{M}_{p,2}\times\mathcal{T}_2$ not equal to the all-one tuple and that satisfies

$$\big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(2)},\hat{l}_{1,b-1},1),U_{1,b}^n(\hat{m}_{p,2,b},\hat{t}_{2,b}|\hat{\mathbf{m}}_{c,b}^{(2)},\hat{l}_{1,b-1},1),\\ \tilde{Y}_{1,b}^n(1,\hat{j}_{1,b}),Y_{2,b}^n\big)\in\mathcal{T}_\varepsilon^n(P_{U_0U_2\tilde{Y}_1Y_2}).$$

By the Packing Lemma, we conclude that $\Pr\left[\mathcal{E}_{4,2,b}|\mathcal{E}_{3,2,b}^c\right]$ tends to zero as $n\to\infty$ if

$$\begin{aligned}
\hat{R}_1 &\le I(U_0,U_2,Y_2;\tilde{Y}_1|U_0)-\delta(\varepsilon)\\
R_{p,2}+R_2' &\le I(U_2;Y_2,\tilde{Y}_1|U_0)-\delta(\varepsilon)\\
R_2+R_{c,1}+\tilde{R}_1+R_2' &\le I(U_0,U_2;Y_2,\tilde{Y}_1)-\delta(\varepsilon)\\
R_2+R_{c,1}+\tilde{R}_1+R_2'+\hat{R}_1 &\le I(U_0,U_2;Y_2,\tilde{Y}_1)+I(Y_2;\tilde{Y}_1)-\delta(\varepsilon)\\
R_{p,2}+R_2'+\hat{R}_1 &\le I(U_2;Y_2,\tilde{Y}_1|U_0)+I(\tilde{Y}_1;Y_2,U_0)-\delta(\varepsilon).
\end{aligned}\tag{6.81}$$

Whenever the event $\mathcal{E}_{b+1}^c$ occurs but none of the events above, then $\mathcal{E}_b^c$. Therefore,

$$\begin{aligned}
&\Pr\big[\mathcal{E}_b|\mathcal{E}_{b+1}^c\big] \\
&\leq \Pr\left[\mathcal{E}_{0,b} \cup \mathcal{E}_{1,b} \cup \bigcup_{k=1}^{2} \Big(\mathcal{E}_{2,k,b} \cup \mathcal{E}_{3,k,b} \cup \mathcal{E}_{4,k,b}\Big) \Big| \mathcal{E}_{b+1}^c\right] \\
&\leq \Pr\big[\mathcal{E}_{0,b}|\mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c, \mathcal{E}_{b+1}^c\big] \\
&\qquad + \Pr\big[\mathcal{E}_{3,1,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,2,b}^c, \mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{3,2,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c, \mathcal{E}_{b+1}^c\big] \\
&\qquad + \sum_{k=1}^{2} \Big(\Pr\big[\mathcal{E}_{2,k,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{4,k,b}|\mathcal{E}_{3,k,b}^c, \mathcal{E}_{b+1}^c\big] \Big) \\
&= \Pr[\mathcal{E}_{0,b}] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\big] \\
&\qquad + \Pr\big[\mathcal{E}_{3,1,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,2,b}^c\big] + \Pr\big[\mathcal{E}_{3,2,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c\big] \\
&\qquad + \sum_{k=1}^{2} \Big(\Pr\big[\mathcal{E}_{2,k,b}|\mathcal{E}_{1,b}^c\big] + \Pr\big[\mathcal{E}_{4,k,b}|\mathcal{E}_{3,k,b}^c\big] \Big), \qquad (6.82)
\end{aligned}$$

where the last equality follows because the channel is memoryless and the codebooks for blocks b and $b+1$ have been generated independently. As explained in the previous paragraphs, each of the terms in the last three lines tends to 0 as $n \to \infty$, if constraints (6.78)–(6.81) are satisfied. Thus, by (6.77) and (6.82) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if constraints (6.78)–(6.81) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error

can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{6.83a}$$
$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1) \tag{6.83b}$$
$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2) \tag{6.83c}$$
$$\hat{R}_1 < I(U_0, U_2, Y_2; \tilde{Y}_1|U_0) \tag{6.83d}$$
$$\hat{R}_2 < I(U_0, U_1, Y_1; \tilde{Y}_2|U_0) \tag{6.83e}$$
$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_2|U_0) \tag{6.83f}$$
$$R_{p,2} + R_2' < I(U_2; Y_2, \tilde{Y}_1|U_0) \tag{6.83g}$$
$$R_1 + R_{c,2} + \tilde{R}_2 + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_2) \tag{6.83h}$$
$$R_2 + R_{c,1} + \tilde{R}_1 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_1) \tag{6.83i}$$
$$R_1 + R_{c,2} + \tilde{R}_2 + R_1' + \hat{R}_2 < I(U_0, U_1; Y_1, \tilde{Y}_2) + I(Y_1; \tilde{Y}_2) \tag{6.83j}$$
$$R_2 + R_{c,1} + \tilde{R}_1 + R_2' + \hat{R}_1 < I(U_0, U_2; Y_2, \tilde{Y}_1) + I(Y_2; \tilde{Y}_1) \tag{6.83k}$$
$$R_{p,1} + R_1' + \hat{R}_2 < I(U_1; Y_1, \tilde{Y}_2|U_0) + I(\tilde{Y}_2; Y_1, U_0) \tag{6.83l}$$
$$R_{p,2} + R_2' + \hat{R}_1 < I(U_2; Y_2, \tilde{Y}_1|U_0) + I(\tilde{Y}_1; Y_2, U_0). \tag{6.83m}$$

Moreover, the feedback-rate constraints (6.1) impose that:

$$\tilde{R}_1 \leq R_{\mathrm{Fb},1} \tag{6.83n}$$
$$\tilde{R}_2 \leq R_{\mathrm{Fb},2}. \tag{6.83o}$$

Applying the Fourier-Motzkin elimination algorithm to these constraints, we obtain the desired result in Theorem 6.2 with the additional constraint that

$$I(U_1; Y_1, \tilde{Y}_2|U_0) + I(U_2; Y_2, \tilde{Y}_1|U_0)$$
$$-\Delta_1 - \Delta_2 - I(U_1; U_2|U_0) \geq 0 \tag{6.84a}$$
$$I(U_1; Y_1, \tilde{Y}_2|U_0) - \Delta_2 \geq 0 \tag{6.84b}$$
$$I(U_2; Y_2, \tilde{Y}_1|U_0) - \Delta_1 \geq 0. \tag{6.84c}$$

We can ignore constraint (6.84a) because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that

violates (6.84a), the region defined by the constraints in Theorem 6.2 is contained in the time-sharing region. Constraint (6.84b) can also be ignored because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (6.84b), the region defined by the constraints in Theorem 6.2 is contained in the region in Theorem 6.2 for the choice $\tilde{Y}_2 = \text{const.}$, for which (6.84b) is always satisfied. Constraint (6.84c) can be ignored by analogous arguments.

6.C Appendix: Analysis of Scheme 2 (Theorem 6.4)

An error occurs whenever

$$\hat{M}_{1,b} \neq M_{1,b} \text{ or } \hat{M}_{2,b} \neq M_{2,b}, \text{ for some } b \in \{1, \ldots, B\}.$$

For each $b \in \{1, \ldots, B+1\}$, let $\mathcal{E}_b$ denote the event that in our coding scheme at least one of the following holds for $k \in \{1, 2\}$:

$$\hat{J}_{k,b} \neq J_{k,b} \tag{6.85}$$

$$\hat{T}_{k,b} \neq T_{k,b} \tag{6.86}$$

$$\hat{L}_{k,b} \neq L_{k,b} \tag{6.87}$$

$$\hat{M}_{p,k,b} \neq M_{p,k,b} \tag{6.88}$$

$$\hat{\mathbf{M}}_{c,b}^{(k)} \neq \mathbf{M}_{c,b}^{(k)} \tag{6.89}$$

or when

$$\hat{N}_{b-1} \neq N_{b-1}. \tag{6.90}$$

Then,

$$P_e^{(n)} \leq \Pr\left[\bigcup_{b=1}^{B+1} \mathcal{E}_b\right] \leq \sum_{b=1}^{B} \Pr\left[\mathcal{E}_b | \mathcal{E}_{b+1}^c\right] + \Pr[\mathcal{E}_{B+1}]\,. \tag{6.91}$$

In the following we analyze the probabilities of these events averaged over the random code construction. In particular, we shall identify conditions such that for each $b \in \{1, \ldots, B\}$, the probability $\Pr\left[\mathcal{E}_b | \mathcal{E}_{b+1}^c\right]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\mathcal{E}_{B+1}] \to 0$ as $n \to \infty$. Using standard arguments one can then conclude that there must exist a deterministic code

for which the probability of error $P_e^{(N)}$ tends to 0 as $N \to \infty$ when the mentioned conditions are satisfied.

Fix $b \in \{1, \ldots, B\}$ and $\varepsilon > 0$. By the symmetry of our code construction, the probability $\Pr\left[\mathcal{E}_b | \mathcal{E}_{b+1}^c\right]$ does not depend on the realizations of N_{b-1}, N_b, or $M_{c,k,b}$, $M_{p,k,b}$, $T_{k,b}$, $J_{k,b}$, $L_{k,b}$, for $k \in \{1, 2\}$. To simplify exposition we therefore assume that for $k \in \{1, 2\}$, $M_{c,k,b} = M_{p,k,b} = T_{k,b} = J_{k,b} = L_{k,b} = 1$, and $N_b = N_{b-1} = 1$.

Define the following events.

- Let $\mathcal{E}_{0,b}$ be the event that there is no pair $(t_{1,b}, t_{2,b}) \in \mathcal{T}_1 \times \mathcal{T}_2$ that satisfies

$$\left(U_{0,b}(\mathbf{1}_{[3]}), U_{1,b}^n(1, t_{1,b} | \mathbf{1}_{[2]}), U_{2,b}^n(1, t_{2,b} | \mathbf{1}_{[3]})\right) \in \mathcal{T}_{\varepsilon/64}^n(P_{U_0 U_1 U_2}).$$

  By the Covering Lemma, $\Pr[\mathcal{E}_{0,b}]$ tends to 0 as $n \to \infty$ if

$$R_1' + R_2' \geq I(U_1; U_2 | U_0) + \delta(\varepsilon), \tag{6.92}$$

  where throughout this section $\delta(\varepsilon)$ stands for some function that tends to 0 as $\varepsilon \to 0$.

- Let $\mathcal{E}_{1,b}$ be the event that

$$\left(U_{0,b}^n(\mathbf{1}_{[3]}), U_{1,b}^n(1, 1 | \mathbf{1}_{[3]}), U_{2,b}^n(1, 1, | \mathbf{1}_{[3]}), Y_{1,b}^n, Y_{2,b}^n\right) \notin \mathcal{T}_{\varepsilon/32}^n(P_{U_0 U_1 U_2 Y_1 Y_2}).$$

  Since the channel is memoryless, according to the law of large numbers, $\Pr\left[\mathcal{E}_{1,b} | \mathcal{E}_{0,b}^c\right]$ tends to 0 as $n \to \infty$.

- For $k \in \{1, 2\}$, let $\mathcal{E}_{2,k,b}$ be the event that there is no pair $(l_{k,b}, j_{k,b}) \in \mathcal{L}_k \times \mathcal{J}_k$ that satisfies

$$\left(\tilde{Y}_{k,b}^n(l_{k,b}, j_{k,b}), Y_{k,b}^n\right) \in \mathcal{T}_{\varepsilon/16}^n(P_{\tilde{Y}_k Y_k}).$$

  By the Covering Lemma, $\Pr\left[\mathcal{E}_{2,k,b} | \mathcal{E}_{1,b}^c\right]$ tends to 0 as $n \to \infty$ if

$$\tilde{R}_k + \hat{R}_k \geq I(\tilde{Y}_k; Y_k) + \delta(\epsilon). \tag{6.93}$$

- Let $\mathcal{E}_{3,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[3]}), U_{1,b}^n(1,1|\mathbf{1}_{[3]}), U_{2,b}^n(1,1|\mathbf{1}_{[3]}), \\ \tilde{Y}_{1,b}^n(1,1), \tilde{Y}_{2,b}^n(1,1), Y_{1,b}^n, Y_{2,b}^n\big) \notin \mathcal{T}_{\varepsilon/6}^n(P_{U_0U_1U_2\tilde{Y}_1\tilde{Y}_2Y_1Y_2}).$$

  By the Markov Lemma, $\Pr\Big[\mathcal{E}_{3,b}|\mathcal{E}_{2,1,b}^c, \mathcal{E}_{2,2,b}^c, \mathcal{E}_{1,b}^c\Big]$ tends to 0 as $n \to \infty$.
- Let $\mathcal{E}_{4,b}$ be the event that there is a pair of indices $\hat{j}_{1,b} \in \mathcal{J}_1$ and $\hat{j}_{2,b} \in \mathcal{J}_2$ not equal to the all-one pair $(1,1)$ and that satisfies

$$\big(U_{0,b}^n(\mathbf{1}_{[3]}), U_{1,b}^n(1,1|\mathbf{1}_{[3]}), U_{2,b}^n(1,1|\mathbf{1}_{[3]}), \\ \tilde{Y}_{1,b}^n(1,\hat{j}_{1,b}), \tilde{Y}_{2,b}^n(1,\hat{j}_{2,b})\big) \in \mathcal{T}_{\varepsilon/4}^n(P_{U_0U_1U_2\tilde{Y}_1\tilde{Y}_2}).$$

  By the Packing Lemma, $\Pr\Big[\mathcal{E}_{4,b}|\mathcal{E}_{3,b}^c\Big]$ tends to 0 as $n \to \infty$, if

$$\hat{R}_1 \le I(U_0, U_1, U_2, \tilde{Y}_2; \tilde{Y}_1) - \delta(\varepsilon) \tag{6.94}$$

$$\hat{R}_2 \le I(U_0, U_1, U_2, \tilde{Y}_1; \tilde{Y}_2) - \delta(\varepsilon) \tag{6.95}$$

$$\hat{R}_1 + \hat{R}_2 \le I(U_0, U_1, U_2; \tilde{Y}_1, \tilde{Y}_2) + I(\tilde{Y}_1; \tilde{Y}_2) - \delta(\varepsilon). \tag{6.96}$$

- Let $\mathcal{E}_{5,b}$ be the event that there is no index $n_b \in \mathcal{N}$ that satisfies

$$\big(U_{0,b}^n(\mathbf{1}_{[3]}), U_{1,b}^n(1,1|\mathbf{1}_{[3]}), U_{2,b}^n(1,1|\mathbf{1}_{[3]}), \\ \tilde{Y}_{1,b}^n(1,1), \tilde{Y}_{2,b}^n(1,1), V_b^n(n_b|1)\big) \in \mathcal{T}_{\varepsilon/2}^n(P_{U_0U_1U_2\tilde{Y}_1\tilde{Y}_2V}).$$

  By the Covering Lemma, $\Pr\Big[\mathcal{E}_{5,b}|\mathcal{E}_{3,b}^c\Big]$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_v \ge I(U_0, U_1, U_2, \tilde{Y}_1, \tilde{Y}_2; V) + \delta(\varepsilon). \tag{6.97}$$

- Let $\mathcal{E}_{6,1,b}$ be the event that

$$\Big(U_{0,b}^n(\mathbf{1}_{[3]}, 1), U_{1,b}^n(1,1|\mathbf{1}_{[3]}, 1), V_b^n(1|1), Y_{1,b}^n, \tilde{Y}_{1,b}^n(1,1)\Big) \in \mathcal{T}_\varepsilon^n(P_{U_0U_1VY_1\tilde{Y}_1}).$$

  By the Markov Lemma $\Pr\Big[\mathcal{E}_{6,1,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{5,b}^c\Big]$ tends to zero as $n \to \infty$.

- Let $\mathcal{E}_{6,2,b}$ be the event that

$$\Big(U_{0,b}^n(\mathbf{1}_{[3]},1), U_{2,b}^n(1,1|\mathbf{1}_{[3]},1), V_b^n(1|1), Y_{2,b}^n, \tilde{Y}_{2,b}^n(1,1)\Big) \in \mathcal{T}_\varepsilon^n(P_{U_0U_2VY_2\tilde{Y}_2}).$$

  By the Markov Lemma $\Pr\Big[\mathcal{E}_{6,2,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{5,b}^c\Big]$ tends to zero as $n \to \infty$.

- Let $\mathcal{E}_{7,1,b}$ be the event that there is a tuple $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}, \hat{m}_{p,1,b}, \hat{t}_{1,b}) \in \mathcal{M}_c \times \mathcal{N} \times \mathcal{M}_{p,1} \times \mathcal{T}_1$ that is not equal to the all-one tuple $(\mathbf{1}_{[3]}, 1, 1, 1)$ and that satisfies

$$\Big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}), U_{1,b}^n(\hat{m}_{p,1,b}, \hat{t}_{1,b}|\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}), V_b^n(1|\hat{n}_{b-1}), Y_{1,b}^n, \tilde{Y}_{1,b}^n(1,1)\Big) \in \mathcal{T}_\varepsilon^n(P_{U_0U_1VY_1\tilde{Y}_1}).$$

  By the Packing Lemma, we conclude that $\Pr\Big[\mathcal{E}_{7,1,b}|\mathcal{E}_{6,1,b}^c\Big]$ tends to zero as $n \to \infty$ if

$$R_1 + R_{c,2} + R_1' \le I(U_0, U_1; Y_1, \tilde{Y}_1, V) - \delta(\varepsilon) \tag{6.98}$$

$$R_1 + R_{c,2} + \tilde{R}_v + R_1' \le I(U_0, U_1; Y_1, \tilde{Y}_1, V) + I(V; \tilde{Y}_1, Y_1) - \delta(\varepsilon) \tag{6.99}$$

$$R_{p,1} + R_1' \le I(U_1; Y_1, \tilde{Y}_1, V|U_0) - \delta(\varepsilon). \tag{6.100}$$

- Let $\mathcal{E}_{7,2,b}$ be the event that there is a tuple $(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{n}_{b-1}, \hat{m}_{p,2,b}, \hat{t}_{2,b}) \in \mathcal{M}_c \times \mathcal{N} \times \mathcal{M}_{p,2} \times \mathcal{T}_2$ that is not equal to the all-one tuple $(\mathbf{1}_{[3]}, 1, 1, 1)$ and that satisfies

$$\Big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{n}_{b-1}), U_{2,b}^n(\hat{m}_{p,2,b}, \hat{t}_{2,b}|\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{n}_{b-1}), V_b^n(1|\hat{n}_{b-1}), Y_{2,b}^n, \tilde{Y}_{2,b}^n(1,1)\Big) \in \mathcal{T}_\varepsilon^n(P_{U_0U_2VY_2\tilde{Y}_2}).$$

  By the Markov Lemma and the Packing Lemma, we conclude that $\Pr\Big[\mathcal{E}_{7,2,b}|\mathcal{E}_{6,2,b}^c\Big]$ tends to zero as $n \to \infty$, if

$$R_2 + R_{c,1} + R_2' \le I(U_0, U_2; Y_2, \tilde{Y}_2, V) - \delta(\varepsilon) \tag{6.101}$$

$$R_2 + R_{c,1} + \tilde{R}_v + R_2' \le I(U_0, U_2; Y_2, \tilde{Y}_2, V) + I(V; \tilde{Y}_2, Y_2) - \delta(\varepsilon) \tag{6.102}$$

$$R_{p,2} + R_2' \le I(U_2; Y_2, \tilde{Y}_2, V|U_0) - \delta(\varepsilon). \tag{6.103}$$

Whenever the event $\mathcal{E}_{b+1}^c$ occurs but none of the events above, then $\mathcal{E}_b^c$. Therefore,

$$
\begin{aligned}
&\Pr\big[\mathcal{E}_b|\mathcal{E}_{b+1}^c\big] \\
&\le \Pr\Big[\mathcal{E}_{0,b} \cup \mathcal{E}_{1,b} \cup \mathcal{E}_{2,1,b} \cup \mathcal{E}_{2,2,b} \cup \mathcal{E}_{3,b} \\
&\qquad\qquad \cup \mathcal{E}_{4,b} \cup \mathcal{E}_{5,b} \cup \mathcal{E}_{6,1,b} \cup \mathcal{E}_{6,2,b} \big| \mathcal{E}_{b+1}^c\Big] \\
&\le \Pr\big[\mathcal{E}_{0,b}\big|\mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c, \mathcal{E}_{b+1}^c\big] \\
&\quad + \sum_{k=1}^{2} \Pr\big[\mathcal{E}_{2,k,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{b+1}^c\big] \\
&\quad + \Pr\big[\mathcal{E}_{3,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c, \mathcal{E}_{2,2,b}^c, \mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{4,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{b+1}^c\big] \\
&\quad + \Pr\big[\mathcal{E}_{5,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{b+1}^c\big] + \sum_{k=1}^{2} \Pr\big[\mathcal{E}_{6,k,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{b+1}^c\big] \\
&= \Pr[\mathcal{E}_{0,b}] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\big] + \sum_{k=1}^{2} \Pr\big[\mathcal{E}_{2,k,b}|\mathcal{E}_{1,b}^c\big] \\
&\quad + \Pr\big[\mathcal{E}_{3,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c, \mathcal{E}_{2,2,b}^c\big] + \Pr\big[\mathcal{E}_{4,b}|\mathcal{E}_{3,b}^c\big] + \Pr\big[\mathcal{E}_{5,b}|\mathcal{E}_{3,b}^c\big] \\
&\quad + \sum_{k=1}^{2} \Pr\big[\mathcal{E}_{6,k,b}|\mathcal{E}_{3,b}^c\big]\,, \qquad (6.104)
\end{aligned}
$$

where the last equality follows because the channel is memoryless and the codebooks in blocks b and $b+1$ have been chosen independently. As explained in the previous paragraphs, each of the terms in the last five lines tends to 0 as $n \to \infty$, if constraints (6.92)–(6.103) are satisfied. Thus, by (6.91) and (6.104) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if constraints (6.92)–(6.103) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error

can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{6.105a}$$
$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1) \tag{6.105b}$$
$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2) \tag{6.105c}$$
$$\hat{R}_1 < I(U_0, U_1, U_2, \tilde{Y}_2; \tilde{Y}_1) \tag{6.105d}$$
$$\hat{R}_2 < I(U_0, U_1, U_2, \tilde{Y}_1; \tilde{Y}_2) \tag{6.105e}$$
$$\hat{R}_1 + \hat{R}_2 < I(U_0, U_1, U_2; \tilde{Y}_1, \tilde{Y}_2) + I(\tilde{Y}_1; \tilde{Y}_2) \tag{6.105f}$$
$$\tilde{R}_v > I(U_0, U_1, U_2, \tilde{Y}_1, \tilde{Y}_2; V) \tag{6.105g}$$
$$R_1 + R_{c,2} + \tilde{R}_v + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V) + I(V; \tilde{Y}_1, Y_1) \tag{6.105h}$$
$$R_1 + R_{c,2} + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V) \tag{6.105i}$$
$$R_{c,1} + R_2 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V) \tag{6.105j}$$
$$R_{c,1} + R_2 + \tilde{R}_v + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V) + I(V; \tilde{Y}_2, Y_2) \tag{6.105k}$$
$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_1, V|U_0) \tag{6.105l}$$
$$R_{p,2} + R_2' < I(U_2; Y_2, \tilde{Y}_2, V|U_0). \tag{6.105m}$$

Moreover, the feedback-rate constraints (6.1) impose that:

$$\tilde{R}_1 \leq R_{\mathrm{Fb},1} \tag{6.105n}$$
$$\tilde{R}_2 \leq R_{\mathrm{Fb},2}. \tag{6.105o}$$

Eliminating the auxiliaries $\tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2, \tilde{R}_v$ from the above (using the Fourier-Motzkin

algorithm), we obtain:

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{6.106a}$$

$$R_1 + R_{c,2} + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V) - I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1) \tag{6.106b}$$

$$R_{c,1} + R_2 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V) - I(V; U_0, U_1, U_2, \tilde{Y}_1|\tilde{Y}_2, Y_2) \tag{6.106c}$$

$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_1, V|U_0) \tag{6.106d}$$

$$R_{p,2} + R_2' < (U_2; Y_2, \tilde{Y}_2, V|U_0) \tag{6.106e}$$

where the feedback-rate constraints have to satisfy

$$I(Y_1; \tilde{Y}_1|U_0, U_1, U_2, \tilde{Y}_2) \leq R_{\text{Fb},1} \tag{6.107a}$$

$$I(Y_2; \tilde{Y}_2|U_0, U_1, U_2, \tilde{Y}_1) \leq R_{\text{Fb},2} \tag{6.107b}$$

$$I(Y_1, Y_2; \tilde{Y}_1, \tilde{Y}_2|U_0, U_1, U_2) \leq R_{\text{Fb},1} + R_{\text{Fb},2}. \tag{6.107c}$$

Applying again the Fourier-Motzkin elimination algorithm to constraints (6.106) and keeping constraints (6.107), we obtain the desired result in Theorem 6.4 with the additional constraint that

$$I(U_1; U_2|U_0) \leq I(U_1; Y_1, \tilde{Y}_1, V|U_0) + (U_2; Y_2, \tilde{Y}_2, V|U_0). \tag{6.108}$$

Finally, this last constraint can be ignored because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (6.108), the region defined by the constraints in Theorem 6.4 is contained in the time-sharing region.

6.D Appendix: Proof of Theorem 6.5

Let $R_{\text{Fb},1} > 0$. Fix a tuple $(U_0^{(\text{M})}, U_1^{(\text{M})}, U_2^{(\text{M})}, X^{(\text{M})})$ and rate pairs $(R_1^{(\text{M})}, R_2^{(\text{M})})$ and $(R_1^{(\text{Enh})}, R_2^{(\text{Enh})}) \in \mathcal{C}_{\text{Enh}}^{(1)}$ as stated in the theorem. Then, by the assumptions in the

theorem,

$$R_1^{(\mathrm{M})} \le I(U_0^{(\mathrm{M})}, U_1^{(\mathrm{M})}; Y_1^{(\mathrm{M})}) \tag{6.109a}$$

$$R_2^{(\mathrm{M})} < I(U_0^{(\mathrm{M})}, U_2^{(\mathrm{M})}; Y_2^{(\mathrm{M})}) \tag{6.109b}$$

$$\begin{aligned} R_1^{(\mathrm{M})} + R_2^{(\mathrm{M})} \le\ & I(U_0^{(\mathrm{M})}, U_1^{(\mathrm{M})}; Y_1^{(\mathrm{M})}) + I(U_2^{(\mathrm{M})}; Y_2^{(\mathrm{M})} | U_0^{(\mathrm{M})}) \\ & - I(U_1^{(\mathrm{M})}; U_2^{(\mathrm{M})} | U_0^{(\mathrm{M})}), \end{aligned} \tag{6.109c}$$

where $Y_1^{(\mathrm{M})}$ and $Y_2^{(\mathrm{M})}$ denote the outputs of the considered DMBC corresponding to input $X^{(\mathrm{M})}$. (Notice the strict inequality of the second constraint.)

By the definition of $\mathcal{C}_{\mathrm{Enh}}^{(1)}$ we can identify random variables $U_0^{(\mathrm{Enh})}$ and $X^{(\mathrm{Enh})}$ such that

$$R_1^{(\mathrm{Enh})} \le I(U_0^{(\mathrm{Enh})}; Y_1^{(\mathrm{Enh})}) \tag{6.110a}$$

$$R_2^{(\mathrm{Enh})} \le I(X^{(\mathrm{Enh})}; Y_1^{(\mathrm{Enh})}, Y_2^{(\mathrm{Enh})} | U_0^{(\mathrm{Enh})}), \tag{6.110b}$$

where $Y_1^{(\mathrm{Enh})}$ and $Y_2^{(\mathrm{Enh})}$ denote the outputs of the considered DMBC corresponding to input $X^{(\mathrm{Enh})}$.

Define further $U_1^{(\mathrm{Enh})}$ =const., $U_2^{(\mathrm{Enh})} = X^{(\mathrm{Enh})}$, $\tilde{Y}_1^{(\mathrm{Enh})} = Y_1^{(\mathrm{Enh})}$, $\tilde{Y}_1^{M}$ =const, and a binary random variable Q independent of all previously defined random variables and of pmf

$$P_Q(q) = \begin{cases} \gamma, & q = \mathrm{Enh} \\ 1-\gamma, & q = \mathrm{M}. \end{cases} \tag{6.111}$$

We show that when γ is sufficiently small, then the random variables

$$\begin{aligned} U_0 := U_0^{(Q)},\ U_1 := U_1^{(Q)},\ U_2 := U_2^{(Q)} \\ X := X^{(Q)},\ \text{and } \tilde{Y}_1 := \tilde{Y}_1^{(Q)} \end{aligned} \tag{6.112}$$

satisfy the feedback rate constraints (6.24) and the rate pair (R_1', R_2'),

$$R_1' := (1-\gamma) R_1^{(\mathrm{M})} + \gamma R_1^{(\mathrm{Enh})} \tag{6.113a}$$

$$R_2' := (1-\gamma) R_2^{(\mathrm{M})} + \gamma R_2^{(\mathrm{Enh})}, \tag{6.113b}$$

satisfies the constraints in (6.23) for the choice in (6.112). The two imply that the rate pair (R_1', R_2') lies in $\mathcal{R}^{(1)}_{\text{relay,hb}}$ and concludes our proof.

Notice that the pmf of the tuple $U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1$ has the desired form

$$P_Q P_{U_0U_1U_2|Q} P_{X|U_0U_1U_2Q} P_{Y_1Y_2|X} P_{\tilde{Y}_1|Y_1Q}. \tag{6.114}$$

where $P_{Y_1Y_2|X}$ denotes the channel law.

For the described choice of random variables (6.112), the feedback-rate constraint (6.24) specializes to

$$\gamma H(Y_1^{(\text{Enh})}|Y_2^{(\text{Enh})}, X^{(\text{Enh})}) \leq R_{\text{Fb},1}, \tag{6.115}$$

which is satisfied for all sufficiently small $\gamma \in (0,1)$. Moreover, for this choice the constraints in (6.23) specialize to

$$\begin{aligned}
R_1 &\leq (1-\gamma) I(U_0^{(\text{M})}, U_1^{(\text{M})}; Y_1^{(\text{M})}) \\
&\quad + \gamma I(U_0^{(\text{Enh})}; Y_1^{(\text{Enh})}) && \text{(6.116a)} \\
R_2 &\leq (1-\gamma) I(U_0^{(\text{M})}, U_2^{(\text{M})}; Y_2^{(\text{M})}) \\
&\quad + \gamma \big(I(X^{(\text{Enh})}; Y_1^{(\text{Enh})}, Y_2^{(\text{Enh})}) - H(Y_1^{(\text{Enh})}|Y_2^{(\text{Enh})}) \big) && \text{(6.116b)} \\
R_1 + R_2 &\leq (1-\gamma) \big(I(U_0^{(\text{M})}, U_1^{(\text{M})}; Y_1^{(\text{M})}) + I(U_2^{(\text{M})}; Y_2^{(\text{M})}|U_0^{(\text{M})}) \\
&\qquad\qquad - I(U_1^{(\text{M})}; U_2^{(\text{M})}|U_0^{(\text{M})}) \big) \\
&\quad + \gamma \big(I(U_0^{(\text{Enh})}; Y_1^{(\text{Enh})}) + I(X^{(\text{Enh})}; Y_1^{(\text{Enh})}, Y_2^{(\text{Enh})}|U_0^{(\text{Enh})}) \big) && \text{(6.116c)} \\
R_1 + R_2 &\leq (1-\gamma) \big(I(U_1^{(\text{M})}; Y_1^{(\text{M})}|U_0^{(\text{M})}) + I(U_0^{(\text{M})}, U_2^{(\text{M})}; Y_2^{(\text{M})}) \\
&\qquad\qquad - I(U_1^{(\text{M})}; U_2^{(\text{M})}|U_0^{(\text{M})}) \big) \\
&\quad + \gamma \big(I(X^{(\text{Enh})}; Y_1^{(\text{Enh})}, Y_2^{(\text{Enh})}) - H(Y_1^{(\text{Enh})}|Y_2^{(\text{Enh})}) \big). && \text{(6.116d)}
\end{aligned}$$

We argue in the following that the rate pair $(R_1 = R_1', R_2 = R_2')$ defined in (6.113) satisfies these constraints for all sufficiently small $\gamma > 0$. Comparing (6.109a), (6.110a), and (6.113a), we see that the first constraint (6.116a) is satisfied for any choice of $\gamma \in [0,1]$. Similarly, comparing (6.109c), (6.110a), (6.110b), and (6.113a) and (6.113b), we note that also the third constraint (6.116c) is satisfied for any $\gamma \in [0,1]$. The second constraint (6.116b) is satisfied when γ is sufficiently small. This can be seen by comparing (6.109b), (6.110b), and (6.113b), and because constraint (6.109b) holds with

strict inequality. The last constraint (6.116d) is not active in view of constraint (6.116c) whenever

$$\gamma H(Y_1^{(\mathrm{Enh})}|Y_2^{(\mathrm{Enh})}) \leq (1-\gamma)\Gamma^{(\mathrm{M})}, \tag{6.117}$$

where $\Gamma^{(\mathrm{M})}$ is defined in (6.30). Thus, also this last constraint is satisfied when γ is sufficiently small. This concludes our proof.

Chapter 7

Conclusion and Outlook

In this dissertation we studied the usefulness of feedback in memoryless BCs and proposed new ways of exploiting feedback to improve communication reliability or increase the non-feedback capacity region.

We investigated the linear-feedback schemes with a message point (in the spirit of Schalkwijk-Kailath coding scheme) for the $K \geq 2$-receiver memoryless Gaussian BCs with only common message. We showed that this type of linear-feedback schemes, even in the presence of perfect feedback, is strictly suboptimal, i.e. it fail to achieve the capacity. Furthermore, as the number of receivers K increases, the largest rate that is achievable this type scheme vanishes. In contrast, we presented an intermittent coding scheme with rate-limited feedback which can achieve the capacity with a L-th order exponential decay in the probability of error when the feedback rate satisfies $R_{\text{Fb}} > (L-1)R$.

we also studied the achievable regions for the two-receiver memoryless BCs with rate-limited/noisy feedback. We proposed two types of coding schemes based on Marton's coding and block-Markov strategy. Our first type of scheme was shown to strictly improve over the non-feedback capacity region for the class of strictly essentially less-noisy BCs, for any positive feedback rate and even when there is only one-sided feedback from the weaker receiver. This result even holds for some more capable BC, e.g. for BSC/BEC-BC. Our second type of scheme can recover all previously known capacity and DoF results for memoryless BCs with feedback when the feedback rates are sufficiently large. When the feedback rates tend to infinity, our scheme improves over a special case

of the Shayevitz-Wigger scheme.

In the following we discuss some related problems that are interesting to be investigate in the future.

- **Reliability on DMBC with general messages and feedback:** In Chapter 5, we investigated the communication reliability on Gaussian BCs with feedback and common message only. We are also interested in how feedback improves communication reliability for DMBC with general messages, including both private messages and common messages.
- **Outer bound on the capacity region of BC with feedback:** Previously, for general DMBC with feedback, only simple (and certainly weak) outer bounds by Cover [139] and by Ozarow&Leung [62] were known, see also in Section 4.2. In the future, we would like to compare these two outer bounds and derive a new better one. This will provide more insights on how far the performances of our coding schemes or other coding schemes lie from the optimal performance that can be achieved with feedback for BCs.
- **Coding scheme for more general network:** In Chapter 6, we have dealt only two-receiver BCs with rate-limited (or noisy) feedback. Considering the case of more than two receivers would be of interest. We also wish to extend our coding schemes to more general network: *discrete memoryless multi-message broadcast network*, in which there are multiple source nodes and each wishes to send more than one message to multiple destination nodes. Such filed is interesting and almost new at the moment.

References

[1] C. E. Shannon. The zero error capacity of a noisy channel. *Information Theory, IEEE Transactions on*, 2(3):8–19, Sept. 1956.

[2] J. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback–I: No bandwidth constraint. *Information Theory, IEEE Transactions on*, 12(2):172–182, Apr. 1966.

[3] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell System Technical Journal, The*, 38(3):611–656, May 1959.

[4] Young-Han Kim, A. Lapidoth, and T. Weissman. The Gaussian channel with noisy feedback. In *Information Theory Proceedings (ISIT), 2007 IEEE International Symposium on*, pages 1416–1420, June 2007.

[5] M. V. Burnashev and H. Yamamoto. On the reliability function for a noisy feedback Gaussian channel: Zero rate. *Problems of Information Transmission*, 48(3):199–216, 2012.

[6] Young-Han Kim, A. Lapidoth, and T. Weissman. Error exponents for the Gaussian channel with active noisy feedback. *Information Theory, IEEE Transactions on*, 57(3):1223–1236, Mar. 2011.

[7] M. S. Pinsker. The probability of error in block transmission in a memoryless Gaussian channel with feedback. *Probl. Peredachi Inf.*, 4(4):3–19, 1968.

[8] A. Kramer. Improving communication reliability by use of an intermittent feedback channel. *Information Theory, IEEE Transactions on*, 15(1):52–60, Jan. 1969.

[9] R. Mirghaderi, A. Goldsmith, and T. Weissman. Achievable error exponents in the Gaussian channel with rate-limited feedback. *Information Theory, IEEE Transactions on*, 59(12):8144–8156, Dec. 2013.

[10] L. H. Ozarow. The capacity of the white Gaussian multiple access channel with feedback. *Information Theory, IEEE Transactions on*, 30(4):623–629, July 1984.

[11] G. Kramer. Feedback strategies for white Gaussian interference networks. *Information Theory, IEEE Transactions on*, 48(6):1423–1438, June 2002.

[12] E. Ardestanizadeh, M. Wigger, Young-Han Kim, and T. Javidi. Linear sum-capacity for Gaussian multiple access channels with feedback. *Information Theory, IEEE Transactions on*, 58(1):224–236, Jan. 2012.

[13] E. Ardestanizadeh, P. Minero, and M. Franceschetti. LQG control approach to Gaussian broadcast channels with feedback. *Information Theory, IEEE Transactions on*, 58(8):5267–5278, Aug. 2012.

[14] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger. Coding schemes and asymptotic capacity for the Gaussian broadcast and interference channels with feedback. *Information Theory, IEEE Transactions on*, 60(1):54–71, Jan. 2014.

[15] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger. New achievable rates for the Gaussian broadcast channel with feedback. In *Wireless Communication Systems (ISWCS), 2011 8th International Symposium on*, pages 579–583, Nov. 2011.

[16] S. B. Amor, Y. Steinberg, and M. Wigger. New achievable rates for the Gaussian broadcast channel with feedback. In *Proc. of 8th ISWCS*, pages 25–28, Zurich, Switzerland, Nov. 2014.

[17] S. B. Amor, Y. Steinberg, and M. Wigger. MIMO MAC-BC duality with linear-feedback coding schemes. submitted to *Information Theory, IEEE Transactions on*, 2014, available at `http://arxiv.org/abs/1404.2584`.

[18] R. L. Dobrushin. Asymptotic bounds on error probability for transmission over dmc with symmetric transition probabilities. *Theory Probab. Applicat*, 7:283–311, 1962.

[19] M. V. Burnashev. Data transmission over a discrete channel with feedback. Random transmission time. *Problemy peredachi informatsii*, 12(4):10–30, 1976.

[20] H. Yamamoto and K. Itoh. Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback. *Information Theory, IEEE Transactions on*, 25(6):729–733, Nov. 1979.

[21] J. M. Ooi and G. W. Wornell. Fast iterative coding techniques for feedback channels. *Information Theory, IEEE Transactions on*, 44(7), Nov. 1998.

[22] A. Tchamkerten and IE. Telatar. Variable length coding over an unknown channel. *Information Theory, IEEE Transactions on*, 52(5):2126–2145, May 2006.

[23] M. Naghshvar, T. Javidi, and M. Wigger. Extrinsic Jensen-Shannon divergence: Applications to variable-length coding. submitted to *Information Theory, IEEE Transactions on*, 2013, available at `http://arxiv.org/abs/1307.0067`.

[24] G. D. Forney. Exponential error bounds for erasure, list, and decision feedback schemes. *Information Theory, IEEE Transactions on*, 14(2):206–220, Mar. 1968.

[25] N. Gaarder and J. K. Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *Information Theory, IEEE Transactions on*, 21(1):100–102, Jan. 1975.

[26] T. Cover and C. Leung. An achievable rate region for the multiple-access channel with feedback. *Information Theory, IEEE Transactions on*, 27(3):292–298, May 1981.

[27] F. M. J. Willems. The feedback capacity region of a class of discrete memoryless multiple access channels. *Information Theory, IEEE Transactions on*, 28(1):93–95, Jan. 1982.

[28] S. I. Bross and A. Lapidoth. An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback. *Information Theory, IEEE Transactions on*, 51(3):811–833, Mar. 2005.

[29] G. Kramer. Capacity results for the discrete memoryless network. *Information Theory, IEEE Transactions on*, 49(1):4–21, Jan. 2003.

[30] G. Kramer. *Directed information for channels with feedback*. PhD thesis, University of Manitoba, Canada, 1998.

[31] H. H Permuter, T. Weissman, and J. Chen. Capacity region of the finite-state multiple-access channel with and without feedback. *Information Theory, IEEE Transactions on*, 55(6):2455–2477, June 2009.

[32] A. Carleial. Multiple-access channels with different generalized feedback signals. *Information Theory, IEEE Transactions on*, 28(6):841–850, Nov. 1982.

[33] D. Shaviv and Y. Steinberg. On the multiple-access channel with common rate-limited feedback. *Information Theory, IEEE Transactions on*, 59(6):3780–3795, June 2013.

[34] A. Lapidoth and M. Wigger. On the AWGN MAC with imperfect feedback. *Information Theory, IEEE Transactions on*, 56(11):5432–5476, Nov. 2010.

[35] C. Heegard and T. Berger. Rate distortion when side information may be absent. *Information Theory, IEEE Transactions on*, 31(6):727–734, Nov. 1985.

[36] R. Knopp and P. A. Humblet. Information capacity and power control in single-cell multiuser communications. In *Communications, 1995. ICC '95 Seattle, 'Gateway to Globalization', 1995 IEEE International Conference on*, volume 1, pages 331–335 vol.1, June 1995.

[37] S. V. Hanly and D. Tse. Multiaccess fading channels. II. delay-limited capacities. *Information Theory, IEEE Transactions on*, 44(7):2816–2831, Nov. 1998.

[38] X. Qin and R. Berry. Exploiting multiuser diversity for medium access control in wireless networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference*

of the IEEE Computer and Communications. IEEE Societies, volume 2, pages 1084–1094, Mar. 2003.

[39] S. Adireddy and L. Tong. Exploiting decentralized channel state information for random access. *Information Theory, IEEE Transactions on*, 51(2):537–561, Feb. 2005.

[40] Chan-Soo Hwang, K. Seong, and J. M. Cioffi. Opportunistic p-persistent CSMA in wireless networks. In *Communications, 2006. ICC '06. IEEE International Conference on*, volume 1, pages 183–188, June 2006.

[41] Y. Cemal and Y. Steinberg. The multiple-access channel with partial state information at the encoders. *Information Theory, IEEE Transactions on*, 51(11):3992–4003, Nov. 2005.

[42] S. A. Jafar. Capacity with causal and noncausal side information: A unified view. *Information Theory, IEEE Transactions on*, 52(12):5468–5474, Dec. 2006.

[43] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: Information-theoretic and communications aspects. *Information Theory, IEEE Transactions on*, 44(6):2619–2692, Oct. 1998.

[44] Chan-Soo Hwang, M. Malkin, A. El Gamal, and J. M. Cioffi. Multiple-access channels with distributed channel state information. In *Information Theory Proceedings (ISIT), 2007 IEEE International Symposium on*, pages 1561–1565, June 2007.

[45] K. Singh, S. R. B. Pillai, and B. K. Dey. On fading MAC channels with asymmetric CSI. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 2626–2630, July 2013.

[46] S. Shamai and T. L. Marzetta. Multiuser capacity in block fading with no channel state information. In *Information Theory, 2000. Proceedings. IEEE International Symposium on*, pages 129–, June 2000.

[47] S. A. Jafar, S. Vishwanath, and .A Goldsmith. Vector MAC capacity region with covariance feedback. In *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, pages 54–, June 2001.

[48] P. Viswanath, D. Tse, and V. Anantharam. Asymptotically optimal water-filling in vector multiple-access channels. *Information Theory, IEEE Transactions on*, 47(1):241–267, Jan. 2001.

[49] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *Selected Areas in Communications, IEEE Journal on*, 21(5):684–702, June 2003.

[50] K. R. Kumar and G. Caire. Channel state feedback over the MIMO-MAC. *Information Theory, IEEE Transactions on*, 57(12):7787–7797, Dec. 2011.

[51] M. Gastpar and G. Kramer. On cooperation via noisy feedback. In *Communications, 2006 International Zurich Seminar on*, pages 146–149, 2006.

[52] R. Tandon and S. Ulukus. Dependence balance based outer bounds for Gaussian networks with cooperation and feedback. *Information Theory, IEEE Transactions on*, 57(7):4063–4086, July 2011.

[53] A. P. Hekstra and F. M. J. Willems. Dependence balance bounds for single-output two-way channels. *Information Theory, IEEE Transactions on*, 35(1):44–53, Jan. 1989.

[54] A. El Gamal. The feedback capacity of degraded broadcast channels. *Information Theory, IEEE Transactions on*, 24(3):379–381, May 1978.

[55] G. Dueck. Partial feedback for two-way and broadcast channels. *Information and Control*, 46:1–15, July 1980.

[56] O. Shayevitz and M. Wigger. On the capacity of the discrete memoryless broadcast channel with feedback. *Information Theory, IEEE Transactions on*, 59(3):1329–1345, Mar. 2013.

[57] H. Kim, Yeow-Khiang Chia, and A. El Gamal. A note on broadcast channels with stale state information at the transmitter. submitted to *Information Theory, IEEE Transactions on*, 2013, available at `http://arxiv.org/abs/1309.7437`.

[58] Chih-Chun Wang. On the capacity of 1-to-K broadcast packet erasure channels with channel output feedback. *Information Theory, IEEE Transactions on*, 58(2):931–956, Feb. 2012.

[59] L. Georgiadis and L. Tassiulas. Broadcast erasure channel with feedback-capacity and algorithms. In *Proc. 5th Workshop on Network Coding, Theory and Applications*, pages 54–61, Lausanne, Switzerland, June 2009.

[60] M. A. Maddah-Ali and D. Tse. Completely stale transmitter channel state information is still very useful. *Information Theory, IEEE Transactions on*, 58(7):4418–4431, July 2012.

[61] R. Venkataramanan and S. S. Pradhan. An achievable rate region for the broadcast channel with feedback. *Information Theory, IEEE Transactions on*, 59(10):6175–6191, Oct. 2013.

[62] L. H. Ozarow and S. Leung-Yan-Cheong. An achievable region and outer bound for the Gaussian broadcast channel with feedback. *Information Theory, IEEE Transactions on*, 30(4):667–671, Jul. 1984.

[63] S. Vishwanath, W. Wu, and A. Arapostathis. Gaussian interference networks with feedback: Duality, sum capacity and dynamic team problems. In *Proc. of the 44th Allerton*, 2005.

[64] N. Elia. When Bode meets Shannon: Control-oriented feedback communication schemes. *Automatic Control, IEEE Transactions on*, 49(9):1477–1488, Sept. 2004.

[65] S. R. Bhaskaran. Gaussian broadcast channel with feedback. *Information Theory, IEEE Transactions on*, 54(11):5252–5257, Nov. 2008.

[66] L. Li and A. J. Goldsmith. Capacity and optimal resource allocation for fading broadcast channels .i. ergodic capacity. *Information Theory, IEEE Transactions on*, 47(3):1083–1102, Mar. 2001.

[67] N. Jindal. MIMO broadcast channels with finite-rate feedback. *Information Theory, IEEE Transactions on*, 52(11):5045–5060, Nov. 2006.

[68] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran. Multiuser MIMO achievable rates with downlink training and channel state feedback. *Information Theory, IEEE Transactions on*, 56(6):2845–2866, June 2010.

[69] M. Kobayashi, N. Jindal, and G. Caire. Training and feedback optimization for multiuser MIMO downlink. *Communications, IEEE Transactions on*, 59(8):2228–2240, Aug. 2011.

[70] N. Ravindran and N. Jindal. Limited feedback-based block diagonalization for the MIMO broadcast channel. *Selected Areas in Communications, IEEE Journal on*, 26(8):1473–1482, Oct. 2008.

[71] A. Lapidoth, S. Shamai, and M. Wigger. On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information. In *Proc. of the 43th Allerton*, 2005.

[72] T. Gou, S. A. Jafar, and C. Wang. On the degrees of freedom of finite state compound wireless networks. *Information Theory, IEEE Transactions on*, 57(6):3286–3308, June 2011.

[73] H. Weingarten, S. Shamai, and G. Kramer. On the compound MIMO broadcast channel. In *Proceedings of Annual Information Theory and Applications Workshop UCSD. 2007.*

[74] M. A. Maddah-Ali. On the degrees of freedom of the compound MIMO broadcast channels with finite states. submitted to *Information Theory, IEEE Transactions on*, 2013, available at `http://arxiv.org/abs/0909.5006`.

[75] A. G. Davoodi and S. A. Jafar. Aligned image sets under channel uncertainty: Settling a conjecture by Lapidoth, Shamai and Wigger on the collapse of degrees of freedom under finite precision CSIT. 2013, available at `http://arxiv.org/abs/1403.1541`.

[76] J. Chen and P. Elia. Can imperfect delayed CSIT be as useful as perfect delayed CSIT? DoF analysis and constructions for the BC. In *Proc. of the 50th Allerton*, pages 1254–1261, Oct. 2012.

[77] J. Chen and P. Elia. Toward the performance versus feedback tradeoff for the two-user MISO broadcast channel. *Information Theory, IEEE Transactions on*, 59(12):8336–8356, Dec. 2013.

[78] J. Chen, R. Knopp, and P. Elia. Interference alignment for achieving both full DoF and full diversity in the broadcast channel with delayed CSIT. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1887–1891, July 2012.

[79] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi. Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT. *Information Theory, IEEE Transactions on*, 59(1):315–328, Jan. 2013.

[80] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi. The degrees of freedom region of temporally correlated MIMO networks with delayed CSIT. *Information Theory, IEEE Transactions on*, 60(1):494–514, Jan. 2014.

[81] R. Tandon, S. A. Jafar, S. Shamai Shitz, and H. V. Poor. On the synergistic benefits of alternating CSIT for the MISO broadcast channel. *Information Theory, IEEE Transactions on*, 59(7):4106–4128, July 2013.

[82] T. Gou and S. A Jafar. Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT. *Communications Letters, IEEE*, 16(7):1084–1087, July 2012.

[83] J. Jiang, Y. Xin, and H. K. Garg. Discrete memoryless interference channels with feedback. In *Information Sciences and Systems, 2007. CISS '07. 41st Annual Conference on*, pages 581–584, Mar. 2007.

[84] G. Kramer. Feedback strategies for white Gaussian interference networks. *Information Theory, IEEE Transactions on*, 48(6):1423–1438, June 2002.

[85] G. Kramer. Correction to "Feedback strategies for white Gaussian interference networks, " and a capacity theorem for Gaussian interference channels with feedback. *Information Theory, IEEE Transactions on*, 50(6):1373–1374, June 2004.

[86] C. Suh and D. Tse. Feedback capacity of the Gaussian interference channel to within 2 bits. *Information Theory, IEEE Transactions on*, 57(5):2667–2685, May 2011.

[87] A. Sahai, V. Aggarwal, M. Yuksel, and A Sabharwal. On channel output feedback in deterministic interference channels. In *Information Theory Workshop, 2009. ITW 2009. IEEE*, pages 298–302, Oct. 2009.

[88] A. Vahid, C. Suh, and A. S. Avestimehr. Interference channels with rate-limited feedback. *Information Theory, IEEE Transactions on*, 58(5):2788–2812, May 2012.

[89] C. Suh, I-Hsiang Wang, and D. Tse. Two-way interference channels. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2801–2805, July 2012.

[90] D. S. Papailiopoulos, C. Suh, and A. G. Dimakis. Feedback in the K-user interference channel. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 3130–3134, July 2012.

[91] S.I Bross, Y. Steinberg, and S. Tinguely. The discrete memoryless interference channel with one-sided generalized feedback. *Information Theory, IEEE Transactions on*, 59(7):4171–4191, July 2013.

[92] A. El Gamal and M. Costa. The capacity region of a class of deterministic interference channels. *Information Theory, IEEE Transactions on*, 28(2):343–346, Mar. 1982.

[93] A. S. Avestimehr, S. N. Diggavi, and D. Tse. Wireless network information flow: A deterministic approach. *Information Theory, IEEE Transactions on*, 57(4):1872–1905, Apr. 2011.

[94] S. A. Jafar and M. J. Fakhereddin. Degrees of freedom for the MIMO interference channel. *Information Theory, IEEE Transactions on*, 53(7):2637–2642, July 2007.

[95] C. Huang and S. A. Jafar. Degrees of freedom of the MIMO interference channel with cooperation and cognition. *Information Theory, IEEE Transactions on*, 55(9):4211–4220, Sept. 2009.

[96] V. R. Cadambe and S. A. Jafar. Interference alignment afnd the degrees of freedom of wireless X networks. *Information Theory, IEEE Transactions on*, 55(9):3893–3908, Sept. 2009.

[97] C. Wang, T. Gou, and S. A. Jafar. Subspace alignment chains and the degrees of freedom of the three-user MIMO interference channel. *Information Theory, IEEE Transactions on*, 60(5):2432–2479, May 2014.

[98] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran. On feasibility of interference alignment in MIMO interference networks. *Signal Processing, IEEE Transactions on*, 58(9):4771–4782, Sept. 2010.

[99] M. Rezaee, R. Couillet, M. Guillaud, and G. Matz. Sum-rate optimization for the MIMO IC under imperfect CSI: A deterministic equivalent approach. In *Signal Processing Advances in Wireless Communications (SPAWC), 2012 IEEE 13th International Workshop on*, pages 259–263, June 2012.

[100] R. T. Krishnamachari and M. K. Varanasi. Interference alignment under limited feedback for MIMO interference channels. *Signal Processing, IEEE Transactions on*, 61(15):3908–3917, Aug. 2013.

[101] M. Ashraphijuo, V. Aggarwal, and X. Wang. On the capacity region and the generalized degrees of freedom region for the MIMO interference channel with feedback. *Information Theory, IEEE Transactions on*, 59(12):8357–8376, Dec. 2013.

[102] H. Maleki, S. A Jafar, and S. Shamai. Retrospective interference alignment over interference networks. *Selected Topics in Signal Processing, IEEE Journal of*, 6(3):228–240, June 2012.

[103] A. Ghasemi, A. S. Motahari, and A. K. Khandani. On the degrees of freedom of X channel with delayed CSIT. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 767–770, July 2011.

[104] M. J. Abdoli, A. Ghasemi, and A. K. Khandani. On the degrees of freedom of K-user SISO interference and X channels with delayed CSIT. *Information Theory, IEEE Transactions on*, 59(10):6542–6561, Oct. 2013.

[105] X. Yi, D. Gesbert, S. Yang, and M. Kobayashi. On the DoF of the multiple-antenna time correlated interference channel with delayed CSIT. In *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, pages 1566–1570, Nov. 2012.

[106] C. S. Vaze and M. K. Varanasi. The degrees of freedom region and interference alignment for the MIMO interference channel with delayed CSIT. *Information Theory, IEEE Transactions on*, 58(7):4396–4417, July 2012.

[107] C. S. Vaze and M. K. Varanasi. The degrees-of-freedom region of the MIMO interference channel with Shannon feedback. *Information Theory, IEEE Transactions on*, 59(8):4798–4810, Aug. 2013.

[108] E. C. Van Der Meulen. Three-terminal communication channels. *Advances in applied Probability*, pages 120–154, Spring 1971.

[109] T. Cover and A. El Gamal. Capacity theorems for the relay channel. *Information Theory, IEEE Transactions on*, 25(5):572–584, Sept. 1979.

[110] S. I. Bross. A converse result for the discrete memoryless relay channel with relay–transmitter feedback. *Information Theory, IEEE Transactions on*, 52(8):3789–3792, Aug. 2006.

[111] Y. Gabbai and S. I. Bross. Achievable rates for the discrete memoryless relay channel with partial feedback configurations. *Information Theory, IEEE Transactions on*, 52(11):4989–5007, Nov. 2006.

[112] S. I. Bross and M. Wigger. On the relay channel with receiver–transmitter feedback. *Information Theory, IEEE Transactions on*, 55(1):275–291, Jan. 2009.

[113] F. M. J. Willems and E. C. Van Der Meulen. Partial feedback for the discrete memoryless multiple access channel. *Information Theory, IEEE Transactions on*, 29(2):287–290, Mar. 1983.

[114] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *Information Theory, IEEE Transactions on*, 51(9):3037–3063, Sept. 2005.

[115] M. H. Yassaee and M. R. Aref. Generalized compress-and-forward strategy for relay networks. In *Information Theory Proceedings (ISIT), 2008 IEEE International Symposium on*, pages 2683–2687, July 2008.

[116] O. Sahin and E. Erkip. Achievable rates for the Gaussian interference relay channel. In *Global Telecommunications Conference, (GLOBECOM 2007) 2007. IEEE*, pages 1627–1631, Nov. 2007.

[117] I. Maric, R. Dabora, and A. Goldsmith. On the capacity of the interference channel with a relay. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 554–558, July 2008.

[118] L. Sankar, N. B. Mandayam, and H. V. Poor. On the sum-capacity of degraded Gaussian multiple-access relay channels. *Information Theory, IEEE Transactions on*, 55(12):5394–5411, Dec. 2009.

[119] J. Hou, R. Koetter, and G. Kramer. Rate regions for multiple access relay channels with relay-source feedback. In *Information Theory Workshop, 2009. ITW 2009. IEEE*, pages 288–292, Oct. 2009.

[120] C. Keong Ho, K. T. Gowda, and S. Sun. Achievable rates for multiple access relay channel with generalized feedback. In *Information Theory and Its Applications, 2008. ISITA 2008. International Symposium on*, pages 1–6, Dec. 2008.

[121] S. I. Bross. On the discrete memoryless partially cooperative relay broadcast channel and the broadcast channel with cooperating decoders. *Information Theory, IEEE Transactions on*, 55(5):2161–2182, May 2009.

[122] Y. Liang and G. Kramer. Rate regions for relay broadcast channels. *Information Theory, IEEE Transactions on*, 53(10):3517–3535, Oct. 2007.

[123] R. Zhang, C. C. Chai, and Ying-Chang Liang. Joint beamforming and power control for multiantenna relay broadcast channel with QoS constraints. *Signal Processing, IEEE Transactions on*, 57(2):726–737, Feb. 2009.

[124] R. Khosravi-Farsani, B. Akhbari, M. Mirmohseni, and M. R. Aref. Cooperative relay-broadcast channels with causal channel state information. In *Information Theory Proceedings (ISIT), 2009 IEEE International Symposium on*, pages 1174–1178, June 2009.

[125] W. Xu, X. Dong, and W. Lu. MIMO relaying broadcast channels with linear precoding and quantized channel state information feedback. *Signal Processing, IEEE Transactions on*, 58(10):5233–5245, Oct. 2010.

[126] B. Zhang, Z. He, K. Niu, and L. Zhang. Robust linear beamforming for MIMO relay broadcast channel with limited feedback. *Signal Processing Letters, IEEE*, 17(2):209–212, Feb. 2010.

[127] L. Liang, W. Xu, and X. Dong. Limited feedback-based multi-antenna relay broadcast channels with block diagonalization. *Wireless Communications, IEEE Transactions on*, 12(8):4092–4101, Aug. 2013.

[128] V. R. Cadambe and S. A Jafar. Degrees of freedom of wireless networks with relays, feedback, cooperation, and full duplex operation. *Information Theory, IEEE Transactions on*, 55(5):2334–2344, May 2009.

[129] D. Zahavi and R. Dabora. Capacity theorems for the fading interference channel with a relay and feedback links. *Information Theory, IEEE Transactions on*, 58(8):5185–5213, Aug. 2012.

[130] M. Plainchault, N. Gresset, and G.R.-B. Othman. Interference relay channel with precoded dynamic decode and forward protocols. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–6, Dec. 2011.

[131] A. El Gamal and Young-Han Kim. *Network information theory.* Cambridge University Press, 2011.

[132] G. Kramer. *Topics in multi-user information theory.* now publishers inc, 2008.

[133] T. Cover. Comments on broadcast channels. *Information Theory, IEEE Transactions on*, 44(6):2524–2530, Oct. 1998.

[134] J. Körner and K. Marton. General broadcast channels with degraded message sets. *Information Theory, IEEE Transactions on*, 23(1):60–64, Jan. 1977.

[135] C. Nair. Capacity regions of two new classes of two-receiver broadcast channels. *Information Theory, IEEE Transactions on*, 56(9):4207–4214, Sept. 2010.

[136] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Probl. Peredachi Inf.*, 10(3):3–14, 1974.

[137] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *Information Theory, IEEE Transactions on*, 19(2):197–207, Mar. 1973.

[138] J. Körner and K. Marton. Comparison of two noisy channels. *Topics in Information Theory (ed. by I. Csiszar and P. Elias), Keszthely, Hungary*, pages 411–423, 1977.

[139] T. Cover and J. A Thomas. *Elements of information theory.* John Wiley & Sons, 2012.

[140] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *Information Theory, IEEE Transactions on*, 25(3):306–311, May 1979.

[141] S. I. Gelfand and M. S. Pinsker. Capacity of a broadcast channel with one deterministic component. *Problems of Information Transmission*, 16(1):17–25, Jan.-Mar. 1980.

[142] Y. Liang, G. Kramer, and H. V. Poor. On the equivalence of two achievable regions for the broadcast channel. *Information Theory, IEEE Transactions on*, 57(1):95–100, Jan. 2011.

[143] Y. Liang, G. Kramer, and S. Shamai. Capacity outer bounds for broadcast channels. In *Information Theory Workshop, 2008. ITW 2008. IEEE*, pages 2–4, May 2008.

[144] C. Nair and A. El Gamal. An outer bound to the capacity region of the broadcast channel. *Information Theory, IEEE Transactions on*, 53(1):350–355, Jan. 2007.

[145] C. Nair. A note on outer bounds for broadcast channel.

[146] Y. Geng, A. Gohari, C. Nair, and Y. Yu. The capacity region for two classes of product broadcast channels. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 1544–1548, July 2011.

[147] R. Gray and A. Wyne. Source coding for a simple network. *Bell System Technical Journal, The*, 48:1681–1721, Nov. 1974.

[148] M. Wigger and M. Gastpar. The pre-log of Gaussian broadcast with feedback can be two. In *Information Theory Proceedings (ISIT), 2008 IEEE International Symposium on*, pages 1258–1262, July 2008.

[149] Young-Han Kim, A. Lapidoth, and T. Weissman. On the reliability of Gaussian channels with noisy feedback. In *Proc. of the 44th Allerton*, pages 364–371, Sept. 2006.

[150] R. G. Gallager and B. Nakiboglu. Variations on a theme by Schalkwijk and Kailath. *Information Theory, IEEE Transactions on*, 56(1):6–17, Jan. 2010.

[151] A. Sahai, S. C. Draper, and M. Gastpar. Boosting reliability over AWGN networks with average power constraints and noiseless feedback. In *Information Theory Proceedings (ISIT), 2005 IEEE International Symposium on*, pages 402–406, Sept. 2005.

[152] P. Bergmans. A simple converse for broadcast channels with additive white Gaussian noise. *Information Theory, IEEE Transactions on*, 20(2):279–280, Mar. 1974.

[153] S. Lim, Young-Han Kim, A. El Gamal, and Sae-Young Chung. Noisy network coding. *Information Theory, IEEE Transactions on*, 57(5):3132–3152, May 2011.

[154] J. Hou and G. Kramer. Short message noisy network coding for multiple sources. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1677–1681, July 2012.

[155] A. El Gamal. The capacity of a class of broadcast channels. *Information Theory, IEEE Transactions on*, 25(2):166–169, Mar. 1979.