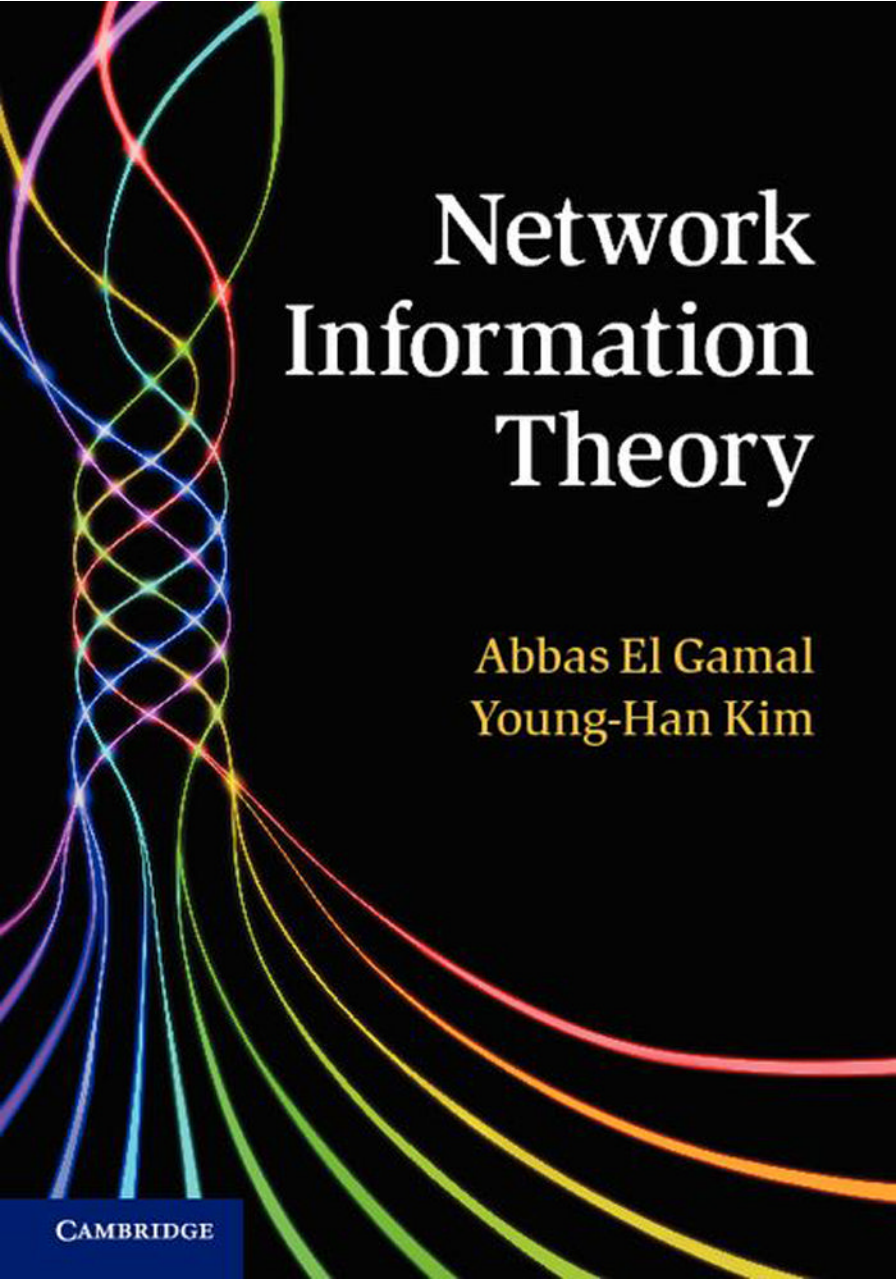
Network Information Theory
Abbas El Gamal
Young-Han Kim
CAMBRIDGE

Network Information Theory

This comprehensive treatment of network information theory and its applications provides the first unified coverage of both classical and recent results. With an approach that balances the introduction of new models and new coding techniques, readers are guided through Shannon's point-to-point information theory, single-hop networks, multihop networks, and extensions to distributed computing, secrecy, wireless communication, and networking. Elementary mathematical tools and techniques are used throughout, requiring only basic knowledge of probability, whilst unified proofs of coding theorems are based on a few simple lemmas, making the text accessible to newcomers. Key topics covered include successive cancellation and superposition coding, MIMO wireless communication, network coding, and cooperative relaying. Also covered are feedback and interactive communication, capacity approximations and scaling laws, and asynchronous and random access channels. This book is ideal for use in the classroom, for self-study, and as a reference for researchers and engineers in industry and academia.

Abbas El Gamal is the Hitachi America Chaired Professor in the School of Engineering and the Director of the Information Systems Laboratory in the Department of Electrical Engineering at Stanford University. In the field of network information theory, he is best known for his seminal contributions to the relay, broadcast, and interference channels; multiple description coding; coding for noisy networks; and energy-efficient packet scheduling and throughput–delay tradeoffs in wireless networks. He is a Fellow of IEEE and the winner of the 2012 Claude E. Shannon Award, the highest honor in the field of information theory.

Young-Han Kim is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of California, San Diego. His research focuses on information theory and statistical signal processing. He is a recipient of the 2008 NSF Faculty Early Career Development (CAREER) Award and the 2009 US–Israel Binational Science Foundation Bergmann Memorial Award.

NETWORK INFORMATION THEORY

Abbas El Gamal
Stanford University

Young-Han Kim
University of California, San Diego

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107008731

First published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

ISBN 978-1-107-00873-1 Hardback

Additional resources for this publication at www.cambridge.org/9781107008731

Contents

Part III Multihop Networks

Appendices

CHAPTER 5

Degraded Broadcast Channels

We introduce the broadcast channel as a model for noisy one-to-many communication, such as the downlink of a cellular system, digital TV broadcasting, or a lecture to a group of people with diverse backgrounds. Unlike the multiple access channel, no single-letter characterization is known for the capacity region of the broadcast channel. However, there are several interesting coding schemes and corresponding inner bounds on the capacity region that are tight in some special cases.

In this chapter, we introduce the technique of superposition coding and use it to establish an inner bound on the capacity region of the broadcast channel. We show that this inner bound is tight for the class of degraded broadcast channels, which includes the Gaussian broadcast channel. The converse proofs involve the new technique of auxiliary random variable identification and the use of Mrs. Gerber's lemma, a symmetrization argument, and the entropy power inequality. We then show that superposition coding is optimal for the more general classes of less noisy and more capable broadcast channels. The converse proof uses the Csiszár sum identity.

We will resume the discussion of the broadcast channel in Chapters 8 and 9. In these chapters, we will present more general inner and outer bounds on the capacity region and show that they coincide for other classes of channels. The reason for postponing the presentation of these results is pedagogical—we will need coding techniques that can be introduced at a more elementary level through the discussion of channels with state in Chapter 7.

5.1 DISCRETE MEMORYLESS BROADCAST CHANNEL

Consider the broadcast communication system depicted in Figure 5.1. The sender wishes to reliably communicate a private message to each receiver and a common message to both receivers. It encodes the message triple (M_0, M_1, M_2) into a codeword X^n and transmits it over the channel. Upon receiving Y_j^n, receiver $j = 1, 2$ finds the estimates $\hat{M}_{0j}$ and $\hat{M}_j$ of the common message and its private message, respectively. A tradeoff arises between the rates of the three messages—when one of the rates is high, the other rates may need to be reduced to ensure reliable communication of all three messages. As before, we study the asymptotic limit on this tradeoff.

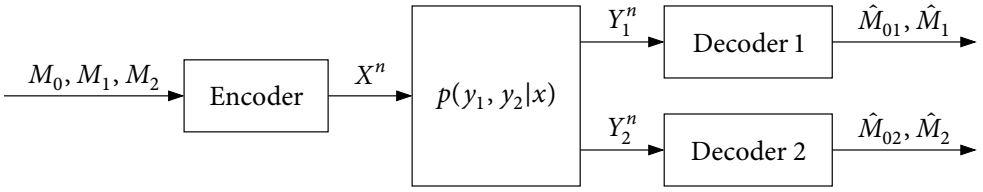

Figure 5.1. Two-receiver broadcast communication system.

We first consider a 2-receiver *discrete memoryless broadcast channel* (DM-BC) model $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ that consists of three finite sets $\mathcal{X}$, $\mathcal{Y}_1$, $\mathcal{Y}_2$, and a collection of conditional pmfs $p(y_1, y_2|x)$ on $\mathcal{Y}_1 \times \mathcal{Y}_2$ (one for each input symbol x).

A $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n)$ code for a DM-BC consists of

- three message sets $[1 : 2^{nR_0}]$, $[1 : 2^{nR_1}]$, and $[1 : 2^{nR_2}]$,
- an encoder that assigns a codeword $x^n(m_0, m_1, m_2)$ to each message triple $(m_0, m_1, m_2) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$, and
- two decoders, where decoder 1 assigns an estimate $(\hat{m}_{01}, \hat{m}_1) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_1}]$ or an error message e to each received sequence y_1^n, and decoder 2 assigns an estimate $(\hat{m}_{02}, \hat{m}_2) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_2}]$ or an error message e to each received sequence y_2^n.

We assume that the message triple (M_0, M_1, M_2) is uniformly distributed over $[1 : 2^{nR_0}] \times [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The average probability of error is defined as

$$P_e^{(n)} = \mathsf{P}\{(\hat{M}_{01}, \hat{M}_1) \neq (M_0, M_1) \text{ or } (\hat{M}_{02}, \hat{M}_2) \neq (M_0, M_2)\}.$$

A rate triple (R_0, R_1, R_2) is said to be achievable for the DM-BC if there exists a sequence of $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$. The capacity region $\mathscr{C}$ of the DM-BC is the closure of the set of achievable rate triples (R_0, R_1, R_2).

The following simple observation, which results from the lack of cooperation between the two receivers, will prove useful.

Lemma 5.1. The capacity region of the DM-BC depends on the channel conditional pmf $p(y_1, y_2|x)$ only through the conditional marginal pmfs $p(y_1|x)$ and $p(y_2|x)$.

Proof. Consider the individual probabilities of error

$$P_{ej}^{(n)} = \mathsf{P}\{(\hat{M}_{0j}, \hat{M}_j) \neq (M_0, M_j)\}, \quad j = 1, 2.$$

Note that each term depends only on its corresponding conditional marginal pmf $p(y_j|x)$, $j = 1, 2$. By the union of events bound, $P_e^{(n)} \leq P_{e1}^{(n)} + P_{e2}^{(n)}$. Also $P_e^{(n)} \geq \max\{P_{e1}^{(n)}, P_{e2}^{(n)}\}$. Hence $\lim_{n\to\infty} P_e^{(n)} = 0$ iff $\lim_{n\to\infty} P_{e1}^{(n)} = 0$ and $\lim_{n\to\infty} P_{e2}^{(n)} = 0$, which implies that the capacity region of a DM-BC depends only on the conditional marginal pmfs. This completes the proof of the lemma.

For simplicity of presentation, we focus the discussion on the *private-message* capacity region, that is, the capacity region when $R_0 = 0$. We then show how the results can be readily extended to the case of both common and private messages.

5.2 SIMPLE BOUNDS ON THE CAPACITY REGION

Consider a DM-BC $p(y_1, y_2|x)$. Let $C_j = \max_{p(x)} I(X; Y_j)$, $j = 1, 2$, be the capacities of the DMCs $p(y_1|x)$ and $p(y_2|x)$. These capacities define the time-division inner bound in Figure 5.2. By allowing full cooperation between the receivers, we obtain the bound on the sum-rate

$$R_1 + R_2 \le C_{12} = \max_{p(x)} I(X; Y_1, Y_2).$$

Combining this bound with the individual capacity bounds gives the outer bound on the capacity region in Figure 5.2.

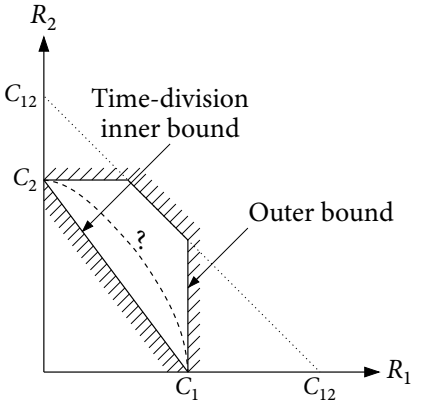

Figure 5.2. Time-division inner bound and an outer bound on the capacity region.

The time-division bound can be tight in some cases.

Example 5.1. Consider a symmetric DM-BC, where $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Y}$ and $p_{Y_1|X}(y|x) = p_{Y_2|X}(y|x) = p(y|x)$. In this example, $C_1 = C_2 = \max_{p(x)} I(X; Y)$. Since the capacity region depends only on the marginals of $p(y_1, y_2|x)$, we can assume that $Y_1 = Y_2 = Y$. Thus

$$R_1 + R_2 \le C_{12} = \max_{p(x)} I(X; Y) = C_1 = C_2.$$

This shows that the time-division inner bound and the outer bound sometimes coincide.

The following example shows that the bounds do not always coincide.

Example 5.2. Consider a DM-BC with orthogonal components, where $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and $p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)$. For this example, the capacity region is the set of rate pairs (R_1, R_2) such that $R_1 \le C_1$ and $R_2 \le C_2$; thus the outer bound is tight, but the time-division inner bound is not.

As we will see, neither bound is tight in general.

5.3 SUPERPOSITION CODING INNER BOUND

The superposition coding technique is motivated by broadcast channels where one receiver is "stronger" than the other, for example, because it is closer to the sender, such that it can always recover the "weaker" receiver's message. This suggests a *layered* coding approach in which the weaker receiver's message is treated as a "public" (common) message and the stronger receiver's message is treated as a "private" message. The weaker receiver recovers only the public message, while the stronger receiver recovers both messages using successive cancellation decoding as for the MAC; see Section 4.5.1. We illustrate this coding scheme in the following.

Example 5.3 (Binary symmetric broadcast channel). The binary symmetric broadcast channel (BS-BC) consists of a BSC(p_1) and a BSC(p_2) as depicted in Figure 5.3. Assume that $p_1 < p_2 < 1/2$.

Using time division, we can achieve the straight line between the capacities $1 - H(p_1)$ and $1 - H(p_2)$ of the individual BSCs. Can we achieve higher rates than time division? To answer this question, consider the following *superposition coding* technique illustrated in Figure 5.4.

For $\alpha \in [0, 1/2]$, let $U \sim \text{Bern}(1/2)$ and $V \sim \text{Bern}(\alpha)$ be independent, and $X = U \oplus V$. Randomly and independently generate 2^{nR_2} sequences $u^n(m_2)$, each i.i.d. Bern(1/2) (cloud centers). Randomly and independently generate 2^{nR_1} sequences $v^n(m_1)$, each i.i.d. Bern(α). The sender transmits $x^n(m_1, m_2) = u^n(m_2) \oplus v^n(m_1)$ (satellite codeword).

To recover m_2, receiver 2 decodes $y_2^n = u^n(m_2) \oplus (v^n(m_1) \oplus z_2^n)$ while treating $v^n(m_1)$ as noise. The probability of decoding error tends to zero as $n \to \infty$ if $R_2 < 1 - H(\alpha * p_2)$, where $\alpha * p_2 = \alpha\bar{p}_2 + \bar{\alpha}p_2$.

Receiver 1 uses successive cancellation decoding—it first decodes $y_1^n = u^n(m_2) \oplus (v^n(m_1) \oplus z_1^n)$ to recover m_2 while treating $v^n(m_1)$ as part of the noise, subtracts off $u^n(m_2)$,

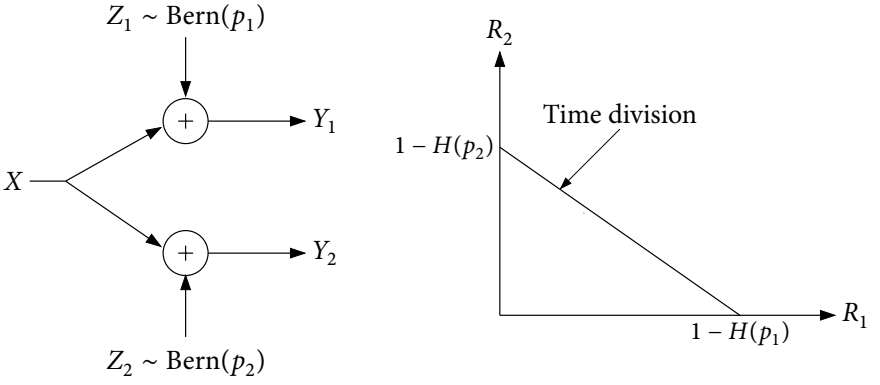

Figure 5.3. Binary symmetric broadcast channel and its time-division inner bound.

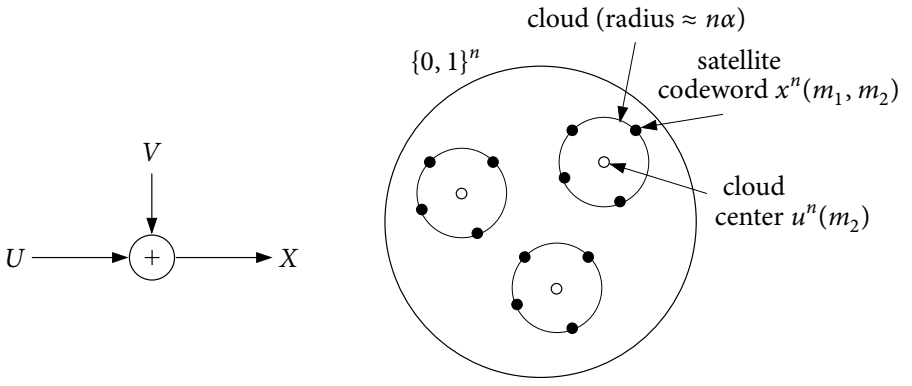

Figure 5.4. Superposition coding for the BS-BC.

and then decodes $v^n(m_1) \oplus z_1^n$ to recover m_1. The probability of decoding error tends to zero as $n \to \infty$ if $R_1 < I(V; V \oplus Z_1) = H(\alpha * p_1) - H(p_1)$ and $R_2 < 1 - H(\alpha * p_1)$. The latter condition is already satisfied by the rate constraint for receiver 2 since $p_1 < p_2$.

Thus, superposition coding leads to an inner bound consisting of the set of rate pairs (R_1, R_2) such that

$$R_1 \le H(\alpha * p_1) - H(p_1),$$
$$R_2 \le 1 - H(\alpha * p_2)$$

for some $\alpha \in [0, 1/2]$. This inner bound is larger than the time-division inner bound as sketched in Figure 5.5. We will show later that this bound is the capacity region of the BS-BC.

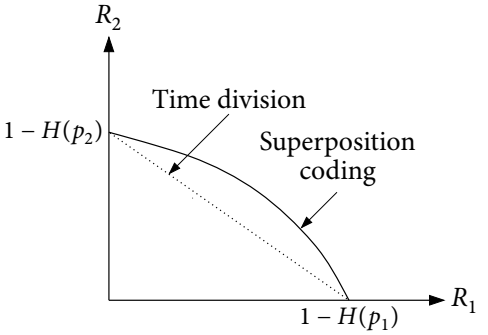

Figure 5.5. Superposition coding inner bound for the BS-BC.

The superposition coding technique illustrated in the above example can be generalized to obtain the following inner bound on the capacity region of the general DM-BC.

Theorem 5.1 (Superposition Coding Inner Bound). A rate pair (R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$ if

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

for some pmf $p(u, x)$.

Remark 5.1. The random variable U in the theorem is an *auxiliary* random variable that does not correspond to any of the channel variables. While the time-sharing random variable encountered in Chapter 4 is also an auxiliary random variable, U actually carries message information and as such plays a more essential role in the characterization of the capacity region.

Remark 5.2. The superposition coding inner bound is evaluated for the joint pmf $p(u, x)$ $p(y_1, y_2|x)$, i.e., $U \to X \to (Y_1, Y_2)$; see Remark 4.7.

Remark 5.3. It can be shown that the inner bound is convex and therefore there is no need for further convexification via a time-sharing random variable.

5.3.1 Proof of the Superposition Coding Inner Bound

We begin with a sketch of achievability; see Figure 5.6. Fix a pmf $p(u, x)$ and randomly generate 2^{nR_2} "cloud centers" $u^n(m_2)$. For each cloud center $u^n(m_2)$, generate 2^{nR_1} "satellite" codewords $x^n(m_1, m_2)$. Receiver 2 decodes for the cloud center $u^n(m_2)$ and receiver 1 decodes for the satellite codeword.

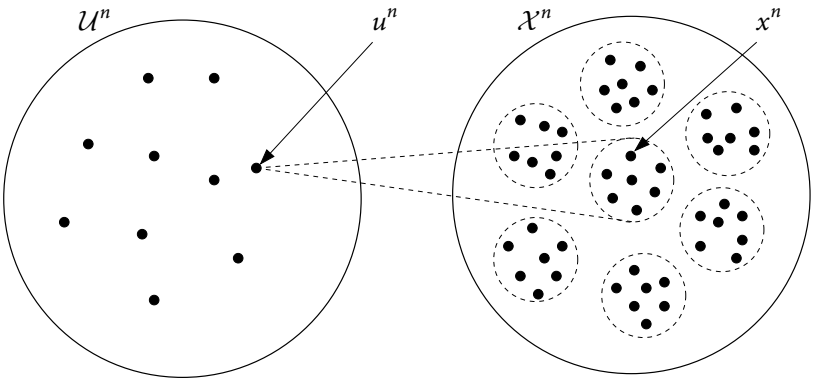

Figure 5.6. Superposition coding.

We now provide details of the proof.

Codebook generation. Fix a pmf $p(u)p(x|u)$. Randomly and independently generate 2^{nR_2} sequences $u^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, each according to $\prod_{i=1}^n p_U(u_i)$. For each $m_2 \in [1 : 2^{nR_2}]$, randomly and conditionally independently generate 2^{nR_1} sequences $x^n(m_1, m_2)$, $m_1 \in [1 : 2^{nR_1}]$, each according to $\prod_{i=1}^n p_{X|U}(x_i|u_i(m_2))$.

Encoding. To send (m_1, m_2), transmit $x^n(m_1, m_2)$.

Decoding. Decoder 2 declares that $\hat{m}_2$ is sent if it is the unique message such that $(u^n(\hat{m}_2), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$; otherwise it declares an error. Since decoder 1 is interested only in recovering m_1, it uses simultaneous decoding without requiring the recovery of m_2, henceforth referred to as *simultaneous nonunique decoding*. Decoder 1 declares that $\hat{m}_1$ is sent if it is the unique message such that $(u^n(m_2), x^n(\hat{m}_1, m_2), y_1^n) \in \mathcal{T}_\epsilon^{(n)}$ for some m_2; otherwise it declares an error.

Analysis of the probability of error. Assume without loss of generality that $(M_1, M_2) = (1, 1)$ is sent. First consider the average probability of error for decoder 2. This decoder makes an error iff one or both of the following events occur:

$$\mathcal{E}_{21} = \{(U^n(1), Y_2^n) \notin \mathcal{T}_\epsilon^{(n)}\},$$
$$\mathcal{E}_{22} = \{(U^n(m_2), Y_2^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}.$$

Thus the probability of error for decoder 2 is upper bounded as

$$\mathsf{P}(\mathcal{E}_2) \le \mathsf{P}(\mathcal{E}_{21}) + \mathsf{P}(\mathcal{E}_{22}).$$

By the LLN, the first term $\mathsf{P}(\mathcal{E}_{21})$ tends to zero as $n \to \infty$. For the second term, since $U^n(m_2)$ is independent of $(U^n(1), Y_2^n)$ for $m_2 \neq 1$, by the packing lemma, $\mathsf{P}(\mathcal{E}_{22})$ tends to zero as $n \to \infty$ if $R_2 < I(U; Y_2) - \delta(\epsilon)$.

Next consider the average probability of error for decoder 1. To analyze this probability, consider all possible pmfs for the triple $(U^n(m_2), X^n(m_1, m_2), Y_1^n)$ as listed in Table 5.1. Since the last case in the table does not result in an error, the decoder makes an error iff

m_1	m_2	Joint pmf
1	1	$p(u^n, x^n)p(y_1^n\|x^n)$
$*$	1	$p(u^n, x^n)p(y_1^n\|u^n)$
$*$	$*$	$p(u^n, x^n)p(y_1^n)$
1	$*$	$p(u^n, x^n)p(y_1^n)$

Table 5.1. The joint pmfs induced by different (m_1, m_2) pairs. The $*$ symbol corresponds to message $m_1 \neq 1$ or $m_2 \neq 1$.

one or more of the following three events occur:

$$\mathcal{E}_{11} = \{(U^n(1), X^n(1,1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\},$$
$$\mathcal{E}_{12} = \{(U^n(1), X^n(m_1,1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\},$$
$$\mathcal{E}_{13} = \{(U^n(m_2), X^n(m_1,m_2), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1,\ m_2 \neq 1\}.$$

Thus the probability of error for decoder 1 is upper bounded as

$$\mathsf{P}(\mathcal{E}_1) \leq \mathsf{P}(\mathcal{E}_{11}) + P(\mathcal{E}_{12}) + \mathsf{P}(\mathcal{E}_{13}).$$

We now bound each term. By the LLN, $\mathsf{P}(\mathcal{E}_{11})$ tends to zero as $n \to \infty$. For the second term, note that if $m_1 \neq 1$, then $X^n(m_1, 1)$ is conditionally independent of $(X^n(1,1), Y_1^n)$ given $U^n(1)$ and is distributed according to $\prod_{i=1}^n p_{X|U}(x_i|u_i(1))$. Hence, by the packing lemma, $\mathsf{P}(\mathcal{E}_{12})$ tends to zero as $n \to \infty$ if $R_1 < I(X; Y_1|U) - \delta(\epsilon)$.

Finally, for the third term, note that for $m_2 \neq 1$ (and any m_1), $(U^n(m_2), X^n(m_1, m_2))$ is independent of $(U^n(1), X^n(1,1), Y_1^n)$. Hence, by the packing lemma, $\mathsf{P}(\mathcal{E}_{13})$ tends to zero as $n \to \infty$ if $R_1 + R_2 < I(U, X; Y_1) - \delta(\epsilon) = I(X; Y_1) - \delta(\epsilon)$ (recall that $U \to X \to Y_1$ form a Markov chain). This completes the proof of achievability.

Remark 5.4. Consider the error event

$$\mathcal{E}_{14} = \{(U^n(m_2), X^n(1, m_2), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}.$$

By the packing lemma, $\mathsf{P}(\mathcal{E}_{14})$ tends to zero as $n \to \infty$ if $R_2 < I(U, X; Y_1) - \delta(\epsilon) = I(X; Y_1) - \delta(\epsilon)$, which is already satisfied. Therefore, the inner bound does not change if we use simultaneous (unique) decoding and require decoder 1 to also recover M_2.

Remark 5.5. We can obtain a second superposition coding inner bound by having receiver 1 decode for the cloud center (which would now represent M_1) and receiver 2 decode for the satellite codeword. This yields the set of rate pairs (R_1, R_2) such that

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

for some pmf $p(u, x)$. The convex closure of the union of the two superposition coding inner bounds constitutes a generally tighter inner bound on the capacity region.

Remark 5.6. Superposition coding is optimal for several classes of broadcast channels for which one receiver is stronger than the other, as we detail in the following sections. It is not, however, optimal in general since requiring one of the receivers to recover both messages (even nonuniquely for the other receiver's message) can unduly constrain the set of achievable rates. In Chapter 8, we present Marton's coding scheme, which can outperform superposition coding by not requiring either receiver to recover both messages.

5.4 DEGRADED DM-BC

In a degraded broadcast channel, one of the receivers is *statistically* stronger than the other. Formally, a DM-BC is said to be *physically degraded* if

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

i.e., $X \to Y_1 \to Y_2$ form a Markov chain.

More generally, a DM-BC $p(y_1, y_2|x)$ is said to be *stochastically degraded* (or simply degraded) if there exists a random variable $\tilde{Y}_1$ such that $\tilde{Y}_1|\{X = x\} \sim p_{Y_1|X}(\tilde{y}_1|x)$, i.e., $\tilde{Y}_1$ has the same conditional pmf as Y_1 (given X), and $X \to \tilde{Y}_1 \to Y_2$ form a Markov chain.

Since the capacity region of a DM-BC depends only on the conditional marginal pmfs, the capacity region of a stochastically degraded DM-BC is the same as that of its corresponding physically degraded channel. This observation will be used later in the proof of the converse.

Note that the BS-BC in Example 5.3 is a degraded DM-BC. Again assume that $p_1 < p_2 < 1/2$. The channel is degraded since we can write $Y_2 = X \oplus \tilde{Z}_1 \oplus \tilde{Z}_2$, where $\tilde{Z}_1 \sim \text{Bern}(p_1)$ and $\tilde{Z}_2 \sim \text{Bern}(\tilde{p}_2)$ are independent, and

$$\tilde{p}_2 = \frac{p_2 - p_1}{1 - 2p_1}.$$

Figure 5.7 shows the physically degraded BS-BC with the same conditional marginal pmfs. The capacity region of the original BS-BC is the same as that of the physically degraded version.

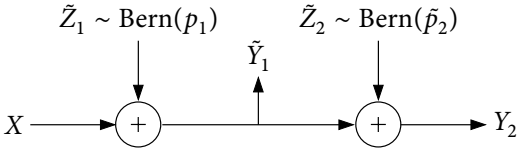

Figure 5.7. Corresponding physically degraded BS-BC.

The superposition coding inner bound is tight for the class of degraded broadcast channels.

Theorem 5.2. The capacity region of the degraded DM-BC $p(y_1, y_2|x)$ is the set of rate pairs (R_1, R_2) such that

$$R_1 \le I(X; Y_1|U),$$
$$R_2 \le I(U; Y_2)$$

for some pmf $p(u, x)$, where the cardinality of the auxiliary random variable U satisfies $|\mathcal{U}| \le \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\} + 1$.

To prove achievability, note that the sum of the first two inequalities in the superposition coding inner bound gives $R_1 + R_2 < I(U; Y_2) + I(X; Y_1|U)$. Since the channel is degraded, $I(U; Y_1) \geq I(U; Y_2)$ for all $p(u, x)$. Hence, $I(U; Y_2) + I(X; Y_1|U) \leq I(U; Y_1) + I(X; Y_1|U) = I(X; Y_1)$ and the third inequality is automatically satisfied.

5.4.1 Proof of the Converse

We need to show that given any sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$, we must have $R_1 \leq I(X; Y_1|U)$ and $R_2 \leq I(U; Y_2)$ for some pmf $p(u, x)$ such that $U \to X \to (Y_1, Y_2)$ form a Markov chain. The key new idea in the converse proof is the identification of the auxiliary random variable U. Every $(2^{nR_1}, 2^{nR_2}, n)$ code (by definition) induces a joint pmf on $(M_1, M_2, X^n, Y_1^n, Y_2^n)$ of the form

$$p(m_1, m_2, x^n, y_1^n, y_2^n) = 2^{-n(R_1+R_2)} p(x^n|m_1, m_2) \prod_{i=1}^{n} p_{Y_1,Y_2|X}(y_{1i}, y_{2i}|x_i).$$

By Fano's inequality,

$$\begin{aligned} H(M_1|Y_1^n) &\leq nR_1 P_e^{(n)} + 1 \leq n\epsilon_n, \\ H(M_2|Y_2^n) &\leq nR_2 P_e^{(n)} + 1 \leq n\epsilon_n \end{aligned}$$

for some ϵ_n that tends to zero as $n \to \infty$. Hence

$$\begin{aligned} nR_1 &\leq I(M_1; Y_1^n) + n\epsilon_n, \\ nR_2 &\leq I(M_2; Y_2^n) + n\epsilon_n. \end{aligned} \tag{5.1}$$

First consider the following intuitive identification of the auxiliary random variable. Since U represents M_2 in the superposition coding scheme, it is natural to choose $U = M_2$, which also satisfies the desired Markov chain condition $U \to X_i \to (Y_{1i}, Y_{2i})$. Using this identification, consider the first mutual information term in (5.1),

$$\begin{aligned} I(M_1; Y_1^n) &\leq I(M_1; Y_1^n|M_2) \\ &= I(M_1; Y_1^n|U) \\ &= \sum_{i=1}^{n} I(M_1; Y_{1i}|U, Y_1^{i-1}) \qquad (5.2) \\ &\leq \sum_{i=1}^{n} I(M_1, Y_1^{i-1}; Y_{1i}|U) \\ &= \sum_{i=1}^{n} I(X_i, M_1, Y_1^{i-1}; Y_{1i}|U) \\ &= \sum_{i=1}^{n} I(X_i; Y_{1i}|U). \qquad (5.3) \end{aligned}$$

Note that this bound has the same structure as the inequality on R_1 in Theorem 5.2. Next, consider the second mutual information term in (5.1),

$$\begin{aligned} I(M_2; Y_2^n) &= \sum_{i=1}^{n} I(M_2; Y_{2i}|Y_2^{i-1}) \\ &= \sum_{i=1}^{n} I(U; Y_{2i}|Y_2^{i-1}). \end{aligned}$$

However, $I(U; Y_{2i}|Y_2^{i-1})$ is not necessarily less than or equal to $I(U; Y_{2i})$. Thus setting $U = M_2$ does not yield an inequality that has the same form as the one on R_2 in Theorem 5.2.

Step (5.2) in the above series of inequalities suggests the identification of the auxiliary random variable as $U_i = (M_2, Y_1^{i-1})$, which also satisfies $U_i \to X_i \to (Y_{1i}, Y_{2i})$. This gives

$$\begin{aligned} I(M_1; Y_1^n) &\le I(M_1; Y_1^n|M_2) \\ &= \sum_{i=1}^{n} I(M_1; Y_{1i}|Y_1^{i-1}, M_2) \\ &= \sum_{i=1}^{n} I(X_i, M_1; Y_{1i}|U_i) \\ &= \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i). \end{aligned} \tag{5.4}$$

Now consider the second term in (5.1)

$$\begin{aligned} I(M_2; Y_2^n) &= \sum_{i=1}^{n} I(M_2; Y_{2i}|Y_2^{i-1}) \\ &\le \sum_{i=1}^{n} I(M_2, Y_2^{i-1}; Y_{2i}) \\ &\le \sum_{i=1}^{n} I(M_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2i}). \end{aligned} \tag{5.5}$$

However, $I(M_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2i})$ is not necessarily equal to $I(M_2, Y_1^{i-1}; Y_{2i})$.

Using the observation that the capacity region of the general degraded DM-BC is the same as that of the corresponding physically degraded BC, we can assume without loss of generality that $X \to Y_1 \to Y_2$ form a Markov chain. This implies that $Y_2^{i-1} \to (M_2, Y_1^{i-1}) \to Y_{2i}$ also form a Markov chain and hence (5.5) implies that

$$I(M_2; Y_2^n) \le \sum_{i=1}^{n} I(U_i; Y_{2i}). \tag{5.6}$$

To complete the proof, define the time-sharing random variable Q to be uniformly distributed over $[1 : n]$ and independent of $(M_1, M_2, X^n, Y_1^n, Y_2^n)$, and identify $U = (Q, U_Q)$,

$X = X_Q$, $Y_1 = Y_{1Q}$, and $Y_2 = Y_{2Q}$. Clearly, $U \to X \to (Y_1, Y_2)$ form a Markov chain. Using these definitions and substituting from (5.4) and (5.6) into (5.1), we have

$$\begin{aligned} nR_1 &\le \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i) + n\epsilon_n \\ &= nI(X_Q; Y_{1Q}|U_Q, Q) + n\epsilon_n \\ &= nI(X; Y_1|U) + n\epsilon_n, \\ nR_2 &\le \sum_{i=1}^{n} I(U_i; Y_{2i}) + n\epsilon_n \\ &= nI(U_Q; Y_{2Q}|Q) + n\epsilon_n \\ &\le nI(U; Y_2) + n\epsilon_n. \end{aligned}$$

The bound on the cardinality of U can be established using the convex cover method in Appendix C. This completes the proof of Theorem 5.2.

Remark 5.7. The proof works also for $U_i = (M_2, Y_2^{i-1})$ or $U_i = (M_2, Y_1^{i-1}, Y_2^{i-1})$; both identifications satisfy the Markov condition $U_i \to X_i \to (Y_{1i}, Y_{2i})$.

5.4.2 Capacity Region of the BS-BC

We show that the superposition coding inner bound for the BS-BC presented in Example 5.3 is tight. Since the BS-BC is degraded, Theorem 5.2 characterizes its capacity region. We show that this region can be simplified to the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le H(\alpha * p_1) - H(p_1), \\ R_2 &\le 1 - H(\alpha * p_2) \end{aligned} \tag{5.7}$$

for some $\alpha \in [0, 1/2]$.

In Example 5.3, we argued that the interior of this rate region is achievable via superposition coding by taking $U \sim \text{Bern}(1/2)$ and $X = U \oplus V$, where $V \sim \text{Bern}(\alpha)$ is independent of U. We now show that the characterization of the capacity region in Theorem 5.2 reduces to (5.7) using two alternative methods.

Proof via Mrs. Gerber's lemma. First recall that the capacity region is the same as that of the physically degraded DM-BC $X \to \tilde{Y}_1 \to Y_2$. Thus we assume that it is physically degraded, i.e., $Y_2 = Y_1 \oplus \tilde{Z}_2$, where $Z_1 \sim \text{Bern}(p_1)$ and $\tilde{Z}_2 \sim \text{Bern}(\tilde{p}_2)$ are independent and $p_2 = p_1 * \tilde{p}_2$.

Consider the second inequality in the capacity region characterization in Theorem 5.2

$$I(U; Y_2) = H(Y_2) - H(Y_2|U) \le 1 - H(Y_2|U).$$

Since $1 \ge H(Y_2|U) \ge H(Y_2|X) = H(p_2)$, there exists an $\alpha \in [0, 1/2]$ such that $H(Y_2|U) = H(\alpha * p_2)$. Next consider

$$I(X; Y_1|U) = H(Y_1|U) - H(Y_1|X) = H(Y_1|U) - H(p_1).$$

Now let $0 \le H^{-1}(v) \le 1/2$ be the inverse of the binary entropy function. By physical degradedness and the scalar MGL in Section 2.1,

$$H(Y_2|U) = H(Y_1 \oplus \tilde{Z}_2|U) \ge H\big(H^{-1}(H(Y_1|U)) * \tilde{p}_2\big).$$

But $H(Y_2|U) = H(\alpha * p_2) = H(\alpha * p_1 * \tilde{p}_2)$, and thus

$$H(Y_1|U) \le H(\alpha * p_1).$$

Proof via a symmetrization argument. Since for the BS-BC $|\mathcal{U}| \le |\mathcal{X}| + 1 = 3$, assume without loss of generality that $\mathcal{U} = \{1, 2, 3\}$. Given any $(U, X) \sim p(u, x)$, define $(\tilde{U}, \tilde{X}) \sim p(\tilde{u}, \tilde{x})$ as

$$p_{\tilde{U}}(u) = p_{\tilde{U}}(-u) = \frac{1}{2} p_U(u), \quad u \in \{1, 2, 3\},$$
$$p_{\tilde{X}|\tilde{U}}(x|u) = p_{\tilde{X}|\tilde{U}}(1 - x| - u) = p_{X|U}(x|u), \quad (u, x) \in \{1, 2, 3\} \times \{0, 1\}.$$

Further define $\tilde{Y}_1$ to be the output of the BS-BC when the input is $\tilde{X}$. Given $\{|\tilde{U}| = u\}$, $u = 1, 2, 3$, the channel from $\tilde{U}$ to $\tilde{X}$ is a BSC with parameter $p(x|u)$ (with input alphabet $\{-u, u\}$ instead of $\{0, 1\}$). Thus, $H(Y_1|U = u) = H(\tilde{Y}_1|\tilde{U} = u) = H(\tilde{Y}_1|\tilde{U} = -u)$ for $u = 1, 2, 3$, which implies that $H(Y_1|U) = H(\tilde{Y}_1|\tilde{U})$ and the first bound in the capacity region is preserved. Similarly, $H(Y_2|U) = H(\tilde{Y}_2|\tilde{U})$. Also note that $\tilde{X} \sim \text{Bern}(1/2)$ and so are the corresponding outputs $\tilde{Y}_1$ and $\tilde{Y}_2$. Hence,

$$H(Y_2) - H(Y_2|U) \le H(\tilde{Y}_2) - H(\tilde{Y}_2|\tilde{U}).$$

Therefore, it suffices to evaluate the capacity region characterization in Theorem 5.2 with the above symmetric input pmfs $p(\tilde{u}, \tilde{x})$.

Next consider the weighted sum

$$\begin{aligned}
&\lambda I(\tilde{X}; \tilde{Y}_1|\tilde{U}) + (1 - \lambda) I(\tilde{U}; \tilde{Y}_2) \\
&\quad = \lambda H(\tilde{Y}_1|\tilde{U}) - (1 - \lambda) H(\tilde{Y}_2|\tilde{U}) - \lambda H(p_1) + (1 - \lambda) \\
&\quad = \sum_{u=1}^{3} \big(\lambda H(\tilde{Y}_1|\tilde{U}, |\tilde{U}| = u) - (1 - \lambda) H(\tilde{Y}_2|\tilde{U}, |\tilde{U}| = u)\big) p(u) - \lambda H(p_1) + (1 - \lambda) \\
&\quad \le \max_{u \in \{1,2,3\}} \big(\lambda H(\tilde{Y}_1|\tilde{U}, |\tilde{U}| = u) - (1 - \lambda) H(\tilde{Y}_2|\tilde{U}, |\tilde{U}| = u)\big) - \lambda H(p_1) + (1 - \lambda)
\end{aligned}$$

for some $\lambda \in [0, 1]$. Since the maximum is attained by a single u (i.e., $p_{\tilde{U}}(u) = p_{\tilde{U}}(-u) = 1/2$) and given $|\tilde{U}| = u$, the channel from $\tilde{U}$ to $\tilde{X}$ is a BSC, the set of rate pairs (R_1, R_2) such that

$$\lambda R_1 + (1 - \lambda) R_2 \le \max_{\alpha} \big[\lambda (H(\alpha * p_1) - H(p_1)) + (1 - \lambda)(1 - H(\alpha * p_2))\big]$$

constitutes an outer bound on the capacity region given by the supporting line corresponding to each $\lambda \in [0, 1]$. Finally, since the rate region in (5.7) is convex (see Problem 5.3), by Lemma A.1 it coincides with the capacity region.

Remark 5.8. A brute-force optimization of the weighted sum $R_{\text{sum}}(\lambda) = \lambda I(\tilde{X}; \tilde{Y}_1|\tilde{U}) + (1-\lambda) I(\tilde{U}; \tilde{Y}_2)$ can lead to the same conclusion. In this approach, it suffices to optimize over all $p(u, x)$ with $|\mathcal{U}| \le 2$. The above symmetrization argument provides a more elegant approach to performing this three-dimensional optimization.

Remark 5.9. The second converse proof does not require physical degradedness. In fact, this symmetrization argument can be used to evaluate the superposition coding inner bound in Theorem 5.1 for *any* nondegraded binary-input DM-BC with $p_{Y_1,Y_2|X}(y_1, y_2|0) = p_{Y_1,Y_2|X}(-y_1, -y_2|1)$ for all y_1, y_2 (after proper relabeling of the symbols) such as the BSC–BEC BC in Example 5.4 of Section 5.6.

5.5 GAUSSIAN BROADCAST CHANNEL

Consider the 2-receiver discrete-time additive white Gaussian noise (Gaussian in short) BC model depicted in Figure 5.8. The channel outputs corresponding to the input X are

$$Y_1 = g_1 X + Z_1,$$
$$Y_2 = g_2 X + Z_2,$$

where g_1 and g_2 are channel gains, and $Z_1 \sim \text{N}(0, N_0/2)$ and $Z_2 \sim \text{N}(0, N_0/2)$ are noise components. Thus in transmission time i, the channel outputs are

$$Y_{1i} = g_1 X_i + Z_{1i},$$
$$Y_{2i} = g_2 X_i + Z_{2i},$$

where $\{Z_{1i}\}$ and $\{Z_{2i}\}$ are WGN$(N_0/2)$ processes, independent of the channel input X^n. Note that only the marginal distributions of $\{(Z_{1i}, Z_{2i})\}$ are relevant to the capacity region and hence we do not need to specify their joint distribution. Assume without loss of generality that $|g_1| \ge |g_2|$ and $N_0/2 = 1$. Further assume an average transmission power constraint

$$\sum_{i=1}^{n} x_i^2(m_1, m_2) \le nP, \quad (m_1, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}].$$

For notational convenience, we consider the equivalent Gaussian BC channel

$$Y_1 = X + Z_1,$$
$$Y_2 = X + Z_2.$$

In this model, the channel gains are normalized to 1 and the *transmitter-referred* noise components Z_1 and Z_2 are zero-mean Gaussian with powers $N_1 = 1/g_1^2$ and $N_2 = 1/g_2^2$, respectively. This equivalence can be seen by first multiplying both sides of the equations of the original channel model by $1/g_1$ and $1/g_2$, respectively, and then scaling the resulting channel outputs by g_1 and g_2. Note that the equivalent broadcast channel is

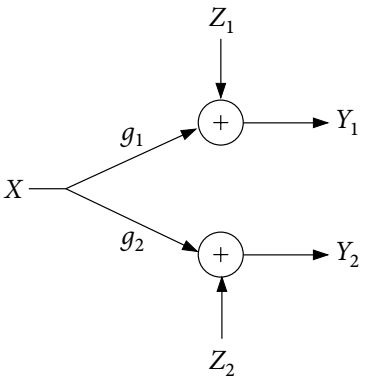

Figure 5.8. Gaussian broadcast channel.

stochastically degraded. Hence, its capacity region is the same as that of the physically degraded Gaussian BC (see Figure 5.9)

$$Y_1 = X + Z_1,$$
$$Y_2 = Y_1 + \tilde{Z}_2,$$

where Z_1 and $\tilde{Z}_2$ are independent Gaussian random variables with average powers N_1 and $(N_2 - N_1)$, respectively. Define the channel received signal-to-noise ratios as $S_1 = P/N_1$ and $S_2 = P/N_2$.

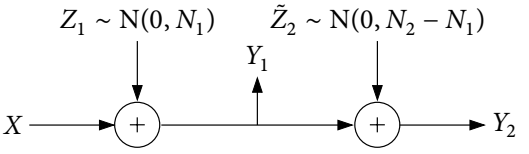

Figure 5.9. Corresponding physically degraded Gaussian BC.

5.5.1 Capacity Region of the Gaussian BC

The capacity region of the Gaussian BC is a function only of the channel SNRs and a power allocation parameter.

Theorem 5.3. The capacity region of the Gaussian BC is the set of rate pairs (R_1, R_2) such that

$$R_1 \le \mathsf{C}(\alpha S_1),$$
$$R_2 \le \mathsf{C}\left(\frac{\bar{\alpha} S_2}{\alpha S_2 + 1}\right)$$

for some $\alpha \in [0, 1]$, where $\mathsf{C}(x)$ is the Gaussian capacity function.

Achievability follows by setting $U \sim \mathrm{N}(0, \bar{\alpha}P)$ and $V \sim \mathrm{N}(0, \alpha P)$, independent of each other, $X = U + V \sim \mathrm{N}(0, P)$ in the superposition coding inner bound. With this choice of (U, X), it can be readily to shown that

$$I(X; Y_1|U) = \mathsf{C}\left(\frac{\alpha P}{N_1}\right) = \mathsf{C}(\alpha S_1),$$
$$I(U; Y_2) = \mathsf{C}\left(\frac{\bar{\alpha}P}{\alpha P + N_2}\right) = \mathsf{C}\left(\frac{\bar{\alpha}S_2}{\alpha S_2 + 1}\right),$$

which are the expressions in the theorem.

As for the BS-BC in Example 5.3, the superposition coding scheme for the Gaussian BC can be described more explicitly. Randomly and independently generate 2^{nR_2} sequences $u^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, each i.i.d. $\mathrm{N}(0, \bar{\alpha}P)$, and 2^{nR_1} sequences $v^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, each i.i.d. $\mathrm{N}(0, \alpha P)$. To send the message pair (m_1, m_2), the encoder transmits $x^n(m_1, m_2) = u^n(m_2) + v^n(m_1)$.

Receiver 2 recovers m_2 from $y_2^n = u^n(m_2) + (v^n(m_1) + z_2^n)$ while treating $v^n(m_1)$ as noise. The probability of decoding error tends to zero as $n \to \infty$ if $R_2 < \mathsf{C}(\bar{\alpha}P/(\alpha P + N_2))$.

Receiver 1 uses successive cancellation—it first decodes $y_1^n = u^n(m_2) + (v^n(m_1) + z_1^n)$ to recover m_2 while treating $v^n(m_1)$ as part of the noise, subtracts off $u^n(m_2)$, and then decodes $v^n(m_1) + z_1^n$ to recover m_1. The probability of decoding error tends to zero as $n \to \infty$ if $R_1 < \mathsf{C}(\alpha P/N_1)$ and $R_2 < \mathsf{C}(\bar{\alpha}P/(\alpha P + N_1))$. Since $N_1 < N_2$, the latter condition is already satisfied by the rate constraint for receiver 2.

Remark 5.10. To make the above arguments rigorous, we can use the discretization procedure detailed in the achievability proof for the point-to-point Gaussian channel in Section 3.4.

Remark 5.11. As for the Gaussian MAC in Section 4.6, every point in the capacity region can be achieved simply using good point-to-point Gaussian channel codes and successive cancellation decoding.

5.5.2 Proof of the Converse

Since the capacity region of the Gaussian BC is the same as that of its corresponding physically degraded Gaussian BC, we prove the converse for the physically degraded Gaussian BC shown in Figure 5.9. By Fano's inequality, we have

$$nR_1 \le I(M_1; Y_1^n|M_2) + n\epsilon_n,$$
$$nR_2 \le I(M_2; Y_2^n) + n\epsilon_n.$$

We need to show that there exists an $\alpha \in [0, 1]$ such that

$$I(M_1; Y_1^n|M_2) \le n\,\mathsf{C}(\alpha S_1) = n\,\mathsf{C}\left(\frac{\alpha P}{N_1}\right)$$

and

$$I(M_2; Y_2^n) \le n\,\mathsf{C}\left(\frac{\bar{\alpha}S_2}{\alpha S_2 + 1}\right) = n\,\mathsf{C}\left(\frac{\bar{\alpha}P}{\alpha P + N_2}\right).$$

Consider

$$I(M_2; Y_2^n) = h(Y_2^n) - h(Y_2^n|M_2) \le \frac{n}{2}\log\big(2\pi e(P+N_2)\big) - h(Y_2^n|M_2).$$

Since

$$\frac{n}{2}\log\big(2\pi e N_2\big) = h(Z_2^n) = h(Y_2^n|M_2, X^n) \le h(Y_2^n|M_2) \le h(Y_2^n) \le \frac{n}{2}\log(2\pi e(P+N_2)),$$

there must exist an $\alpha \in [0,1]$ such that

$$h(Y_2^n|M_2) = \frac{n}{2}\log(2\pi e(\alpha P + N_2)). \tag{5.8}$$

Next consider

$$\begin{aligned} I(M_1; Y_1^n|M_2) &= h(Y_1^n|M_2) - h(Y_1^n|M_1, M_2) \\ &= h(Y_1^n|M_2) - h(Y_1^n|M_1, M_2, X^n) \\ &= h(Y_1^n|M_2) - h(Y_1^n|X^n) \\ &= h(Y_1^n|M_2) - \frac{n}{2}\log(2\pi e N_1). \end{aligned}$$

Now using the conditional EPI in Section 2.1, we obtain

$$\begin{aligned} h(Y_2^n|M_2) &= h(Y_1^n + \tilde{Z}_2^n|M_2) \\ &\ge \frac{n}{2}\log\big(2^{2h(Y_1^n|M_2)/n} + 2^{2h(\tilde{Z}_2^n|M_2)/n}\big) \\ &= \frac{n}{2}\log\big(2^{2h(Y_1^n|M_2)/n} + 2\pi e(N_2 - N_1)\big). \end{aligned}$$

Combining this inequality with (5.8) implies that

$$2\pi e(\alpha P + N_2) \ge 2^{2h(Y_1^n|M_2)/n} + 2\pi e(N_2 - N_1).$$

Thus, $h(Y_1^n|M_2) \le (n/2)\log(2\pi e(\alpha P + N_1))$ and hence

$$I(M_1; Y_1^n|M_2) \le \frac{n}{2}\log\big(2\pi e(\alpha P + N_1)\big) - \frac{n}{2}\log(2\pi e N_1) = n\,\mathsf{C}\left(\frac{\alpha P}{N_1}\right).$$

This completes the proof of Theorem 5.3.

Remark 5.12. The converse can be proved directly starting with the single-letter characterization $R_1 \le I(X; Y_1|U)$, $R_2 \le I(U; Y_2)$ with the power constraint $\mathsf{E}(X^2) \le P$. The proof then follows similar steps to the above converse proof and uses the scalar version of the conditional EPI.

Remark 5.13. The similarity between the above converse proof and the converse proof for the BS-BC hints at some form of duality between Mrs. Gerber's lemma for binary random variables and the entropy power inequality for continuous random variables.

5.6 LESS NOISY AND MORE CAPABLE BROADCAST CHANNELS

Superposition coding is optimal for the following two classes of broadcast channels, which are more general than the class of degraded broadcast channels.

Less noisy DM-BC. A DM-BC $p(y_1, y_2|x)$ is said to be *less noisy* if $I(U; Y_1) \geq I(U; Y_2)$ for all $p(u, x)$. In this case we say that receiver 1 is less noisy than receiver 2. The private-message capacity region of the less noisy DM-BC is the set of rate pairs (R_1, R_2) such that

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

for some pmf $p(u, x)$, where $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_2|\} + 1$. Note that the less noisy condition guarantees that in the superposition coding scheme, receiver 1 can recover the message intended for receiver 2.

More capable DM-BC. A DM-BC $p(y_1, y_2|x)$ is said to be *more capable* if $I(X; Y_1) \geq I(X; Y_2)$ for all $p(x)$. In this case we say that receiver 1 is more capable than receiver 2. The private-message capacity region of the more capable DM-BC is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(X; Y_1|U), \\ R_2 &\leq I(U; Y_2), \\ R_1 + R_2 &\leq I(X; Y_1) \end{aligned} \tag{5.9}$$

for some pmf $p(u, x)$, where $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1| \cdot |\mathcal{Y}_2|\} + 1$.

It can be easily shown that if a DM-BC is degraded then it is less noisy, and that if a DM-BC is less noisy then it is more capable. The converse to each of these statements does not hold in general as illustrated in the following example.

Example 5.4 (A BSC and a BEC). Consider a DM-BC with $X \in \{0, 1\}$, $Y_1 \in \{0, 1\}$, and $Y_2 \in \{0, 1, \mathrm{e}\}$, where the channel from X to Y_1 is a BSC(p), $p \in (0, 1/2)$, and the channel from X to Y_2 is a BEC(ϵ), $\epsilon \in (0, 1)$. Then it can be shown that the following hold:

1. For $0 < \epsilon \leq 2p$, Y_1 is a *degraded* version of Y_2.
2. For $2p < \epsilon \leq 4p(1 - p)$, Y_2 is *less noisy* than Y_1, but Y_1 is not a degraded version of Y_2.
3. For $4p(1 - p) < \epsilon \leq H(p)$, Y_2 is *more capable* than Y_1, but not less noisy.
4. For $H(p) < \epsilon < 1$, the channel does not belong to *any* of the three classes.

The capacity region for each of these cases is achieved using superposition coding. The converse proofs for the first three cases follow by evaluating the capacity expressions for the degraded BC, less noisy BC, and more capable BC, respectively. The converse proof for the last case will be given in Chapter 8.

5.6.1 Proof of the Converse for the More Capable DM-BC

It is difficult to find an identification of the auxiliary random variable that satisfies the desired properties for the capacity region characterization in (5.9). Instead, we prove the converse for the equivalent region consisting of all rate pairs (R_1, R_2) such that

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

for some pmf $p(u, x)$.

The converse proof for this alternative region involves a tricky identification of the auxiliary random variable and the application of the Csiszár sum identity in Section 2.3.

By Fano's inequality, it is straightforward to show that

$$\begin{aligned} nR_2 &\le I(M_2; Y_2^n) + n\epsilon_n, \\ n(R_1 + R_2) &\le I(M_1; Y_1^n|M_2) + I(M_2; Y_2^n) + n\epsilon_n, \\ n(R_1 + R_2) &\le I(M_1; Y_1^n) + I(M_2; Y_2^n|M_1) + n\epsilon_n. \end{aligned}$$

Consider the mutual information terms in the second inequality.

$$\begin{aligned} I(M_1; Y_1^n|M_2) + I(M_2; Y_2^n) &= \sum_{i=1}^n I(M_1; Y_{1i}|M_2, Y_1^{i-1}) + \sum_{i=1}^n I(M_2; Y_{2i}|Y_{2,i+1}^n) \\ &\le \sum_{i=1}^n I(M_1, Y_{2,i+1}^n; Y_{1i}|M_2, Y_1^{i-1}) + \sum_{i=1}^n I(M_2, Y_{2,i+1}^n; Y_{2i}) \\ &= \sum_{i=1}^n I(M_1, Y_{2,i+1}^n; Y_{1i}|M_2, Y_1^{i-1}) + \sum_{i=1}^n I(M_2, Y_{2,i+1}^n, Y_1^{i-1}; Y_{2i}) \\ &\quad - \sum_{i=1}^n I(Y_1^{i-1}; Y_{2i}|M_2, Y_{2,i+1}^n) \\ &= \sum_{i=1}^n I(M_1; Y_{1i}|M_2, Y_1^{i-1}, Y_{2,i+1}^n) + \sum_{i=1}^n I(M_2, Y_{2,i+1}^n, Y_1^{i-1}; Y_{2i}) \\ &\quad - \sum_{i=1}^n I(Y_1^{i-1}; Y_{2i}|M_2, Y_{2,i+1}^n) + \sum_{i=1}^n I(Y_{2,i+1}^n; Y_{1i}|M_2, Y_1^{i-1}) \\ &\overset{(a)}{=} \sum_{i=1}^n \big(I(M_1; Y_{1i}|U_i) + I(U_i; Y_{2i})\big) \\ &\le \sum_{i=1}^n \big(I(X_i; Y_{1i}|U_i) + I(U_i; Y_{2i})\big), \end{aligned}$$

where $Y_1^0, Y_{2,n+1}^n = \emptyset$ and (a) follows by the Csiszár sum identity and the auxiliary random variable identification $U_i = (M_2, Y_1^{i-1}, Y_{2,i+1}^n)$.

Next consider the mutual information term in the first inequality

$$\begin{aligned} I(M_2; Y_2^n) &= \sum_{i=1}^{n} I(M_2; Y_{2i} | Y_{2,i+1}^n) \\ &\le \sum_{i=1}^{n} I(M_2, Y_1^{i-1}, Y_{2,i+1}^n; Y_{2i}) = \sum_{i=1}^{n} I(U_i; Y_{2i}). \end{aligned}$$

For the third inequality, define $V_i = (M_1, Y_1^{i-1}, Y_{2,i+1}^n)$. Following similar steps to the bound for the second inequality, we have

$$\begin{aligned} I(M_1; Y_1^n) + I(M_2; Y_2^n | M_1) &\le \sum_{i=1}^{n} \big(I(V_i; Y_{1i}) + I(X_i; Y_{2i} | V_i) \big) \\ &\overset{(a)}{\le} \sum_{i=1}^{n} \big(I(V_i; Y_{1i}) + I(X_i; Y_{1i} | V_i) \big) = \sum_{i=1}^{n} I(X_i; Y_{1i}), \end{aligned}$$

where (a) follows by the more capable condition, which implies $I(X; Y_2|V) \le I(X; Y_1|V)$ whenever $V \to X \to (Y_1, Y_2)$ form a Markov chain.

The rest of the proof follows by introducing a time-sharing random variable $Q \sim \text{Unif}[1 : n]$ independent of $(M_1, M_2, X^n, Y_1^n, Y_2^n)$ and defining $U = (Q, U_Q)$, $X = X_Q$, $Y_1 = Y_{1Q}$, and $Y_2 = Y_{2Q}$. The bound on the cardinality of U can be proved using the convex cover method in Appendix C. This completes the proof of the converse.

Remark 5.14. The converse proof for the more capable DM-BC also establishes the converse for the less noisy and degraded DM-BCs.

5.7 EXTENSIONS

We extend the results in the previous sections to the setup with common message and to channels with more than two receivers.

Capacity region with common and private messages. Our discussion so far has focused on the private-message capacity region. As mentioned in Remark 5.4, the superposition coding inner bound is still achievable if we require the "stronger" receiver to also recover the message intended for the "weaker" receiver. Therefore, if the rate pair (R_1, R_2) is achievable for private messages, then the rate triple $(R_0, R_1, R_2 - R_0)$ is achievable for private and common messages.

Using this observation, we can readily show that the capacity region with common message of the more capable DM-BC is the set of rate triples (R_0, R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X; Y_1|U), \\ R_0 + R_2 &\le I(U; Y_2), \\ R_0 + R_1 + R_2 &\le I(X; Y_1) \end{aligned}$$

for some pmf $p(u, x)$. Unlike achievability, the converse is not implied automatically by

the converse for the private-message capacity region. The proof, however, follows similar steps to that for the private-message capacity region.

k-Receiver degraded DM-BC. Consider a k-receiver degraded DM-BC $p(y_1, \ldots, y_k|x)$, where $X \to Y_1 \to Y_2 \to \cdots \to Y_k$ form a Markov chain. The private-message capacity region is the set of rate tuples $(R_1, \ldots, R_k)$ such that

$$\begin{aligned} R_1 &\le I(X; Y_1|U_2), \\ R_j &\le I(U_j; Y_j|U_{j+1}), \quad j \in [2:k], \end{aligned}$$

for some pmf $p(u_k, u_{k-1})p(u_{k-2}|u_{k-1}) \cdots p(x|u_2)$ and $U_{k+1} = \emptyset$. The capacity region for the 2-receiver Gaussian BC also extends to more than two receivers.

Remark 5.15. The capacity region is not known in general for the less noisy DM-BC with $k > 3$ receivers and for the more capable DM-BC with $k > 2$ receivers.

SUMMARY

- Discrete memoryless broadcast channel (DM-BC)
- Capacity region depends only on the channel marginal pmfs
- Superposition coding
- Simultaneous nonunique decoding
- Physically and stochastically degraded BCs
- Capacity region of degraded BCs is achieved by superposition coding
- Identification of the auxiliary random variable in the proof of the converse
- Bounding the cardinality of the auxiliary random variable
- Proof of the converse for the BS-BC:
  - Mrs. Gerber's lemma
  - Symmetrization argument
- Gaussian BC is always degraded
- Use of EPI in the proof of the converse for the Gaussian BC
- Less noisy and more capable BCs:
  - Degraded $\Rightarrow$ less noisy $\Rightarrow$ more capable
  - Superposition coding is optimal
- Use of Csiszár sum identity in the proof of the converse for more capable BCs

- **Open problems:**

  5.1. What is the capacity region of less noisy BCs with four or more receivers?

  5.2. What is the capacity region of more capable BCs with three or more receivers?

BIBLIOGRAPHIC NOTES

The broadcast channel was first introduced by Cover (1972), who demonstrated superposition coding through the BS-BC and Gaussian BC examples, and conjectured the characterization of the capacity region of the stochastically degraded broadcast channel using the auxiliary random variable U. Bergmans (1973) proved achievability of the capacity region of the degraded DM-BC. Subsequently, Gallager (1974) proved the converse by providing the nonintuitive identification of the auxiliary random variable discussed in Section 5.4. He also provided a bound on the cardinality of U. Wyner (1973) proved the converse for the capacity region of the BS-BC using Mrs. Gerber's lemma. The symmetrization argument in the alternative converse proof for the BS-BC is due to Nair (2010), who also applied it to binary-input symmetric-output BCs. Bergmans (1974) established the converse for the capacity region of the Gaussian BC using the entropy power inequality. A strong converse for the capacity region of the DM-BC was proved by Ahlswede and Körner (1975) for the maximal probability of error. A technique by Willems (1990) can be used to extend this strong converse to the average probability of error; see also Csiszár and Körner (1981b) for an indirect proof based on the correlation elimination technique by Ahlswede (1978).

The classes of less noisy and more capable DM-BCs were introduced by Körner and Marton (1977a), who provided operational definitions of these classes and established the capacity region for the less noisy case. The capacity region for the more capable case was established by El Gamal (1979). Nair (2010) established the classification the BSC–BEC broadcast channel and showed that superposition coding is optimal for all classes. The capacity region of the 3-receiver less noisy DM-BC is due to Wang and Nair (2010), who showed that superposition coding is optimal for this case as well. Surveys of the literature on the broadcast channel can be found in van der Meulen (1977, 1981) and Cover (1998).

PROBLEMS

5.1. Show that the converse for the degraded DM-BC can be proved with the auxiliary random variable identification $U_i = (M_2, Y_2^{i-1})$ or $U_i = (M_2, Y_1^{i-1}, Y_2^{i-1})$.

5.2. Prove the converse for the capacity region of the Gaussian BC by starting from the single-letter characterization

$$R_1 \le I(X; Y_1|U),$$
$$R_2 \le I(U; Y_2)$$

for some cdf $F(u, x)$ such that $\mathsf{E}(X^2) \le P$.

5.3. Verify that the characterizations of the capacity region for the BS-BC and the Gaussian BC in (5.7) and in Theorem 5.3, respectively, are convex.

5.4. Show that if a DM-BC is degraded, then it is also less noisy.

5.5. Show that if a DM-BC is less noisy, then it is also more capable.

5.6. Given a DM-BC $p(y_1, y_2|x)$, let $D(p(x)) = I(X; Y_1) - I(X; Y_2)$. Show that $D(p(x))$ is concave in $p(x)$ iff Y_1 is less noisy than Y_2.

5.7. Prove the classification of the BSC–BEC broadcast channel in Example 5.4.

5.8. Show that the two characterizations of the capacity region for the more capable DM-BC in Section 5.6 are equivalent. (Hint: One direction is trivial. For the other direction, show that the corner points of these regions are the same.)

5.9. *Another simple outer bound.* Consider the DM-BC with three messages.

(a) Show that if a rate triple (R_0, R_1, R_2) is achievable, then it must satisfy the inequalities

$$R_0 + R_1 \le I(X; Y_1),$$
$$R_0 + R_2 \le I(X; Y_2)$$

for some pmf $p(x)$.

(b) Show that this outer bound on the capacity region can be strictly tighter than the simple outer bound in Figure 5.2. (Hint: Consider two antisymmetric Z channels.)

(c) Set $R_1 = R_2 = 0$ in the outer bound in part (a) to show that the *common-message capacity* is

$$C_0 = \max_{p(x)} \min\{I(X; Y_1),\ I(X; Y_2)\}.$$

Argue that C_0 is in general strictly smaller than $\min\{C_1, C_2\}$.

5.10. *Binary erasure broadcast channel.* Consider a DM-BC $p(y_1, y_2|x)$ where the channel from X to Y_1 is a BEC(p_1) and the channel from X to Y_2 is a BEC(p_2) with $p_1 \le p_2$. Find the capacity region in terms of p_1 and p_2.

5.11. *Product of two degraded broadcast channels.* Consider two degraded DM-BCs $p(y_{11}|x_1)p(y_{21}|y_{11})$ and $p(y_{12}|x_2)p(y_{22}|y_{12})$. The product of these two degraded DM-BCs depicted in Figure 5.10 is a DM-BC with $X = (X_1, X_2)$, $Y_1 = (Y_{11}, Y_{12})$, $Y_2 = (Y_{21}, Y_{22})$, and $p(y_1, y_2|x) = p(y_{11}|x_1)p(y_{21}|y_{11})p(y_{12}|x_2)p(y_{22}|y_{12})$. Show that the private-message capacity region of the product DM-BC is the set of rate pairs (R_1, R_2) such that

$$R_1 \le I(X_1; Y_{11}|U_1) + I(X_2; Y_{12}|U_2),$$
$$R_2 \le I(U_1; Y_{21}) + I(U_2; Y_{22})$$

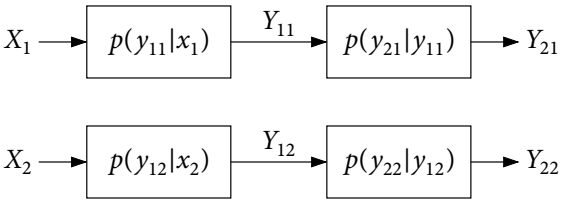

Figure 5.10. Product of two degraded broadcast channels.

for some pmf $p(u_1, x_1)p(u_2, x_2)$. Thus, the capacity region is the Minkowski sum of the capacity regions of the two component DM-BCs.
Remark: This result is due to Poltyrev (1977).

5.12. *Product of two Gaussian broadcast channels.* Consider two Gaussian BCs

$$\begin{aligned} Y_{1j} &= X_j + Z_{1j}, \\ Y_{2j} &= Y_{1j} + \tilde{Z}_{2j}, \quad j = 1, 2, \end{aligned}$$

where $Z_{11} \sim \mathrm{N}(0, N_{11})$, $Z_{12} \sim \mathrm{N}(0, N_{12})$, $\tilde{Z}_{21} \sim \mathrm{N}(0, \tilde{N}_{21})$, and $\tilde{Z}_{22} \sim \mathrm{N}(0, \tilde{N}_{22})$ are independent noise components. Assume the average transmission power constraint

$$\sum_{i=1}^{n} \big(x_{1i}^2(m_1, m_2) + x_{2i}^2(m_1, m_2)\big) \le nP.$$

Find the private-message capacity region of the product Gaussian BC with $X = (X_1, X_2)$, $Y_1 = (Y_{11}, Y_{12})$, and $Y_2 = (Y_{21}, Y_{22})$, in terms of the noise powers, P, and power allocation parameters $\alpha_1, \alpha_2 \in [0, 1]$ for each subchannel and $\beta \in [0, 1]$ between two channels.
Remark: This result is due to Hughes-Hartogs (1975).

5.13. *Minimum-energy-per-bit region.* Consider the Gaussian BC with noise powers N_1 and N_2. Find the minimum-energy-per-bit region, that is, the set of all energy pairs $(E_1, E_2) = (P/R_1, P/R_2)$ such that the rate pair (R_1, R_2) is achievable with average code power P.

5.14. *Reversely degraded broadcast channels with common message.* Consider two reversely degraded DM-BCs $p(y_{11}|x_1)p(y_{21}|y_{11})$ and $p(y_{22}|x_2)p(y_{12}|y_{22})$. The product of these two degraded DM-BCs depicted in Figure 5.11 is a DM-BC with $X = (X_1, X_2)$, $Y_1 = (Y_{11}, Y_{12})$, $Y_2 = (Y_{21}, Y_{22})$, and $p(y_1, y_2|x) = p(y_{11}|x_1)p(y_{21}|y_{11}) \cdot p(y_{22}|x_2)p(y_{12}|y_{22})$. A common message $M_0 \in [1 : 2^{nR_0}]$ is to be communicated to both receivers. Show that the common-message capacity is

$$C_0 = \max_{p(x_1)p(x_2)} \min\{I(X_1; Y_{11}) + I(X_2; Y_{12}),\ I(X_1; Y_{21}) + I(X_2; Y_{22})\}.$$

Remark: This channel is studied in more detail in Section 9.4.

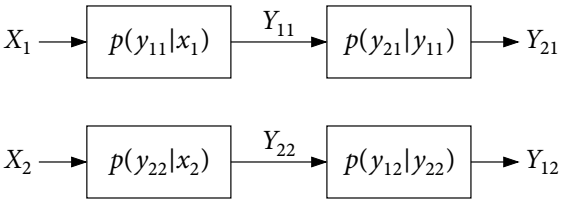

Figure 5.11. Product of reversely degraded broadcast channels.

5.15. *Duality between Gaussian broadcast and multiple access channels.* Consider the following Gaussian BC and Gaussian MAC:

- Gaussian BC: $Y_1 = g_1 X + Z_1$ and $Y_2 = g_2 X + Z_2$, where $Z_1 \sim \mathrm{N}(0, 1)$ and $Z_2 \sim \mathrm{N}(0, 1)$. Assume average power constraint P on X.
- Gaussian MAC: $Y = g_1 X_1 + g_2 X_2 + Z$, where $Z \sim \mathrm{N}(0, 1)$. Assume the average sum-power constraint

$$\sum_{i=1}^{n} \left(x_{1i}^2(m_1) + x_{2i}^2(m_2)\right) \le nP, \quad (m_1, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}].$$

(a) Characterize the (private-message) capacity regions of these two channels in terms of P, g_1, g_2, and power allocation parameter $\alpha \in [0, 1]$.

(b) Show that the two capacity regions are equal.

(c) Show that every point (R_1, R_2) on the boundary of the capacity region of the above Gaussian MAC is achievable using random coding and successive cancellation decoding. That is, time sharing is not needed in this case.

(d) Argue that the sequence of codes that achieves the rate pairs (R_1, R_2) on the boundary of the Gaussian MAC capacity region can be used to achieve the same point on the capacity region of the above Gaussian BC.

Remark: This result is a special case of a general duality result between the Gaussian vector BC and MAC presented in Chapter 9.

5.16. *k-receiver Gaussian BC.* Consider the k-receiver Gaussian BC

$$\begin{aligned} Y_1 &= X + Z_1, \\ Y_j &= Y_{j-1} + \tilde{Z}_j, \quad j \in [2 : k], \end{aligned}$$

where $Z_1, \tilde{Z}_2, \ldots, \tilde{Z}_k$ are independent Gaussian noise components with powers $N_1, \tilde{N}_2, \ldots, \tilde{N}_k$, respectively. Assume average power constraint P on X. Provide a characterization of the private-message capacity region in terms of the noise powers, P, and power allocation parameters $\alpha_1, \ldots, \alpha_k \ge 0$ with $\sum_{j=1}^{k} \alpha_j = 1$.

5.17. *Three-receiver less noisy BC.* Consider the 3-receiver DM-BC $p(y_1, y_2, y_3|x)$ such that $I(U; Y_1) \ge I(U; Y_2) \ge I(U; Y_3)$ for all $p(u, x)$.

(a) Suppose that $W \to X^n \to (Y_1^n, Y_2^n)$ form a Markov chain. Show that

$$I(Y_2^{i-1}; Y_{1i}|W) \le I(Y_1^{i-1}; Y_{1i}|W), \quad i \in [2:n].$$

(b) Suppose that $W \to X^n \to (Y_2^n, Y_3^n)$ form a Markov chain. Show that

$$I(Y_3^{i-1}; Y_{3i}|W) \le I(Y_2^{i-1}; Y_{3i}|W), \quad i \in [2:n].$$

(c) Using superposition coding and parts (a) and (b), show that the capacity region is the set of rate triples (R_1, R_2, R_3) such that

$$\begin{aligned} R_1 &\le I(X; Y_1|U, V), \\ R_2 &\le I(V; Y_2|U), \\ R_3 &\le I(U; Y_3) \end{aligned}$$

for some pmf $p(u, v, x)$. (Hint: Identify $U_i = (M_3, Y_2^{i-1})$ and $V_i = M_2$.)

Remark: This result is due to Wang and Nair (2010).

5.18. *MAC with degraded message sets.* Consider a DM-MAC $p(y|x_1, x_2)$ with message pair (M_0, M_1) uniformly distributed over $[1 : 2^{nR_0}] \times [1 : 2^{nR_1}]$. Sender 1 encodes (M_0, M_1), while sender 2 encodes only M_1. The receiver wishes to recover both messages. The probability of error, achievability, and the capacity region are defined as for the DM-MAC with private messages.

(a) Show that the capacity region is the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2), \\ R_0 + R_1 &\le I(X_1, X_2; Y) \end{aligned}$$

for some pmf $p(x_1, x_2)$.

(b) Characterize the capacity region of the Gaussian MAC with noise power 1, channel gains g_1 and g_2, and average power constraint P on each of X_1 and X_2.

5.19. *MAC with common message.* Consider a DM-MAC $p(y|x_1, x_2)$ with message triple (M_0, M_1, M_2) uniformly distributed over $[1 : 2^{nR_0}] \times [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. Sender 1 encodes (M_0, M_1) and sender 2 encodes (M_0, M_2). Thus, the common message M_0 is available to both senders, while the private messages M_1 and M_2 are available only to the respective senders. The receiver wishes to recover all three messages. The probability of error, achievability, and the capacity region are defined as for the DM-MAC with private messages.

(a) Show that the capacity region is the set of rate triples (R_0, R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X_1; Y|U, X_2), \\ R_2 &\le I(X_2; Y|U, X_1), \\ R_1 + R_2 &\le I(X_1, X_2; Y|U), \\ R_0 + R_1 + R_2 &\le I(X_1, X_2; Y) \end{aligned}$$

for some pmf $p(u)p(x_1|u)p(x_2|u)$.

(b) Show that the capacity region of the Gaussian MAC with average power constraint P on each of X_1 and X_2 is the set of rate triples (R_0, R_1, R_2) such that

$$\begin{aligned} R_1 &\le \mathsf{C}(\alpha_1 S_1), \\ R_2 &\le \mathsf{C}(\alpha_2 S_2), \\ R_1 + R_2 &\le \mathsf{C}(a_1 S_1 + \alpha_2 S_2), \\ R_0 + R_1 + R_2 &\le \mathsf{C}\Big(S_1 + S_2 + 2\sqrt{\bar{\alpha}_1 \bar{\alpha}_2 S_1 S_2}\Big) \end{aligned}$$

for some $\alpha_1, \alpha_2 \in [0, 1]$.

Remark: The capacity region of the DM-MAC with common message is due to Slepian and Wolf (1973b). The converse for the Gaussian case is due to Bross, Lapidoth, and Wigger (2008).

5.20. *MAC with a helper.* Consider the 3-sender DM-MAC $p(y|x_1, x_2, x_3)$ with message pair (M_1, M_2) uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ as depicted in Figure 5.12. Sender 1 encodes M_1, sender 2 encodes M_2, and sender 3 encodes (M_1, M_2). The receiver wishes to recover both messages. Show that the capacity region is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X_1, X_3; Y | X_2, Q), \\ R_2 &\le I(X_2, X_3; Y | X_1, Q), \\ R_1 + R_2 &\le I(X_1, X_2, X_3; Y | Q) \end{aligned}$$

for some pmf $p(q)p(x_1|q)p(x_2|q)p(x_3|x_1, x_2, q)$ with $|\mathcal{Q}| \le 2$.

Remark: This is a special case of the capacity region of the general l-message k-sender DM-MAC established by Han (1979).

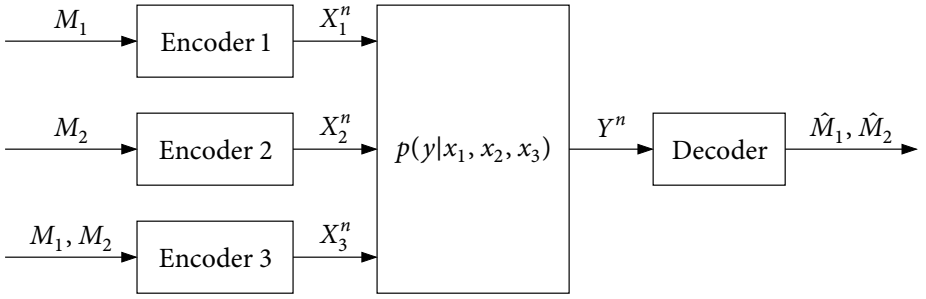

Figure 5.12. Three-sender DM-MAC with two messages.

CHAPTER 8

General Broadcast Channels

We resume the discussion of broadcast channels started in Chapter 5. Again consider the 2-receiver DM-BC $p(y_1, y_2|x)$ with private and common messages depicted in Figure 8.1. The definitions of a code, achievability, and capacity regions are the same as in Chapter 5. As mentioned before, the capacity region of the DM-BC is not known in general. In Chapter 5, we presented the superposition coding scheme and showed that it is optimal for several classes of channels in which one receiver is stronger than the other. In this chapter, we study coding schemes that can outperform superposition coding and present the tightest known inner and outer bounds on the capacity region of the general broadcast channel.

We first show that superposition coding is optimal for the 2-receiver DM-BC with degraded message sets, that is, when either $R_1 = 0$ or $R_2 = 0$. We then show that superposition coding is not optimal for BCs with more than two receivers. In particular, we establish the capacity region of the 3-receiver multilevel BC. The achievability proof involves the new idea of indirect decoding, whereby a receiver who wishes to recover only the common message still uses satellite codewords in decoding for the cloud center.

We then present Marton's inner bound on the private-message capacity region of the 2-receiver DM-BC and show that it is optimal for the class of semideterministic BCs. The coding scheme involves the multicoding technique introduced in Chapter 7 and the new idea of joint typicality codebook generation to construct dependent codewords for independent messages without the use of a superposition structure. The proof of the inner bound uses the mutual covering lemma, which is a generalization of the covering lemma in Section 3.7. Marton's coding scheme is then combined with superposition coding to establish an inner bound on the capacity region of the DM-BC that is tight for all classes of DM-BCs with known capacity regions. Next, we establish the Nair–El Gamal outer bound on the capacity region of the DM-BC. We show through an example that there is

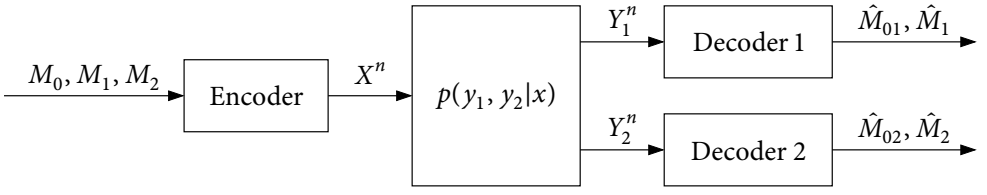

Figure 8.1. Two-receiver broadcast communication system.

a gap between these inner and outer bounds. Finally, we discuss extensions of the aforementioned coding techniques to broadcast channels with more than two receivers and with arbitrary messaging requirements.

8.1 DM-BC WITH DEGRADED MESSAGE SETS

A broadcast channel with *degraded message sets* is a model for a *layered* communication setup in which a sender wishes to communicate a common message to *all* receivers, a first private message to a first subset of the receivers, a second private message to a second subset of the first subset, and so on. Such a setup arises, for example, in video or music broadcasting over a wireless network at varying levels of quality. The common message represents the lowest quality description of the data to be sent to all receivers, and each private message represents a refinement over the common message and previous private messages.

We first consider the 2-receiver DM-BC with degraded message sets, i.e., with $R_2 = 0$. The capacity region for this case is known.

Theorem 8.1. The capacity region of the 2-receiver DM-BC $p(y_1, y_2|x)$ with degraded message sets is the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le I(U; Y_2), \\ R_1 &\le I(X; Y_1|U), \\ R_0 + R_1 &\le I(X; Y_1) \end{aligned}$$

for some pmf $p(u, x)$ with $|\mathcal{U}| \le \min\{|\mathcal{X}|,\ |\mathcal{Y}_1| + |\mathcal{Y}_2|\} + 1$.

This capacity region is achieved using superposition coding (see Section 5.3) by noting that receiver 1 can recover both messages M_0 and M_1. The converse proof follows similar lines to that for the more capable BC in Section 5.6. It is proved by first considering the alternative characterization of the capacity region consisting of all rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le I(U; Y_2), \\ R_0 + R_1 &\le I(X; Y_1), \\ R_0 + R_1 &\le I(X; Y_1|U) + I(U; Y_2) \end{aligned} \tag{8.1}$$

for some pmf $p(u, x)$. The Csiszár sum identity and other standard techniques are then used to complete the proof. The cardinality bound on $\mathcal{U}$ can be proved using the convex cover method in Appendix C.

It turns out that we can prove achievability for the alternative region in (8.1) directly. The proof involves an (unnecessary) *rate splitting* step. We divide the private message M_1 into two independent messages M_{10} at rate R_{10} and M_{11} at rate R_{11}. We use superposition coding whereby the cloud center U represents the message pair (M_0, M_{10}) and

the satellite codeword X represents the message triple (M_0, M_{10}, M_{11}). Following similar steps to the achievability proof of the superposition coding inner bound, we can show that (R_0, R_{10}, R_{11}) is achievable if

$$\begin{aligned} R_0 + R_{10} &< I(U; Y_2), \\ R_{11} &< I(X; Y_1|U), \\ R_0 + R_1 &< I(X; Y_1) \end{aligned}$$

for some pmf $p(u, x)$. Substituting $R_{11} = R_1 - R_{10}$, we have the conditions

$$\begin{aligned} R_{10} &\geq 0, \\ R_{10} &< I(U; Y_2) - R_0, \\ R_{10} &> R_1 - I(X; Y_1|U), \\ R_0 + R_1 &< I(X; Y_1), \\ R_{10} &\leq R_1. \end{aligned}$$

Eliminating R_{10} by the Fourier–Motzkin procedure in Appendix D yields the desired characterization.

Rate splitting turns out to be crucial when the DM-BC has more than 2 receivers.

8.2 THREE-RECEIVER MULTILEVEL DM-BC

The degraded message set capacity region of the DM-BC with more than two receivers is not known in general. We show that the straightforward extension of the superposition coding inner bound in Section 5.3 to more than two receivers is not optimal in general.

Consider the 3-receiver *multilevel* DM-BC $(\mathcal{X}, p(y_1, y_3|x)p(y_2|y_1), \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3)$ depicted in Figure 8.2, which is a 3-receiver DM-BC where Y_2 is a degraded version of Y_1. We consider the case of two degraded message sets, where a common message $M_0 \in [1 : 2^{nR_0}]$ is to be communicated to all receivers, and a private message $M_1 \in [1 : 2^{nR_1}]$ is to be communicated only to receiver 1. The definitions of a code, probability of error, achievability, and capacity region for this setting are as before.

A straightforward extension of the superposition coding inner bound to the 3-receiver multilevel DM-BC, where receivers 2 and 3 decode for the cloud center and receiver 1

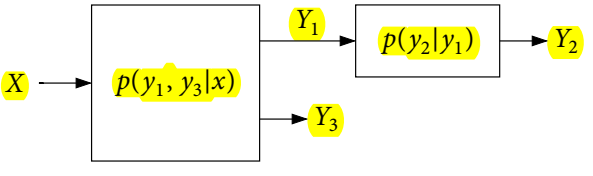

Figure 8.2. Three-receiver multilevel broadcast channel.

decodes for the satellite codeword, gives the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &< \min\{I(U;Y_2), I(U;Y_3)\}, \\ R_1 &< I(X;Y_1|U) \end{aligned} \tag{8.2}$$

for some pmf $p(u,x)$. Note that the last inequality $R_0 + R_1 < I(X;Y_1)$ in the superposition coding inner bound in Theorem 5.1 drops out by the assumption that Y_2 is a degraded version of Y_1.

This region turns out not to be optimal in general.

Theorem 8.2. The capacity region of the 3-receiver multilevel DM-BC $p(y_1, y_3|x)$ $p(y_2|y_1)$ is the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le \min\{I(U;Y_2), I(V;Y_3)\}, \\ R_1 &\le I(X;Y_1|U), \\ R_0 + R_1 &\le I(V;Y_3) + I(X;Y_1|V) \end{aligned}$$

for some pmf $p(u,v)p(x|v)$ with $|\mathcal{U}| \le |\mathcal{X}| + 4$ and $|\mathcal{V}| \le (|\mathcal{X}| + 1)(|\mathcal{X}| + 4)$.

The proof of the converse uses steps from the converse proofs for the degraded BC and the 2-receiver BC with degraded message sets. The cardinality bounds on $\mathcal{U}$ and $\mathcal{V}$ can be proved using the extension of the convex cover method to multiple random variables in Appendix C.

8.2.1 Proof of Achievability

The achievability proof of Theorem 8.2 uses the new idea of *indirect decoding*. We divide the private message M_1 into two messages M_{10} and M_{11}. We generate a codebook of u^n sequences for the common message M_0, and use superposition coding to generate a codebook of v^n sequences for the message pair (M_0, M_{10}). We then use superposition coding again to generate a codebook of x^n sequences for the message triple $(M_0, M_{10}, M_{11}) = (M_0, M_1)$. Receiver 1 finds (M_0, M_1) by decoding for x^n, receiver 2 finds M_0 by decoding for u^n, and receiver 3 finds M_0 indirectly by simultaneous decoding for (u^n, v^n). We now provide details of the proof.

Rate splitting. Divide the private message M_1 into two independent messages M_{10} at rate R_{10} and M_{11} at rate R_{11}. Hence $R_1 = R_{10} + R_{11}$.

Codebook generation. Fix a pmf $p(u,v)p(x|v)$. Randomly and independently generate 2^{nR_0} sequences $u^n(m_0)$, $m_0 \in [1 : 2^{nR_0}]$, each according to $\prod_{i=1}^n p_U(u_i)$. For each m_0, randomly and conditionally independently generate $2^{nR_{10}}$ sequences $v^n(m_0, m_{10})$, $m_{10} \in [1 : 2^{nR_{10}}]$, each according to $\prod_{i=1}^n p_{V|U}(v_i|u_i(m_0))$. For each pair (m_0, m_{10}), randomly and conditionally independently generate $2^{nR_{11}}$ sequences $x^n(m_0, m_{10}, m_{11})$, $m_{11} \in [1 : 2^{nR_{11}}]$, each according to $\prod_{i=1}^n p_{X|V}(x_i|v_i(m_0, m_{10}))$.

Encoding. To send the message pair $(m_0, m_1) = (m_0, m_{10}, m_{11})$, the encoder transmits $x^n(m_0, m_{10}, m_{11})$.

Decoding and analysis of the probability of error for decoders 1 and 2. Decoder 2 declares that $\hat{m}_{02} \in [1 : 2^{nR_0}]$ is sent if it is the unique message such that $(u^n(\hat{m}_{02}), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and the packing lemma, the probability of error tends to zero as $n \to \infty$ if $R_0 < I(U; Y_2) - \delta(\epsilon)$. Decoder 1 declares that $(\hat{m}_{01}, \hat{m}_{10}, \hat{m}_{11})$ is sent if it is the unique message triple such that $(u^n(\hat{m}_{01}), v^n(\hat{m}_{01}, \hat{m}_{10}), x^n(\hat{m}_{01}, \hat{m}_{10}, \hat{m}_{11}), y_1^n) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and the packing lemma, the probability of error tends to zero as $n \to \infty$ if

$$\begin{aligned} R_{11} &< I(X; Y_1|V) - \delta(\epsilon), \\ R_{10} + R_{11} &< I(X; Y_1|U) - \delta(\epsilon), \\ R_0 + R_{10} + R_{11} &< I(X; Y_1) - \delta(\epsilon). \end{aligned}$$

Decoding and analysis of the probability of error for decoder 3. If receiver 3 decodes for m_0 directly by finding the unique message $\hat{m}_{03}$ such that $(u^n(\hat{m}_{03}), y_3^n) \in \mathcal{T}_\epsilon^{(n)}$, we obtain the condition $R_0 < I(U; Y_3) - \delta(\epsilon)$, which together with the previous conditions gives the extended superposition coding inner bound in (8.2).

To achieve the capacity region, receiver 3 decodes for m_0 *indirectly*. It declares that $\hat{m}_{03}$ is sent if it is the unique message such that $(u^n(\hat{m}_{03}), v^n(\hat{m}_{03}, m_{10}), y_3^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $m_{10} \in [1 : 2^{nR_{10}}]$. Assume that $(M_0, M_{10}) = (1, 1)$ is sent and consider all possible pmfs for the triple $(U^n(m_0), V^n(m_0, m_{10}), Y_3^n)$ as listed in Table 8.1.

m_0	m_{10}	Joint pmf
1	1	$p(u^n, v^n)p(y_3^n\|v^n)$
1	$*$	$p(u^n, v^n)p(y_3^n\|u^n)$
$*$	$*$	$p(u^n, v^n)p(y_3^n)$
$*$	1	$p(u^n, v^n)p(y_3^n)$

Table 8.1. The joint distribution induced by different (m_0, m_{10}) pairs.

The second case does not result in an error and the last two cases have the same pmf. Thus, the decoding error occurs iff one or both of the following events occur:

$$\begin{aligned} \mathcal{E}_{31} &= \{(U^n(1), V^n(1, 1), Y_3^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \\ \mathcal{E}_{32} &= \{(U^n(m_0), V^n(m_0, m_{10}), Y_3^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_0 \neq 1, m_{10}\}. \end{aligned}$$

Then, the probability of error for decoder 3 averaged over codebooks $\mathsf{P}(\mathcal{E}_3) \le \mathsf{P}(\mathcal{E}_{31}) + \mathsf{P}(\mathcal{E}_{32})$. By the LLN, $\mathsf{P}(\mathcal{E}_{31})$ tends to zero as $n \to \infty$. By the packing lemma (with $\mathcal{A} = [2 : 2^{n(R_0+R_{10})}]$, $X \leftarrow (U, V)$, and $U = \emptyset$), $\mathsf{P}(\mathcal{E}_{32})$ tends to zero as $n \to \infty$ if $R_0 + R_{10} < I(U, V; Y_3) - \delta(\epsilon) = I(V; Y_3) - \delta(\epsilon)$. Combining the bounds, substituting $R_{10} + R_{11} = R_1$, and eliminating R_{10} and R_{11} by the Fourier–Motzkin procedure completes the proof of achievability.

Indirect decoding is illustrated in Figures 8.3 and 8.4. Suppose that $R_0 > I(U;Y_3)$ as shown in Figure 8.3. Then receiver 3 cannot decode for the cloud center $u^n(1)$ directly. Now suppose that $R_0 + R_{10} < I(V;Y_3)$ in addition. Then receiver 3 can decode for the cloud center *indirectly* by finding the unique message $\hat{m}_0$ such that $(u^n(\hat{m}_0), v^n(\hat{m}_0, m_{10}), y_3^n) \in \mathcal{T}_\epsilon^{(n)}$ for some m_{10} as shown in Figure 8.4. Note that the condition $R_0 + R_{10} < I(V;Y_3)$ suffices in general (even when $R_0 \le I(U;Y_3)$).

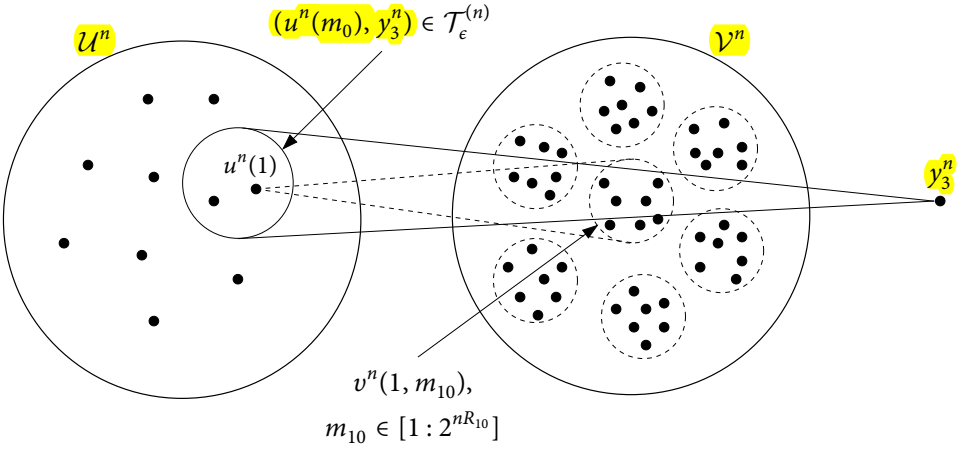

Figure 8.3. Direct decoding for the cloud center $u^n(1)$.

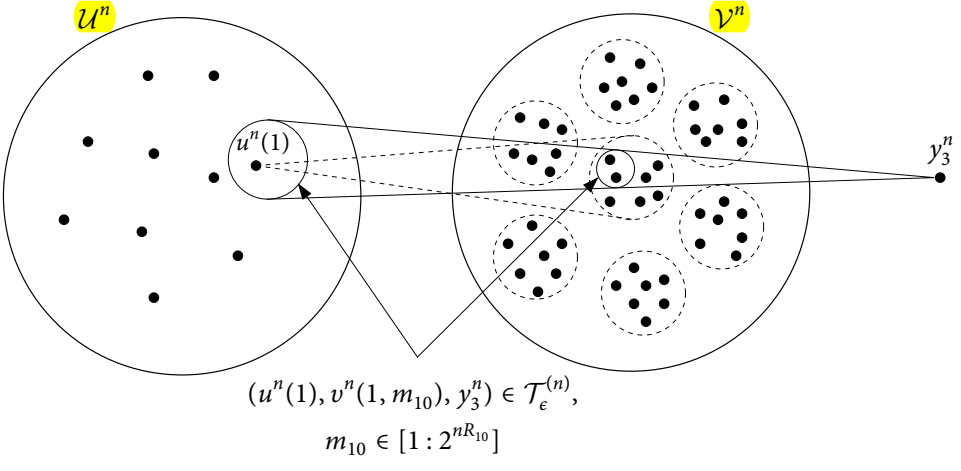

Figure 8.4. Indirect decoding for $u^n(1)$ via $v^n(1, m_{10})$.

Remark 8.1. Although it seems surprising that higher rates can be achieved by having receiver 3 recover more than it needs to, the reason we can do better than superposition coding can be explained by the observation that for a 2-receiver BC $p(y_2, y_3|x)$, the conditions $I(U;Y_2) < I(U;Y_3)$ and $I(V;Y_2) > I(V;Y_3)$ can hold simultaneously for

some $U \to V \to X$; see discussions on less noisy and more capable BCs in Section 5.6. Now, considering our 3-receiver BC scenario, suppose we have a choice of U such that $I(U;Y_3) \le I(U;Y_2)$. In this case, requiring receivers 2 and 3 to directly decode for u^n necessitates that $R_0 \le I(U;Y_3)$. From the above observation, a V may exist such that $U \to V \to X$ and $I(V;Y_3) > I(V;Y_2)$, in which case the rate of the common message can be increased to $I(U;Y_2)$ and receiver 3 can still find u^n indirectly by decoding for (u^n, v^n). Thus, the additional "degree-of-freedom" introduced by V helps increase the rates in spite of the fact that receiver 3 may need to recover more than just the common message.

Remark 8.2. If we require receiver 3 to recover M_{10}, we obtain the region consisting of all rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le \min\{I(U;Y_2), I(V;Y_3)\}, \\ R_1 &\le I(X;Y_1|U), \\ R_0 + R_1 &\le I(V;Y_3) + I(X;Y_1|V), \\ R_1 &\le I(V;Y_3|U) + I(X;Y_1|V) \end{aligned} \tag{8.3}$$

for some pmf $p(u)p(v|u)p(x|v)$. While this region involves one more inequality than the capacity region in Theorem 8.2, it can be shown by optimizing the choice of V for each given U that it coincides with the capacity region. However, requiring decoder 3 to recover M_{10} is unnecessary and leads to a region with more inequalities for which the converse is difficult to establish.

8.2.2 Multilevel Product DM-BC

We show via an example that the extended superposition coding inner bound in (8.2) can be strictly smaller than the capacity region in Theorem 8.2 for the 3-receiver multilevel DM-BC. Consider the product of two 3-receiver BCs specified by the Markov chains

$$\begin{aligned} X_1 &\to Y_{31} \to Y_{11} \to Y_{21}, \\ X_2 &\to Y_{12} \to Y_{22}. \end{aligned}$$

For this channel, the extended superposition coding inner bound in (8.2) reduces to the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le I(U_1;Y_{21}) + I(U_2;Y_{22}), \\ R_0 &\le I(U_1;Y_{31}), \\ R_1 &\le I(X_1;Y_{11}|U_1) + I(X_2;Y_{12}|U_2) \end{aligned} \tag{8.4}$$

for some pmf $p(u_1, x_1)p(u_2, x_2)$. Similarly, it can be shown that the capacity region in Theorem 8.2 reduces to the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le I(U_1;Y_{21}) + I(U_2;Y_{22}), \\ R_0 &\le I(V_1;Y_{31}), \\ R_1 &\le I(X_1;Y_{11}|U_1) + I(X_2;Y_{12}|U_2), \\ R_0 + R_1 &\le I(V_1;Y_{31}) + I(X_1;Y_{11}|V_1) + I(X_2;Y_{12}|U_2) \end{aligned} \tag{8.5}$$

for some pmf $p(u_1, v_1)p(x_1|v_1)p(u_2, x_2)$.

In the following, we compare the extended superposition coding inner bound in (8.4) to the capacity region in (8.5).

Example 8.1. Consider the multilevel product DM-BC depicted in Figure 8.5, where X_1, X_2, Y_{12}, and Y_{31} are binary, and Y_{11}, Y_{21}, $Y_{22} \in \{0, 1, \mathrm{e}\}$.

The extended superposition coding inner bound in (8.4) can be simplified to the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le \min\left\{\frac{\alpha}{6} + \frac{\beta}{2}, \alpha\right\}, \\ R_1 &\le \frac{\bar{\alpha}}{2} + \bar{\beta} \end{aligned} \tag{8.6}$$

for some $\alpha, \beta \in [0, 1]$. It is straightforward to show that $(R_0, R_1) = (1/2, 5/12)$ lies on the boundary of this region. By contrast, the capacity region in (8.5) can be simplified to the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le \min\left\{\frac{r}{6} + \frac{s}{2}, t\right\}, \\ R_1 &\le \frac{1-r}{2} + 1 - s, \\ R_0 + R_1 &\le t + \frac{1-t}{2} + 1 - s \end{aligned} \tag{8.7}$$

for some $0 \le r \le t \le 1$, $0 \le s \le 1$. Note that substituting $r = t$ in (8.7) yields the extended superposition coding inner bound in (8.6). By setting $r = 0$, $s = 1$, $t = 1$, it can be readily checked that $(R_0, R_1) = (1/2, 1/2)$ lies on the boundary of the capacity region. For $R_0 = 1/2$, however, the maximum achievable R_1 in the extended superposition coding inner bound in (8.6) is 5/12. Thus the capacity region is strictly larger than the extended superposition coding inner bound.

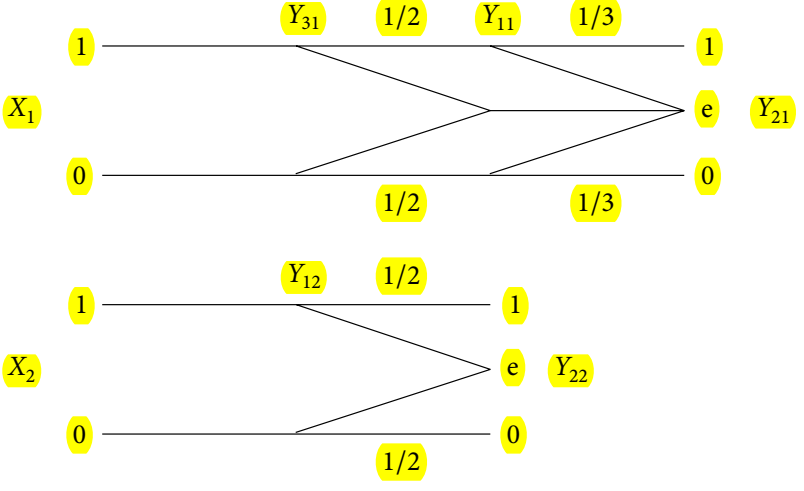

Figure 8.5. Binary erasure multilevel product DM-BC.

8.3 MARTON'S INNER BOUND

We now turn our attention to the 2-receiver DM-BC with only private messages, i.e., when $R_0 = 0$. First we show that a rate pair (R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$ if

$$\begin{aligned} R_1 &< I(U_1; Y_1), \\ R_2 &< I(U_2; Y_2) \end{aligned} \tag{8.8}$$

for some pmf $p(u_1)p(u_2)$ and function $x(u_1, u_2)$. The proof of achievability for this inner bound is straightforward.

Codebook generation. Fix a pmf $p(u_1)p(u_2)$ and $x(u_1, u_2)$. Randomly and independently generate 2^{nR_1} sequences $u_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, each according to $\prod_{i=1}^n p_{U_1}(u_{1i})$, and 2^{nR_2} sequences $u_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, each according to $\prod_{i=1}^n p_{U_2}(u_{2i})$.

Encoding. To send (m_1, m_2), transmit $x_i(u_{1i}(m_1), u_{2i}(m_2))$ at time $i \in [1 : n]$.

Decoding and analysis of the probability of error. Decoder $j = 1, 2$ declares that $\hat{m}_j$ is sent if it is the unique message such that $(U_j^n(\hat{m}_j), Y_j^n) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and the packing lemma, the probability of decoding error tends to zero as $n \to \infty$ if $R_j < I(U_j; Y_j) - \delta(\epsilon)$ for $j = 1, 2$. This completes the proof of achievability.

Marton's inner bound allows U_1, U_2 in (8.8) to be arbitrarily dependent (even though the messages themselves are independent). This comes at an apparent penalty term in the sum-rate.

Theorem 8.3 (Marton's Inner Bound). A rate pair (R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$ if

$$\begin{aligned} R_1 &< I(U_1; Y_1), \\ R_2 &< I(U_2; Y_2), \\ R_1 + R_2 &< I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) \end{aligned}$$

for some pmf $p(u_1, u_2)$ and function $x(u_1, u_2)$.

Before proving the theorem, we make a few remarks and show that Marton's inner bound is tight for the semideterministic BC.

Remark 8.3. Note that Marton's inner bound reduces to the inner bound in (8.8) if U_1 and U_2 are restricted to be independent.

Remark 8.4. As for the Gelfand–Pinsker theorem in Section 7.6, Marton's inner bound does not become larger if we evaluate it using general conditional pmfs $p(x|u_1, u_2)$. To show this, fix a pmf $p(u_1, u_2, x)$. By the functional representation lemma in Appendix B, there exists a random variable V independent of (U_1, U_2) such that X is a function of

(U_1, U_2, V). Now defining $U_1' = (U_1, V)$, we have

$$I(U_1; Y_1) \le I(U_1'; Y_1),$$
$$I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) \le I(U_1'; Y_1) + I(U_2; Y_2) - I(U_1'; U_2),$$

and X is a function of (U_1', U_2). Thus there is no loss of generality in restricting X to be a deterministic function of (U_1, U_2).

Remark 8.5. Marton's inner bound is not convex in general, but can be readily convexified via a time-sharing random variable Q.

8.3.1 Semideterministic DM-BC

Marton's inner bound is tight for the class of *semideterministic* BCs for which Y_1 is a function of X, i.e., $Y_1 = y_1(X)$. Hence, we can set the auxiliary random variable $U_1 = Y_1$, which simplifies Marton's inner bound to the set of rate pairs (R_1, R_2) such that

$$R_1 \le H(Y_1),$$
$$R_2 \le I(U; Y_2),$$
$$R_1 + R_2 \le H(Y_1|U) + I(U; Y_2)$$

for some pmf $p(u, x)$. This region turns out to be the capacity region. The converse follows by the general outer bound on the capacity region of the broadcast channel presented in Section 8.5.

For the special case of fully deterministic DM-BCs, where $Y_1 = y_1(X)$ and $Y_2 = y_2(X)$, the capacity region further simplifies to the set of rate pairs (R_1, R_2) such that

$$R_1 \le H(Y_1),$$
$$R_2 \le H(Y_2),$$
$$R_1 + R_2 \le H(Y_1, Y_2)$$

for some pmf $p(x)$. This region is evaluated for the following simple channel.

Example 8.2 (Blackwell channel). Consider the deterministic BC depicted in Figure 8.6. The capacity region, which is plotted in Figure 8.7, is the union of the two regions

$$\mathscr{R}_1 = \{(R_1, R_2)\colon R_1 \le H(\alpha),\ R_2 \le H(\alpha/2),\ R_1 + R_2 \le H(\alpha) + \alpha \text{ for } \alpha \in [1/2, 2/3]\},$$
$$\mathscr{R}_2 = \{(R_1, R_2)\colon R_1 \le H(\alpha/2),\ R_2 \le H(\alpha),\ R_1 + R_2 \le H(\alpha) + \alpha \text{ for } \alpha \in [1/2, 2/3]\}.$$

The first region is attained by setting $p_X(0) = p_X(2) = \alpha/2$ and $p_X(1) = \bar{\alpha}$, while the second region is attained by setting $p_X(0) = \bar{\alpha}$ and $p_X(1) = p_X(2) = \alpha/2$. It can be shown that this capacity region is strictly larger than both the superposition coding inner bound and the inner bound in (8.8).

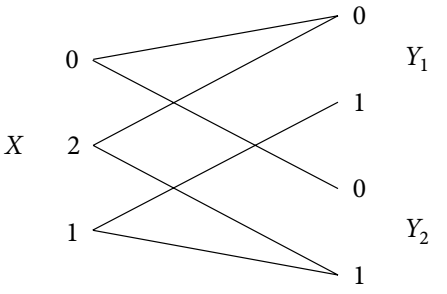

Figure 8.6. Blackwell channel. When $X = 0$, $Y_1 = Y_2 = 0$. When $X = 1$, $Y_1 = Y_2 = 1$. However, when $X = 2$, $Y_1 = 0$ while $Y_2 = 1$.

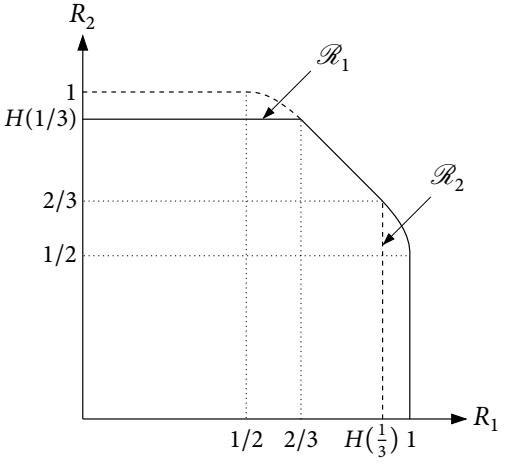

Figure 8.7. The capacity region of the Blackwell channel: $\mathscr{C} = \mathscr{R}_1 \cup \mathscr{R}_2$. Note that the sum-rate $H(1/3) + 2/3 = \log 3$ is achievable.

8.3.2 Achievability of Marton's Inner Bound

The proof of achievability uses multicoding and joint typicality codebook generation. The idea is illustrated in Figure 8.8. For each message m_j, $j = 1, 2$, we generate a subcodebook $\mathcal{C}_j(m_j)$ consisting of independently generated u_j^n sequence. For each message pair (m_1, m_2), we find a jointly typical sequence pair (u_1^n, u_2^n) in the product subcodebook $\mathcal{C}_1(m_1) \times \mathcal{C}_2(m_2)$. Note that although these codeword pairs are dependent, they represent independent messages! To send (m_1, m_2), we transmit a symbol-by-symbol function $x(u_{1i}, u_{2i})$, $i \in [1 : n]$, of the selected sequence pair (u_1^n, u_2^n). Each receiver decodes for its intended codeword using joint typicality decoding.

A crucial requirement for Marton's coding scheme to succeed is the existence of at least one jointly typical sequence pair (u_1^n, u_2^n) in the chosen product subcodebook $\mathcal{C}_1(m_1) \times \mathcal{C}_2(m_2)$. A sufficient condition on the subcodebook sizes to guarantee the existence of such a sequence pair is provided in the following.

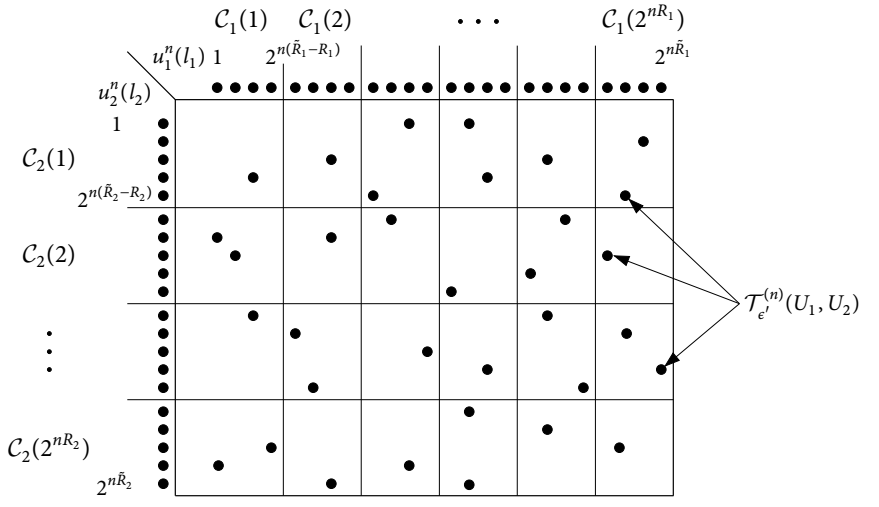

Figure 8.8. Marton's coding scheme.

Lemma 8.1 (Mutual Covering Lemma). Let $(U_0, U_1, U_2) \sim p(u_0, u_1, u_2)$ and $\epsilon' < \epsilon$. Let $U_0^n \sim p(u_0^n)$ be a random sequence such that $\lim_{n\to\infty} \mathsf{P}\{U_0^n \in \mathcal{T}_{\epsilon'}^{(n)}\} = 1$. Let $U_1^n(m_1)$, $m_1 \in [1 : 2^{nr_1}]$, be pairwise conditionally independent random sequences, each distributed according to $\prod_{i=1}^n p_{U_1|U_0}(u_{1i}|u_{0i})$. Similarly, let $U_2^n(m_2)$, $m_2 \in [1 : 2^{nr_2}]$, be pairwise conditionally independent random sequences, each distributed according to $\prod_{i=1}^n p_{U_2|U_0}(u_{2i}|u_{0i})$. Assume that $(U_1^n(m_1)\colon\ m_1 \in [1 : 2^{nr_1}])$ and $(U_2^n(m_2)\colon\ m_2 \in [1 : 2^{nr_2}])$ are conditionally independent given U_0^n. Then, there exists $\delta(\epsilon)$ that tends to zero as $\epsilon \to 0$ such that

$$\lim_{n\to\infty} \mathsf{P}\{(U_0^n, U_1^n(m_1), U_2^n(m_2)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } (m_1, m_2) \in [1 : 2^{nr_1}] \times [1 : 2^{nr_2}]\} = 0,$$

if $r_1 + r_2 > I(U_1; U_2|U_0) + \delta(\epsilon)$.

The proof of this lemma is given in Appendix 8A. Note that the mutual covering lemma extends the covering lemma in two ways:

- By considering a single U_1^n sequence ($r_1 = 0$), we obtain the same rate requirement $r_2 > I(U_1; U_2|U_0) + \delta(\epsilon)$ as in the covering lemma.
- The lemma requires only pairwise conditional independence among the sequences $U_1^n(m_1)$, $m_1 \in [1 : 2^{nr_1}]$ (and among the sequences $U_2^n(m_2)$, $m_2 \in [1 : 2^{nr_2}]$). This implies, for example, that it suffices to use linear codes for lossy source coding of a binary symmetric source with Hamming distortion.

We are now ready to present the details of the proof of Marton's inner bound in Theorem 8.3.

Codebook generation. Fix a pmf $p(u_1, u_2)$ and function $x(u_1, u_2)$ and let $\tilde{R}_1 \geq R_1, \tilde{R}_2 \geq R_2$. For each message $m_1 \in [1 : 2^{nR_1}]$ generate a *subcodebook* $\mathcal{C}_1(m_1)$ consisting of $2^{n(\tilde{R}_1 - R_1)}$ randomly and independently generated sequences $u_1^n(l_1)$, $l_1 \in [(m_1 - 1)2^{n(\tilde{R}_1 - R_1)} + 1 : m_1 2^{n(\tilde{R}_1 - R_1)}]$, each according to $\prod_{i=1}^n p_{U_1}(u_{1i})$. Similarly, for each message $m_2 \in [1 : 2^{nR_2}]$ generate a subcodebook $\mathcal{C}_2(m_2)$ consisting of $2^{n(\tilde{R}_2 - R_2)}$ randomly and independently generated sequences $u_2^n(l_2)$, $l_2 \in [(m_2 - 1)2^{n(\tilde{R}_2 - R_2)} + 1 : m_2 2^{n(\tilde{R}_2 - R_2)}]$, each according to $\prod_{i=1}^n p_{U_2}(u_{2i})$.

For each $(m_1, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$, find an index pair (l_1, l_2) such that $u_1^n(l_1) \in \mathcal{C}_1(m_1)$, $u_2^n(l_2) \in \mathcal{C}_2(m_2)$, and $(u_1^n(l_1), u_2^n(l_2)) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is more than one such pair, choose an arbitrary one among those. If no such pair exists, choose $(l_1, l_2) = (1, 1)$. Then generate $x^n(m_1, m_2)$ as $x_i(m_1, m_2) = x(u_{1i}(l_1), u_{2i}(l_2))$, $i \in [1 : n]$.

Encoding. To send the message pair (m_1, m_2), transmit $x^n(m_1, m_2)$.

Decoding. Let $\epsilon > \epsilon'$. Decoder 1 declares that $\hat{m}_1$ is sent if it is the unique message such that $(u_1^n(l_1), y_1^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $u_1^n(l_1) \in \mathcal{C}_1(\hat{m}_1)$; otherwise it declares an error. Similarly, decoder 2 finds the unique message $\hat{m}_2$ such that $(u_2^n(l_2), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $u_2^n(l_2) \in \mathcal{C}_2(\hat{m}_2)$.

Analysis of the probability of error. Assume without loss of generality that $(M_1, M_2) = (1, 1)$ and let (L_1, L_2) be the index pair of the chosen sequences $(U_1^n(L_1), U_2^n(L_2)) \in \mathcal{C}_1^n(1) \times \mathcal{C}_2^n(1)$. Then decoder 1 makes an error only if one or more of the following events occur:

$$\begin{aligned}
\mathcal{E}_0 &= \{(U_1^n(l_1), U_2^n(l_2)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } (U_1^n(l_1), U_2^n(l_2)) \in \mathcal{C}_1(1) \times \mathcal{C}_2(1)\},\\
\mathcal{E}_{11} &= \{(U_1^n(L_1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\},\\
\mathcal{E}_{12} &= \{(U_1^n(l_1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)}(U_1, Y_1) \text{ for some } U_1^n(l_1) \notin \mathcal{C}_1(1)\}.
\end{aligned}$$

Thus the probability of error for decoder 1 is upper bounded as

$$\mathsf{P}(\mathcal{E}_1) \leq \mathsf{P}(\mathcal{E}_0) + \mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{11}) + \mathsf{P}(\mathcal{E}_{12}).$$

To bound $\mathsf{P}(\mathcal{E}_0)$, we note that the subcodebook $\mathcal{C}_1(1)$ consists of $2^{n(\tilde{R}_1 - R_1)}$ i.i.d. $U_1^n(l_1)$ sequences and the subcodebook $\mathcal{C}_2(1)$ consists of $2^{n(\tilde{R}_2 - R_2)}$ i.i.d. $U_2^n(l_2)$ sequences. Hence, by the mutual covering lemma (with $U_0 = \emptyset$, $r_1 = \tilde{R}_1 - R_1$, and $r_2 = \tilde{R}_2 - R_2$), $\mathsf{P}(\mathcal{E}_0)$ tends to zero as $n \to \infty$ if $(\tilde{R}_1 - R_1) + (\tilde{R}_2 - R_2) > I(U_1; U_2) + \delta(\epsilon')$.

To bound $\mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{11})$, note that since $(U_1^n(L_1), U_2^n(L_2)) \in \mathcal{T}_{\epsilon'}^{(n)}$ and $\epsilon' < \epsilon$, then by the conditional typicality lemma in Section 2.5, $\mathsf{P}\{(U_1^n(L_1), U_2^n(L_2), X^n, Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\}$ tends to zero as $n \to \infty$. Hence, $\mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{11})$ tends to zero as $n \to \infty$.

To bound $\mathsf{P}(\mathcal{E}_{12})$, note that since $U_1^n(l_1) \sim \prod_{i=1}^n p_{U_1}(u_{1i})$ and Y_1^n is independent of every $U_1^n(l_1) \notin \mathcal{C}(1)$, then by the packing lemma, $\mathsf{P}(\mathcal{E}_{12})$ tends to zero as $n \to \infty$ if $\tilde{R}_1 < I(U_1; Y_1) - \delta(\epsilon)$. Similarly, the average probability of error $\mathsf{P}(\mathcal{E}_2)$ for decoder 2 tends to zero as $n \to \infty$ if $\tilde{R}_2 < I(U_2; Y_2) + \delta(\epsilon)$ and $(\tilde{R}_1 - R_1) + (\tilde{R}_2 - R_2) > I(U_1; U_2) + \delta(\epsilon')$.

Thus, the average probability of error $\mathsf{P}(\mathcal{E})$ tends to zero as $n \to \infty$ if the rate pair (R_1, R_2) satisfies the inequalities

$$\begin{aligned} R_1 &\le \tilde{R}_1, \\ R_2 &\le \tilde{R}_2, \\ \tilde{R}_1 &< I(U_1; Y_1) - \delta(\epsilon), \\ \tilde{R}_2 &< I(U_2; Y_2) - \delta(\epsilon), \\ R_1 + R_2 &< \tilde{R}_1 + \tilde{R}_2 - I(U_1; U_2) - \delta(\epsilon') \end{aligned}$$

for some $(\tilde{R}_1, \tilde{R}_2)$, or equivalently, if

$$\begin{aligned} R_1 &< I(U_1; Y_1) - \delta(\epsilon), \\ R_2 &< I(U_2; Y_2) - \delta(\epsilon), \\ R_1 + R_2 &< I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) - \delta(\epsilon'). \end{aligned}$$

This completes the proof of achievability.

8.3.3 Relationship to Gelfand–Pinsker Coding

Marton's coding scheme is closely related to the Gelfand–Pinsker coding scheme for the DMC with DM state when the state information is available noncausally at the encoder; see Section 7.6. Fix a pmf $p(u_1, u_2)$ and function $x(u_1, u_2)$ in Marton's inner bound and consider the achievable rate pair $R_1 < I(U_1; Y_1) - I(U_1; U_2)$ and $R_2 < I(U_2; Y_2)$. As shown in Figure 8.9, the coding scheme for communicating M_1 to receiver 1 is identical to that for communicating M_1 over the channel $p(y_1|u_1, u_2) = p(y_1|x(u_1, u_2))$ with state U_2 available noncausally at the encoder.

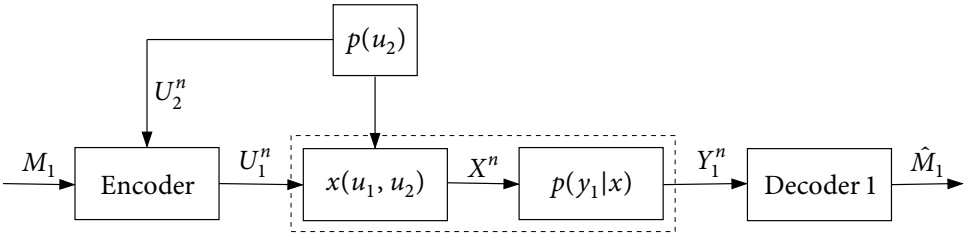

Figure 8.9. An interpretation of Marton's coding scheme at a corner point.

Application to the Gaussian Broadcast Channel. We revisit the Gaussian BC studied in Section 5.5, with outputs

$$\begin{aligned} Y_1 &= X + Z_1, \\ Y_2 &= X + Z_2, \end{aligned}$$

where $Z_1 \sim \mathrm{N}(0, N_1)$ and $Z_2 \sim \mathrm{N}(0, N_2)$. As before, assume average power constraint P on X and define the received SNRs as $S_j = P/N_j$, $j = 1, 2$. We show that a rate pair (R_1, R_2) is achievable if

$$R_1 < \mathsf{C}(\alpha S_1),$$
$$R_2 < \mathsf{C}\left(\frac{\bar{\alpha} S_2}{\alpha S_2 + 1}\right)$$

for some $\alpha \in [0, 1]$ without using superposition coding and successive cancellation decoding, and even when $N_1 > N_2$.

Decompose the channel input X into the sum of two independent parts $X_1 \sim \mathrm{N}(0, \alpha P)$ and $X_2 \sim \mathrm{N}(0, \bar{\alpha} P)$ as shown in Figure 8.10. To send M_2 to Y_2, consider the channel $Y_2 = X_2 + X_1 + Z_2$ with input X_2, Gaussian interference signal X_1, and Gaussian noise Z_2. Treating the interference signal X_1 as noise, M_2 can be communicated reliably to receiver 2 if $R_2 < \mathsf{C}(\bar{\alpha} S_2/(\alpha S_2 + 1))$. To send M_1 to receiver 1, consider the channel $Y_1 = X_1 + X_2 + Z_1$ with input X_1, additive Gaussian state X_2, and Gaussian noise Z_1, where the state $X_2^n(M_2)$ is known noncausally at the encoder. By the writing on dirty paper result (which is a special case of Gelfand–Pinsker coding and hence Marton coding), M_1 can be communicated reliably to receiver 1 if $R_1 < \mathsf{C}(\alpha S_1)$.

The above heuristic argument can be made rigorous by considering Marton's inner bound with input cost, setting $U_2 = X_2$, $U_1 = \beta U_2 + X_1$, and $X = X_1 + X_2 = (U_1 - \beta U_2) + U_2$, where $X_1 \sim \mathrm{N}(0, \alpha P)$ and $X_2 \sim \mathrm{N}(0, \bar{\alpha} P)$ are independent, and $\beta = \alpha P/(\alpha P + N_1)$, and using the discretization procedure in Section 3.4.

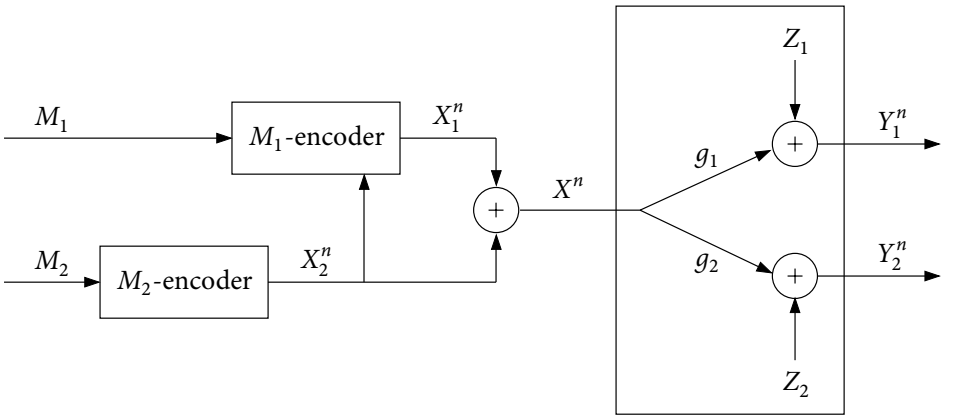

Figure 8.10. Writing on dirty paper for the Gaussian BC.

In Section 9.3, we extend this scheme to the multiple-input multiple-output (MIMO) Gaussian BC, which is not degraded in general, and show that a vector version of writing on dirty paper achieves the capacity region.

8.4 MARTON'S INNER BOUND WITH COMMON MESSAGE

Marton's inner bound can be extended to the case of common and private messages by using superposition coding and rate splitting in addition to multicoding and joint typicality codebook generation. This yields the following inner bound on the capacity of the DM-BC.

Theorem 8.4 (Marton's Inner Bound with Common Message). A rate triple (R_0, R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$ if

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

for some pmf $p(u_0, u_1, u_2)$ and function $x(u_0, u_1, u_2)$, where $|\mathcal{U}_0| \le |\mathcal{X}| + 4$, $|\mathcal{U}_1| \le |\mathcal{X}|$, and $|\mathcal{U}_2| \le |\mathcal{X}|$.

This inner bound is tight for all classes of DM-BCs with known capacity regions. In particular, the capacity region of the deterministic BC $(Y_1 = y_1(X), Y_2 = y_2(X))$ with common message is the set of rate triples (R_0, R_1, R_2) such that

$$\begin{aligned}
R_0 &\le \min\{I(U; Y_1),\ I(U; Y_2)\},\\
R_0 + R_1 &< H(Y_1),\\
R_0 + R_2 &< H(Y_2),\\
R_0 + R_1 + R_2 &< H(Y_1) + H(Y_2|U, Y_1),\\
R_0 + R_1 + R_2 &< H(Y_1|U, Y_2) + H(Y_2)
\end{aligned}$$

for some pmf $p(u, x)$. It is not known, however, if this inner bound is tight in general.

Proof of Theorem 8.4 (outline). Divide M_j, $j = 1, 2$, into two independent messages M_{j0} at rate R_{j0} and M_{jj} at rate R_{jj}. Hence $R_j = R_{j0} + R_{jj}$. Randomly and independently generate $2^{n(R_0+R_{10}+R_{20})}$ sequences $u_0^n(m_0, m_{10}, m_{20})$. For each $(m_0, m_{10}, m_{20}, m_{11})$, generate a subcodebook $\mathcal{C}_1(m_0, m_{10}, m_{20}, m_{11})$ consisting of $2^{n(\tilde{R}_{11}-R_{11})}$ independent sequences $u_1^n(m_0, m_{10}, m_{20}, l_{11})$, $l_{11} \in [(m_{11} - 1)2^{n(\tilde{R}_{11}-R_{11})} + 1 : m_{11}2^{n(\tilde{R}_{11}-R_{11})}]$. Similarly, for each $(m_0, m_{10}, m_{20}, m_{22})$, generate a subcodebook $\mathcal{C}_2(m_0, m_{10}, m_{20}, m_{22})$ consisting of $2^{n(\tilde{R}_{22}-R_{22})}$ independent sequences $u_1^n(m_0, m_{10}, m_{20}, l_{22})$, $l_{22} \in [(m_{22} - 1)2^{n(\tilde{R}_{22}-R_{22})} + 1 : m_{22}2^{n(\tilde{R}_{22}-R_{22})}]$. As in the codebook generation step in Marton's coding scheme for the private-message case, for each $(m_0, m_{10}, m_{11}, m_{20}, m_{22})$, find an index pair (l_{11}, l_{22}) such that $u_j^n(m_0, m_{10}, m_{20}, l_{jj}) \in \mathcal{C}_j(m_0, m_{10}, m_{20}, m_{jj})$, $j = 1, 2$, and

$$\big(u_1^n(m_0, m_{10}, m_{20}, l_{11}), u_2^n(m_0, m_{10}, m_{20}, l_{22})\big) \in \mathcal{T}_{\epsilon'}^{(n)},$$

and generate $x^n(m_0, m_{10}, m_{11}, m_{20}, m_{22})$ as

$$x_i = x\big(u_{0i}(m_0, m_{10}, m_{20}), u_{1i}(m_0, m_{10}, m_{20}, l_{11}), u_{2i}(m_0, m_{10}, m_{20}, l_{22})\big)$$

for $i \in [1 : n]$. To send the message triple $(m_0, m_1, m_2) = (m_0, m_{10}, m_{11}, m_{20}, m_{22})$, transmit $x^n(m_0, m_{10}, m_{20}, m_{11}, m_{22})$.

Receiver $j = 1, 2$ uses joint typicality decoding to find the unique message triple $(\hat{m}_{0j}, \hat{m}_{j0}, \hat{m}_{jj})$. Following similar steps to the proof for the private-message case and using the Fourier–Motzkin elimination procedure, it can be shown that the probability of decoding error tends to zero as $n \to \infty$ if the inequalities in Theorem 8.4 are satisfied. The cardinality bounds of the auxiliary random variables can be proved using the *perturbation method* in Appendix C.

Remark 8.6. The inner bound in Theorem 8.4 can be equivalently characterized as the set of rate triples (R_0, R_1, R_2) such that

$$\begin{aligned}
R_0 &< \min\{I(U_0; Y_1),\ I(U_0; Y_2)\},\\
R_0 + R_1 &< I(U_0, U_1; Y_1),\\
R_0 + R_2 &< I(U_0, U_2; Y_2),\\
R_0 + R_1 + R_2 &< I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0),\\
R_0 + R_1 + R_2 &< I(U_1; Y_1|U_0) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0)
\end{aligned} \tag{8.9}$$

for some pmf $p(u_0, u_1, u_2)$ and function $x(u_0, u_1, u_2)$. This region can be achieved *without* rate splitting.

Marton's private-message inner bound with U_0. By setting $R_0 = 0$ in Theorem 8.4, we obtain the following inner bound.

Proposition 8.1. A rate pair (R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$ if

$$\begin{aligned}
R_1 &< I(U_0, U_1; Y_1),\\
R_2 &< I(U_0, U_2; Y_2),\\
R_1 + R_2 &< I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0),\\
R_1 + R_2 &< I(U_1; Y_1|U_0) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0)
\end{aligned}$$

for some pmf $p(u_0, u_1, u_2)$ and function $x(u_0, u_1, u_2)$.

This inner bound is tight for all classes of broadcast channels with known private-message capacity regions. Furthermore, it can be shown that the special case of $U_0 = \emptyset$ (that is, Marton's inner bound in Theorem 8.3) is not tight in general for the degraded BC.

8.5 OUTER BOUNDS

First consider the following outer bound on the private-message capacity region of the DM-BC.

Theorem 8.5. If a rate pair (R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$, then it must satisfy the inequalities

$$\begin{aligned} R_1 &\le I(U_1; Y_1), \\ R_2 &\le I(U_2; Y_2), \\ R_1 + R_2 &\le \min\{I(U_1; Y_1) + I(X; Y_2|U_1),\ I(U_2; Y_2) + I(X; Y_1|U_2)\} \end{aligned}$$

for some pmf $p(u_1, u_2, x)$.

The proof of this theorem follows similar steps to the converse proof for the more capable DM-BC in Section 5.6. The outer bound in the theorem is tight for all broadcast channels that we presented so far in this chapter and in Chapter 5, but not tight in general. In the following, we discuss two applications of this outer bound.

8.5.1 A BSC and A BEC

We revisit Example 5.4 in Section 5.6, where the channel from X to Y_1 is a BSC(p) and the channel from X to Y_2 is a BEC(ϵ). As mentioned in the example, when $H(p) \le \epsilon < 1$, this DM-BC does not belong to any class with known capacity region. Nevertheless, the private-message capacity region for this range of parameter values is still achievable using superposition coding and is given by the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_2 &\le I(U; Y_2), \\ R_1 + R_2 &\le I(U; Y_2) + I(X; Y_1|U) \end{aligned} \tag{8.10}$$

for some pmf $p(u, x)$ with $X \sim \text{Bern}(1/2)$.

Proof of achievability. Recall the superposition coding inner bound that consists of all rate pairs (R_1, R_2) such that

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

for some pmf $p(u, x)$. Now it can be easily shown that if $H(p) < \epsilon < 1$ and $X \sim \text{Bern}(1/2)$, then $I(U; Y_2) \le I(U; Y_1)$ for all $p(u|x)$. Hence, the third inequality in the superposition coding inner bound is inactive and any rate pair in the capacity region is achievable.

Proof of the converse. We relax the inequalities in Theorem 8.5 by replacing the first

inequality with $R_1 \le I(X;Y_1)$ and dropping the first term inside the minimum in the third inequality. Now setting $U_2 = U$, we obtain an outer bound on the capacity region that consists of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_2 &\le I(U;Y_2), \\ R_1 + R_2 &\le I(U;Y_2) + I(X;Y_1|U), \\ R_1 &\le I(X;Y_1) \end{aligned} \tag{8.11}$$

for some pmf $p(u,x)$. We first show that it suffices to consider only joint pmfs $p(u,x)$ with $X \sim \text{Bern}(1/2)$. Given a pair $(U,X) \sim p(u,x)$, let $W' \sim \text{Bern}(1/2)$ and $(U', \tilde{X})$ be defined as

$$\mathsf{P}\{U' = u, \tilde{X} = x \,|\, W' = w\} = p_{U,X}(u, x \oplus w).$$

Let $(\tilde{Y}_1, \tilde{Y}_2)$ be the channel output pair corresponding to the input $\tilde{X}$. Then, by the symmetries in the input construction and the channels, it is easy to check that $\tilde{X} \sim \text{Bern}(1/2)$ and

$$\begin{aligned} I(U;Y_2) = I(U';\tilde{Y}_2|W') &\le I(U', W';\tilde{Y}_2) = I(\tilde{U};\tilde{Y}_2), \\ I(X;Y_1|U) = I(\tilde{X};\tilde{Y}_1|U', W') &= I(\tilde{X};\tilde{Y}_1|\tilde{U}), \\ I(X;Y_1) = I(\tilde{X};\tilde{Y}_1|W') &\le I(\tilde{X};\tilde{Y}_1), \end{aligned}$$

where $\tilde{U} = (U', W')$. This proves the sufficiency of $X \sim \text{Bern}(1/2)$. As mentioned in the proof of achievability, this implies that $I(U;Y_2) \le I(U;Y_1)$ for all $p(u|x)$ and hence the outer bound in (8.11) reduces to the characterization of the capacity region in (8.10).

8.5.2 Binary Skew-Symmetric Broadcast Channel

The binary skew-symmetric broadcast channel consists of two symmetric Z channels as depicted in Figure 8.11.

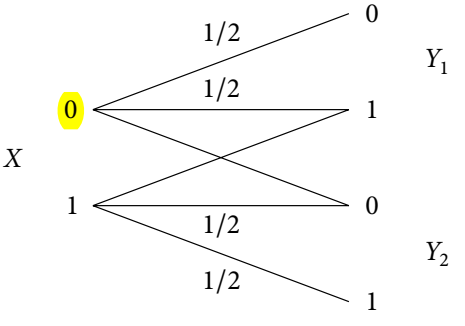

Figure 8.11. Binary skew-symmetric broadcast channel.

We show that the private-message sum-capacity $C_{\text{sum}} = \max\{R_1 + R_2 : (R_1, R_2) \in \mathscr{C}\}$ for this channel is bounded as

$$0.3616 \le C_{\text{sum}} \le 0.3726. \tag{8.12}$$

The lower bound is achieved by the following *randomized time-sharing* technique. We divide message M_j, $j = 1, 2$, into two independent messages M_{j0} and M_{jj}.

Codebook generation. Randomly and independently generate $2^{n(R_{10}+R_{20})}$ sequences $u^n(m_{10}, m_{20})$, each i.i.d. Bern(1/2). For each (m_{10}, m_{20}), let $k(m_{10}, m_{20})$ be the number of locations where $u_i(m_{10}, m_{20}) = 1$. Randomly and conditionally independently generate $2^{nR_{11}}$ sequences $x^{k(m_{10},m_{20})}(m_{10}, m_{20}, m_{11})$, each i.i.d. Bern($\alpha$). Similarly, randomly and conditionally independently generate $2^{nR_{22}}$ sequences $x^{n-k(m_{10},m_{20})}(m_{10}, m_{20}, m_{22})$, each i.i.d. Bern($\bar{\alpha}$).

Encoding. To send the message pair (m_1, m_2), represent it by the message quadruple $(m_{10}, m_{20}, m_{11}, m_{22})$. Transmit $x^{k(m_{10},m_{20})}(m_{10}, m_{20}, m_{11})$ in the locations where $u_i(m_{10}, m_{20}) = 1$ and $x^{n-k(m_{10},m_{20})}(m_{10}, m_{20}, m_{22})$ in the locations where $u_i(m_{10}, m_{20}) = 0$. Thus the messages m_{11} and m_{22} are transmitted using time sharing with respect to $u^n(m_{10}, m_{20})$.

Decoding. Each decoder $j = 1, 2$ first decodes for $u^n(m_{10}, m_{20})$ and then proceeds to recover m_{jj} from the output subsequence that corresponds to the respective symbol locations in $u^n(m_{10}, m_{20})$. It is straightforward to show that a rate pair (R_1, R_2) is achievable using this scheme if

$$\begin{aligned} R_1 &< \min\{I(U; Y_1), I(U; Y_2)\} + \frac{1}{2} I(X; Y_1|U = 1), \\ R_2 &< \min\{I(U; Y_1), I(U; Y_2)\} + \frac{1}{2} I(X; Y_2|U = 0), \\ R_1 + R_2 &< \min\{I(U; Y_1), I(U; Y_2)\} + \frac{1}{2}\big(I(X; Y_1|U = 1) + I(X; Y_2|U = 0)\big) \end{aligned} \tag{8.13}$$

for some $\alpha \in [0, 1]$. Taking the maximum of the sum-rate bound over α establishes the sum-capacity lower bound $C_{\text{sum}} \geq 0.3616$.

Note that by considering $U_0 \sim \text{Bern}(1/2)$, $U_1 \sim \text{Bern}(\alpha)$, and $U_2 \sim \text{Bern}(\alpha)$, independent of each other, and setting $X = U_0U_1 + (1 - U_0)(1 - U_2)$, Marton's inner bound with U_0 in Proposition 8.1 reduces to the randomized time-sharing rate region in (8.13). Furthermore, it can be shown that the best achievable sum-rate using Marton's coding scheme is 0.3616, which is the rate achieved by randomized time-sharing.

The upper bound in (8.12) follows by Theorem 8.5, which implies that

$$\begin{aligned} C_{\text{sum}} &\leq \max_{p(u_1,u_2,x)} \min\{I(U_1; Y_1) + I(U_2; Y_2|U_1), I(U_2; Y_2) + I(U_1; Y_1|U_2)\} \\ &\leq \max_{p(u_1,u_2,x)} \frac{1}{2}\big(I(U_1; Y_1) + I(U_2; Y_2|U_1) + I(U_2; Y_2) + I(U_1; Y_1|U_2)\big). \end{aligned}$$

It can be shown that the latter maximum is attained by $X \sim \text{Bern}(1/2)$ and binary U_1 and U_2, which gives the upper bound on the sum-capacity $C_{\text{sum}} \leq 0.3726$.

8.5.3 Nair–El Gamal Outer Bound

We now consider the following outer bound on the capacity region of the DM-BC with common and private messages.

Theorem 8.6 (Nair–El Gamal Outer Bound). If a rate triple (R_0, R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$, then it must satisfy the inequalities

$$\begin{aligned} R_0 &\le \min\{I(U_0; Y_1), I(U_0; Y_2)\}, \\ R_0 + R_1 &\le I(U_0, U_1; Y_1), \\ R_0 + R_2 &\le I(U_0, U_2; Y_2), \\ R_0 + R_1 + R_2 &\le I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0, U_1), \\ R_0 + R_1 + R_2 &\le I(U_1; Y_1|U_0, U_2) + I(U_0, U_2; Y_2) \end{aligned}$$

for some pmf $p(u_0, u_1, u_2, x) = p(u_1)p(u_2)p(u_0|u_1, u_2)$ and function $x(u_0, u_1, u_2)$.

The proof of this outer bound is quite similar to the converse proof for the more capable DM-BC in Section 5.6 and is given in Appendix 8B.

The Nair–El Gamal bound is tight for all broadcast channels with known capacity regions that we discussed so far. Note that the outer bound coincides with Marton's inner bound with common message in Theorem 8.4 (or more directly, the alternative characterization in (8.9)) if $I(U_1; U_2|U_0, Y_1) = I(U_1; U_2|U_0, Y_2) = 0$ for every pmf $p(u_0, u_1, u_2, x)$ that attains a boundary point of the outer bound. This bound is not tight in general, however.

Remark 8.7. It can be shown that the Nair–El Gamal outer bound with no common message, i.e., with $R_0 = 0$, simplifies to the outer bound in Theorem 8.5, that is, no U_0 is needed when evaluating the outer bound in Theorem 8.6 with $R_0 = 0$. This is in contrast to Marton's inner bound, where the private-message region without U_0 in Theorem 8.3 can be strictly smaller than that with U_0 in Proposition 8.1.

8.6 INNER BOUNDS FOR MORE THAN TWO RECEIVERS

Marton's inner bound can be easily extended to more than two receivers. For example, consider a 3-receiver DM-BC $p(y_1, y_2, y_3|x)$ with three private messages $M_j \in [1 : 2^{nR_j}]$, $j = 1, 2, 3$. A rate triple (R_1, R_2, R_3) is achievable for the 3-receiver DM-BC if

$$\begin{aligned} R_1 &< I(U_1; Y_1), \\ R_2 &< I(U_2; Y_2), \\ R_3 &< I(U_3; Y_3), \\ R_1 + R_2 &< I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2), \\ R_1 + R_3 &< I(U_1; Y_1) + I(U_3; Y_3) - I(U_1; U_3), \\ R_2 + R_3 &< I(U_2; Y_2) + I(U_3; Y_3) - I(U_2; U_3), \\ R_1 + R_2 + R_3 &< I(U_1; Y_1) + I(U_2; Y_2) + I(U_3; Y_3) - I(U_1; U_2) - I(U_1, U_2; U_3) \end{aligned}$$

for some pmf $p(u_1, u_2, u_3)$ and function $x(u_1, u_2, u_3)$. This region can be readily extended to any number of receivers k.

It can be readily shown that the extended Marton inner bound is tight for deterministic DM-BCs with an arbitrary number of receivers, where we substitute $U_j = Y_j$ for $j \in [1:k]$. The proof of achievability follows similar steps to the case of two receivers using the following extension of the mutual covering lemma.

Lemma 8.2 (Multivariate Covering Lemma). Let $(U_0, U_1, \ldots, U_k) \sim p(u_0, u_1, \ldots, u_k)$ and $\epsilon' < \epsilon$. Let $U_0^n \sim p(u_0^n)$ be a random sequence with $\lim_{n\to\infty} \mathsf{P}\{U_0^n \in \mathcal{T}_{\epsilon'}^{(n)}\} = 1$. For each $j \in [1:k]$, let $U_j^n(m_j)$, $m_j \in [1:2^{nr_j}]$, be pairwise conditionally independent random sequences, each distributed according to $\prod_{i=1}^n p_{U_j|U_0}(u_{ji}|u_{0i})$. Assume that $(U_1^n(m_1)\colon m_1 \in [1:2^{nr_1}])$, $(U_2^n(m_2)\colon m_2 \in [1:2^{nr_2}])$, $\ldots$, $(U_k^n(m_k)\colon m_k \in [1:2^{nr_k}])$ are mutually conditionally independent given U_0^n. Then, there exists $\delta(\epsilon)$ that tends to zero as $\epsilon \to 0$ such that

$$\lim_{n\to\infty} \mathsf{P}\{(U_0^n, U_1^n(m_1), U_2^n(m_2), \ldots, U_k^n(m_k)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } (m_1, m_2, \ldots, m_k)\} = 0,$$

if $\sum_{j\in\mathcal{J}} r_j > \sum_{j\in\mathcal{J}} H(U_j|U_0) - H(U(\mathcal{J})|U_0) + \delta(\epsilon)$ for all $\mathcal{J} \subseteq [1:k]$ with $|\mathcal{J}| \geq 2$.

The proof of this lemma is a straightforward extension of that for the $k = 2$ case presented in Appendix 8A. When $k = 3$ and $U_0 = \emptyset$, the conditions for joint typicality become

$$\begin{aligned} r_1 + r_2 &> I(U_1; U_2) + \delta(\epsilon), \\ r_1 + r_3 &> I(U_1; U_3) + \delta(\epsilon), \\ r_2 + r_3 &> I(U_2; U_3) + \delta(\epsilon), \\ r_1 + r_2 + r_3 &> I(U_1; U_2) + I(U_1, U_2; U_3) + \delta(\epsilon). \end{aligned}$$

In general, combining Marton's coding with superposition coding, rate splitting, and indirect decoding, we can construct an inner bound for DM-BCs with more than two receivers for any given messaging requirement. We illustrate this construction via the 3-receiver DM-BC $p(y_1, y_2, y_3|x)$ with two degraded message sets, where a common message $M_0 \in [1:2^{nR_0}]$ is to be communicated to receivers 1, 2, and 3, and a private message $M_1 \in [1:2^{nR_1}]$ is to be communicated only to receiver 1.

Proposition 8.2. A rate pair (R_0, R_1) is achievable for the 3-receiver DM-BC $p(y_1, y_2, y_3|x)$ with 2 degraded message sets if

$$\begin{aligned} R_0 &< \min\{I(V_2; Y_2),\ I(V_3; Y_3)\}, \\ 2R_0 &< I(V_2; Y_2) + I(V_3; Y_3) - I(V_2; V_3|U), \\ R_0 + R_1 &< \min\{I(X; Y_1),\ I(V_2; Y_2) + I(X; Y_1|V_2),\ I(V_3; Y_3) + I(X; Y_1|V_3)\}, \\ 2R_0 + R_1 &< I(V_2; Y_2) + I(V_3; Y_3) + I(X; Y_1|V_2, V_3) - I(V_2; V_3|U), \\ 2R_0 + 2R_1 &< I(V_2; Y_2) + I(V_3; Y_3) + I(X; Y_1|V_2) + I(X; Y_1|V_3) - I(V_2; V_3|U), \\ 2R_0 + 2R_1 &< I(V_2; Y_2) + I(V_3; Y_3) + I(X; Y_1|U) + I(X; Y_1|V_2, V_3) - I(V_2; V_3|U) \end{aligned}$$

for some pmf $p(u, v_2, v_3, x) = p(u)p(v_2|u)p(x, v_3|v_2) = p(u)p(v_3|u)p(x, v_2|v_3)$.

It can be shown that this inner bound is tight when Y_1 is less noisy than Y_2, which is a generalization of the class of multilevel DM-BCs discussed earlier in Section 8.2.

Proof of Proposition 8.2 (outline). Divide M_1 into four independent messages M_{10} at rate R_{10}, M_{11} at rate R_{11}, M_{12} at rate R_{12}, and M_{13} at rate R_{13}. Let $\tilde{R}_{12} \geq R_{12}$ and $\tilde{R}_{13} \geq R_{13}$. Randomly and independently generate $2^{n(R_0+R_{10})}$ sequences $u^n(m_0, m_{10})$, $(m_0, m_{10}) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_{10}}]$. As in the achievability proof of Theorem 8.4, for each (m_0, m_{10}), we use the codebook generation step in the proof of Marton's inner bound in Theorem 8.3 to randomly and conditionally independently generate $2^{n\tilde{R}_{12}}$ sequences $(v_2^n(m_0, m_{10}, l_2)$, $l_2 \in [1 : 2^{n\tilde{R}_{12}}]$, and $2^{n\tilde{R}_{13}}$ sequences $v_3^n(m_0, m_{10}, l_3))$, $l_3 \in [1 : 2^{n\tilde{R}_{13}}]$. For each $(m_0, m_{10}, m_{12}, m_{13})$, find a jointly typical sequence pair $(v_2^n(m_0, m_{10}, l_2), v_3^n(m_0, m_{10}, l_3))$, where $l_2 \in [(m_{12} - 1)2^{n(\tilde{R}_{12}-R_{12})} + 1 : m_{12}2^{n(\tilde{R}_{12}-R_{12})}]$ and $l_3 \in [(m_{13} - 1)2^{n(\tilde{R}_{13}-R_{13})} + 1 : m_{13}2^{n(\tilde{R}_{13}-R_{13})}]$, and randomly and conditionally independently generate $2^{nR_{11}}$ sequences $x^n(m_0, m_{10}, m_{11}, m_{12}, m_{13})$. To send the message pair (m_0, m_1), transmit $x^n(m_0, m_{10}, m_{11}, m_{12}, m_{13})$.

Receiver 1 uses joint typicality decoding to find the unique tuple $(\hat{m}_{01}, \hat{m}_{10}, \hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13})$. Receiver $j = 2, 3$ uses indirect decoding to find the unique $\hat{m}_{0j}$ through (u^n, v_j^n). Following a standard analysis of the probability of error and using the Fourier–Motzkin elimination procedure, it can be shown that the probability of error tends to zero as $n \to \infty$ if the inequalities in Proposition 8.2 are satisfied.

SUMMARY

- Indirect decoding
- Marton's inner bound:
  - Multidimensional subcodebook generation
  - Generating correlated codewords for independent messages
  - Tight for all BCs with known capacity regions
  - A common auxiliary random variable is needed even when coding only for private messages
- Mutual covering lemma and its multivariate extension
- Connection between Marton coding and Gelfand–Pinsker coding
- Writing on dirty paper achieves the capacity region of the Gaussian BC
- Randomized time sharing for the binary skew-symmetric BC
- Nair–El Gamal outer bound:
  - Does not always coincide with Marton's inner bound
  - Not tight in general

- **Open problems:**

  8.1. What is the capacity region of the general 3-receiver DM-BC with one common message to all three receivers and one private message to one receiver?

  8.2. Is superposition coding optimal for the general 3-receiver DM-BC with one message to all three receivers and another message to two receivers?

  8.3. What is the sum-capacity of the binary skew-symmetric broadcast channel?

  8.4. Is Marton's inner bound tight in general?

BIBLIOGRAPHIC NOTES

The capacity region of the 2-receiver DM-BC with degraded message sets was established by Körner and Marton (1977b), who proved the strong converse using the technique of images of sets (Körner and Marton 1977c). The multilevel BC was introduced by Borade, Zheng, and Trott (2007), who conjectured that the straightforward generalization of Theorem 8.1 to three receivers in (8.2) is optimal. The capacity region of the multilevel BC in Theorem 8.2 was established using indirect decoding by Nair and El Gamal (2009). They also established the general inner bound for the 3-receiver DM-BC with degraded message sets in Section 8.6.

The inner bound on the private-message capacity region in (8.8) is a special case of a more general inner bound by Cover (1975b) and van der Meulen (1975), which consists of all rate pairs (R_1, R_2) that satisfy the inequalities in Proposition 8.1 for some pmf $p(u_0)p(u_1)p(u_2)$ and $x(u_0, u_1, u_2)$. The deterministic broadcast channel in Example 8.2 is attributed to D. Blackwell. The capacity region of the Blackwell channel was established by Gelfand (1977). The capacity region of the deterministic broadcast channel was established by Pinsker (1978). The capacity region of the semideterministic broadcast channel was established independently by Marton (1979) and Gelfand and Pinsker (1980a). The inner bounds in Theorem 8.3 and in (8.9) are due to Marton (1979). The mutual covering lemma and its application in the proof of achievability are due to El Gamal and van der Meulen (1981). The inner bound on the capacity region of the DM-BC with common message in Theorem 8.4 is due to Liang (2005). The equivalence to Marton's characterization in (8.9) was shown by Liang, Kramer, and Poor (2011). Gohari, El Gamal, and Anantharam (2010) showed that Marton's inner bound without U_0 is not optimal even for a degraded DM-BC with no common message.

The outer bound in Theorem 8.5 is a direct consequence of the converse proof for the more capable BC by El Gamal (1979). Nair and El Gamal (2007) showed through the binary skew-symmetric BC example that it is strictly tighter than the earlier outer bounds by Sato (1978a) and Körner and Marton (Marton 1979). The outer bound in Theorem 8.6 is due to Nair and El Gamal (2007). Other outer bounds can be found in Liang, Kramer, and Shamai (2008), Gohari and Anantharam (2008), and Nair (2008). These bounds coincide with the bound in Theorem 8.5 when there is no common message. It is not known,

however, if they are strictly tighter than the Nair–El Gamal outer bound. The application of the outer bound in Theorem 8.5 to the BSC–BEC broadcast channel is due to Nair (2010). The binary skew-symmetric channel was first studied by Hajek and Pursley (1979), who showed through randomized time sharing that U_0 is needed for the Cover–van der Meulen inner bound even when there is no common message. Gohari and Anantharam (2009), Nair and Wang (2008), and Jog and Nair (2010) showed that the maximum sum-rate of Marton's inner bound (with U_0) coincides with the randomized time-sharing lower bound, establishing a strict gap between Marton's inner bound with U_0 and the outer bound in Theorem 8.5. Geng, Gohari, Nair, and Yu (2011) established the capacity region of the product of reversely semideterministic or more capable DM-BCs and provided an ingenuous counterexample that shows that the Nair–El Gamal outer bound is not tight.

PROBLEMS

8.1. Prove achievability of the extended superposition coding inner bound in (8.2) for the 3-receiver multilevel DM-BC.

8.2. Consider the binary erasure multilevel product DM-BC in Example 8.1. Show that the extended superposition coding inner bound in (8.2) can be simplified to (8.6).

8.3. Prove the converse for the capacity region of the 3-receiver multilevel DM-BC in Theorem 8.2. (Hint: Use the converse proof for the degraded DM-BC for Y_1 and Y_2 and the converse proof for the 2-receiver DM-BC with degraded message sets for Y_1 and Y_3.)

8.4. Consider the BSC–BEC broadcast channel in Section 8.5.1. Show that if $H(p) < \epsilon < 1$ and $X \sim \text{Bern}(1/2)$, then $I(U; Y_2) \le I(U; Y_1)$ for all $p(u|x)$.

8.5. Complete the proof of the mutual covering lemma for the case $U_0 \neq \emptyset$.

8.6. Show that the inner bound on the capacity region of the 3-receiver DM-BC with two degraded message sets in Proposition 8.2 is tight when Y_1 is less noisy that Y_2.

8.7. *Complete decoding for the 3-receiver multilevel DM-BC.* Show that the fourth inequality in (8.3) is redundant and thus that the region simplifies to the capacity region in Theorem 8.2. (Hint: Given the choice of U, consider two cases $I(U; Y_2) \le I(U; Y_3)$ and $I(U; Y_2) > I(U; Y_3)$ separately, and optimize V.)

8.8. *Sato's outer bound.* Show that if a rate pair (R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$, then it must satisfy

$$\begin{aligned} R_1 &\le I(X; Y_1), \\ R_2 &\le I(X; Y_2), \\ R_1 + R_2 &\le \min_{\tilde{p}(y_1, y_2|x)} I(X; Y_1, Y_2) \end{aligned}$$

for some pmf $p(x)$, where the minimum is over all conditional pmfs $\tilde{p}(y_1, y_2|x)$

with the same conditional marginal pmfs $p(y_1|x)$ and $p(y_2|x)$ as the given channel conditional pmf $p(y_1, y_2|x)$.

8.9. *Write-once memory.* Consider a memory that can store bits without any noise. But unlike a regular memory, each cell of the memory is permanently programmed to zero once a zero is stored. Thus, the memory can be modeled as a DMC with state $Y = XS$, where S is the previous state of the memory, X is the input to the memory, and Y is the output of the memory.

Suppose that the initial state of the memory is $S_i = 1$, $i \in [1 : n]$. We wish to use the memory twice at rates R_1 and R_2, respectively. For the first write, we store data with input X_1 and output $Y_1 = X_1$. For the second write, we store data with input X_2 and output $Y_2 = X_1X_2$.

(a) Show that the capacity region for two writes over the memory is the set of rate pairs (R_1, R_2) such that

$$R_1 \le H(\alpha),$$
$$R_2 \le 1 - \alpha$$

for some $\alpha \in [0, 1/2]$.

(b) Find the sum-capacity.

(c) Now suppose we use the memory k times at rates $R_1, \ldots, R_k$. The memory at stage k has $Y_k = X_1X_2 \cdots X_k$. Find the capacity region and sum-capacity.

Remark: This result is due to Wolf, Wyner, Ziv, and Körner (1984).

8.10. *Marton inner bound with common message.* Consider the Marton inner bound with common message in Theorem 8.4. Provide the details of the coding scheme outlined below.

(a) Using the packing lemma and the mutual covering lemma, show that the probability of error tends to zero as $n \to \infty$ if

$$\begin{aligned} R_0 + R_{01} + R_{02} + \tilde{R}_{11} &< I(U_0, U_1; Y_1), \\ R_0 + R_{01} + R_{02} + \tilde{R}_{22} &< I(U_0, U_2; Y_2), \\ \tilde{R}_{11} &< I(U_1; Y_1|U_0), \\ \tilde{R}_{22} &< I(U_2; Y_2|U_0), \\ \tilde{R}_{11} + \tilde{R}_{22} - R_{11} - R_{22} &> I(U_1; U_2|U_0). \end{aligned}$$

(b) Establish the inner bound by using the Fourier–Motzkin elimination procedure to reduce the set of inequalities in part (a).

(c) Show that the inner bound is convex.

8.11. *Maximal probability of error.* Consider a $(2^{nR_1}, 2^{nR_2}, n)$ code $\mathcal{C}$ for the DM-BC $p(y_1, y_2|x)$ with average probability of error $P_e^{(n)}$. In this problem, we show that there exists a $(2^{nR_1}/n^2, 2^{nR_2}/n^2, n)$ code with maximal probability of error less than

$(P_e^{(n)})^{1/2}$. Consequently, the capacity region with maximal probability of error is the same as that with average probability of error. This is in contrast to the DM-MAC case, where the capacity region for maximal probability of error can be strictly smaller than that for average probability of error; see Problem 4.3.

(a) A codeword $x^n(m_1, m_2)$ is said to be "bad" if its probability of error $\lambda_{m_2 m_2} = \mathsf{P}\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2) \,|\, (M_1, M_2) = (m_1, m_2)\} > (P_e^{(n)})^{1/2}$. Show that there are at most $2^{n(R_1+R_2)}(P_e^{(n)})^{1/2}$ "bad" codewords $x^n(m_1, m_2)$.

(b) Randomly and independently permute the message indices m_1 and m_2 to generate a new codebook $\bar{\mathcal{C}}$ consisting of codewords $x^n(\sigma_1(m_1), \sigma(m_2))$, where σ_1 and σ_2 denote the random permutations. Then partition the codebook $\bar{\mathcal{C}}$ into subcodebooks $\bar{\mathcal{C}}(m_1', m_2')$, $(m_1', m_2') \in [1 : 2^{nR_1}/n^2] \times [1 : 2^{nR_2}/n^2]$, each consisting of $n^2 \times n^2$ sequences. Show that the probability that all n^4 codewords in the subcodebook $\bar{\mathcal{C}}(m_1', m_2')$ are "bad" is upper bounded by $(P_e^{(n)})^{n^2/2}$. (Hint: This probability is upper bounded by the probability that all n^2 "diagonal" codewords are "bad." Upper bound the latter probability using part (a) and the independence of the permutations.)

(c) Suppose that a subcodebook $\bar{\mathcal{C}}(m_1', m_2')$ is said to be "bad" if all sequences in the subcodebook are "bad." Show that the expected number of "bad" subcodebooks is upper bounded by

$$\frac{2^{n(R_1+R_2)}}{n^4}\left(P_e^{(n)}\right)^{n^2/2}.$$

Further show that this upper bound tends to zero as $n \to \infty$ if $P_e^{(n)}$ tends to zero as $n \to \infty$.

(d) Argue that there exists at least one permutation pair (σ_1, σ_2) such that there is no "bad" subcodebook. Conclude that there exists a $(2^{nR_1}/n^2, 2^{nR_2}/n^2, n)$ code with maximal probability of error less than or equal to $(P_e^{(n)})^{1/2}$ for n sufficiently large, if $P_e^{(n)}$ tends to zero as $n \to \infty$.

Remark: This argument is due to İ. E. Telatar, which is a simplified version of the original proof by Willems (1990).

APPENDIX 8A PROOF OF THE MUTUAL COVERING LEMMA

We prove the case $U_0 = \emptyset$ only. The proof for the general case follows similarly.

Let $\mathcal{A} = \{(m_1, m_2) \in [1 : 2^{nr_1}] \times [1 : 2^{nr_2}] \colon (U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2)\}$. Then by the Chebyshev lemma in Appendix B, the probability of the event of interest can be bounded as

$$\mathsf{P}\{|\mathcal{A}| = 0\} \le \mathsf{P}\{(|\mathcal{A}| - \mathsf{E}\,|\mathcal{A}|)^2 \ge (\mathsf{E}\,|\mathcal{A}|)^2\} \le \frac{\mathrm{Var}(|\mathcal{A}|)}{(\mathsf{E}\,|\mathcal{A}|)^2}.$$

We now show that $\mathrm{Var}(|\mathcal{A}|)/(\mathsf{E}\,|\mathcal{A}|)^2$ tends to zero as $n \to \infty$ if

$$\begin{aligned} r_1 &> 3\delta(\epsilon), \\ r_2 &> 3\delta(\epsilon), \\ r_1 + r_2 &> I(U_1; U_2) + \delta(\epsilon). \end{aligned}$$

Using indicator random variables, we can express $|\mathcal{A}|$ as

$$|\mathcal{A}| = \sum_{m_1=1}^{2^{nr_1}} \sum_{m_2=1}^{2^{nr_2}} E(m_1, m_2),$$

where

$$E(m_1, m_2) = \begin{cases} 1 & \text{if } (U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, \\ 0 & \text{otherwise} \end{cases}$$

for each $(m_1, m_2) \in [1 : 2^{nr_1}] \times [1 : 2^{nr_2}]$. Let

$$\begin{aligned} p_1 &= \mathsf{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}\}, \\ p_2 &= \mathsf{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(1), U_2^n(2)) \in \mathcal{T}_\epsilon^{(n)}\}, \\ p_3 &= \mathsf{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(2), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}\}, \\ p_4 &= \mathsf{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(2), U_2^n(2)) \in \mathcal{T}_\epsilon^{(n)}\} = p_1^2. \end{aligned}$$

Then

$$\begin{aligned} \mathsf{E}(|\mathcal{A}|) &= \sum_{m_1, m_2} \mathsf{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}\} = 2^{n(r_1+r_2)} p_1, \\ \mathsf{E}(|\mathcal{A}|^2) &= \sum_{m_1, m_2} \mathsf{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}\} \\ &\quad + \sum_{m_1, m_2} \sum_{m_2' \neq m_2} \mathsf{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(m_1), U_2^n(m_2')) \in \mathcal{T}_\epsilon^{(n)}\} \\ &\quad + \sum_{m_1, m_2} \sum_{m_1' \neq m_1} \mathsf{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(m_1'), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}\} \\ &\quad + \sum_{m_1, m_2} \sum_{m_1' \neq m_1} \sum_{m_2' \neq m_2} \mathsf{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(m_1'), U_2^n(m_2')) \in \mathcal{T}_\epsilon^{(n)}\} \\ &\le 2^{n(r_1+r_2)} p_1 + 2^{n(r_1+2r_2)} p_2 + 2^{n(2r_1+r_2)} p_3 + 2^{2n(r_1+r_2)} p_4. \end{aligned}$$

Hence

$$\mathrm{Var}(|\mathcal{A}|) \le 2^{n(r_1+r_2)} p_1 + 2^{n(r_1+2r_2)} p_2 + 2^{n(2r_1+r_2)} p_3.$$

Now by the joint typicality lemma, for sufficiently large n, we have

$$\begin{aligned} p_1 &\ge 2^{-n(I(U_1;U_2)+\delta(\epsilon))}, \\ p_2 &\le 2^{-n(2I(U_1;U_2)-\delta(\epsilon))}, \\ p_3 &\le 2^{-n(2I(U_1;U_2)-\delta(\epsilon))}, \end{aligned}$$

and hence

$$p_2/p_1^2 \le 2^{3n\delta(\epsilon)},$$
$$p_3/p_1^2 \le 2^{3n\delta(\epsilon)}.$$

Therefore

$$\frac{\mathrm{Var}(|\mathcal{A}|)}{(\mathsf{E}\,|\mathcal{A}|)^2} \le 2^{-n(r_1+r_2-I(U_1;U_2)-\delta(\epsilon))} + 2^{-n(r_1-3\delta(\epsilon))} + 2^{-n(r_2-3\delta(\epsilon))},$$

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

$$r_1 > 3\delta(\epsilon),$$
$$r_2 > 3\delta(\epsilon),$$
$$r_1 + r_2 > I(U_1; U_2) + \delta(\epsilon).$$

It can be similarly shown that $\mathsf{P}\{|\mathcal{A}| = 0\}$ tends to zero as $n \to \infty$ if $r_1 = 0$ and $r_2 > I(U_1; U_2) + \delta(\epsilon)$, or if $r_1 > I(U_1; U_2) + \delta(\epsilon)$ and $r_2 = 0$. Combining these three sets of inequalities, we have shown that $\mathsf{P}\{|\mathcal{A}| = 0\}$ tends to zero as $n \to \infty$ if $r_1 + r_2 > I(U_1; U_2) + 4\delta(\epsilon)$.

APPENDIX 8B PROOF OF THE NAIR–EL GAMAL OUTER BOUND

By Fano's inequality and following standard steps, we have

$$nR_0 \le \min\{I(M_0; Y_1^n), I(M_0; Y_2^n)\} + n\epsilon_n,$$
$$n(R_0 + R_1) \le I(M_0, M_1; Y_1^n) + n\epsilon_n,$$
$$n(R_0 + R_2) \le I(M_0, M_2; Y_2^n) + n\epsilon_n,$$
$$n(R_0 + R_1 + R_2) \le I(M_0, M_1; Y_1^n) + I(M_2; Y_2^n | M_0, M_1) + n\epsilon_n,$$
$$n(R_0 + R_1 + R_2) \le I(M_1; Y_1^n | M_0, M_2) + I(M_0, M_2; Y_2^n) + n\epsilon_n.$$

We bound the mutual information terms in the above bounds. First consider the terms in the fourth inequality

$$I(M_0, M_1; Y_1^n) + I(M_2; Y_2^n | M_0, M_1) = \sum_{i=1}^n I(M_0, M_1; Y_{1i} | Y_1^{i-1}) + \sum_{i=1}^n I(M_2; Y_{2i} | M_0, M_1, Y_{2,i+1}^n).$$

Now consider

$$\sum_{i=1}^n I(M_0, M_1; Y_{1i} | Y_1^{i-1}) \le \sum_{i=1}^n I(M_0, M_1, Y_1^{i-1}; Y_{1i})$$
$$= \sum_{i=1}^n I(M_0, M_1, Y_1^{i-1}, Y_{2,i+1}^n; Y_{1i}) - \sum_{i=1}^n I(Y_{2,i+1}^n; Y_{1i} | M_0, M_1, Y_1^{i-1}).$$

Also

$$\sum_{i=1}^{n} I(M_2; Y_{2i}|M_0, M_1, Y_{2,i+1}^n) \le \sum_{i=1}^{n} I(M_2, Y_1^{i-1}; Y_{2i}|M_0, M_1, Y_{2,i+1}^n)$$
$$= \sum_{i=1}^{n} I(Y_1^{i-1}; Y_{2i}|M_0, M_1, Y_{2,i+1}^n)$$
$$+ \sum_{i=1}^{n} I(M_2; Y_{2i}|M_0, M_1, Y_{2,i+1}^n, Y_1^{i-1}).$$

Combining the above results and identifying auxiliary random variables as $U_{0i} = (M_0, Y_1^{i-1}, Y_{2,i+1}^n)$, $U_{1i} = M_1$, and $U_{2i} = M_2$, we obtain

$$I(M_0, M_1; Y_1^n) + I(M_2; Y_2^n|M_0, M_1)$$
$$\le \sum_{i=1}^{n} I(M_0, M_1, Y_1^{i-1}, Y_{2,i+1}^n; Y_{1i}) - \sum_{i=1}^{n} I(Y_{2,i+1}^n; Y_{1i}|M_0, M_1, Y_1^{i-1})$$
$$+ \sum_{i=1}^{n} I(Y_1^{i-1}; Y_{2i}|M_0, M_1, Y_{2,i+1}^n) + \sum_{i=1}^{n} I(M_2; Y_{2i}|M_0, M_1, Y_{2,i+1}^n, Y_1^{i-1})$$
$$\overset{(a)}{=} \sum_{i=1}^{n} I(M_0, M_1, Y_1^{i-1}, Y_{2,i+1}^n; Y_{1i}) + \sum_{i=1}^{n} I(M_2; Y_{2i}|M_0, M_1, Y_{2,i+1}^n, Y_1^{i-1})$$
$$= \sum_{i=1}^{n} I(U_{0i}, U_{1i}; Y_{1i}) + \sum_{i=1}^{n} I(U_{2i}; Y_{2i}|U_{0i}, U_{1i}),$$

where (a) follows by the Csiszár sum identity. Similarly

$$I(M_1; Y_1^n|M_0, M_2) + I(M_0, M_2; Y_2^n) \le \sum_{i=1}^{n} I(U_{1i}; Y_{1i}|U_{0i}, U_{2i}) + \sum_{i=1}^{n} I(U_{0i}, U_{2i}; Y_{2i}).$$

It can be also easily shown that

$$I(M_0; Y_j^n) \le \sum_{i=1}^{n} I(U_{0i}; Y_{ji}),$$
$$I(M_0, M_j; Y_j^n) \le \sum_{i=1}^{n} I(U_{0i}, U_{ji}; Y_{ji}), \quad j = 1, 2.$$

The rest of the proof follows by introducing a time-sharing random variable Q independent of $(M_0, M_1, M_2, X^n, Y_1^n, Y_2^n)$, and uniformly distributed over $[1 : n]$, and defining $U_0 = (Q, U_{0Q})$, $U_1 = U_{1Q}$, $U_2 = U_{2Q}$, $X = X_Q$, $Y_1 = Y_{1Q}$, and $Y_2 = Y_{2Q}$. Using arguments similar to Remark 8.4, it can be easily verified that taking a function $x(u_0, u_1, u_2)$ suffices. Note that the independence of the messages M_1 and M_2 implies the independence of the auxiliary random variables U_1 and U_2 as specified.

CHAPTER 9

Gaussian Vector Channels

Gaussian vector channels are models for multiple-input multiple-output (MIMO) wireless communication systems in which each transmitter and each receiver can have more than a single antenna. The use of multiple antennas provides several benefits in a wireless multipath medium, including higher received power via beamforming, higher channel capacity via spatial multiplexing without increasing bandwidth or transmission power, and improved transmission robustness via diversity coding. In this chapter, we focus on the limits on spatial multiplexing of MIMO communication.

We first establish the capacity of the Gaussian vector point-to-point channel. We show that this channel is equivalent to the Gaussian product channel discussed in Section 3.4 and thus the capacity is achieved via water-filling. We then establish the capacity region of the Gaussian vector multiple access channel as a straightforward extension of the scalar case and show that the sum-capacity is achieved via iterative water-filling.

The rest of the chapter is dedicated to studying Gaussian vector broadcast channels. We first establish the capacity region with common message for the special case of the Gaussian product broadcast channel, which is not in general degraded. Although the capacity region of this channel is achieved via superposition coding with a product input pmf on the channel components, the codebook generation for the common message and decoding must each be performed jointly across the channel components. For the special case of private messages only, the capacity region can be expressed as the intersection of the capacity regions of two enhanced degraded broadcast channels. Next, we turn our attention to the general Gaussian vector broadcast channel and establish the private-message capacity region. We first describe a vector extension of writing on dirty paper for the Gaussian BC discussed in Section 8.3. We show that this scheme is optimal by constructing an enhanced degraded Gaussian vector broadcast channel for every corner point on the boundary of the capacity region. The proof uses Gaussian vector BC–MAC duality, convex optimization techniques (Lagrange duality and the KKT condition), and the entropy power inequality.

In Chapter 23, we show that in the presence of fading, the capacity of the vector Gaussian channel grows linearly with the number of antennas.

9.1 GAUSSIAN VECTOR POINT-TO-POINT CHANNEL

Consider the point-to-point communication system depicted in Figure 9.1. The sender

SUMMARY

- Gaussian vector channel models
- Reciprocity via singular value decomposition
- Water-filling and iterative water-filling
- Vector writing on dirty paper:
  - Marton's coding scheme
  - Optimal for the Gaussian vector BC with private messages
- BC–MAC duality
- Proof of the converse for the Gaussian vector BC:
  - Converse for each boundary corner point
  - Characterization of the corner point via Lagrange duality and KKT condition
  - Construction of an enhanced degraded BC
  - Use of the vector EPI for the enhanced degraded BC
- **Open problems:**

  9.1. What is the capacity region of the Gaussian product BC with more than 2 receivers?

  9.2. Can the converse for the Gaussian vector BC be proved directly by optimizing the Nair–El Gamal outer bound?

  9.3. What is the capacity region of the 2-receiver Gaussian vector BC with common message?

BIBLIOGRAPHIC NOTES

The Gaussian vector channel was first considered by Telatar (1999) and Foschini (1996). The capacity region of the Gaussian vector MAC in Theorem 9.2 was established by Cheng and Verdú (1993). The iterative water-filling algorithm in Section 9.2 is due to Yu, Rhee, Boyd, and Cioffi (2004). The capacity region of the degraded Gaussian product broadcast channel was established by Hughes-Hartogs (1975). The more general Theorem 9.3 and Proposition 9.1 appeared in El Gamal (1980). The private-message capacity region in (9.2) is due to Poltyrev (1977).

Caire and Shamai (2003) devised a writing on dirty paper scheme for the Gaussian vector broadcast channel and showed that it achieves the sum-capacity for $r = 2$ and $t = 1$ antennas. The sum-capacity for an arbitrary number of antennas was established by Vishwanath, Jindal, and Goldsmith (2003) and Viswanath and Tse (2003) using the BC–MAC

APPENDIX C

Cardinality Bounding Techniques

We introduce techniques for bounding the cardinalities of auxiliary random variables.

Convex Cover Method

The convex cover method is based on the following lemma (Ahlswede and Körner 1975, Wyner and Ziv 1976), which is a direct consequence of the Fenchel–Eggleston–Carathéodory theorem in Appendix A.

Support Lemma. Let $\mathcal{X}$ be a finite set and $\mathcal{U}$ be an arbitrary set. Let $\mathcal{P}$ be a connected compact subset of pmfs on $\mathcal{X}$ and $p(x|u) \in \mathcal{P}$, indexed by $u \in \mathcal{U}$, be a collection of (conditional) pmfs on $\mathcal{X}$. Suppose that $g_j(\pi)$, $j = 1, \ldots, d$, are real-valued continuous functions of $\pi \in \mathcal{P}$. Then for every $U \sim F(u)$ defined on $\mathcal{U}$, there exist a random variable $U' \sim p(u')$ with $|\mathcal{U}'| \le d$ and a collection of conditional pmfs $p(x|u') \in \mathcal{P}$, indexed by $u' \in \mathcal{U}'$, such that for $j = 1, \ldots, d$,

$$\int_{\mathcal{U}} g_j(p(x|u))\, dF(u) = \sum_{u' \in \mathcal{U}'} g_j(p(x|u'))p(u').$$

We now show how this lemma is used to bound the cardinality of auxiliary random variables. For concreteness, we focus on bounding the cardinality of the auxiliary random variable U in the characterization of the capacity region of the (physically) degraded DM-BC $p(y_1|x)p(y_2|y_1)$ in Theorem 5.2.

Let $U \sim F(u)$ and $X|\{U = u\} \sim p(x|u)$, where U takes values in an arbitrary set $\mathcal{U}$. Let $\mathcal{R}(U, X)$ be the set of rate pairs (R_1, R_2) such that

$$R_1 \le I(X; Y_1|U),$$
$$R_2 \le I(U; Y_2).$$

We prove that given any (U, X), there exists (U', X) with $|\mathcal{U}'| \le \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\} + 1$ such that $\mathcal{R}(U, X) = \mathcal{R}(U', X)$, which implies that it suffices to consider auxiliary random variables with $|\mathcal{U}| \le \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\} + 1$.

We first show that it suffices to take $|\mathcal{U}| \le |\mathcal{X}| + 1$. Assume without loss of generality that $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$. Given (U, X), consider the set $\mathcal{P}$ of all pmfs on $\mathcal{X}$ (which is

connected and compact) and the following $|\mathcal{X}| + 1$ continuous functions on $\mathscr{P}$:

$$g_j(\pi) = \begin{cases} \pi(j) & j = 1, \ldots, |\mathcal{X}| - 1, \\ H(Y_1) & j = |\mathcal{X}|, \\ H(Y_2) & j = |\mathcal{X}| + 1. \end{cases}$$

Clearly, the first $|\mathcal{X}| - 1$ functions are continuous. The last two functions are also continuous in π by continuity of the entropy function and linearity of $p(y_1) = \sum_x p(y_1|x)\pi(x)$ and $p(y_2) = \sum_x p(y_2|x)\pi(x)$ in π. Now by the support lemma, we can find a random variable U' taking at most $|\mathcal{X}| + 1$ values such that

$$H(Y_1|U) = \int_{\mathscr{U}} H(Y_1|U = u)\, dF(u) = \sum_{u'} H(Y_1|U' = u')p(u') = H(Y_1|U'),$$
$$H(Y_2|U) = \int_{\mathscr{U}} H(Y_2|U = u)\, dF(u) = \sum_{u'} H(Y_2|U' = u')p(u') = H(Y_2|U'),$$
$$\int_{\mathscr{U}} p(x|u)\, dF(u) = p(x) = \sum_{u'} p_{X|U}(x|u')p(u')$$

for $x = 1, \ldots, |\mathcal{X}| - 1$ and hence for $x = |\mathcal{X}|$. Since $p(x)$ determines $p(x, y_1) = p(x) \cdot p(y_1|x)$ and $p(y_2) = \sum_x p(x)p(y_2|x)$, $H(Y_1|X)$ and $H(Y_2)$ are also preserved, and

$$I(X; Y_1|U) = H(Y_1|U) - H(Y_1|X) = H(Y_1|U') - H(Y_1|X) = I(X; Y_1|U'),$$
$$I(U; Y_2) = H(Y_2) - H(Y_2|U) = H(Y_2) - H(Y_2|U') = I(U'; Y_2).$$

Thus we have shown that $\mathscr{R}(U, X) = \mathscr{R}(U', X)$ for some U' with $|\mathcal{U}'| \le |\mathcal{X}| + 1$.

Now we derive the bound $|\mathcal{U}'| \le |\mathcal{Y}_2| + 1$. Again assume that $\mathcal{Y}_2 = \{1, 2, \ldots, |\mathcal{Y}_2|\}$. Consider the following $|\mathcal{Y}_2| + 1$ continuous functions on $\mathscr{P}$:

$$g_j(\pi) = \begin{cases} p_{Y_2}(j) = \sum_x p_{Y_2|X}(j|x)\pi(x) & j = 1, \ldots, |\mathcal{Y}_2| - 1, \\ H(Y_2) & j = |\mathcal{Y}_2|, \\ I(X; Y_1) & j = |\mathcal{Y}_2| + 1. \end{cases}$$

Again by the support lemma, there exists U' (not necessarily the same as the one above) with $|\mathcal{U}'| \le |\mathcal{Y}_2| + 1$ such that $p(y_2) = \sum_{u',x} p(y_2|x)p(x|u')p(u')$, $H(Y_2)$, $H(Y_2|U')$, and $I(X; Y_1|U')$ stay the same as with the original U. Hence $\mathscr{R}(U, X) = \mathscr{R}(U', X)$. Note that the pmf $p(x)$ of X itself is not necessarily preserved under U', which, however, does not affect the quantities of interest.

In the same manner, we can show that $\mathscr{R}(U, X) = \mathscr{R}(U', X)$ for some U' with $|\mathcal{U}'| \le |\mathcal{Y}_1| + 1$ by preserving $p(y_1)$ instead of $p(y_2)$. In this case, the physical degradedness of the channel guarantees that preserving $p(y_1)$ also preserves $H(Y_2)$.

Combining the above three steps, we have shown that for every U there exists a random variable $U' \sim p(u')$ with $|\mathcal{U}'| \le \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\} + 1$ such that $\mathscr{R}(U, X) = \mathscr{R}(U', X)$. This establishes the desired cardinality bound on $\mathcal{U}$ in the capacity region characterization of the degraded broadcast channel in Theorem 5.2.

Remark C.1. We can apply the same technique to bound the cardinality of the time-sharing random variable Q appearing in the characterization of the DM-MAC capacity region. Recall that the capacity region of the DM-MAC is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2, Q), \\ R_2 &\le I(X_2; Y|X_1, Q), \\ R_1 + R_2 &\le I(X_1, X_2; Y|Q) \end{aligned}$$

for some pmf $p(q)p(x_1|q)p(x_2|q)$. By considering the set $\mathscr{P}$ of all *product* pmfs on $\mathcal{X}_1 \times \mathcal{X}_2$ (which is a connected compact subset of all pmfs on $\mathcal{X}_1 \times \mathcal{X}_2$) and the following three continuous functions:

$$\begin{aligned} g_1(p(x_1|q)p(x_2|q)) &= I(X_1; Y|X_2, Q = q), \\ g_2(p(x_1|q)p(x_2|q)) &= I(X_2; Y|X_1, Q = q), \\ g_3(p(x_1|q)p(x_2|q)) &= I(X_1, X_2; Y|Q = q), \end{aligned}$$

we can easily establish the cardinality bound $|\mathcal{Q}| \le 3$. Although this is weaker than the bound $|\mathcal{Q}| \le 2$ derived in Chapter 4, the technique described here is more general and can be easily extended to other scenarios encountered in the book.

Extension to Multiple Random Variables

The above technique can be extended to give cardinality bounds for capacity regions with two or more auxiliary random variables (Csiszár and Körner 1978, Nair and El Gamal 2009).

As an example, consider the 3-receiver degraded DM-BC $p(y_1, y_2, y_3|x)$. The capacity region is the set of rate triples (R_1, R_2, R_3) such that

$$\begin{aligned} R_1 &\le I(X; Y_1|V), \\ R_2 &\le I(V; Y_2|U), \\ R_3 &\le I(U; Y_3) \end{aligned}$$

for some pmf $p(u)p(v|u)p(x|v)$. We show that it suffices to take $|\mathcal{U}| \le |\mathcal{X}| + 2$ and $|\mathcal{V}| \le (|\mathcal{X}| + 1)(|\mathcal{X}| + 2)$.

First fix $p(x|v)$ and consider the following $|\mathcal{X}| + 2$ continuous functions of $p(v|u)$:

$$\begin{aligned} &p(x|u) = \sum_v p(x|v)p(v|u), \quad x = 1, \ldots, |\mathcal{X}| - 1, \\ &I(X; Y_1|V, U = u) = I(X; Y_1|U = u) - I(V; Y_1|U = u), \\ &I(V; Y_2|U = u), \\ &H(Y_3|U = u). \end{aligned}$$

As in the 2-receiver DM-BC example (with the bound $|\mathcal{U}| \le |\mathcal{Y}_2| + 1$), there exists U' with

$|\mathcal{U}'| \le |\mathcal{X}| + 2$ such that $p(x)$, $I(X; Y_1|V) = I(X; Y_1|V, U)$, $I(V; Y_2|U)$, and $I(U; Y_3)$ are preserved. Let V' denote the corresponding random variable.

Now for each $u' \in \mathcal{U}'$, consider the following $|\mathcal{X}| + 1$ continuous functions of $p(x|v', u')$: $p(x|v', u')$, $H(Y_1|V' = v', U' = u')$, and $H(Y_2|V' = v', U' = u')$. Then as in the 2-receiver DM-BC example (with the bound $|\mathcal{U}| \le |\mathcal{X}| + 1$), for each u', there exists $V''|\{U' = u'\} \sim p(v''|u')$ such that the cardinality of the support of $p(v''|u')$ is upper bounded by $|\mathcal{X}| + 1$ and the quantities $p(x|u')$, $I(X; Y_1|V', U' = u')$, and $I(V'; Y_2|U' = u')$ are preserved. By relabeling the support of $p(v''|u')$ for each $u' \in \mathcal{U}'$ and redefining $p(x|v'', u')$ accordingly, we can construct V'' with $|\mathcal{V}''| \le |\mathcal{X}| + 1$. However, this choice does not satisfy the Markov chain condition $U' \to V'' \to X$ in general!

Instead, consider $V''' = (U', V'')$. Then $|\mathcal{V}'''| \le |\mathcal{U}'| \cdot |\mathcal{V}''| \le (|\mathcal{X}| + 1)(|\mathcal{X}| + 2)$ and $U' \to V''' \to X \to (Y_1, Y_2, Y_3)$ form a Markov chain. Furthermore,

$$\begin{aligned} I(X; Y_1|V''') &= I(X; Y_1|U') - I(V'''; Y_1|U') \\ &= I(X; Y_1|U') - I(V''; Y_1|U') \\ &= I(X; Y_1|U') - I(V'; Y_1|U') \\ &= I(X; Y_1|V') \\ &= I(X; Y_1|V). \end{aligned}$$

Similarly

$$I(V'''; Y_2|U') = I(V''; Y_2|U') = I(V'; Y_2|U') = I(V; Y_2|U),$$
$$I(U'; Y_3) = I(U; Y_3).$$

This completes the proof of the cardinality bound.

Perturbation Method

The technique described above does not provide cardinality bounds for all capacity/rate–distortion regions. Most notably, the cardinality bounds on U_0, U_1, and U_2 in Marton's inner bound for the DM-BC in Theorem 8.4 require a different technique based on a perturbation method introduced by Gohari and Anantharam (2009) and further simplified by Jog and Nair (2010).

To be concrete, we consider the maximum sum-rate for Marton's inner bound in Theorem 8.3,

$$\max_{p(u_1,u_2),\, x(u_1,u_2)} \big(I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2)\big), \tag{C.1}$$

and show that it suffices to take $|\mathcal{U}_1|, |\mathcal{U}_2| \le |\mathcal{X}|$.

Let (U_1, U_2, X) be the random variables that attain the maximum in (C.1) and let $(U_1', U_2', X') \sim p_\epsilon(u_1', u_2', x') = p_{U_1,U_2,X}(u_1', u_2', x')(1 + \epsilon\phi(u_1'))$ be its *perturbed* version. We assume that $1 + \epsilon\phi(u_1) \ge 0$ for all u_1 and $\mathsf{E}(\phi(U_1)) = \sum_{u_1} p(u_1)\phi(u_1) = 0$, so that $p_\epsilon(u_1', u_2', x')$ is a valid pmf. We further assume that

$$\mathsf{E}(\phi(U_1)|X = x) = \sum_{u_1,u_2} p(u_1, u_2|x)\phi(u_1) = 0$$

for all $x \in \mathcal{X}$, which implies that $p_\epsilon(x') = p_X(x')$. In other words, the perturbation preserves the pmf of X and hence the pmf of (Y_1, Y_2). Note that there exists a nonzero perturbation that satisfies this condition (along with $\mathsf{E}(\phi(U_1)) = 0$) as long as $|\mathcal{U}_1| \geq |\mathcal{X}| + 1$, since there are $|\mathcal{X}| + 1$ linear constraints. Note also that X' continues to be a function of (U_1', U_2'), since $p_{U_1,U_2,X}(u_1', u_2', x') = 0$ implies that $p_\epsilon(u_1', u_2', x') = 0$. Now consider

$$\begin{aligned}
&I(U_1'; Y_1') + I(U_2'; Y_2') - I(U_1'; U_2') \\
&\quad = H(Y_1') + H(Y_2') + H(U_1', U_2') - H(U_1', Y_1') - H(U_2', Y_2') \\
&\quad = H(Y_1) + H(Y_2) + H(U_1', U_2') - H(U_1', Y_1') - H(U_2', Y_2') \\
&\quad = H(Y_1) + H(Y_2) + H(U_1, U_2) - H(U_1, Y_1) + \epsilon H_\phi(U_1, U_2) - \epsilon H_\phi(U_1, Y_1) - H(U_2', Y_2'),
\end{aligned}$$

where

$$\begin{aligned}
H_\phi(U_1, U_2) &= -\sum_{u_1,u_2} p(u_1, u_2)\phi(u_1) \log p(u_1, u_2), \\
H_\phi(U_1, Y_1) &= -\sum_{u_1,y_1} p(u_1, y_1)\phi(u_1) \log p(u_1, y_1).
\end{aligned}$$

Since $p(u_1, u_2, x)$ attains the maximum in (C.1),

$$\frac{\partial^2}{\partial \epsilon^2}\left(I(U_1'; Y_1') + I(U_2'; Y_2') - I(U_1'; U_2')\right)\Big|_{\epsilon=0} = -\frac{\partial^2}{\partial \epsilon^2} H(U_2', Y_2')\Big|_{\epsilon=0} \leq 0.$$

It can be shown by simple algebra that this is equivalent to $\mathsf{E}(\mathsf{E}(\phi(U_1)|U_2, Y_2)^2) \leq 0$. In particular, $\mathsf{E}(\phi(U_1)|U_2 = u_2, Y_2 = y_2) = 0$ whenever $p(u_2, y_2) > 0$. Hence $p_\epsilon(u_2', y_2') = p_{U_2,Y_2}(u_2', y_2')$ and $H(U_2', Y_2') = H(U_2, Y_2)$. Therefore

$$\begin{aligned}
&I(U_1'; Y_1') + I(U_2'; Y_2') - I(U_1'; U_2') \\
&\quad = H(Y_1) + H(Y_2) + H(U_1, U_2) - H(U_1, Y_1) - H(U_2, Y_2) + \epsilon H_\phi(U_1, U_2) - \epsilon H_\phi(U_1, Y_1).
\end{aligned}$$

Once again by the optimality of $p(u_1, u_2)$, we have

$$\frac{\partial}{\partial \epsilon} I(U_1'; Y_1') + I(U_2'; Y_2') - I(U_1'; U_2') = H_\phi(U_1, U_2) - H_\phi(U_1, Y_1) = 0,$$

which, in turn, implies that

$$I(U_1'; Y_1') + I(U_2'; Y_2') - I(U_1'; U_2') = I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2),$$

that is, $(U_1', U_2', X') \sim p_\epsilon(u_1', u_2', x')$ also attains the maximum in (C.1). Finally, we choose the largest $\epsilon > 0$ such that $1 + \epsilon\phi(u_1) \geq 0$, that is, $1 + \epsilon\phi(u_1^*) = 0$ for some $u_1^* \in \mathcal{U}_1$. Then $p_\epsilon(u_1^*) = 0$, i.e., $|\mathcal{U}_1'| \leq |\mathcal{U}_1| - 1$, and the maximum is still attained.

We can repeat the same argument as long as $|\mathcal{U}_1'| \geq |\mathcal{X}| + 1$. Hence by induction, we can take $|\mathcal{U}_1'| = |\mathcal{X}|$ while preserving $I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2)$. Using the same argument for U_2, we can take $|\mathcal{U}_2'| = |\mathcal{X}|$ as well.